

Children's understanding of nonliteral criticism:

A pilot study of speaker belief and attitude

Marianna Szabóné Habók 

University of Debrecen, Doctoral School of Linguistics
4002 Debrecen, Pf. 400.
mannahabok@gmail.com

Abstract

This study describes a pilot experiment investigating children's recognition of speaker belief and their evaluation of speaker attitude in literal and nonliteral utterances. Following the logic of Glenwright and Pexman's (2010) paradigm, a set of puppet shows was presented to preschool children aged 5–6 years ($N = 18$) and first-grade children aged 6–7 years ($N = 20$) and an adult control group ($N = 9$). The nonliteral stories included person-directed sarcastic criticism and situation-directed ironic criticism.

Preschool children showed lower and more variable accuracy in nonliteral items, while first-grade children performed more accurately and consistently. Literal praise received the most positive attitude ratings, literal criticism the most negative, and the two nonliteral criticism types intermediate ratings. The difference between person-directed sarcastic criticism and situation-directed ironic criticism was not statistically significant. These findings suggest that earlier developmental patterns in nonliteral criticism comprehension can also be observed among Hungarian-speaking children.

Keywords: children, nonliteral criticism, irony comprehension, sarcasm, speaker belief, speaker attitude

1 Introduction

1.1 Irony, sarcasm and nonliteral criticism

The terms sarcasm and irony are intertwined. To date, there has been no unanimously accepted distinction between them in linguistics. Some theorists consider sarcasm a synonym for irony, while others treat it as a subtype of irony (e.g. Gibbs & Colston 2007; Kreuz 2020). This study considers them from a pragmatic point of view, concentrating on speaker belief and attitude. It treats irony as an umbrella term and sarcasm as a subtype of irony (Gibbs & Colston 2007).

Verbal irony may serve critical and non-critical functions. It may express criticism, but it may also be used for ironic praise or humour. The present study examines only ironic criticism (Gibbs & Colston 2007; Kreuz 2020). Verbal irony is a form of nonliteral language in which speakers say the opposite or something very different from what they mean. Sarcasm is defined as “the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticise something in a humorous way” (Cambridge Dictionary, cited in Garmendia 2018: 10). Thus, sarcastic utterances typically involve saying something that is overtly false to communicate a negative but truthful evaluation.

In the operational distinction adopted here, sarcasm refers to ironic criticism directed at a particular person's performance or possession, while irony refers to situation-directed ironic criticism (Glenwright & Pexman 2010). Both have an object of criticism; the difference is whether the criticism is directed at a particular person or at a general situation. These conditions are referred to as person-directed sarcastic criticism and situation-directed ironic criticism.

1.2 *Development of irony comprehension*

When it comes to understanding irony and sarcasm, both require recognising speaker meaning beyond literal semantics (Sperber & Wilson 1981, 1995). Comprehension involves both recognition of speaker belief and evaluation of speaker attitude, which are related but distinct aspects of irony comprehension (Happé 1993; Filippova & Astington 2008; Schnell & Varga 2012). Recognition of speaker belief concerns whether the listener understands that the speaker's belief differs from the literal meaning of the utterance, while evaluation of speaker attitude concerns whether the listener evaluates the speaker as nice or mean. The relevant mental processes include abilities linked to developmental changes in mental-state reasoning (Perner & Wimmer 1985; Frith & Frith 2010). Both first-order and second-order mental-state reasoning are components of Theory of Mind. According to Perner and Wimmer (1985), there are two levels of belief attribution that play a central role in children's comprehension of social interactions and emerge at different stages of development. Firstly, around the age of four years, first-order beliefs appear, which involve representing another person's beliefs about the world. Secondly, a few years later, children develop second-order beliefs that include representing another person's beliefs about someone else's thoughts (Miller 2009; Frith & Frith 2010).

If distinguishing irony and sarcasm requires intention attribution, developmental differences should be expected (Winner & Leekam 1991; Filippova & Astington 2008; Schnell & Varga 2012). Previous research indicates that adults can distinguish sarcasm from verbal irony (Kreuz & Glucksberg 1989). Adults have been shown to recognise that sarcasm typically concerns a specific person, whereas verbal irony may instead concern a situation or state of affairs (Kreuz 2000). They have also been shown to perceive sarcastic utterances more negatively than ironic utterances. For this reason, adult competence can provide a benchmark for mature pragmatic processing (Kreuz & Glucksberg 1989; Kreuz 2000). The key developmental question concerns when children begin to show similar sensitivity (Harris & Pexman 2003; Glenwright & Pexman 2010). Comparing children's responses with adult patterns can provide a benchmark for examining group-level differences (Glenwright & Pexman 2010).

In the study by Glenwright and Pexman (2010), the distinction between verbal irony and sarcasm was systematically investigated. Their experiment served as a starting point for the pilot study described here. The main aim of their study was to explore whether children aged 5–6 and 9–10 years could differentiate between verbal irony and sarcasm. Their sample consisted of 71 participants in each age group from middle-class Canadian families with English as their primary language. They had two independent variables: utterance type and topic type. The utterance types were nonliteral criticism, literal criticism and literal compliments, while the topic types were the target's performance, the target's possession and a general situation. Children were shown nine puppet-show items. In each set, two person-directed sarcastic items concerned a particular child's failed performance or faulty possession, while one situation-directed ironic item concerned a general situation. The dialogues were prerecorded. Sarcastic and ironic criticisms were delivered with prolonged articulation and exaggerated pitch, literal criticisms with lower pitch and a blunt tone, and literal compliments with higher pitch and a warm tone. During questioning, the experimenter repeated the intonation consistently.

