

Exploring the role of conventionality in children's interpretation of ironic remarks*

DEBRA L. BURNETT

Kansas State University

(Received 13 January 2014 – Revised 25 July 2014 – Accepted 27 October 2014 –
First published online 11 December 2014)

ABSTRACT

Irony comprehension in seven- and eight-year-old children with typically developing language skills was explored under the framework of the graded salience hypothesis. Target ironic remarks, either conventional or novel/situation-specific, were presented following brief story contexts. Children's responses to comprehension questions were used to determine their understanding of the components of irony: speaker meaning, speaker attitude, and speaker intent. It was hypothesized that conventional remarks would be easier to comprehend than novel/situation-specific remarks because they are more likely to be familiar to the children. Results indicated that children demonstrated better comprehension of speaker meaning for conventional remarks than for novel/situation-specific remarks but no significant differences were found for inferring speaker attitude or speaker intent.

INTRODUCTION

To arrive at the speaker's meaning for any utterance encountered in verbal discourse, the listener must integrate information from multiple sources (e.g., word meanings, speaker's tone of voice, situational context). For ironic utterances, or those utterances when what is said indirectly conveys the speaker's intended message, listeners are required to integrate incongruent cues to infer meaning. In some studies, the terms 'irony' and 'sarcasm' have been used interchangeably (Attardo, Eisterhold, Hay, & Poggi, 2003;

[*] The author would like to thank Linda Milosky, Benita Blachman, Mary Louise Edwards, Janet Ford, and Annette Jenner-Matthews for their considerable support and mentoring in the completion of this project, Caroline Moore for her wonderful illustrations, the undergraduate and graduate students who worked in the language lab on various stages of the project, and the children who participated. Thanks also to Ann Bosma Smit and two anonymous reviewers for valuable feedback on an earlier version of this paper. Address for correspondence: Debra L. Burnett, Program in Communication Sciences and Disorders, School of Family Studies and Human Services, Kansas State University, 238 Campus Creek Complex, Manhattan, Kansas 66506. tel: (785) 532-0802; fax: (785) 532-6523; e-mail: deburnet@ksu.edu

Eisterhold, Attardo, & Boxer, 2006), but they are distinct. Sarcastic remarks are directed at another person with a cutting, bitter, or caustic delivery intended to convey a negative attitude indirectly (Gibbs, 1986b; McDonald, 2000). Ironic remarks may be used to criticize, but ironic intentions also include humor, understatements, circumlocutions, and rhetorical questions (Gibbs, 1986a, 1986b; Utsumi, 2004). Ironic criticisms directed at another person (i.e., remarks with both ironic and sarcastic intents) are the most widely used in everyday discourse (Dews *et al.*, 1996), and they were the focus of the current study exploring children's comprehension of irony.

The divergence between stated and intended meaning creates a comprehension challenge for the listener (Eisterhold *et al.*, 2006). A speaker may use paralinguistic, non-linguistic, and/or linguistic cues to signal that the listener should infer intent beyond the literal meaning of an utterance. Paralinguistic cues such as prosody (e.g., pitch, stress) occur as part of the utterance, while non-linguistic cues such as gestures or facial expressions (e.g., frowning while using a semantically positive utterance) occur simultaneously with the utterance. The role of prosodic cues has been more widely researched (Anolli, Ciceri, & Infantino, 2000, 2002; Attardo *et al.*, 2003; Bryant & Fox Tree, 2005; Dews *et al.*, 1996; Kreuz & Roberts, 1995; Milosky & Ford, 1997; Rockwell, 2000) than non-linguistic cues such as facial expression (Attardo *et al.*, 2003; Utsumi, 2000) or gesture (Pexman, 2005; Utsumi, 2000). However, when these types of cues occur in ironic discourse they appear to be somewhat idiosyncratic and difficult to generalize across speakers. Linguistic cues, such as the wording of the remark, may signal irony in a more consistent way. Furthermore, a remark may be so widely used that the ironic meaning is considered conventional (e.g., *Big deal.*) and thus as familiar to the listener as the literal interpretation (Giora & Fein, 1999a; Giora, Fein, & Schwartz, 1998). Alternatively, the ironic meaning of an utterance may be a novel one if it is unique to a situation (e.g., looking at a lopsided cake, one says, *That's a lovely cake.*). Therefore, the current study examined how the wording of the remark related to comprehension in children, given literal and ironic utterances.

While research to date has explored topics related to children's irony comprehension skills such as the integration of conflicting cues (Bugental, Kaswan, & Love, 1970) and the ability to make inferences about intent (Bishop & Adams, 1992), the role of word meanings (i.e., how familiar the usage might be) is an area requiring further study. Inferring meaning is of particular interest in the development of the ability to recognize and interpret ironic remarks. This skill appears to follow a distinct and somewhat protracted course when compared to literal language comprehension (Hancock, Dunham, & Purdy, 2000).

Some would argue that irony comprehension is no more difficult than literal language comprehension (Gibbs, 1986; Jorgensen, Miller, & Sperber, 1984), but there is evidence to suggest otherwise. Three types of processing have been proposed to explain how listeners recognize and understand irony: multi-step, direct access, and graded salience. The multi-step approach is engaged when non-literal utterances are encountered (Grice, 1975; Searle, 1979). The lexical meaning is accessed and then that meaning is compared to the context and found to be inappropriate. Following that process, non-literal meanings are accessed and the initial literal interpretation is discarded in favor of the context-appropriate non-literal or ironic interpretation. Such a multi-step approach is activated each time a listener encounters a non-literal use of an utterance (Clark & Lucy, 1975; Honeck, Welge, & Temple, 1998).

In contrast, several theories fall under a direct access view of irony comprehension: echoic mention theory (Sperber, 1984; Sperber & Wilson, 1981), pretense/allusional pretense theory (Clark & Gerrig, 1984; Kumon-Nakamura, Glucksberg, & Brown, 1995), and implicit display theory (Utsumi, 2000). Within this view, the comprehension of irony is made possible by several contextual cues that allow the listener to infer ironic intent without noticing the literal meaning of the message. Unlike multi-step processing, context drives meaning selection from the beginning so that only the appropriate interpretation is activated and an inappropriate meaning is not accessed at all. Therefore, a speaker's intended meaning is directly accessed by a listener without first processing all possible semantic meanings of the utterance.

