CANADIAN PSYCHOLOGICAL ASSOCIATION
SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale

ISSN: 1196-1961

2021, Vol. 75, No. 2, 211–220
https://doi.org/10.1037/cep0000223

Tracking Nonliteral Language Processing Using Audiovisual Scenarios

Kathrin Rothermich[1], Elizabeth Schoen Simmons[2], Pavitra Rao Makarla[3], Lauren Benson[4, 5], Emma Plyler[1], Hana Kim[1, 7], and Gitte Henssel Joergensen[6, 8]

[1] Department of Communication Sciences and Disorders, East Carolina University
[2] Department of Communication Disorders, Sacred Heart University
[3] Department of Cognitive Science, University of Connecticut
[4] School of Informatics and Computing, Indiana University
[5] Department of Cognitive Science, Indiana University
[6] Department of Psychological Sciences, University of Connecticut
[7] Department of Neurology, Johns Hopkins University School of Medicine
[8] Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut

Recognizing sarcasm and jocularity during face-to-face communication requires the integration of verbal, paralinguistic, and nonverbal cues, yet most previous research on nonliteral language processing has been carried out using written or static stimuli. In the current study, we examined the processing of dynamic literal and nonliteral intentions using eye tracking. Participants ($N = 37$) viewed short, ecologically valid video vignettes and were asked to identify the speakers' intention. Participants had greater difficulty identifying jocular statements as insincere in comparison to sarcastic statements and spent significantly more time looking at faces during nonliteral versus literal social interactions. Finally, participants took longer to shift their attention from one talker to the other talker during interactions that conveyed literal positive intentions compared with jocular and literal negative intentions. These findings currently support the *Standard Pragmatic Model* and the *Parallel-Constraint-Satisfaction Model* of nonliteral language processing.

Public Significance Statement
By using dynamic, ecologically valid stimuli during an eye tracking task, our data add to the understanding of speaker intentions such as sarcasm or jocularity. For example, participants spent more time watching nonliteral compared with literal interactions, and they had a harder time identifying the cues used to express jocularity compared with sarcasm. Our findings confirm existing theories of nonliteral language perception by showing differentiated processing of the two forms of speaker intentions.

Keywords: pragmatics, social perception, sarcasm, irony, prosody

Sarcasm and jocularity are nonliteral forms of language that each serve a set of social pragmatic functions (Jorgensen, 1996; Toplak & Katz, 2000). The processing of these complex processes requires the interpretation of the intended meaning, integration of contextual and nonverbal cues, and use of Theory of Mind (ToM). Sarcasm (sometimes called ironic criticism) is considered an aggressive form of language that is usually directed towards a target (Attardo, 2000) and used to indirectly convey criticism (Shany-Ur et al., 2012) or to express humor (Caucci & Kreuz, 2012). Similarly, jocularity (sometimes called an ironic compliment) serves as a means to convey humor (Katz et al., 2004), and often combines provocation with nonseriousness (Haugh, 2016). There is limited work on processing both sarcasm (e.g., Katz et al., 2004) and jocularity despite the complex nature of these forms of nonliteral language (Pexman & Olineck, 2002).

Sarcasm is a positive statement that conveys a negative intention, whereas jocularity is a negative statement that conveys a positive intention. Jocularity may also be described as "positive" sarcasm, or teasing, and is considered a risky form of communication because it can be interpreted as an insult if listeners miss crucial contextual, auditory, or visual cues required for a successful interpretation (Gibbs, 2000; Seckman & Couch, 1989). A sarcastic comment, such as "Sure, I'd love to," paired with an eye-roll, in response to a request to attend a less than desirable event implies a negative intention despite using affirmative

This article was published Online First April 1, 2021.
Kathrin Rothermich https://orcid.org/0000-0002-7586-3842
Pavitra Rao Makarla https://orcid.org/0000-0001-7169-3877
Lauren Benson https://orcid.org/0000-0002-9930-4033
Emma Plyler https://orcid.org/0000-0001-8580-1007
We would like to thank Maria Lattanzi, Julia Mocciola, and Tom Pietruszewski for help with data acquisition, as well as Gerry Altmann and Eiling Yee for sharing their laboratory equipment. We would also like to thank Vicky Tzuyin Lai for their helpful comments and suggestions.
Correspondence concerning this article should be addressed to Kathrin Rothermich, Department of Communication Science and Disorders, East Carolina University, 3310 AD Health Sciences Bldg. 600 Moye Blvd., Mail Stop 668, Greenville, NC 27858, United States. Email: ROTHERMICHK17@ECU.EDU

language. Jocularity often functions in the reverse; an example of this effect could be responding with "No, this is disgusting!" when asked if enjoying a clearly delicious homemade pastry. If this comment is paired with obvious enjoyment of the pastry, the intent and meaning are both nonserious and positive, despite the negative language use. When jocularity is perceived as intended, it is thought to increase the affiliation between interlocutors because it is often accompanied by shared laughter (Gibbs, 2000). It has been debated in the literature if sarcasm and jocularity increase or decrease the positive or negative effect of a statement (Filik et al., 2017). On the one hand, Dews and Winner (Dews & Winner, 1999) showed that sarcastic statements were judged less hurtful when compared with literal negative statements, supporting the *Tinge Hypothesis*. This hypothesis suggests that the superficial positive structure of a sarcastic utterance "tinges" an overall more positive interpretation, compared with a literal negative (i.e., blunt) utterance (Dews et al., 1995). While some studies have confirmed the *Tinge Hypothesis* (e.g., Dews & Winner, 1999; Pomareda et al., 2019), other data show that sarcasm can be evaluated as more critical in comparison with literal negative statements (e.g., Kreuz et al., 1991; Pexman & Olineck, 2002). A related phenomenon is the so-called "asymmetry of affect" (Clark & Gerrig, 1984), which refers to the asymmetric findings regarding the use and perception of sarcasm and jocularity. It states that sarcasm generally occurs more frequently than jocularity (Clark & Gerrig, 1984; Matthews et al., 2006; Sperber & Wilson, 1981) and there is additional data showing that sarcasm is often recognized earlier and more accurately compared with jocularity (Caillies et al., 2019; Hancock et al., 2000; Harris & Pexman, 2003).