After each story, children were asked four questions. The first question focused on speaker belief. Participants had to indicate whether the speaker considered the performance, possession or situation good or bad. The second question concerned speaker attitude. Children had to rate the speaker's attitude on the Nice/Mean scale. The third question dealt with identification, while the final question asked children to justify their choice.

Glenwright and Pexman found that younger children aged 5–6 years could identify the non-literal meaning but not the pragmatic difference between sarcasm and irony. Children aged 9–10 years showed strong comprehension of nonliteral meanings and distinguished the speakers' in-

tentions in sarcasm and irony at a higher rate. Their speaker-attitude ratings also showed developmental differences, as older children rated person-directed sarcastic criticism meaner than situation-directed ironic criticism. Children aged 5–6 years rated all nonliteral criticism as equally negative. These results supported a two-step developmental pattern, in line with earlier accounts of nonliteral language comprehension (e.g. Happé 1993; Filippova & Astington 2008).

First, children distinguish sarcasm from literal language by inferring what the sarcastic speaker believes. At an early stage, however, they do not show consistent sensitivity to differences in speaker attitude. Thus, recognition of speaker belief may precede more fine-grained evaluation of speaker attitude (Glenwright & Pexman 2010). This two-step pattern may be related to the development of mental-state reasoning. Recognising speaker belief may draw mainly on first-order mental-state reasoning, while evaluating more complex communicative intentions may place additional demands on second-order mental-state reasoning (Perner & Wimmer 1985; Miller 2009; Frith & Frith 2010).

2 The Present Pilot Study

Hungarian pragmatics has addressed irony comprehension from both theoretical and cognitive perspectives. For instance, Szücs & Babarczy (2017) investigated the role of Theory of Mind, grammatical competence, and metapragmatic awareness in children's irony comprehension. Schnell (2012, 2016, 2021) examined the development of pragmatic competence and the cognitive aspects of irony comprehension. These studies examine different aspects of irony comprehension, but they do not focus specifically on the distinction between recognition of speaker belief and evaluation of speaker attitude.

This pilot study applies Glenwright and Pexman's (2010) paradigm to Hungarian-speaking participants. The original study compared English-speaking children aged 5–6 and 9–10 years, whereas the present study compares preschool children aged 5–6 years and first-grade children aged 6–7 years. By using Hungarian-language materials, it provides evidence from another linguistic and cultural context. It examines whether earlier developmental patterns also appear in Hungarian-speaking participants around school entry. The adult control group provides a pragmatic benchmark. The study does not directly compare English and Hungarian or examine language-specific or cultural differences.

The study aims to explore Hungarian children's recognition of speaker belief and their evaluation of speaker attitude in literal and nonliteral utterances. It addresses the following research questions:

RQ1: How accurately do Hungarian-speaking preschoolers and schoolchildren identify speaker belief in literal and nonliteral utterances?

RQ2: How do preschoolers and schoolchildren evaluate speaker attitude in person-directed sarcastic criticism, situation-directed ironic criticism, literal criticism and literal praise?

Based on previous research, schoolchildren are expected to show more consistent comprehension of nonliteral utterances than preschool children (Filippova & Astington 2008; Glenwright & Pexman 2010). The comparison between person-directed sarcastic criticism and situation-directed ironic criticism remains exploratory because the task contains two sarcastic items and only one ironic item.

2.1 Participants

All participants were monolingual speakers of Hungarian without diagnosed developmental or language delays. The experiment involved three groups (see Table 1). First, a control group of 9 adults whose results served as a benchmark for adult-like comprehension. Second, a preschool

group of 18 preschool-attending children. Third, 20 first-grade students. It must be noted here that group assignment was determined by schooling status and not chronological age, which resulted in partial age overlap between the children's groups. Due to this, they were not treated as two clearly separated developmental age groups.

Groups	Preschool children	Schoolchildren	Control group
Number of participants (males/females)	18 (9/9)	20 (9/11)	9 (5/4)
Average age (years; months)	5;9	6; 6	41;6
Range of age (years; months)	4;11 – 6;8	6;2 – 7;0	18 – 70
Standard Deviation (months)	4.90	2.68	N/A

Table 1. Participants' demographic data

2.2 Methodology

Testing of the control group took place in home environments while children were tested in a quiet room in their preschool and school individually. The tests lasted 12–15 minutes and unfolded in a single session facilitated by the same experimenter. The study received ethical approval from the relevant institution, and all parents provided written informed consent prior to the study. Participation was voluntary, and confidentiality was fully guaranteed. As a token of appreciation, a sticker was given to the children.

The study drew on the logic of Glenwright and Pexman's (2010) paradigm as outlined above. Accordingly, nine puppet-show items involving common children's activities were presented to each participant: three literal criticism items, three literal praise items, two person-directed sarcastic criticism items and one situation-directed ironic criticism item. Figure 1 shows one of these items as used during testing. The operationalisation involved the same three topic types as the original design: a target's performance, a target's possession and a general situation (see Table 2). The unequal numbers of person-directed sarcastic and situation-directed ironic items followed the original paradigm but represent a methodological limitation. For this reason, their direct comparison was treated as exploratory.