A third approach to irony processing, the graded salience hypothesis, has more recently posited that the most salient meaning of an utterance will be accessed first regardless of whether the utterance is literal or non-literal. The semantic meaning of an utterance is coded in the lexicon (i.e., one's mental dictionary, including word definitions and associations) based on its frequency and familiarity. 'Salience' is the term applied to how strong a meaning representation is in the lexicon, with the most salient meanings being the most coded or prominently stored and therefore the most easily retrieved (Giora, Balaban, Fein, & Alkabets, 2005). Once this initial salient meaning is activated, it is then compared to the context, and the process ends there if that meaning is deemed appropriate. The appropriate meaning is accessed directly but, unlike in the direct access view, the driving force is meaning salience of the utterance and not context. In cases when the initially activated meaning is not appropriate given the context, then additional meanings are activated until the contextually appropriate meaning is found. When this occurs, meaning selection follows a process similar to the multi-step approach.

In some instances of figurative language (e.g., idioms, conventional ironies), the non-literal meaning has been lexicalized and is as salient as

the literal meaning and so is accessed immediately. Since most ironic utterances do not have fixed meanings, the literal meaning is often the most salient given the lexical content (Giora & Fein, 1999a, 1999b). However, there are a number of ironic utterances that are used with enough frequency and familiarity to have become conventional and as salient as their literal counterparts (e.g., *That's just great.*). Giora and colleagues (Giora & Fein, 1999a; Giora, Fein, & Schwartz, 1998) have demonstrated, using lexical decision tasks, that meaning activation for conventional irony is similar to literal statements.

In each type of processing outlined in the above theoretical models, the listener must ultimately derive the meaning of what was said as it relates to the situation or context created by prior utterances. There is evidence about the emergence of these skills around six years of age from a variety of studies (Andrews, Rosenblatt, Malkus, Gardner, & Winner, 1986; Dews & Winner, 1997; Filippova & Astington, 2008, 2010; Pexman & Glenwright, 2007; Winner & Leekam, 1991). However, the research remains inconclusive about what skills develop first: the ability to detect a discrepancy between the semantic (literal) meaning and context or the ability to infer speaker intent. When presented with stories, younger children who have not yet begun to make ironic interpretations often mistake ironic utterances for true statements by interpreting the remark literally (Ackerman, 1981; Hancock *et al.*, 2000; Winner & Leekam, 1991). In addition to a literal bias, Filippova and Astington (2010) found that the form of the ironic utterance may affect how easily children can detect the discrepancy (e.g., counterfactual statements are easier than hyperbole). Hancock *et al.* (2000) also found that the function affects the ease with which children recognized irony (e.g., ironic criticisms were easier than ironic compliments).

Speaker attitude is another important component for discrepancy detection since the attitude a speaker is conveying is incongruent with the semantic meaning of the statement. Attitudes reflect the emotion a speaker has in response to the situation. Similar to recognizing the semantic discrepancy between the words and the situation, speaker attitude can be inferred by six years of age (Andrews *et al.*, 1986). However, this ability did not seem to assist children in correctly inferring intent. Children who could recognize speaker attitude by responding to questions about different possible attitudes (e.g., *Was Karen disappointed with the children? Was Karen upset with the children? Was Karen happy with the children?*) did so without necessarily being able to infer ironic intent correctly (Andrews *et al.*, 1986).

The age at which children begin to infer intent differs depending on the tasks used in each study, which range in difficulty from answering yes/no questions to paraphrasing. In addition to task demands, researchers have

imputed different skills to the children from the same or similar tasks. Hancock *et al.* (2000) used yes/no questions to examine five- and six-year-olds' ability to understand belief (e.g., *Did B really think that A was a good basketball player?*) and forced-choice questions to examine inference of speaker intent given ironic criticisms and ironic compliments (e.g., *Was B being mean or nice?*). However, the modality for assessing speaker intent allowed the child to respond verbally or non-verbally (i.e., point to either a happy face or a sad face). It would seem that what was called intent could have been confused with speaker attitude, particularly since Pexman and Glenwright (2007) employed similar facial expressions to address speaker attitude.

Capelli, Nakagawa, and Madden (1990) questioned children about the speaker's intent in using a target remark, either ironic or literal, following a story context by asking an open-ended question (e.g., *Why did Laura say that?*). Examples given for responses describing speaker intent that were coded as sarcastic included labeling an emotion (e.g., *Laura was angry.*), stating an opposite meaning (e.g., *What a jerk.*), or describing speaker intent (e.g., *to be mean, to be sarcastic*). In a similar fashion, Keenan and Quigley (1999) asked six-, eight-, and ten-year-old children to paraphrase the meaning of a sarcastic statement (e.g., *When Lucy said, 'Oh great, now I'll really look pretty', what do you think she meant?*). Correct responses included describing either meaning or intent (e.g., *stating the speaker meant the opposite or the speaker was teasing*) so it is difficult to parse out the development of each skill from their findings.

Examining speaker meaning, speaker attitude, and speaker intent as separate skills was not the focus of these prior studies *per se*; however, by aggregating these findings and employing definitions of speaker meaning, speaker attitude, and speaker intent as they have been described herein, children have some understanding of the levels of speaker intent by nine years of age (Ackerman, 1981). Given what is known about development of irony comprehension, the current study examined these skills in seven- and eight-year-olds. Within this age group, children have had some exposure and success in detecting discrepancy between stated and intended meaning but have not mastered that skill. Rather, their ability to infer pragmatic intent for ironic remarks is in its earliest developmental stages.

One aspect of the development of irony comprehension that has yet to be examined is how the lexicalization of conventional instances of irony affects the ability to infer a speaker's meaning, attitude, and intent. The prior work on the development of irony comprehension has typically employed situation-specific ironic remarks (e.g., *I see you won again.* in Ackerman, 1981; *You always bake great cookies.* in Dews *et al.*, 1996; *You really are good at lifting weights.* in Hancock *et al.*, 2000; *That was a great play.* in Pexman & Glenwright, 2007). Many of the conclusions about how

children understand ironic remarks were based on those types of remarks, which may be more difficult than conventional ironies because ironic meanings are less likely to be lexicalized. Ackerman (1983) implied that certain ironic remarks may have become idiomatic, since children interpreted them ironically with limited support from other cues, but he did not examine this empirically. Given the theoretical framework of the graded salience hypothesis, both the literal and figurative meanings can become coded as entire phrases in the lexicon based on how conventional the meanings are (e.g., *Thanks a lot.* is conventionally used with both sincere and ironic intent). Therefore, the figurative meanings of conventionally used ironic utterances may be more readily available in the lexicon and thus easier to comprehend than situation-specific ironic utterances. The latter may require additional cues or processing to be interpreted correctly.