Both sarcasm and jocularity are considered complex language functions because comprehension depends on several mental operations, including the inhibition of the literal meaning (Angeleri & Airenti, 2014). Thus, several models have been proposed to describe the nature of sarcasm and jocularity processing. For example, the *Standard Pragmatic Model* assumes that sarcasm and jocularity are harder to process compared with literal statements, due to the extra effort caused by rejection of the literal meaning, with no influence of a situational context (Gibbs, 1999; Grice, 1989). The *Graded Salience Hypothesis* predicts that the most salient meaning is processed first, regardless of the context (Fein et al., 2015; Giora, 1997; Giora et al., 2007). This is usually the literal meaning, but for common sarcastic remarks (e.g., " Yeah, right"), both the sarcastic and literal interpretation are activated in parallel (Ţurcan & Filik, 2016); it also suggests that familiarity with the nonliteral statement predicts the nature of processing. The *Parallel-Constraint-Satisfaction Model* (Katz et al., 2004; Katz & Ferretti, 2001; Pexman, 2008) assumes that listeners utilize various cues, such as situational context, familiarity, and prosody, to process nonliteral language. The model predicts that all cues are integrated as soon as they are relevant in order to infer the speaker's intention, and a nonliteral interpretation of a statement is considered as soon as there is adequate information available (Spotorno et al., 2013). While the *Parallel-Constraint-Satisfaction* model makes assumptions about the influence of nonverbal cues on the interpretation of nonliteral language, many frameworks and empirical studies do not address the multimodal nature of everyday nonliteral language.

Nonliteral language processing has been traditionally investigated using offline measures, such as questionnaires (Creusere, 2000; Dews et al., 1995; Leggitt & Gibbs, 2000; Milanowicz et al., 2017). Recently, eye tracking has been used to examine how individuals process nonliteral statements over time in order to provide empirical support for the theoretical frameworks. A common finding of reading studies shows that nonliteral language increases the rereading of sentences (Filik & Moxey, 2010). For example, Kaakinen et al. (2014) used eye tracking to examine how readers process sentences embedded within either nonliteral or literal contexts. They found that fixation times were longer when the participants read the nonliteral context, and the readers often reread the sentence before moving on to the next sentence. Similarly, Olkoniemi et al. (2016) observed that sarcastic utterances resulted in longer reading times, suggesting that sarcastic utterances were more difficult to process compared with literal utterances. Many of these findings are compatible with the *Standard Pragmatic Model* (Grice, 1989). While these studies provide emerging insight into the nature of nonliteral language processing, less is known about how processing unfolds during social interactions that require rapid integration of linguistic and paralinguistic input.

For these reasons, Deliens et al. (2018) developed a series of videos including literal and ironic messages exchanged between two interlocutors to simulate authentic communicative interactions. They tracked participants' eye movements and discovered that nonverbal cues, such as prosody and facial expressions, impact irony comprehension. Reaction times were shorter in the presence of prosody and facial expression, and accuracy was higher in the presence of prosody. However, there was no cumulative effect of prosody and facial expression on accuracy. In summary, the precise role of noncontextual cues in processing nonliteral language still remains understudied, especially the role of attention toward the communication partners in a scene, as well as the processing of sarcasm and jocularity over time. Similarly to Deliens and colleagues, we employed an ecologically valid, dynamic set of stimuli that included two people engaging in brief discourse. This provided the opportunity to conduct an analysis on the time course of nonliteral language processing including which face captures their locus of attention.

Current Study

The current study will help us better understand the processing of literal or nonliteral exchanges in the context of social interactions. Our audio-visual stimuli were taken from the Relational Inference in Social Communication (RISC) database (Rothermich & Pell, 2015) which was developed to study the perception of talker intentions. Each video includes a short conversation between two adults, accompanied by prosody, facial expressions, gestures, and body language. Using the RISC database has two main advantages: (a) the actors utter the same content in matching literal and nonliteral scenes, and (b) the person receiving the sarcastic or jocular comment is visible and responds contingently. The presence of the asker (asking a question) and the responder (responding in a literal or nonliteral fashion) in each video allows us to compare the allocation of attention over time between the two interactants, which we hypothesize may differ between response types (literal or nonliteral).

We employed a behavioral task that tested the ability of young adults to judge the sincerity of literal and nonliteral video vignettes while monitoring their eye movements as they viewed the two actors engage in conversation. Our eye tracking analyses focused on fixation proportions to the actors' faces in the video exchanges as the social pragmatic message unfolded over time. We extracted fixation duration and then computed mean fixation proportion to the actor asking a question (asker) and the actor responding to the asker in a literal or nonliteral manner (responder). The analysis of mean fixation proportion to the asker and responder allows us to evaluate differences in processing between talkers, language types (literal vs. nonliteral), and nonliteral language functions (sarcasm vs. jocularity). We also conducted a peak analysis to test the timing differences for peak mean fixation proportions during the interactions in order to better understand the time course of processing information provided by the asker and the responder. While fixation duration and spatial fixation proportion have been widely used as measures of visual attention, cognitive processing, and sentence processing, we did not find studies using these metrics in the context of nonliteral language processing. In general, these measures have been considered to reflect cognitive load (e.g., Geisen & Bergstrom, 2017; Rosch & Vogel-Walcutt, 2013). If nonliteral expressions are more effortful to process, participants are likely to spend a greater amount of time evaluating the underlying meaning of these expressions. As a result, we may observe greater fixation durations and fixation proportions along with increased latency of fixation proportion peaks toward the face of the responder compared with the remainder of the scene when presented with nonliteral expressions compared with literal expressions.

Predictions

Based on the models of nonliteral language processing, the following a priori predictions for both the behavioral task and eye tracking data are presented below.