Sample Puppet-Show Scenarios: Recorder Story Set

Nonliteral and Literal Criticisms

Performance

Kati is learning to play the recorder, but it is not going well; the recorder produces very unpleasant sounds in her hands.

1. **Sarcastic criticism**

Zoli says, 'You really play the recorder beautifully.'

2. **Literal criticism**

Zoli says, 'You're not playing the recorder beautifully.'

Possession

Kati has a new recorder, but when Zoli tries it, the instrument produces very poor sounds.

3. **Sarcastic criticism**

Zoli says, 'This recorder sounds wonderful.'

4. **Literal criticism**

Zoli says, 'This recorder sounds very bad.'

Situation

Kati and Zoli receive recorders at the music school to practise on, but the instruments produce very unpleasant sounds.

5. **Ironie criticism**

Kati says, 'These recorders sound wonderful.'

6. **Literal criticism**

Kati says, 'These recorders sound very bad.'

Literal Praises

Performance

Kati is learning to play the recorder, and it is going very well.

7. Zoli says, 'You really play the recorder beautifully.'

Possession

Kati has a new recorder and proudly shows Zoli how beautifully it sounds.

8. Zoli says, 'This recorder sounds wonderful.'

Situation

Kati and Zoli receive recorders at the music school to practise on, and the instruments produce very beautiful sounds.

9. Kati says, 'These recorders sound wonderful.'

Table 2. Sample puppet show items



Figure 1. Example of puppet show set-up used during testing
(Scenario roller: performance, sarcastic criticism)

Before testing, participants were trained on the Nice/Mean rating scale as shown in Figure 2 adapted from Glenwright and Pexman (2010). Children were informed that the six faces on the scale mirror the following evaluations, from left to right: very nice, nice, a little bit nice, a little bit mean, mean and very mean. Examples were also provided for better understanding. The six-point format was retained to follow Glenwright and Pexman's (2010) procedure and to allow comparison with their findings. The use of six response categories may have increased the demands on the preschool participants.

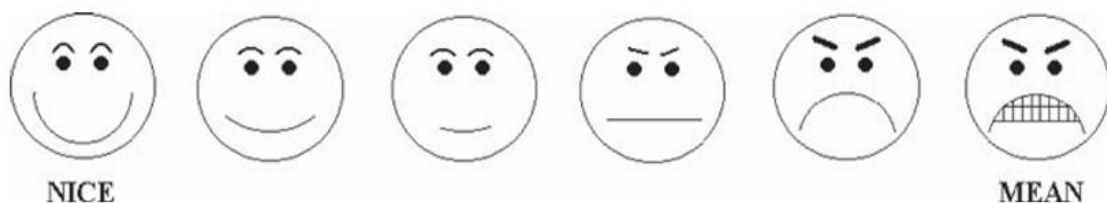


Figure 2. Facial rating scale used to assess perceived meanness

Following each puppet show, participants had to answer four questions. The first question assessed recognition of the speaker's belief about the literal proposition. Participants had to indicate whether the speaker considered the performance, possession or situation good or bad. It did not directly assess communicative intention. For example, in the Trampoline Scenario (target's performance, sarcastic criticism), participants were asked: 'When Zoli said: "This was a great jump", did he think that it was a great jump or a bad jump?' The second question focused

on speaker attitude. Participants had to rate the utterances on the six-point scale ranging from nice (1) to mean (6) presented above: 'Point to one of these faces to show me how mean or nice Zoli was when he said "This was a great jump."' Up to this point the design was the same as Glenwright and Pexman's (2010).

Their third question dealt with identification, that is, which puppet the participants identified with most: 'Which of these puppets acts most like you?' In the present pilot study, the first three adult participants consistently chose the puppet matching their own gender, independently of utterance type. As a result, the question was revised to: 'Would you say something like that?' The first three adult participants answered both versions of the question. The revised question was then used with all subsequent participants. Responses to the revised version were available for every adult and were analysed together. Responses to the original version were treated as pilot data and excluded from the combined analysis. The final question asked participants to justify their response: 'When would you say something like that and why?' These responses were treated as exploratory and were not regarded as a separate measure of communicative intention.

Responses to the first question were used to assess recognition of speaker belief about the literal proposition. They were not treated as a direct measure of communicative intention. The second question assessed evaluation of speaker attitude. Responses to the third and fourth questions were treated as exploratory.

The design included two factors: utterance type (nonliteral criticism, literal criticism and literal praise) and topic type (performance, possession and situation). The distribution of item types described above mirrored Glenwright and Pexman's (2010) original paradigm. For the sake of full counterbalancing, nine versions of the materials were created. In this way, each puppet show appeared with different utterance types and topic types across participants. Within each version, the order of the items was randomised.

The puppet shows were prerecorded to ensure consistency. In line with commonly reported prosodic features of sarcasm and verbal irony, sarcastic and ironic criticisms were voiced with a prolonged articulation and exaggerated pitch (e.g. Ackerman 1983; Nakassis & Snedeker 2002). Literal criticisms and literal praise utterances were voiced at the pace of regular speech. During questioning, the experimenter matched the intonation of the original recording. It follows that prosody was included in the materials but was not manipulated independently.