Given that conventional remarks have not been the focus, or perhaps even present, in these prior studies examining irony comprehension in children, and that lexicalization of conventional ironic meanings may ease comprehension, the current study examined comprehension for both conventional and situation-specific/novel remarks. Further exploration of children's understanding of conventional remarks would allow researchers to better understand if the wording of the remark itself eases the comprehension challenge inherent in the use of irony. Using the theoretical framework provided by the graded salience hypothesis, the component skills involved in irony comprehension (i.e., meaning, attitude, and intent) were explored with young children, whose skills are emerging, in relation to conventional remarks and situation-specific remarks. The graded salience hypothesis predicts that the non-literal meanings of conventional ironic remarks are more readily available in the lexicon than the non-literal meanings of novel/situation-specific ironic remarks. In the current study, children were presented with conventional and novel/situation-specific remarks and asked questions about speaker meaning, speaker attitude, and speaker intent. If the use of conventional remarks eases the comprehension challenge created by irony, children's responses to those questions should be more accurate than when the remarks are novel/situation-specific.

METHOD

Participants

Thirty English-speaking seven- and eight-year-old children ($M = 8;2$, $SD = 6.95$ months) with typically developing cognitive and linguistic skills participated in this study. Children aged seven and eight years were chosen for the study based on prior research that demonstrated that they

were likely to have emerging skills for inferring speaker meaning and intent for ironic remarks (Ackerman, 1981; Dews *et al.*, 1996; Keenan & Quigley, 1999; Pexman, Glenwright, Hala, Kowbel, & Jungen, 2006) and to have encountered a variety of ironic remarks. Participants were recruited through the use of flyers posted in the community, advertisements in local newspapers, and internal university news announcements. There were ten boys with an average age of 8;2 ($SD = 0.4$ years) and twenty girls with an average age of 7;8 ($SD = 0.6$ years). The socioeconomic status (SES) of the participants, measured using the Hollingshead scale (Hollingshead, 1975), averaged Upper Middle Class ($M = 57.15$, $SD = 10.84$, range = 22–73.5). Eligibility requirements for inclusion in the study were as follows: typical hearing (as measured by a hearing screening), age-appropriate non-verbal intelligence and language ability (as measured by standardized assessments), and no parental report of a history of language impairment or neurological disorders.

Materials and design

Language and cognitive measures. The hearing screening followed the American Speech-Language-Hearing Association guidelines for screening children: 20 dB at 500, 1000, 2000, 4000, and 6000 Hz (ASHA, 1997). Hearing screenings were performed in either a double-walled sound-treated room using a GSI-16 audiometer (Grason-Stadler; Madison, WI) or a portable audiometer (MAICO MA 27) in a quiet location, and behavioral responses were required (i.e., the child had to raise a hand when the tone was presented). The Special Nonverbal Composite of the Differential Ability Scales-Second Edition (DAS-II; Elliott, 2007) was used to assess non-verbal cognitive ability; standard scores above 85 indicate typical ability. Language skills were assessed using the EpiSLI diagnostic battery established by Tomblin, Records, and Zhang (1996). The EpiSLI battery creates composite scores based on five subtests of the Test of Language Development-Primary, Third Edition (TOLD-P:3; Newcomer & Hammill, 1997) and narrative tasks from Culatta, Page, and Ellis (1983): vocabulary, grammar, narrative, comprehension, and expression. Typical language ability is indicated by performance within 1.25 Standard Deviations of the mean on all composites. The Competing Language Processing Task (CLPT; Gaulin & Campbell, 1994) was administered to obtain a measure of verbal working memory capacity in typically developing children; its results are not relevant to testing the graded salience hypothesis and are not included in the current study.

Experimental task materials. The experimental task consisted of story contexts involving a gender-neutral child, Pat, that were either experimenter-generated or

TABLE 1. *Example story used in the experimental task – Spill context and target remarks*

Pat asks a friend to carry a vase of water to the table.
Pat's friend is not being very careful and slips.
The vase breaks and water spills everywhere. (<i>Negative outcome</i>)
Pat says,
'Smart move.' (<i>Conventionally ironic remark</i>)
Or:
'Great save.' (<i>Novel/situation-specific ironic remark</i>)
Pat asks a friend to carry a vase of water to the table.
Pat's friend is not being very careful and slips.
Pat's friend makes a great catch before the vase falls. (<i>Positive outcome</i>)
Pat says,
'Smart move.' (<i>Conventionally ironic remark used literally</i>)
Or:
'Great save.' (<i>Novel/situation-specific literal remark</i>)

adapted from prior literature (Pexman & Glenwright, 2007). Each context depicted an event that was likely to be familiar to children, had equally plausible positive and negative outcomes, and met the rules for ironic environment: an event, an expectation for that event, and an outcome that was either congruent or incongruent (Utsumi, 2000). The story contexts included a variety of relationships to Pat (e.g., sister, mother), as prior research has indicated that a child's ability to infer mental states may be affected by the speaker's role in the family (Massaro, Valle, & Marchetti, 2013).

Each of the story contexts was adapted into positive and negative versions by altering the outcome in the third sentence (see Table 1 for example). An iterative process was used to develop story contexts that would be perceived as clearly negative or positive. A total of 186 adult volunteers rated the stories on a 5-point scale (1 = very negative to 5 = very positive); however, each rater saw one version of the story context (e.g., negative outcome for the Spill story). A finalized set of story contexts, in which all versions of each context was present, was then rated for situational negativity in a manner similar to procedures used by Ivanko and Pexman (2003). Twenty-two adult participants rated the degree of negativity for each version of the story contexts. The ratings of the final set of story contexts were significantly different between the negative and positive versions. Since some situations may be inherently more negative than others, in addition to the ratings for positive and negative versions within one story, the ratings were also compared across contexts so that there were no large disparities among the different stories (e.g., the negative versions of the Spill story and the Softball story were perceived as equivalently negative).

In order to test the graded salience hypothesis, the target remarks that followed each story context were either conventional (higher salience) or

novel/situation-specific (lower salience). A list of potential remarks was gathered from prior literature, young adult fiction, television shows, Internet searches, and observations of children. The list was rated by 264 university students using a 5-point rating scale (1 = extremely unlikely to be said sarcastically to 5 = extremely likely to be said sarcastically). The term 'sarcastically' was used rather than 'ironically' as it was more likely to be understood by the raters. In addition, the current study was exploring verbal irony and thus the raters were instructed to think about phrases used verbally, rather than terms such as 'used sarcastically' or 'encountered' which may have invited raters to think about other contexts such as written discourse. All students were seated in a large lecture hall and given a response device (i.e., clicker) with 5 keys; one key for each point on the rating scale. Verbal instructions for how to complete the task were provided and then the remarks were presented one at a time for 5 seconds each on a large screen. Similar to the procedure used by Giora and Fein (1999b), a mean rating was computed for each remark and the following criterion was used: phrases rated as 1.0–3.4 were characterized as novel/situation-specific and phrases rated as 3.5–5.0 were characterized as conventionally ironic.