Behavioral Task

The *Standard Pragmatic Model Hypothesis* would predict that participants will be better at identifying literal language as sincere compared with nonliteral language. We also expect higher accuracy scores for literal videos due to their unmarkedness, as being sincere is often seen as the default intention (Rankin et al., 2009) (Hypothesis 1; H1). We also hypothesize that participants will perform better on the identification of sarcastic compared with jocular scenes as insincere (but see also Rothermich & Pell, 2015) based on previous studies using the same materials and similar tasks (Jakobson et al., 2018; Rothermich et al., 2019) (Hypothesis 2; H2).

Eye Movements

Based on the *Standard Pragmatic Model* and the *Parallel-Constraint-Satisfaction Model*, we posit that participants will fixate faces longer, as measured by higher mean fixation proportions, during nonliteral scenes as both models stress the importance of contextual input for nonliteral language processing (Hypothesis 3; H3). We also predict that participants will spend more time fixating on the face of the responder compared with the face of the asker as the responder will provide the critical information to accurately identify the sincerity of the response. Finally, we expect fixation proportions to the face of the responder to peak later in the time course during sarcastic and jocular statements compared to literal statements. This is predicted by the *Standard Pragmatic Model* given this framework posits that nonliteral language first requires the listener to reject the default, literal interpretation and then weigh the contextual evidence to process the nonliteral interpretation (Hypothesis 4; H4).

Method

Participants

Forty-six native English speakers with typical or corrected to normal hearing and vision were recruited from the University of Connecticut Psychology Participant Pool. All participants were undergraduate college students at the time of participation and were compensated with psychology course credit. This study was approved by the University of Connecticut Institutional Review Board. Nine participants were excluded due to eye tracker calibration issues, missing data, or fatigue. The final sample consisted of 37 participants (24 female, 13 male, mean age = 18.89 years, SD = 1.05 years).

Dynamic Stimuli

Our stimuli consisted of 192 short videos (4 dyads × 4 intentions × 12 scenes) derived from the RISC database (Rothermich & Pell, 2015). Each video depicts a dyadic social interaction between two out of the four possible actors, each who assumed a unique fictional identity they consistently portrayed throughout all the videos. Four types of paired relationships were exhibited between dyads, including a set of female friends (Lisa and Anna), a mixed-sex couple (Paul and Lisa), a male employer with their employee (Paul and Peter), and a mixed pair of colleagues (Peter and Anna). The length of the interactions varied from 5 to 10 s. Each video included an actor asking a question (asker) and a different actor responding to the question (responder). The responder communicated one of the following intentions when answering the question: literal positive, literal negative, jocularity, or sarcasm. For example, in one of the scenarios, an actor held a plate of cookies and asked if the other actor in the scene wanted one. The responder replied with one of the following statements: a *literal positive* response that was expressed as a sincere positive opinion ("Mmm, they look good!"); a *sarcastic* response which was lexically identical to the literal positive statement but expressed an insincere positive opinion ("Mmm, they look so good!" coordinated with a disgusted facial expression and/or tone); a *literal negative* (blunt) response that was expressed as a sincere negative opinion ("Honestly, they don't look very appetizing"); or a jocular response which was lexically identical to the literal negative statement and expressed a negative opinion that is intended to be interpreted as positive ("Honestly, they don't look very appetizing," but proceeds to laugh and take a cookie anyway; see Figure 1). Although participants are presented with the same scenario (e.g., topic content of cookies) more than once, the pseudo randomization of trial order assures that similar scenes in terms of content were presented maximally distant from one another. Participants were never presented with duplicate trials.

Figure 1
Screenshots of an Example Scene

Note. See the online article for the color version of this figure.

Procedure

The experiment was conducted using an Eyelink 1000 Plus eye-tracker (SR Research, Ltd., Ontario, Canada), which sampled at 500 Hz from the right eye, but viewing was binocular. Participants were seated in front of a 24" LED monitor with their eyes approximately 60 cm from the display. Each participant was presented with 192 colored videos from the RISC database (4 dyads × 4 intentions × 12 scenes). Each video was followed by a yes or no question ("Was the response sincere?") and participants were required to press a button indicating either yes or no. Trials were presented in a fixed randomized order with four counter-balanced lists to minimize inadvertent order effects. Participants were informed when they had completed 50% of the trials and were encouraged to take a short break before completing the second half.

Data Analysis Eye Tracking

We used Data Viewer (SR Research, Ltd.) to extract fixation duration to preselected regions of interest, i.e., faces of the asker and responder in each video interaction. The fixation analyses were restricted to the response segment of the video. We included fixations from the onset of the response segment through the end of the response in our analyses. Length of responses varied between videos. Fixations were then binned in 50 ms segments to obtain the mean fixation proportion for each region of interest (asker/responder) for each stimulus. The duration for each fixation was analyzed over the length of the response and served as the second dependent variable. Since the length of the response in each video varies (M = 1,961 ms; SD = 621 ms), we decided to normalize the time window by dividing the time point for each fixation by the total length of the video using a custom R script. This changed the time points to percentages and allowed us to compare the fixation measures and peak timing using the same analysis window. Peak timing served as the third dependent variable and was measured by extracting the time of the peak fixation proportion in the normalized time window.

Statistical Analysis

The data was statistically analyzed using R (R Development Core Team, 2016) and linear mixed-effects (LME) using lme4 in R (Bates et al., 2013). To compute post hoc p values, we applied the Satterthwaite (1946) approximation, which is implemented in the lmerTest package version 2.0-6 (Kuznetsova et al., 2017). Separate LME models were built for dependent variables of accuracy as well as mean fixation proportion, fixation duration, and peak timing of the fixation proportion. To compare models, we first defined a base model, which included only one random effect (subject). Further models were then identified by performing comparisons using the ANOVA function in R, and we systematically compared the full model with the model reduced by random and fixed effects in turn (see a similar approach in Valuch et al., 2015). Our fixed effects were intention (literal positive, literal negative, sarcasm, jocularity) and asker/responder for the eye tracking analysis. Random effects included intercepts for subjects and items. Models were compared based on χ^2, Akaike information criterion (AIC; Hu, 2007), and p values. For post hoc comparisons we report β estimates, z values, and p values.