2.3 Data analysis

Responses to the belief question were scored as correct or incorrect. In nonliteral items, responses were correct when participants recognised the mismatch between the utterance and the events in the story. In literal items, responses consistent with the literal meaning were correct. Each correct response received one point. The maximum score was 6 for literal items and 3 for nonliteral items. Accuracy responses were analysed at item level using a binomial logistic mixed-effects model. The model examined participant group, literalness (literal or nonliteral) and their interaction. It also accounted for repeated responses from the same participant.

Speaker attitude was rated on a six-point scale from nice (1) to mean (6). Responses to the third and fourth questions were analysed descriptively. Attitude ratings were analysed using a linear mixed-effects model. The model examined utterance type, participant group and their interaction and accounted for repeated responses from the same participant. Utterance type had four levels: person-directed sarcastic criticism, situation-directed ironic criticism, literal criticism and literal praise. Pairwise comparisons were Holm-corrected, and effect sizes were reported as Cohen's *d*.

2.4 Results

2.4.1 Accuracy in literal and nonliteral stories

Concerning accuracy data, the control group performed at ceiling across all the item types. All adult participants scored 9 out of 9, indicating that they answered all speaker-belief questions correctly. They rejected the literal interpretation in nonliteral items as expected. With their results the baseline was established against which children's accuracy could be measured.

Figure 3 presents preschoolers' accuracy results. Performance in the literal items was very high and close to adult performance. 15 out of 18 children scored 6 out of 6 points, two scored 5, and one child scored 2. Regarding nonliteral items, accuracy was substantially lower. Eight out of 18 children scored 3 out of 3 points, three scored 2, two scored 1, and five scored 0. The visual pattern reflects the asymmetry between literal and nonliteral comprehension. Performance also varied on the ironic item. However, as only one ironic item was included, no firm conclusion can be drawn about irony as a category.

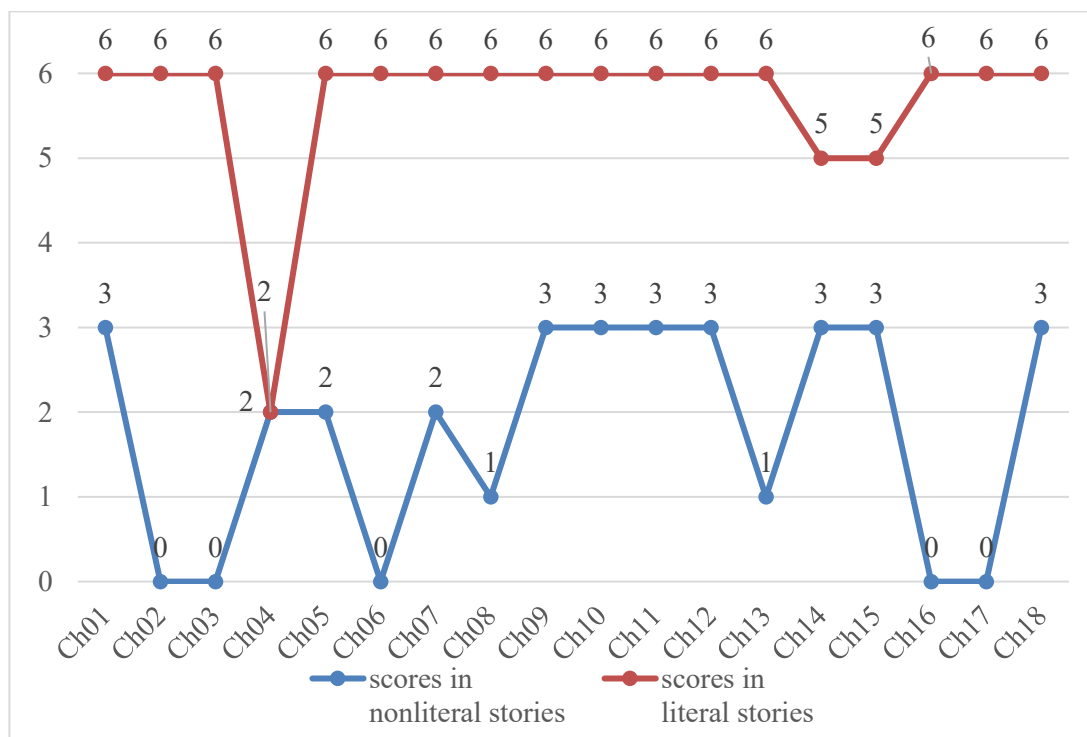


Figure 3. Preschool children's accuracy scores across story types

Figure 4 displays individual accuracy scores for schoolchildren. In the literal condition, performance was consistently high ($M = 5.95$ out of 6, $SD = 0.22$). All but one child achieved 6 points. In the nonliteral condition, accuracy was also high overall ($M = 2.75$ out of 3, $SD = 0.44$), exceeding preschoolers' performance. 15 out of 20 schoolchildren scored 3 out of 3 points. All in all, compared to preschoolers, schoolchildren showed higher accuracy than preschoolers in the nonliteral condition.