Following this process, a conventional or novel/situation-specific target remark that met the criterion was added to each context to create the final set of stimulus stories. As a final phase in the development of the experimental stories, plausibility ratings were gathered to determine that the contexts and remarks made sense once paired together. In a procedure similar to Ivanko and Pexman (2003), adults were asked to rate a set of stories on a 5-point scale related to the target remark and the context (1 = does not fit at all to 5 = fits very well). Four additional non-plausible filler stories were added to the potential experimental stories rated by each adult; each adult rated one version of an experimental story (e.g., Spill story with negative outcome and conventionally ironic remark). A total of seventy-two adults completed the plausibility ratings, and stories with average ratings of 3.0 or higher were included in the experimental task.

So that each child would hear one version of each story context (e.g., one version of the Spill story), rather than all possible stimulus stories, this set was divided into balanced lists; each list contained four negative contexts with conventionally ironic remarks, four negative contexts with novel/situation-specific ironic remarks, and eight filler stories (i.e., positive contexts with ironic remarks, negative contexts with literal criticisms, and neutral contexts and remarks). Two practice stories were also created to familiarize the child with the procedure.

Two original black and white illustrations were drawn to accompany each story and characterize the main events of the story. Prior literature examining irony comprehension in children has often used visual assistance such as simple illustrations (Filippova & Astington, 2008, 2010) or a

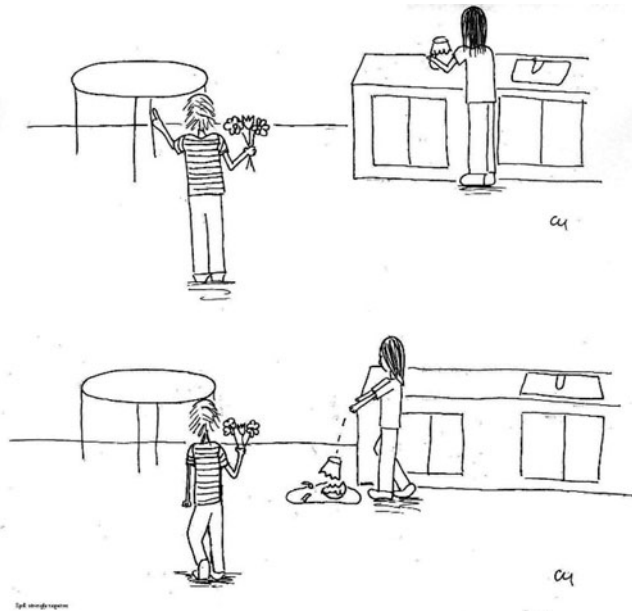


Fig. 1. Sample illustrations for negative Spill story.

sequence of several illustrations (Keenan & Quigley, 1999). The illustrations in the current study contained relevant aspects of the story event (e.g., cookies being placed in the oven; a tray of burned cookies coming out of the oven). While illustrations used in some prior studies have had several elements (e.g., Winner & Leekam's (1991) study used illustrations that depicted an entire house with four rooms), the current study used simple illustrations. The situation in which the remark occurred was depicted, but facial expressions were not visible; only the character's back or a slight profile could be seen. While some studies have used facial expression as an additional cue for irony (Winner & Leekam, 1991), it remains a relatively idiosyncratic feature and was not the focus of the current study. Others have used a neutral facial expression by using a drawing with a straight horizontal line to represent the mouth (Filippova & Astington, 2010), but it is unclear if the perception of the expression is indeed a neutral one. Therefore, making facial information unavailable removed a potential confound in the interpretation of children's understanding of irony. Illustrations were validated by a set of five adult raters, four females and one male, who described the event as it was depicted in the set of two illustrations and provided any relevant feedback. Illustrations were then modified until they appropriately depicted the story context (see Figure 1 for example).

TABLE 2. *Comprehension questions*

Fact question: <i>Can you tell me what happened in the story?</i>
Speaker meaning: <i>What did Pat mean by [target remark]?</i>
Speaker attitude: <i>How did Pat feel when [event outcome]?</i>
Speaker intent open-ended: <i>Why did Pat say, [target remark]?</i>
Speaker intent forced-choice: <i>Did Pat want to make Pat's friend feel [bad/good]?</i>

A series of questions was created for each story context and target remark to assess the child's comprehension of speaker meaning, speaker attitude, and speaker intent (see Table 2). Open-ended questions were used rather than multiple-choice or yes/no formats to allow for the child's interpretation of the event and inference of the speaker meaning and intent (Capelli *et al.*, 1990; Demorest, Silberstein, Gardner, & Winner, 1983; Keenan & Quigley, 1999). However, during piloting it was noted that children tended to repeat the answer from the speaker meaning question for the speaker intent question. Therefore, a forced-choice intent question with a yes/no possible response was developed to address speaker intent. In a procedure similar to Capelli *et al.* (1990), the open-ended question was asked first and the forced-choice question was asked in situations where the child's initial answer was incorrect, vague (i.e., could not be scored online), a repetition of the speaker meaning response, or did not clearly address intent. For the eighteen stories following the two practice stories, up to nine forced-choice questions with correct 'yes' answers and nine forced-choice questions with correct 'no' answers (random ordering of correct yes/no responses was used) could have been presented.

The stories, remarks, and questions were audio-recorded for presentation to the children by a female speaker using Praat software (Boersma & Weenik, 2009). The story context was recorded one sentence at a time and then spliced together to create each trial. The target utterances were audio-recorded using prosody consistent with ironic or literal remarks (Hancock *et al.*, 2000). The set of twenty-four target utterances was validated as containing literal or ironic prosody using an objective method and a subjective method. Recordings of target utterances were measured for fundamental frequency (Fo) and duration, and they were used if Fo and duration were consistent with the intended literal or ironic version (Rockwell, 2000). That is, ironic remarks were recorded with lower mean Fo than literal remarks (187.09 and 252.27 Hz, respectively). Ironic remarks were also longer in duration than their literal counterparts (average duration was 1.66 and 1.17 seconds, respectively). For subjective ratings, each utterance was low-pass filtered at 500 Hz to remove semantic information but retain the patterns of intonation. Five adults listened to the set of forty-eight utterances (each of the 24 remarks was presented with sincere

prosody and with ironic prosody), presented in random order, and rated if they thought the speaker sounded sincere or sarcastic using a 3-point scale (1 = sincere, 2 = not sure, and 3 = sarcastic). Sincere utterances averaged 1–1.8 and ironic utterances averaged 2.2–3. Any recording that failed to meet criterion was re-recorded until it did so and then re-rated. Following this procedure, story contexts and target utterances were digitally spliced together for the final set of stimulus recordings.