Results

Behavioral Results

Participants demonstrated a high accuracy level for identifying speaker intentions (M = 88.51%, SD = 18.35%); they were most

accurate at identifying sarcasm (96% correct), followed by literal positive intentions (90% correct) and literal negative intentions (88% correct). Participants were least accurate at identifying jocularity (79% correct). A linear mixed-effects model was used to examine how speaker intention accounted for variability in participant accuracy. Intention was included as a fixed effect and subject, item, and speakers' relationship were included as random effects. Intention in the accuracy model turned out as a main effect as it improved the model significantly. Post hoc comparisons reveal significantly higher accuracy for sarcastic versus literal positive ($\beta = -.79$, $z = -28.50$, $p < .0001$), literal negative ($\beta = -1.47$, $z = -34.66$, $p < .0001$), and jocular ($\beta = -2.26$, $z = -55.47$, $p < .0001$) items. We also found higher accuracy for literal negative ($\beta = -.79$, $z = -28.50$, $p < .0001$) and literal positive ($\beta = -1.18$, $z = -38.11$, $p < .0001$) compared with jocular scenes.

Eye Movement Results

Fixation Proportion

We analyzed fixation proportion by intention and asker/responder (see Figure 2 and Table 1 for details) and found a main effect of intention when comparing the base model with the model including this fixed factor. Jocularity received a greater number of fixations when compared with literal negative, literal positive, and sarcastic vignettes. The differences were significant between all conditions with the exception of the literal positive and literal negative items. We also found a main effect of asker/responder, as revealed by a better fit for the model containing asker/responder as a fixed factor compared with the base model. Thereby, the responder in the video received significantly more fixations when compared with the asker. Finally, there was a significant interaction between asker/responder and intention. Overall, we found significant differences between all intentions when they fixate on the asker (with the exception of literal negative vs. sarcasm), and similarly, for when they fixate on the responder (for detailed statistics, see Table 1).

Fixation Duration

We also analyzed fixation duration by intention and asker/responder (see Figure 3 and Table 2 for details). There was a main effect of intention when comparing the base model with the fixed factor model. Jocularity received the longest fixations, followed by sarcasm, literal positive, and literal negative (for details, see Table 2). Significant differences were found between all conditions but not between sarcasm and jocularity or between literal positive and literal negative.

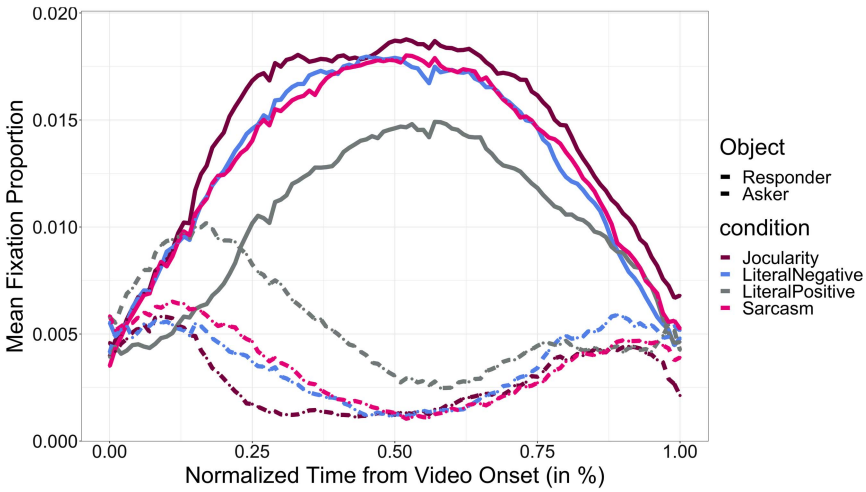
There was also a main effect of asker/responder, due to the model containing asker/responder as a fixed factor being a better fit, when compared with the base model. Therefore, the participants appeared to fixate on the responders longer ($M = 1{,}048$ ms) in comparison with the asker ($M = 472$ ms). Last, there was a significant interaction between intention and asker/responder. In general, there was no significant difference between the jocularity and literal negative conditions, the jocularity and literal positive conditions, and the literal negative and literal positive conditions for the asker. For the asker, the duration was highest for sarcasm, with literal positive, literal negative, and jocularity following it (see Table 2 for details). In addition, there was a significant difference between all intentions when they fixate on the responder, except for jocularity and sarcasm and except for literal negative and literal positive. For the responder, the duration was highest for jocularity, with sarcasm, literal negative, and literal positive following in suit (see Table 2 for details).

Peak Timing Analysis

Finally, we were interested in differences in peak timing of the fixation proportions. The analysis revealed a main effect of intention for the asker; jocularity fixation proportions peaked earlier than literal negative proportions, and literal positive peaked earlier than literal negative ones (see Table 3 for details). No significant results were found for the responder region of interest.

Figure 2

Mean Fixation Proportion of Asker/Responder for All Intentions Averaged Over All Videos



Note. See the online article for the color version of this figure.