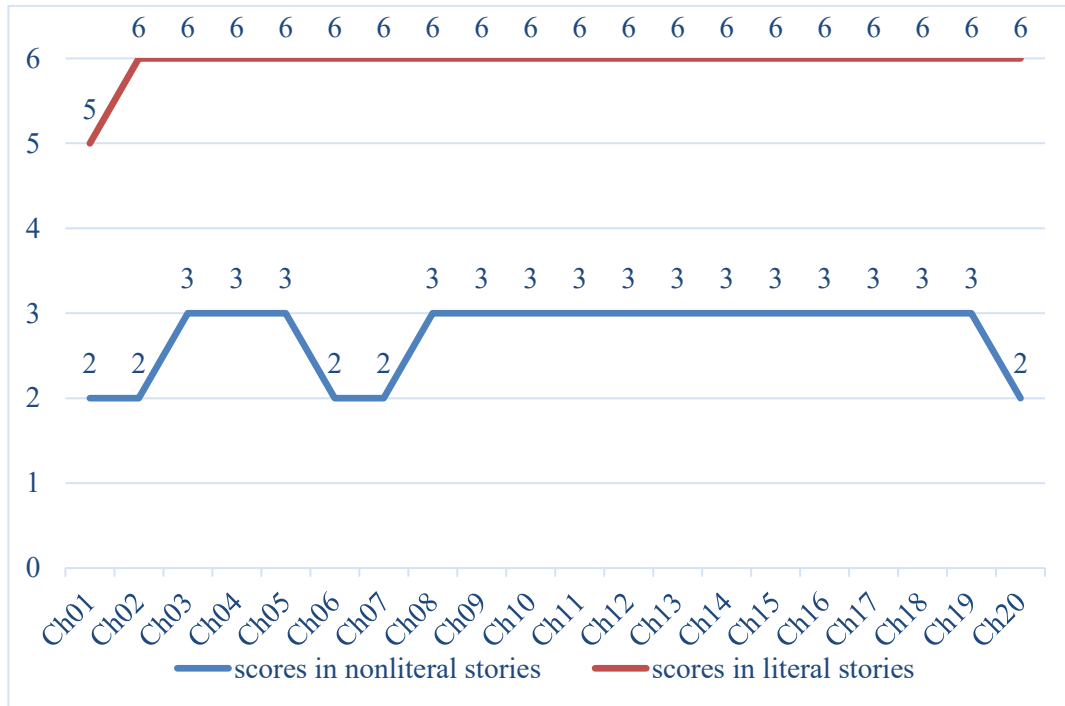


Figure 4. Schoolchildren's accuracy scores across story types

Table 3 summarises group-level accuracy means and standard deviations for literal, nonliteral, sarcastic, and ironic items. It allows for a direct comparison between preschoolers and schoolchildren.

Utterance type	Preschoolers (M)	Preschoolers (SD)	Schoolchildren (M)	Schoolchildren (SD)	Max. score that they could earn
Literal	5.67	0.97	5.95	0.22	6
Nonliteral	1.78	1.31	2.75	0.44	3
Sarcastic	1.11	0.90	1.85	0.37	2
Ironic	0.67	0.49	0.90	0.31	1

Table 3. Accuracy means and standard deviations for preschoolers and schoolchildren

Figure 5 shows the nonliteral criticism boxplot comparing preschoolers and schoolchildren. It refines the group-level differences reported in Table 3. Preschoolers' results in this condition show a large vertical spread with high variability. Their scores ranged from 0 to 3, with the lower quartile near 0–1, the upper quartile near 2–3 and a median around 2. By contrast, schoolchildren's scores were concentrated at the upper end of the scale: 15 scored 3 and five scored 2.

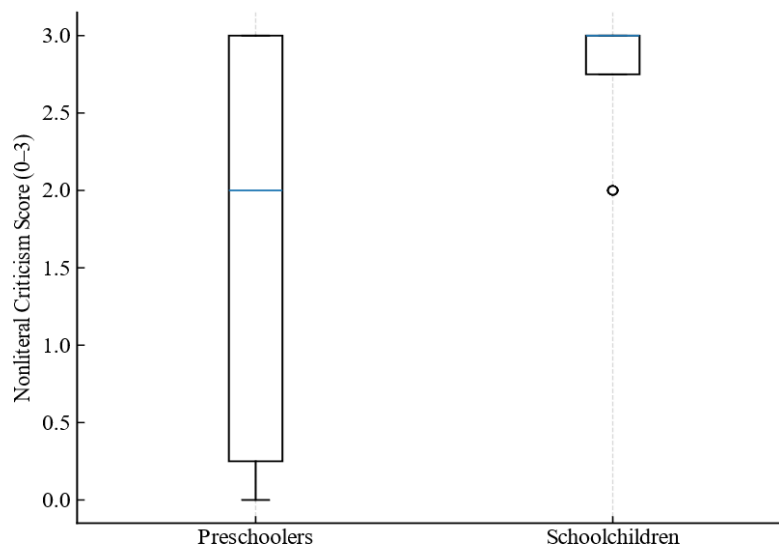


Figure 5. Nonliteral criticism performance by groups

Item-level accuracy responses from the two child groups were analysed using a binomial logistic mixed-effects model. The interaction between participant group and literalness was not significant, $\chi^2(1) = 0.13$, $p = .721$. After the interaction was removed, there were significant main effects of participant group, $\chi^2(1) = 11.30$, $p < .001$, and literalness, $\chi^2(1) = 41.04$, $p < .001$. Holm-corrected comparisons showed that schoolchildren had higher odds of a correct response than preschoolers in the nonliteral items, $OR = 11.84$, 95% CI [2.53, 55.40], $p = .005$. The group difference in the literal items was not significant, $OR = 7.40$, 95% CI [0.67, 81.80], $p = .103$.