Procedure

Children participated in five tasks, administered individually, during one approximately 2-hour session at the language development laboratory (two children were seen at home). Parents observed the testing session through a one-way mirror from an observation room (or from an adjacent room if at home). Children were given breaks as needed to minimize fatigue and to maintain attention to tasks. Children had to pass the hearing screening in both the left and right ears to continue participating in the study. The EpiSLI battery, consisting of the TOLD-P:3 subtests and the Culatta narrative tasks, and the SNC subtests were administered following the hearing screening, with the order counterbalanced across participants. The last two tasks, also with their order counterbalanced, were the CLPT and the experimental task.

For the experimental task, children were seated at a table, and prior to audio presentation of each story, both of the accompanying illustrations were presented and remained visible until the child had completed answering the questions following the story, a procedure similar to that used in the Test of Narrative Language (TNL, Gillam & Pearson, 2004). Two practice stories were presented first to familiarize the child with the procedure and to be sure the questions were understood. The experimental stories and filler stories were then presented to each participant in random order. There was no repetition of story contexts within participants (e.g., each child heard one version of the ‘Spill’ story). After the story, the examiner played the audio-recorded questions to assess factual comprehension of the events of the story and the child’s comprehension of speaker meaning, speaker attitude, and speaker intent. A fixed order was used as prior research has predominantly used a fixed order (Capelli *et al.*, 1990; Hancock *et al.*, 2000; Keenan & Quigley, 1999; Pexman *et al.*, 2006) rather than a varied order (Ackerman, 1981, 1983; Dews *et al.*, 1996) for speaker meaning and intent questions. Testing sessions were video- and audio-recorded. All responses were transcribed verbatim from recordings by undergraduate students in Communication Sciences and Disorders familiar with transcription procedures but unfamiliar with the conditions of the experiment.

Scoring and reliability

Responses to the experimental questions were scored from the transcript and the audio-recorded sessions. For each comprehension question, the child received a score of 0 (incorrect response) or 1 (correct response) as follows. For the fact question, a score of 1 was given when the child provided a retell or paraphrase of the main aspects of the event; inclusion of the final remark was not required. In order to receive correct scores on the meaning, attitude, and intent questions, a child had to answer the fact question correctly. For the speaker meaning question, the response was correct when the child provided the literal meaning for literal remarks or an opposite meaning for ironic remarks. For the speaker attitude question, a correct response reflected the appropriate valence (i.e., positive for positive outcomes and negative for negative outcomes). For speaker intent, the child had to either explain the literal or ironic intent of the speaker or correctly respond to a yes/no question about speaker intent. Participants were given a score of 1 (indicating they understood irony) when they answered all four questions correctly (i.e., fact, meaning, attitude, and intent). Given that there were four stories per condition, each participant could receive a maximum score of 4 for any question. For the experimental task, an independent second coder, blind to the research questions, was trained in the experimental scoring procedure and scored 20% of the participants' responses from each question category. All reliability measures were within acceptable ranges: fact question (0.97), speaker meaning (0.90), speaker attitude (1.0), speaker intent open-ended (0.87), and speaker intent forced-choice (0.94).

To obtain reliability measures for the EpiSLI battery (TOLD:P3 and Culatta tasks), 20% of participants were scored from video-recorded sessions by a second scorer, blind to the research questions but trained in standardized testing administration. For subtests of the TOLD:P3, the following inter-rater reliability was obtained: oral vocabulary (0.95), picture vocabulary (1.0), sentence imitation (0.97), grammar completion (0.98), and grammatic understanding (0.98). For the Culatta tasks, reliability was obtained for story retell (0.95) and comprehension (1.0). Inter-rater reliability was obtained for subtests of the SNC of the DAS-II using test protocols and video-recordings of the sessions: pattern construction (0.81), recall of designs (0.87), matrices (0.99), and sequential and quantitative reasoning (0.99).

RESULTS

All children correctly answered the fact question, which indicated that they recalled the main event depicted in the story. No significant differences ($\alpha = .05$) were found based on gender or age (i.e., seven-year-olds and

TABLE 3. Mean response scores for irony questions ($n = 30$)

Question type		Remark type	
		Conventional	Novel
Speaker meaning*	<i>M</i>	3.60	3.07
	<i>SD</i>	0.62	1.08
Speaker attitude	<i>M</i>	3.90	3.67
	<i>SD</i>	0.31	0.71
Speaker intent	<i>M</i>	2.57	2.40
	<i>SD</i>	1.48	1.54
Irony comprehension score	<i>M</i>	2.27	1.97
	<i>SD</i>	1.40	1.40

NOTE: * indicates $p < .05$ for conventional vs. novel remarks.

eight-year-olds) for inference of speaker meaning, speaker attitude, or speaker intent, so results of all thirty children were combined for further analyses. Children were each assigned to one of six lists of experimental stories; an analysis of variance (ANOVA) showed no significant main effect for list ($\alpha = .05$) so all results were combined.

Results for speaker meaning, attitude, and intent are shown in Table 3; possible scores for each ranged from 0 to 4. Paired two-tailed t -tests ($\alpha = .05$) were used to compare responses for conventional and novel/situation-specific remarks. Conventional remarks were easier for children to understand than novel/situation-specific remarks for the inference of speaker meaning ($t_{(29)} = 2.44$, $p = .0209$, Cohen's $d = 0.602$). No significant differences were found between remark type for speaker attitude ($t_{(29)} = 1.65$, $p = .1094$), speaker intent ($t_{(29)} = 0.82$, $p = .4203$), or overall irony comprehension ($t_{(29)} = 1.61$, $p = .1188$).

The responses for speaker attitude had a ceiling effect which limited the variability among responses. For the positive story contexts, children's responses were typically one of two emotion labels: happy or glad. Eleven of the children used the label 'good' to label Pat's emotion (a total of 27 occurrences). For negative story contexts, labels were more varied: 'sad', 'mad', 'scared', 'angry', 'disappointed', 'worried', 'upset', 'frustrated', or 'lonely'. Five of the children used the word 'bad' to describe Pat's emotion for negative situations (a total of 21 occurrences). However, removing the speaker affect question from the overall irony score calculation, so that it was based on a child's response to speaker meaning and speaker intent, did not have an effect on the statistical significance.