Table 1
LMER Models and Post Hoc Comparisons for Mean Fixation Proportion. Italics Indicate Significant Results P *Values Were Adjusted Using the Tukey Method*

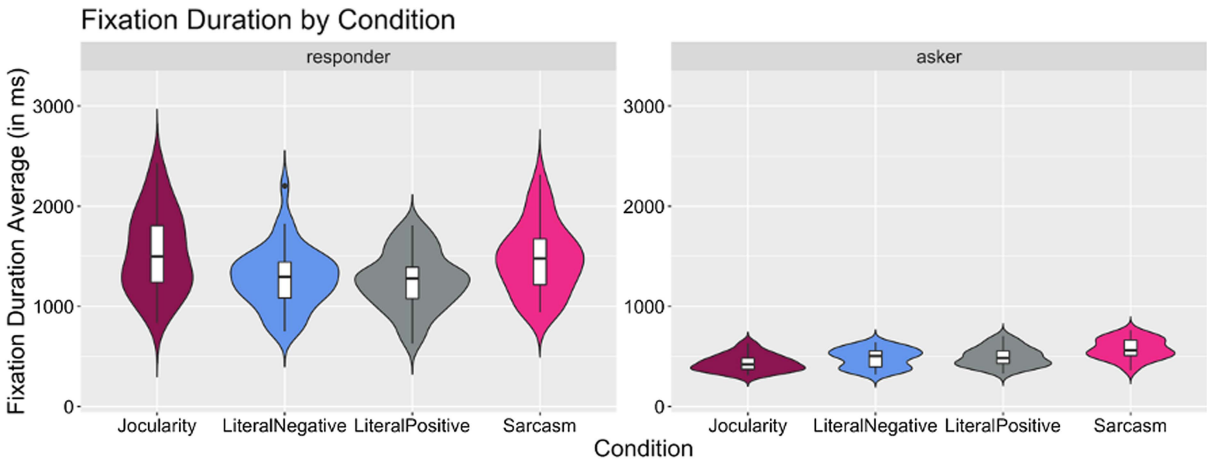
Models	AIC	χ^2	*p*
model.1 <- lmer(fixprop ~ 1 + (1|subject), data=data_fix_prop)	–1,902,440		
*model.1 <- lmer(*fixprop ~ *intention + (1|subject), data=data_fix_prop)*	*–1,902,708*	*274.82*	*<.0001*
*model.1 <- lmer(*fixprop ~ *intention * asker/responder + (1|subject), data=data_fix_prop)*	*–2,022,696*	*119,996*	*<.0001*
Post hoc comparisons	β	*z*	*p*
Jocularity—literal negative	*.0007*	*16.282*	*<.0001*
Jocularity—literal positive	*.0002*	*5.505*	*<.0001*
Jocularity—sarcasm	*.0004*	*10.476*	*<.0001*
Literal negative—literal positive	–.000006	–.14	.999
Literal negative—sarcasm	*–.00049*	*–10.737*	*<.0001*
Literal positive—sarcasm	*.00025*	*5.579*	*<.0001*
Post hoc comparisons—asker	β	*z*	*p*
Literal negative—literal positive	*–.0020770*	*–38.61*	*<.0001*
Literal negative—sarcasm	–.0000220	–.41	.9759
Literal negative—jocularity	*.0005788*	*–11.03*	*<.0001*
Literal positive—sarcasm	*.0020550*	*38.64*	*<.0001*
Literal positive—jocularity	*.0026558*	*50.49*	*<.0001*
Sarcasm—jocularity	*.0006008*	*11.592*	*<.0001*
Post hoc comparisons—responder	β	*z*	*p*
Literal negative—literal positive	*.0030612*	*56.91*	*<.0001*
Literal negative—sarcasm	*.0000089*	*.168*	.9983
Literal negative—jocularity	*–.0041513*	*–20.78*	*<.0001*
Literal positive—sarcasm	*–.0030523*	*–57.39*	*<.0001*
Literal positive—jocularity	*–.0041513*	*–78.93*	*<.0001*
Sarcasm—jocularity	*.0010991*	*21.20*	*<.0001*

Discussion

The comprehension and processing of nonliteral language has been widely studied, especially using written materials, while fewer studies have used dynamic, multimodal stimuli to evaluate these complex language functions during daily social interactions. Therefore, we conducted an eye tracking study to examine the processing of literal and nonliteral intentions using videos of dyadic interactions (RISC video database; Rothermich & Pell, 2015). We monitored the fixations to faces of two actors during social scenes while participants viewed the dynamic video stimuli.

Participants performed well above chance level when identifying the sincerity of literal and nonliteral intentions (between 79% and 96% accurate). Based on the *Standard Pragmatic Model*, we predicted that participants would be more accurate at identifying literal language as sincere compared with nonliteral language and expected higher accuracy scores for literal videos (H1). However, our data failed to support this hypothesis. It is possible that the cues used to signal sarcastic

Figure 3
Mean Fixation Duration of Asker/Responder for All Intentions Averaged Over All Videos



Note. See the online article for the color version of this figure.

Table 2

LMER Models and Post Hoc Comparisons for Mean Fixation Duration. Italics Indicate Significant Results P *Values Were Adjusted Using the Tukey Method*

Models	**AIC**	χ^2	***p***
model.1 <- lmer(fixduration ~ 1 + (1|subject), data=data_fix_prop)	179,525		
*model.1 <- lmer(*fixduration ~ *intention + (1|subject), data=data_fix_prop)*	*179,454*	*76.84*	*<.0001*
*model.1 <- lmer(*fixduration ~ *intention * asker/responder + (1|subject), data=data_fix_prop)*	*176,721*	*2740.8*	*<.0001*
Post hoc comparisons	β	*z*	***p***
Jocularity—literal negative	*158.58*	*6.355*	*<.0001*
Jocularity—literal positive	*159.13*	*6.255*	*<.0001*
Jocularity—sarcasm	6.77	.271	.9931
Literal negative—literal positive	.54	.022	1
Literal negative—sarcasm	*–151.82*	*–6.164*	*<.0001*
Literal positive—sarcasm	*–152.36*	*–6.063*	*<.0001*
Jocularity—literal negative	*158.58*	*6.355*	*<.0001*
Post hoc comparisons—asker	β	*z*	***p***
Literal negative—literal positive	–10.9	–.32	.9883
Literal negative—sarcasm	*–88.5*	*–2.69*	*.0356*
Literal negative—jocularity	45.4	1.34	.5395
Literal positive—sarcasm	–77.6	–2.30	.0988
Literal positive—jocularity	56.3	1.61	.3704
Sarcasm—jocularity	*133.9*	*3.93*	*.0005*
Post hoc comparisons—responder	β	*z*	***p***
Literal negative—literal positive	*27.6*	*.94*	*.7821*
Literal negative—sarcasm	*–193.0*	*–6.68*	<.0001
Literal negative—jocularity	*–249.6*	*8.64*	<.0001
Literal positive—sarcasm	*–220.5*	*–7.52*	*<.0001*
Literal positive—jocularity	*–277.2*	*9.46*	*<.0001*
Sarcasm—jocularity	*–56.7*	*1.96*	*.2034*