2.4.2 Speaker-attitude ratings

Adult participants rated literal praise as nice and the three criticism types as relatively mean. Across the groups, literal praise was evaluated more positively and literal criticism more negatively. Person-directed sarcastic criticism and situation-directed ironic criticism received intermediate attitude ratings.

Apart from the comprehension of the utterances, children's evaluative judgments were also analysed. Figure 6 presents preschoolers' speaker-attitude ratings across the four story types. The two nonliteral criticism types showed considerable overlap. Ratings were given on a 6-point scale (1 = nice, 6 = mean), which, as mentioned above, had been introduced and explained to the children prior to the task. Preschoolers' ratings of literal praise clustered tightly around 1–2 with a low median of ~ 1.3 suggesting that they tended to evaluate praise as nice. On the other hand, the median of literal criticism was around 4.3–4.5 (between a “little mean” and “mean”), with a moderate spread. Thus, regarding evaluation of speaker attitude, preschoolers showed sensitivity to the nice–mean polarity when it came to the difference between praise and criticism.

Concerning sarcasm, the median was approximately 3.7 with considerable variability (1.5–5.5 range). Some children rated sarcastic criticism as less mean than literal criticism, others as equally mean. The median of irony ratings was 4.5, similar to literal criticism, with considerable variability. Irony tended to be evaluated as negative. Still, as only a single irony item was included, conclusions regarding irony as a category should be interpreted with caution. On the whole, the results suggested limited differentiation between sarcasm and irony, as both sarcastic and ironic criticisms were generally rated as mean.

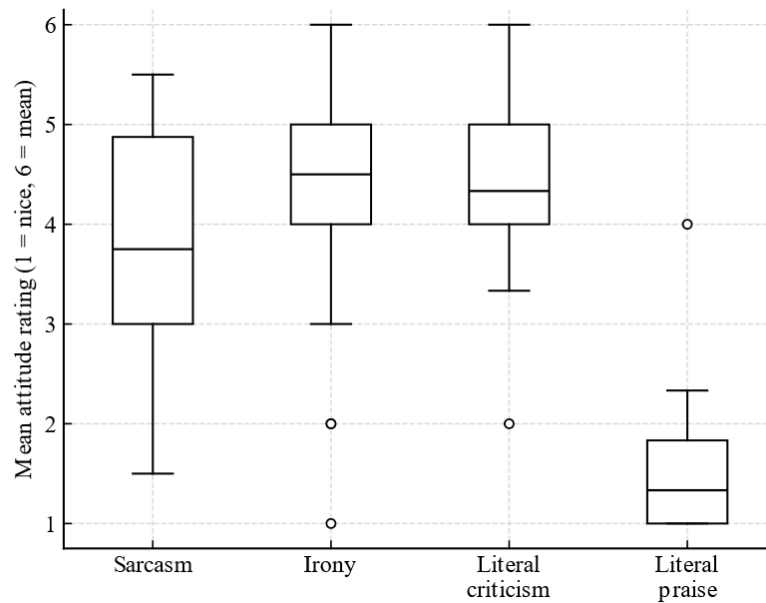


Figure 6. Preschoolers' attitude ratings by story type

Figure 7 shows the boxplot of schoolchildren's attitude ratings by story type. While the ratings of literal praise had an extremely tight cluster at the ceiling of "niceness", literal criticism was close to the ceiling of meanness. Hence, schoolchildren tended to evaluate literal praise as nice and literal criticism as mean. In comparison, sarcasm and irony ratings showed substantial overlap with both conditions receiving intermediate evaluations between literal praise and literal criticism. Sarcasm ratings appeared somewhat more softened than irony ratings. In addition, irony ratings showed more spread. Children varied in how mean they thought it was. These patterns reflect descriptive tendencies in median ratings. The contrast between sarcasm and irony was not statistically significant.

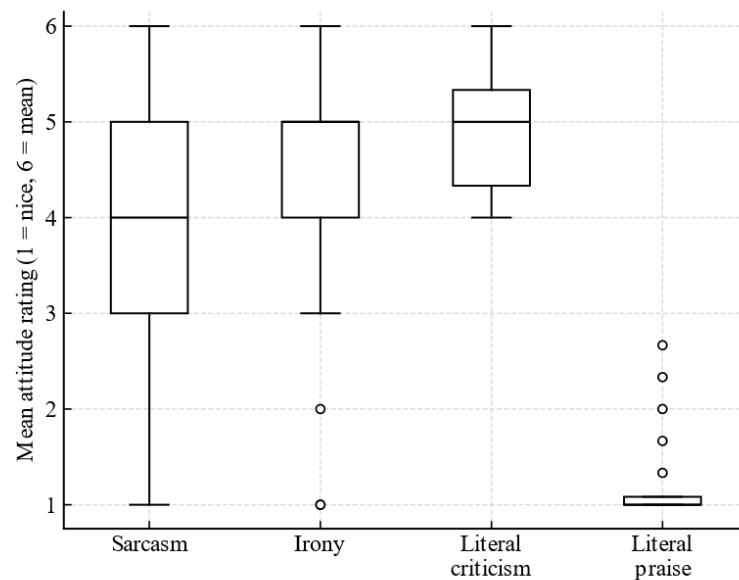


Figure 7. Schoolchildren's attitude ratings by story type

The interaction between participant group and utterance type was not significant, $\chi^2(6) = 11.72$, $p = .068$. After the interaction was removed, the linear mixed-effects model showed a significant main effect of utterance type, $\chi^2(3) = 404.40$, $p < .001$. The main effect of participant group was not significant, $\chi^2(2) = 0.77$, $p = .681$. Thus, group differences in attitude ratings were descriptive and were not statistically significant.