Since the intent question was potentially answered by either an open-ended or a forced-choice (yes/no) response, further exploration of the use of each question type was warranted. Of the 360 trials (i.e., 12 trials for each of 30 participants, including the positive and negative contexts), the forced-choice intent question was employed 296 times. The number of

yes/no forced-choice intent questions asked varied by child (depending upon responses to the open-ended questions), so the percentage of correct responses to yes/no questions was calculated for each child. A one-sample t -test ($H_0: \mu = 0.5$ for yes/no questions) indicated that the percentage of correct responses was significantly different from chance ($t_{(29)} = 7.041$, $p < .0001$).

Individual stories were examined to determine if the situation depicted affected the children's ability to infer meaning, attitude, or intent. For example, would a story about getting a ride to school differ in the ease of comprehension from a story about doing math homework? An ANOVA revealed no significant differences in performance ($\alpha = .05$) depending on the story context for meaning ($F_{(11,24)} = 1.16$, $p = .36$), attitude ($F_{(11,24)} = 1.09$, $p = .41$), intent ($F_{(11,24)} = 0.14$, $p = .99$), or overall irony ($F_{(11,24)} = 0.38$, $p = .95$).

Even though the positive contexts were intended for use as filler stories and not for addressing the core research questions in the current study, some children incorrectly answered the speaker meaning question when a conventionally ironic remark followed a positive (literally biasing) context, despite having recalled the story accurately. In these positive contexts, the remarks are literal remarks. However, the conventional ironic meaning may have affected children's interpretation of those remarks. For example, the remark *Very funny* that followed the positive context in which Pat's classmate shares a funny toy was still interpreted ironically even though it was intended literally (i.e., child responded, *That's not funny.*). Since each child had four trials of conventional ironies used in negative situations and two trials of conventional ironies used in positive situations, a proportion of correct responses was calculated for each type (89% and 88%, respectively). For novel/situation-specific ironic remarks, the proportion of correct responses for speaker meaning in negative situations and positive situations was calculated (78% and 90%, respectively). Given this finding, that children were performing lower than expected for these positive contexts, a 2 (Remark type: Conventional irony and situation-specific/novel irony) \times 2 (Intent: Ironic or Literal) ANOVA was used to explore children's responses for speaker meaning. Since there were eight negative contexts and four positive contexts, a subset of four negative contexts, those rated most strongly during the development of stimulus stories, were used for this analysis. Thus, each child could earn a score between 0 and 4 for remark type or intent. Descriptive statistics were calculated as follows for inferring speaker meaning: remark type was conventional ($M = 3.53$, $SD = 0.63$) or novel/situation-specific ($M = 3.43$, $SD = 0.73$), and intent was ironic ($M = 3.40$, $SD = 0.77$) or literal ($M = 3.57$, $SD = 0.57$). The exploratory ANOVA ($\alpha = .05$) revealed no significant interaction between remark type and intent ($F_{(3,116)} = 0.79$, $p = .038$), no main effect of intent

($F_{(3,116)} = 0.79$, $p = 0.38$), and no main effect of remark type ($F_{(3,116)} = 0.29$, $p = 0.59$). These results demonstrate that when positive story contexts and literal remarks were compared with negative story contexts and ironic remarks, the children did not differ in their ability to infer speaker meaning based on whether the remark was conventional or novel/situation-specific nor whether the intent of the speaker was ironic or literal. However, given the small number of trials for each condition, the statistical power of the analysis was limited.

DISCUSSION

The current study examined the graded salience hypothesis for irony comprehension in typically developing children by utilizing conventional and novel/situation-specific ironic remarks. According to the graded salience hypothesis, the non-literal meanings of conventional remarks are stored in the lexicon with enough salience to be accessed directly; thus, it was hypothesized that conventional remarks would be easier to understand than novel/situation-specific remarks, as measured by children's responses to comprehension questions. The hypothesis was supported because children were better able to infer the speaker's intended meaning for phrases that were conventionally ironic than for those that were novel/situation-specific (e.g., *What did Pat mean by 'That's just great?'*). The current study provided further evidence that when verbal discourse was encountered, literal or non-literal, it was meaning salience that drove comprehension. According to the graded salience hypothesis, salient meanings do not have to be literal; salient meanings can also be non-literal ones such as is the case with conventional ironies (e.g., *That's just perfect.*). This suggests a role for experience and familiarity and that these lexicalized meanings may function idiomatically. Future studies exploring irony comprehension in children should characterize the type of remark used because conventional remarks appear to be easier for children to infer an opposite meaning.

While the number of correct responses for speaker meaning differed between the conventional and novel/situation-specific remarks, performance on speaker attitude and intent questions did not vary by remark type. For speaker attitude, one possible reason that no difference was observed was that scores demonstrated a ceiling effect, since most children answered that question correctly. The high proportion of correct responses may have had several causes. First, the story contexts were events designed to be familiar to the children (e.g., getting a new toy, being dropped off at school) and so the children may have been easily able to draw on their own emotions if put in the situation (i.e., activating relevant background knowledge). For example, empathizing with Pat, a child would also feel mad if his or her

sibling broke a new toy, or perhaps the child recalled a personal event of a broken toy and a sibling. Children did not necessarily have to rely solely on the information from the context or the intonation in Pat's final remark to infer the correct emotion/attitude. The use of simple story contexts, both in terms of the linguistic demands and the situation depicted, may have allowed for easier processing of the speaker's attitude in relation to the remark.

Second, unlike some of the prior studies, where participants read stories silently, these children heard the intonation that Pat used in the final remark (as demonstrated by the female speaker narrating the story). The children may have relied on the prosodic cues to infer speaker attitude over the story outcome. Prosodic cues were consistent with Pat's attitude and intended meaning (i.e., sarcastic intonation was used for sarcastic remarks and literal intonation used for literal remarks). In studies examining irony comprehension, a precedent exists for the use of prosody that is consistent with the type of remark (ironic or sincere; Capelli *et al.*, 1990; Keenan & Quigley, 1999; Winner & Leekam, 1991). Since no comparison condition was used (i.e., without prosody), it is not possible to parse out that possibility for the current study.

Third, the visual cues provided from using illustrations could have eased the level of difficulty of the task. The picture that was presented may have helped the child to think about how s/he would feel in that situation. It served to reinforce the story context presented in the audio-recording and help keep it present in the child's mind while answering questions. Providing a visual representation of the context may have made it easier for children to evaluate the event, rather than relying solely on a verbal mental representation. In turn, it would have been easier to think about how Pat might feel in that situation, and thus the type of attitude the remark conveyed. However, the use of illustrations is often a part of studies such as this one (Filippova & Astington, 2010; Keenan & Quigley, 1999; Winner & Leekam, 1991) and other tasks involving children's narrative (e.g., TNL). They may serve a necessary role within the task to help keep the child's attention while talking about the story and characters' remarks.