intentions were exaggerated and thus most salient for our group of participants. Previous studies have shown that exaggerated cues are often used to signal sarcasm (Attardo et al., 2003). The *Parallel-Constraint-Satisfaction Model* assumes that all features of situational context have a facilitatory effect on interpreting nonliteral language (Katz et al., 2004); thus, multiple cues can be activated in parallel and help to narrow down possible interpretations during online processing (Kaakinen et al., 2014). While the *Parallel-Constraint-Satisfaction Model* stresses the importance of cues that other models overlook, such as the multimodal and social input that occurs when conveying nonliteral intentions, it does not offer specific predictions (Pexman et al., 2019). In addition, our sample of young adults may have been particularly familiar with sarcasm, and potentially employ sarcasm in daily life, which could lead to interpretation advantages (Ivanko et al., 2004). Finally, given the evidence that highly familiar sarcastic utterances are often processed faster than their literal counterparts (Filik et al., 2014; Ţurcan & Filik, 2016), our sarcasm tokens may have been very familiar, leading to higher accuracies. While we did not systematically vary familiarity in the current experiment, future iterations of this work will do so.

Table 3

LMER Models and Post Hoc Comparisons for Peak Timing. Italics Indicate Significant Results. P *Values Were Adjusted Using the Tukey Method*

Models—asker	**AIC**	χ^2	***p***
model.1 <- lmer(peak ~ 1 + (1|subject), data=data_fix_prop)	156.51		
*model.1 <- lmer(*peak ~ *intention + (1|subject), data=data_fix_prop)*	*145.59*	*16.92*	*<.001*
Post hoc comparisons—asker	β	**z**	***p***
Literal negative—literal positive	*.3254*	*3.66*	*.0022*
Literal negative—sarcasm	.1484	1.67	.3456
Literal negative—jocularity	*.3073*	*3.45*	*.0043*
Literal positive—sarcasm	–.1770	–1.99	.1980
Literal positive—jocularity	–.0181	–.20	.9970
Sarcasm—jocularity	.1589	1.79	.2854
Models—responder	**AIC**	χ^2	***p***
model.1 <- lmer(peak ~ 1 + (1|subject), data=data_fix_prop)	–156.22		
model.1 <- lmer(peak ~ intention + (1|subject), data=data_fix_prop)	–155.68	5.46	.1408

Our second hypothesis (H2) predicted that participants would be more accurate at identifying sarcastic compared with jocular scenes as insincere. This prediction was confirmed, as accuracy was higher for sarcasm compared with jocularity. These results are similar to those of other studies using the RISC database (Jakobson et al., 2018; Rothermich et al., 2019), static images (Baptista et al., 2018), written vignettes (Dews & Winner, 1999), and the visual world paradigm (Kowatch et al., 2013). Jocularity and sarcasm are associated with both positive and negative paralinguistic markers (e.g., laughter or eye-rolling), which signal a nonliteral interpretation. Although jocularity is considered to play a crucial role in everyday conversation (Gibbs, 2000), participants in the current study had a seemingly harder time identifying the cues used to express such intentions. This result is predicted by the *asymmetry of affect* which suggests that sarcasm is easier to interpret than jocularity (Clark &, Gerrig, 1984). The *asymmetry of affect* has influenced more complex frameworks of irony processing, such as the *Allusional Pretense Model* (*Kumon-Nakamura et al., 1995*). In this context, the processing of ironic forms of language such as sarcasm and jocularity depend on their allusion to a violation of social norms and expectations. As Matthews et al. (2006) point out, sarcasm can be considered as a less serious violation of politeness norms because it uses superficial positive language. Similarly, Pexman and Olineck (2002) state that sarcasm implicitly adheres to politeness norms by being positive on the lexical surface, which is not the case for jocularity; they propose that it is, therefore, easier to process sarcasm. As Mauchand et al. (2018) suggest, it appears to be more challenging for participants to form a social impression of a speaker when confronted with jocularity. Future studies will need to disentangle the circumstances under which subjects' ability to identify these forms of intentions may vary, especially in relation to individual differences (Pexman & Zvaigzne, 2004).

Based on the *Standard Pragmatic Model*, our third hypothesis (H3) predicted longer fixations and higher fixation proportions for nonliteral versus literal scenes, especially when participants focus on the responder in the videos. This hypothesis was supported by our data. Participants spent more time attending to the responder when the intention was nonliteral, as compared with literal intentions as revealed by a higher mean fixation proportion. It appears that the faces of the actors during jocularity videos received the most attention, as shown by the longer fixation duration and greater mean fixation proportions. This finding aligns well with previous eye tracking studies, demonstrating that sarcasm comprehension takes longer than literal language comprehension (e.g., Filik & Moxey, 2010; Kaakinen et al., 2014) and provides evidence for the *Standard Pragmatic Model*. However, we find a large variance in fixation durations and mean proportions of fixations, which suggests that individual differences and contextual constraints may influence our results. This variation supports the *Parallel-Constraint-Satisfaction Model* (Pexman, 2008) that suggests individual traits or demographic factors may impact the detection and processing of available cues. In addition, it is possible that since the actors in the videos used laughter to signal jocularity, participants are more likely to attend to these stimuli as laughter is a particularly salient cue (Pinheiro et al., 2017).