Holm-corrected pairwise comparisons across groups showed that literal criticism, sarcasm and irony were rated meaner than literal praise ($t = 24.98$, $d = 2.98$; $t = 17.84$, $d = 2.38$; and $t = 15.73$, $d = 2.65$, respectively; all $p < .001$). Literal criticism was rated meaner than sarcasm ($t = 4.51$, $p < .001$, $d = 0.60$). The difference between literal criticism and irony was not significant ($t = 1.94$, $p = .107$, $d = 0.33$), nor was the difference between sarcasm and irony ($t = -1.53$, $p = .126$, $d = -0.27$).

2.4.3 Exploratory responses and explanations

Responses to the third question ('Would you say something like that?') varied in relation to utterance type in all groups. On one hand, literal praise was accepted in all cases in the control group and by the vast majority of children in the present sample. On the other hand, nonliteral criticism was accepted considerably less frequently.

Control adults appeared sensitive to interpersonal norms. Participants of the control group frequently rejected sarcastic and ironic criticism. In their responses to the final question ('When would you say something like that and why?'), adults frequently referred to potential harm (e.g. "it is mean", "I would not want to hurt the other person") or mockery.

Across both child groups, responses to the last question were mostly short and heterogeneous. Preschoolers and schoolchildren differed mainly in the basis of their justifications. In the preschool group ($N = 18$), the majority of justifications referred to concrete situational or object-related factors (e.g. broken objects, unsuccessful actions, specific play situations). In the nonliteral criticism items four preschoolers used explicit evaluative terms such as "bad" or "unkind" to explain why they would not say it. Six preschoolers' justifications did not clearly align with the acceptance or rejection given in the third question.

Half of the participants of the school-age group ($N = 20$) used evaluative language (e.g. "nasty", "not nice", "I would not like it") more frequently. Five children referred to interpersonal considerations or harm. Although situation-directed explanations remained common, their alignment between responses to the third and fourth questions was more consistent, with few mismatches (only 4 cases altogether).

2.5 Discussion

In the present sample, schoolchildren identified speaker belief in nonliteral stories more consistently than preschoolers, while preschoolers showed greater variability. Across the groups, literal praise was evaluated more positively and literal criticism more negatively. Person-directed sarcastic criticism and situation-directed ironic criticism received intermediate attitude ratings.

The findings of the present study are broadly comparable to those of Glenwright and Pexman (2010): preschoolers showed variable nonliteral comprehension and did not consistently differentiate sarcasm and irony. The present study adds a first-grade group aged 6–7 years. These children were more accurate and less variable than the preschoolers. Still, they did not show a statistically significant difference between the two criticism types. The first-grade group showed more consistent comprehension of nonliteral meaning, but no stable differentiation between person-directed sarcastic criticism and situation-directed ironic criticism. Since the study did not directly compare English- and Hungarian-speaking participants, no conclusion can be drawn about language-specific differences.

These findings have clear boundaries. The first question assessed recognition of speaker belief about the literal proposition, and communicative intention was not measured directly. Consequently, the study cannot establish a developmental distinction between person-directed sarcastic criticism and situation-directed ironic criticism. As only one situation-directed ironic item was included, responses may reflect the characteristics of that individual story. The preschool and school groups overlapped in age and were defined by schooling status. This means that group differences cannot be attributed specifically to age. Finally, the Hungarian findings provide preliminary evidence relevant to the generalisability of the earlier pattern but do not demonstrate cross-linguistic or cross-cultural differences.

3 Conclusion

The pilot study presented above investigated the recognition of speaker belief and evaluation of speaker attitude among Hungarian preschool children and schoolchildren in literal and nonliteral utterances. The adult control group answered all speaker-belief questions correctly. In the child groups, preschool children displayed lower and more variable accuracy across nonliteral items, while schoolchildren were more accurate and consistent. The child groups did not show a statistically reliable distinction between person-directed sarcastic criticism and situation-directed ironic criticism.

Concerning attitude ratings, schoolchildren tended to rate situation-directed ironic criticism meaner than person-directed sarcastic criticism. However, this difference was descriptive and not statistically significant. On the whole, the findings indicate group-level differences within the present sample but cannot be attributed specifically to age. The main finding is therefore that the higher and more consistent nonliteral comprehension found in the first-grade group was not accompanied by a statistically reliable distinction between person-directed sarcastic criticism and situation-directed ironic criticism.

Nevertheless, the pilot study had several limitations that necessitate careful interpretation of the findings. It relied on modest sample sizes, which constrain generalisability. The item number was restricted. For this reason, results can be regarded as exploratory and not definitive. It is also important to note that other variables not assessed here (e.g. Theory of Mind) might have played a significant role. Moreover, a methodological limitation concerns the use of marked prosody in the sarcastic and ironic conditions. Exaggerated pitch and prolonged articulation are commonly used to mark sarcasm and verbal irony in adult speech (Ackerman 1983; Nakassis & Snedeker 2002). These prosodic features may have influenced the children's responses. Hence, participants may have relied on intonational cues as indicators of nonliteral meaning instead of inferring nonliteral meaning from contextual information alone. Also, the six-point Nice/Mean scale may have been difficult for preschoolers to use. This difficulty may reflect performance limitations and does not necessarily indicate a lack of competence in evaluating speaker attitude (Eskritt et al. 2008). Future studies should use a scale with fewer response options.