Fourth, the use of the forced-choice follow-up question related to speaker intent (since it was used in the two practice stories) may have influenced the children's responses to the speaker attitude question. By giving a choice (i.e., *Did Pat want to make x feel good/bad?*), some of the children may have gone on to use those terms in describing Pat's emotions in later trials. However, since fewer than half of the children used the terms 'good' or 'bad', and most of the children used a variety of labels, it does not appear that the ceiling effect was due to the use of the terms 'good' and 'bad' in the forced-choice intent question.

For speaker intent, the scores were obtained by allowing the child to answer either an open-ended question or a forced-choice (yes/no) question. If the open-ended question alone had been used, a floor effect would have likely resulted since only 75 out of 360 trials were answered correctly (21%). Adding the forced-choice question allowed children to increase their accuracy. This finding differs from Capelli *et al.* (1990), who also asked an open-ended intent question and then followed up with a forced-choice question; however, since the responses to the open-ended questions provided them with enough information, the forced-choice responses were not analyzed. It is important to note that their use of intent in the open-ended questions also encompassed attitude (e.g., *Laura was angry*) and meaning (e.g., *What a jerk.*). The forced-choice question they used also addressed attitude rather than intent (e.g., *Did Kevin mean that he was scared or not scared?*). In the current study, the forced-choice question was more closely related to intent per se (e.g., *Did Pat want to make Pat's sister feel good?*).

While the finding that speaker intent was the most difficult type of inference for children in the current study may initially appear to contradict previous literature, upon closer examination the finding may be congruent. Winner and Leekam (1991) concluded that the inference of speaker attitude was more difficult than the inference of speaker intent (i.e., second-order intention or what the speaker wants the listener to know). However, when asking children about attitude, researchers asked if the speaker was being mean or nice rather than inferring the speaker's emotional state. In the current study, asking questions about the speaker's motives was termed 'intent' (i.e., the speaker wanted listener to feel bad/good), and so the conclusion that intent was more a difficult task appears to support Winner and Leekam's finding.

Hancock *et al.* (2000) also concluded that the ability to infer intent precedes the ability to infer attitude. Speaker intent was queried (i.e., *Was B being mean or nice?*), and children could respond verbally or point to one of two pictures (a mean, angry face or a nice, happy face). However, children may have relied on the attitude (angry or happy) to answer the question, rather than what was termed 'intent' in the current study. Therefore, it is unclear if the current study supports or refutes their finding. This conflation of attitude and intent makes it difficult to compare findings and further supports the use of the terms in more consistent ways.

Given the task used in the current study where there was one intent for a given remark (either sarcastic or literal), and the child was asked a series of questions about that intent, it is also unclear if children can recognize both types of intent. Future research would help support the graded salience hypothesis if children could recognize conventional ironic meanings without also attending to the literal meanings in a manner similar to

idioms. Some children indicated the speaker was being sarcastic (e.g., child responded that Pat was 'being sarcastic'), which points to their understanding that one has to pay attention to those types of remarks in a certain way that is distinct from other remarks.

Children may have been reluctant to answer questions where the answer was that Pat wanted to make the other character feel bad. While this did not seem to be a pattern, it could have affected the choices made by individual children on individual responses. Hancock *et al.* (2000) reported that children inferred that the speaker was being mean 44% of the time following an ironic criticism. That result did not seem to be associated with the children's reluctance to state that the speaker was mean, since in the literal condition, they did so 81% of the time. Therefore, the interpretation of intent seemed to be related to the literal vs. non-literal conditions rather than reluctance to state that someone would be mean. Therefore, it is likely that children in the current study were similarly not deterred from assigning a negative intent to the speaker.

In addition to the findings related to the research questions, there was an additional finding for positive contexts paired with ironic remarks. Those remarks were intended literally, but some children made an inaccurate (ironic) interpretation of the remark. For example, in the story about Pat's brother helping to clean a rug, a literally biasing context, Pat said, *That's just perfect*. When asked what Pat meant, the child stated, *Now I have more cleaning to do*. In exploring those findings, there was a similar proportion of correct responses for conventional ironies in both positive and negative contexts and the ANOVA did not support a significant statistical difference between the children's performance in the exploratory analysis of positive and negative contexts for either remark type or intent. Since all of the children were able to complete a factual retell of the story, the difference in performance does not appear related to a misunderstanding of the events. It would be expected that the children would have no difficulty inferring speaker meaning for the literal remarks. But they did have difficulty inferring the meaning of conventionally ironic remarks used literally, which provides further evidence for the graded salience hypothesis because the literal context did not override the non-literal meaning of the conventionally ironic remark in some instances. In the current study design, the children did not have many opportunities to demonstrate their understanding of literal remarks, which limited the statistical power of the analysis. Further exploration of conventional and novel/situation-specific ironic remarks within literally biasing contexts would be of interest.

The finding that children were better able to infer the speaker's intended meaning for conventional remarks over novel/situation-specific remarks adds new information to what is understood about children's ability to understand

irony. In prior research, phrases that were situation-specific/novel appeared to be much more frequently utilized than phrases that were conventional. When both types were used (Capelli *et al.*, 1990), researchers did not make a distinction between the two in findings related to children's ability or the development of component skills. The coded meanings that children may have for conventionally ironic remarks, based on how frequently they are exposed to them, influences how likely they are to recognize that speakers may not mean exactly what they say. Therefore, the development of the ability to infer speaker meaning (i.e., that the speaker means the opposite) needs to be further elucidated within a framework of conventionality. Such a framework can be provided by the graded salience hypothesis for future research of children's comprehension of ironic compliments. A set of remarks used conventionally with ironic intent can be developed for this type of irony in a manner similar to the development of a set of conventional ironic criticisms.

CONCLUSION

The current study sought to explore the role of conventionality in how children understand irony. The graded salience hypothesis predicts that conventionally ironic remarks should ease comprehension. The finding that children were able to infer speaker meaning (i.e., that the speaker meant the opposite of what was said) with more accuracy given conventional rather than situation-specific remarks provided support for the graded salience hypothesis. However, no difference was found for two of the other components of irony comprehension: speaker attitude and speaker intent. These findings contribute to the work in irony comprehension in children by demonstrating the importance of more closely examining the types of ironic remarks used in empirical studies. Use of conventional remarks may allow an investigation of the role of familiarity/experience in irony comprehension.