In order to study the time course of nonliteral language processing, we analyzed the timing of peak fixations during each of the four different intentions (H4). We found that literal positive fixation proportions peaked later than both jocularity and literal negative items when participants focus on the asker, which was contrary to our prediction. Our peak timing results suggest that participants shift their attention toward the asker who receives the specific literal or nonliteral comment. We speculate that participants use the nonverbal cues produced by the receiver of the critical comment to inform their decision on statement sincerity. Interestingly, this shift in attention appears to peak earlier for the lexically negative responses (literal negative, jocularity) compared with the positive responses (literal positive, sarcasm). This may be due to the fact that negative responses are more difficult to interpret, thus the locus of attention arrives earlier on the comment receiver in order to take full advantage of nonverbal information. Due to the highly exploratory nature of the peak timing analysis, additional studies will need to be conducted in order to verify the utility of this measure.

Conclusions

Eye tracking can be used to test and subsequently provide empirical support for the theoretical frameworks on the nature of how nonliteral language is processed. The findings in this report suggest that participants spend the majority of time looking at the talker who produces the nonliteral language and more time looking at the talker when processing jocular and sarcastic statements. Results such as these demonstrate the utility of using naturalistic and dynamic stimuli to examine nonliteral language during social conversations. Finally, it provides the foundation for subsequent analyses that can evaluate the nuances of visual attention during nonliteral language processing.

Résumé

La reconnaissance du sarcasme et de la jocularité durant des communications en personne requiert l'intégration d'indices verbaux, paralinguistiques et non verbaux. Pourtant, la plupart des recherches antérieures sur le traitement du langage non littéral ont été réalisées au moyen de stimuli écrits ou statiques. Dans la présente étude, nous avons examiné le traitement d'intentions dynamiques littérales et non littérales au moyen du monitorage oculaire. Après avoir regardé de courtes vignettes vidéo valides sur le plan écologique, on a demandé aux participants ($N = 37$) de déterminer l'intention des locuteurs. Les participants ont eu plus de difficulté à établir le caractère insincère des énoncés humoristiques que celui des énoncés sarcastiques, et ont regardé considérablement plus longtemps les visages durant les propos non littéraux que durant les échanges sociaux littéraux. Enfin, les participants ont pris plus de temps à passer d'un locuteur à l'autre durant les échanges aux intentions littérales positives, comparativement aux intentions humoristiques et aux intentions littérales négatives. Ces résultats appuient le modèle pragmatique standard ainsi que le modèle des processus parallèles de satisfaction des contraintes pour le traitement du langage non littéral.

Mots-clés : pragmatique, perception sociale, sarcasme, ironie, prosodie

References

Angeleri, R., & Airenti, G. (2014). The development of joke and irony understanding: A study with 3- to 6-year-old children. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *68*(2), 133–146. https://doi.org/10.1037/cep0000011

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, *32*, 793–826. https://doi.org/10.1016/S0378-2166(99)00070-3

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor: International Journal of Humor Research*, *16*(2), 243–260. https://doi.org/10.1515/humr.2003.012

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

Bates, D., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes*. R package version0.999999–2.

Caillies, S., Gobin, P., Obert, A., Terrien, S., Coutté, A., Iakimova, G., & Besche-Richard, C. (2019). Asymmetry of affect in verbal irony understanding: What about the N400 and P600 components?. *Journal of Neurolinguistics*, *51*, 268–277.

Caucci, G. M., & Kreuz, R. J. (2012). *Social and paralinguistic cues to sarcasm*. *25*(1), 1–22. https://doi.org/10.1515/humor-2012-0001

Clark, H. H., & Gerrig, R. J. (1984). *On the pretense theory of irony*.

Creusere, M. A. (2000). A developmental test of theoretical perspectives on the understanding of verbal irony: Children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbol*, *15*(1–2), 29–45. https://doi.org/10.1080/10926488.2000.9678863

Deliens, G., Antoniou, K., Clin, E., Ostashchenko, E., & Kissine, M. (2018). Context, facial expression and prosody in irony processing. *Journal of Memory and Language*, *99*, 35–48. https://doi.org/10.1016/j.jml.2017.10.001

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes*, *19*(3), 347–367. https://doi.org/10.1080/01638539509544922

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics*, *31*, 1579–1599. https://doi.org/10.1016/S0378-2166(99)00005-3

Fein, O., Yeari, M., & Giora, R. (2015). On the priority of salience-based interpretations: The case of sarcastic irony. *Intercultural Pragmatics*, *12*(1), 1–32. https://doi.org/10.1515/ip-2015-0001

Filik, R., Brightman, E., Gathercole, C., & Leuthold, H. (2017). The emotional impact of verbal irony: Eye-tracking evidence for a two-stage process. *Journal of Memory and Language*, *93*, 193–202. https://doi.org/10.1016/j.jml.2016.09.006

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.

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, *116*(3), 421–436. https://doi.org/10.1016/j.cognition.2010.06.005

Geisen, E., & Bergstrom, J. R. (2017). *Usability testing for survey research*. Morgan Kaufmann.

Gibbs, R. W. (1999). Interpreting What Speakers Say and Implicate, *Brain and language*, *68*(3), 466–485.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. [Includes Cognitive Linguistic Bibliography]. *Cognitive Linguistics*, *8*(3), 183–206. https://doi.org/10.1515/cogl.1997.8.3.183

Giora, R., Fein, O., Laadan, D., Wolfson, J., Zeituny, M., Kidron, R., Kaufman, R., & Shaham, R. (2007). Expecting irony: Context versus salience-based effects. *Metaphor and Symbol*, *22*(2), 119–146. https://doi.org/10.1080/10926480701235346

Grice, H. P. (1989). Logic and conversation. In *Studies in the way of words*. Harvard University Press.

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.

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

Haugh, M. (2016). "Just kidding": Teasing and claims to non-serious intent. *Journal of Pragmatics*, *95*, 120–136. https://doi.org/10.1016/j.pragma.2015.12.004

Hu, S. (2007). *Akaike information criterion*. Center for Research in Scientific Computation, 93.