Thus, the present study provides useful pilot data and a basis for a larger-scale investigation using balanced numbers of sarcastic and ironic items, more clearly separated participant groups and a simpler attitude scale. Since English- and Hungarian-speaking participants were not directly compared, the findings do not demonstrate language-specific or cultural differences. Future studies should include multiple irony items per utterance type as well as independent Theory of Mind measures. Longitudinal and observational designs would be especially useful in obtaining a more fine-grained picture of the development of irony comprehension.

References

- Ackerman, B. P. (1983): Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology* 35.3, 487–508. [https://doi.org/10.1016/0022-0965\(83\)90023-1](https://doi.org/10.1016/0022-0965(83)90023-1)
- Eskritt, M., Whalen, J. & Lee, K. (2008): Preschoolers can recognize violations of the Gricean maxims. *British Journal of Developmental Psychology* 26.3, 435–443. <https://doi.org/10.1348/026151007X253260>
- Filippova, E. & Astington, J. W. (2008): Further development in social reasoning revealed in discourse irony understanding. *Child Development* 79.1, 126–138. <https://doi.org/10.1111/j.1467-8624.2007.01115.x>
- Frith, C. D. & Frith, U. (2010): The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1537, 165–176. <https://doi.org/10.1098/rstb.2009.0160>
- Garmendia, J. (2018): *Irony*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316136218>
- Gibbs, R. W., Jr. & Colston, H. L. (eds.) (2007): *Irony in Language and Thought: A Cognitive Science Reader*. Mahwah, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410616685>
- Glenwright, M. & Pexman, P. M. (2010): Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language* 37.2, 429–451. <https://doi.org/10.1017/S0305000909009520>
- Happé, F. G. E. (1993): Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition* 48.2, 101–119. [https://doi.org/10.1016/0010-0277\(93\)90026-R](https://doi.org/10.1016/0010-0277(93)90026-R)
- Harris, M. & Pexman, P. M. (2003): Children's perceptions of the social functions of verbal irony. *Discourse Processes* 36.3, 147–165. https://doi.org/10.1207/S15326950DP3603_1
- Kreuz, R. J. (2000): The production and processing of verbal irony. *Metaphor and Symbol* 15.1–2, 99–107. <https://doi.org/10.1080/10926488.2000.9678867>
- Kreuz, R. J. (2020): *Irony and Sarcasm*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/12503.001.0001>
- Kreuz, R. J. & Glucksberg, S. (1989): How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General* 118.4, 374–386. <https://doi.org/10.1037/0096-3445.118.4.374>
- Miller, S. A. (2009): Children's understanding of second-order mental states. *Psychological Bulletin* 135.5, 749–773. <https://doi.org/10.1037/a0016854>
- Nakassis, C. & Snedeker, J. (2002): Beyond sarcasm: Intonation and context as relational cues in children's recognition of irony. In: Skarabela, B., Fish, S. & Do, A. H.-J. (eds.): *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 429–440.
- Perner, J. & Wimmer, H. (1985): “John thinks that Mary thinks that...”: Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology* 39.3, 437–471. [https://doi.org/10.1016/0022-0965\(85\)90051-7](https://doi.org/10.1016/0022-0965(85)90051-7)
- Schnell, Zs. (2012): The development of humour competence in Hungarian children – a cognitive approach. In: Litovkina, A. T., Brzozowska, D., Sollosy, J. & Medgyes, P. (eds.): *Hungarian Humour*. Kraków: Tertium, 235–251.
- Schnell, Zs. (2016): *Az elme nyelve. Társalgás és nyelvfejlődés*. Budapest: Akadémiai Kiadó.
- Schnell, Zs. (2021): Az irónia és a metafora kognitív fejlődési vonatkozásai: Kísérleti pragmatikai vizsgálódások. *Argumentum* 17, 511–531. <https://doi.org/10.34103/ARGUMENTUM/2021/26>

- Schnell, Zs. & Varga, E. (2012): Humour, irony, and social cognition. In: Litovkina, A. T., Brzozowska, D., Sollosy, J. & Medgyes, P. (eds.): *Hungarian Humour*. Kraków: Tertium, 253–270.
- Sperber, D. & Wilson, D. (1981): Irony and the use-mention distinction. In: Cole, P. (ed.): *Radical Pragmatics*. New York: Academic Press, 295–318.
- Sperber, D. & Wilson, D. (1995): *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Szücs, M. & Babarczy, A. (2017): The role of Theory of Mind, grammatical competence and metapragmatic awareness in irony comprehension. In: Assimakopoulos, S. (ed.): *Pragmatics at its Interfaces*. Berlin & Boston: De Gruyter Mouton, 129–148.
<https://doi.org/10.1515/9781501505089-008>
- Winner, E. & Leekam, S. (1991): Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology* 9.2, 257–270.
<https://doi.org/10.1111/j.2044-835X.1991.tb00875.x>