REFERENCES

- Ackerman, B. (1981). Young children's understanding of a speaker's use of a false utterance. *Developmental Psychology*, **17**, 472–480.
- Ackerman, B. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, **35**, 487–508.
- Andrews, J., Rosenblatt, E., Malkus, U., Gardner, H., & Winner, E. (1986). Children's abilities to distinguish metaphorical and ironic utterances from mistakes and lies. *Communication & Cognition*, **19**, 281–298.
- Anolli, L., Ciceri, R., & Infantino, M. G. (2000). Irony as a game of implicitness: acoustic profiles of ironic communication. *Journal of Psycholinguistic Research*, **29**, 275–311.
- Anolli, L., Ciceri, R., & Infantino, M. G. (2002). From 'blame by praise' to 'praise by blame': analysis of vocal patterns in ironic communication. *International Journal of Psychology*, **37**, 266–276.

- ASHA (American Speech-Language Hearing Association) (1997). Guidelines for audiologic screening, online: <www.asha.org/policy>.
- Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, **16**, 243–260.
- Bishop, D. V. M. & Adams, C. (1992). Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech and Hearing Research*, **35**, 119–129.
- Boersma, P. & Weenik, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program], online: <www.praat.org>.
- Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, **48**, 257–277.
- Bugental, D., Kaswan, J., & Love, L. (1970). Perception of contradictory meanings conveyed by verbal and nonverbal channels. *Journal of Personality and Social Psychology*, **16**, 647–655.
- Capelli, C., Nakagawa, N., & Madden, C. (1990). How children understand sarcasm: the role of context and intonation. *Child Development*, **61**, 1824–1841.
- Clark, H., & Gerrig, R. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, **113**, 121–126.
- Clark, H., & Lucy, P. (1975). Understanding what is meant from what is said: a study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, **14**, 56–72.
- Culatta, B., Page, J., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools*, **14**, 66–74.
- Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't: children's understanding of figurative language. *British Journal of Developmental Psychology*, **1**, 121–134.
- Dews, S., & Winner, E. (1997). Attributing meaning to deliberate false utterances: the case of irony. In C. Mandel and A. McCabe (Eds.), *The problem of meaning: behavioral and cognitive perspectives* (pp. 377–414). Amsterdam: Elsevier Science Publishers.
- Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development*, **67**, 3071–3085.
- Eisterhold, J., Attardo, S., & Boxer, D. (2006). Reactions to irony in discourse: evidence for the least disruption principle. *Journal of Pragmatics*, **38**, 1239–1256.
- Elliott, C. (2007). *Differential ability scales-second edition*. San Antonio, TX: Harcourt Assessment, Inc.
- Filippova, E., & Astington, J. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, **79**, 126–138.
- Filippova, E., & Astington, J. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development*, **81**, 913–928.
- Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal school-age children: some preliminary data. *Perceptual and Motor Skills*, **79**, 55–64.
- Gibbs, R. (1986a). Comprehension and memory for nonliteral utterances: the problem of sarcastic indirect requests. *Acta Psychologica*, **62**, 41–57.
- Gibbs, R. (1986b). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology*, **115**, 3–15.
- Gillam, R., & Pearson, N. (2004). *Test of narrative language*. East Moline, IL: Linguisticsystems.
- Giora, R., Balaban, N., Fein, O., & Alkabetz, I. (2005). Explicit negation as positivity in disguise. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: social and cultural influences* (pp. 233–258). Mahwah, NJ: Lawrence Erlbaum Associates.
- Giora, R., & Fein, O. (1999a). Irony comprehension: the graded salience hypothesis. *Humor: International Journal of Humor Research*, **12**, 425–436.
- Giora, R., & Fein, O. (1999b). Irony: context and salience. *Metaphor and Symbol*, **14**, 241–257.
- Giora, R., Fein, O., & Schwartz, T. (1998). Irony: graded salience and indirect negation. *Metaphor and Symbol*, **13**, 83–101.

- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, vol. 3, speech acts* (pp. 41–50). New York: Academic Press.
- Hancock, J., Dunham, P., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development, 1*, 227–248.
- Hollingshead, A. B. (1975). Four factor index of social status. Unpublished manuscript, Department of Sociology, Yale University, New Haven, CT.
- Honeck, R., Welge, J., & Temple, J. (1998). The symmetry control in tests of the standard pragmatic model: the case of proverb comprehension. *Metaphor and Symbol, 13*, 257–273.
- Ivanko, S., & Pexman, P. (2003). Context incongruity and irony processing. *Discourse Processes, 35*, 241–279.
- Jorgensen, J., Miller, G., & Sperber, D. (1984). Test of the mention theory of irony. *Journal of Experimental Psychology, 113*, 112–120.
- Keenan, T., & Quigley, K. (1999). Do young children use echoic information in their comprehension of sarcastic speech? A test of echoic mention theory. *British Journal of Developmental Psychology, 17*, 83–96.
- Kreuz, R., & Roberts, R. (1995). Two cues for verbal irony: hyperbole and ironic tone of voice. *Metaphor and Symbolic Activity, 10*, 21–31.
- Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of the pie? The allusional pretense theory of discourse irony. *Journal of Experimental Psychology, 124*, 3–21.
- Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false belief in children: What changes when mothers rather than siblings speak? *European Journal of Developmental Psychology, 10*, 301–317.
- McDonald, S. (2000). Neuropsychological studies of sarcasm. *Metaphor and Symbol, 115*, 85–98.
- Milosky, L., & Ford, J. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes, 23*, 47–61.
- Newcomer, P., & Hammill, D. (1997). *Test of language development: primary*, 3rd ed. Austin, TX: Pro-Ed.
- Pexman, P. (2005). Social factors in the interpretation of verbal irony: the roles of speaker and listener characteristics. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: social and cultural influences* (pp. 209–232). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pexman, P., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? *Journal of Neurolinguistics, 20*, 178–196.
- Pexman, P., Glenwright, M., Hala, S., Kowbel, S., & Jungen, S. (2006). Children's use of trait information in understanding verbal irony. *Metaphor and Symbol, 21*, 39–60.
- Rockwell, P. (2000). Lower, slower, louder: vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29*, 483–495.
- Searle, J. (1979). *Expression and meaning: studies in the theory of speech acts*. New York: Cambridge University Press.
- Sperber, D. (1984). Verbal irony: pretense or echoic mention? *Journal of Experimental Psychology, 113*, 130–136.
- Sperber, D., & Wilson, D. (1981). Pragmatics. *Cognition, 10*, 281–286.
- Tomblin, J. B., Records, N., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in Kindergarten children. *Journal of Speech and Hearing Research, 39*, 1284–1294.
- Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *Journal of Pragmatics, 32*, 1777–1806.
- Utsumi, A. (2004). Stylistic and contextual effects in irony processing. Paper presented at 26th Annual Meeting of the Cognitive Science Society, Chicago, IL.
- Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: understanding the speaker's second-order intention. *British Journal of Developmental Psychology, 9*, 257–270.