Ivanko, S. L., Pexman, P. M., & Olineck, K. M. (2004). How sarcastic are you? Individual differences and verbal irony. *Journal of Language and Social Psychology*, *23*(3), 244–271. https://doi.org/10.1177/0261927X04266809

Jakobson, L. S., Pearson, P. M., Kozub, Z., Hare, C., & Rigby, S. N. (2018). Links between traits associated with the broad autism phenotype and empathy and young adults' ability to decode speaker intentionality. *Research in Autism Spectrum Disorders*, *50*(August 2017), 11–21. https://doi.org/10.1016/j.rasd.2018.03.001

Jorgensen, J. (1996). The functions of sarcastic irony in speech. *Journal of Pragmatics*, *26*(5), 613–634. https://doi.org/10.1016/0378-2166(95)00067-4

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes*, *51*(4), 287–311. https://doi.org/10.1080/0163853X.2013.870024

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. https://doi.org/10.1111/j.0963-7214.2004.00304.x

Katz, A. N., & Ferretti, T. R. (2001). Moment-by-moment reading of proverbs in literal and nonliteral contexts. *Metaphor and Symbol*, *16*(3–4), 193–221. https://doi.org/10.1080/10926488.2001.9678895

Kowatch, K., Whalen, J. M., & Pexman, P. M. (2013). Irony comprehension in action: A new test of processing for verbal irony. *Discourse Processes*, *50*(5), 301–315. . https://doi.org/10.1080/0163853X.2013.799934

Kreuz, R. J., Long, D. L., & Church, M. B. (1991). On being ironic: Pragmatic and mnemonic implications. *Metaphor and symbol*, *6*(3), 149–162.

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*(1), 3.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). Advance online publication. https://doi.org/10.18637/jss.v082.i13

Leggitt, J. S., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse Processes*, *29*(1), 1–24. https://doi.org/10.1207/S15326950dp2901_1

Matthews, J. K., Hancock, J. T., & Dunham, P. J. (2006). The roles of politeness and humor in the asymmetry of affect in verbal irony. *Discourse Processes*, *41*(1), 3–24.

Mauchand, M., Vergis, N., & Pell, M. D. (2018, June). *Ironic tones of voices* [Conference session]. 9th International Conference on Speech Prosody 2018 (pp. 443–447).

Milanowicz, A., Tarnowski, A., & Bokus, B. (2017). When sugar-coated words taste dry: The relationship between gender, anxiety, and response to irony. *Frontiers in Psychology*, *8*, Article 2215. https://doi.org/10.3389/fpsyg.2017.02215

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433–450. https://doi.org/10.1037/xlm0000176

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, *17*(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

Pexman, P. M., & Olineck, K. M. (2002). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Processes*, *33*(3), 199–217. https://doi.org/10.1207/S15326950DP3303_1

Pexman, P. M., Reggin, L., & Lee, K. (2019). Addressing the challenge of verbal irony: Getting serious about sarcasm training. *Languages 2019*, *4*(2), 23. https://doi.org/10.3390/languages4020023

Pexman, P. M., & Zvaigzne, M. T. (2004). Does irony go better with friends? *Metaphor and Symbol*, *19*(2), 143–163. https://doi.org/10.1207/s15327868ms1902

Pinheiro, A. P., Barros, C., Dias, M., & Kotz, S. A. (2017). Laughter catches attention! *Biological Psychology*, *130*, 11–21. https://doi.org/10.1016/j.biopsycho.2017.09.012

Pomareda, C., Simkute, A., & Phillips, L. H. (2019). Age-Related differences in the ability to decode intentions from non-literal language. *Acta psychologica*, *198*, 102865.

Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Sollberger, M., Wilson, S. M., Pavlic, D., Stanley, C. M., Glenn, S., Weiner, M. W., & Miller, B. L. (2009). Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage*, *47*(4), 2005–2015.

R Development Core Team. (2016). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. R Foundation for Statistical Computing.

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, technology & work*, *15*(3), 313–327.

Rothermich, K., Caivano, O., Knoll, L. J., & Talwar, V. (2019). Do they really mean it? Children's inference of speaker intentions and the role of age and gender. *Language and Speech*, *63*, 689–712. https://doi.org/10.1177/0023830919878742

Rothermich, K., & Pell, M. D. (2015). Introducing RISC: A new video inventory for testing social perception. *PLOS ONE*, *10*(7), Article e0133902. https://doi.org/10.1371/journal.pone.0133902

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://doi.org/10.2307/3002019

Seckman, M., & Couch, C. J. (1989). JOCULARITY, SARCASM, AND RELATIONSHIPS: An Empirical Study. *Journal of Contemporary Ethnography*, *18*(3), 327–344. https://doi.org/10.1177/089124189018003004

Shany-Ur, T., Poorzand, P., Grossman, S. N., Growdon, M. E., Jang, J. Y., Ketelle, R. S., Miller, B. L., & Rankin, K. P. (2012). Comprehension of insincere communication in neurodegenerative disease: Lies, sarcasm, and theory of mind. *Cortex*, *48*(10), 1329–1341. https://doi.org/10.1016/j.cortex.2011.08.003

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. *Philosophy*, *3*, 143–184.

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. (2013). What's behind a P600? Integration operations during irony processing. *PLOS ONE*, *8*(6), Article e66839. https://doi.org/10.1371/journal.pone.0066839

Toplak, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of Pragmatics*, *32*(10), 1467–1488. https://doi.org/10.1016/S0378-2166(99)00101-0

Ţurcan, A., & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1867–1893. https://doi.org/10.1037/xlm0000285

Valuch, C., Pflüger, L. S., Wallner, B., Laeng, B., & Ansorge, U. (2015). Using eye tracking to test for individual differences in attention to attractive faces. *Frontiers in Psychology*, *6*, 42. https://doi.org/10.3389/fpsyg.2015.00042

Received May 11, 2020
Revision received December 24, 2020
Accepted December 30, 2020 ■