LANGUAGE ACQUISITION
2026, VOL. 33, NO. 2, 236–257
https://doi.org/10.1080/10489223.2025.2479497

Routledge
Taylor & Francis Group

Production and comprehension of rhetorical questions in the acquisition of Italian as a first language

Maria Francesca Ferin [a] and Tanja Kupisch [a,b]

[a]University of Konstanz; [b]University of Lund

ABSTRACT

This paper investigates rhetorical questions (RhQs) in monolingual children acquiring Italian, comparing the comprehension and production of RhQs based on their optional lexical, morphosyntactic, and prosodic cues. We show that 6-to-9-year-old children are in the process of acquiring the use of optional lexical and morphosyntactic markers in production (e.g., discourse particles, information structure), but only 9-year-olds show adult-like patterns. In comprehension, children from age 7 were able to consistently recognize RhQs through a combination of prosodic and lexico-syntactic cues, but only some children were able to do so based on prosody alone. Furthermore, the production frequency of lexico-syntactic markers correlated with accuracy in the comprehension task when a lexico-syntactic marker was involved. Our results suggest that the acquisition of RhQs and their pragmatic marking is still ongoing throughout the primary school years, in line with other phenomena at the interface between different modules of language, pragmatics, and communicative competence.

ARTICLE HISTORY
Received 19 August 2024
Accepted 03 March 2025

1. Introduction

This paper investigates the acquisition of rhetorical questions in Italian. Rhetorical questions (RhQs) are a type of noncanonical question, because, unlike canonical ("normal") questions, RhQs are not uttered to request information (Dayal 2016). A person asking a question like (1) does not expect an answer, because the answer is already in the common ground of speaker and addressee—no one likes parking tickets (Banuazizi & Creswell 1999, Han 2002, Caponigro & Sprouse 2007, Biezma & Rawlins 2017).

(1) Who likes parking tickets?!

In order to interpret a RhQ, we need to understand the literal meaning (the interrogative semantics) and, in addition, interpret the intended meaning (the pragmatic interpretation: the answer is considered obvious by the speaker). In this respect, RhQs bear similarities to ironic and sarcastic statements. However, understanding and producing RhQs involves more than the complicated process of decoding and using literal and nonliteral language at the same time, as there is a multitude of linguistic cues[1] speakers make use of when conveying the rhetorical meaning. While the example in (1) illustrates a RhQ that may, at first sight, look like a "normal", information-seeking question, it is likely to exhibit phonological cues that signal rhetoricity, such as special intonation. In

CONTACT Maria Francesca Ferin maria.ferin@uni-konstanz.de University of Siena, DISPOC Via Roma, 56, Siena 53100, Italy

[1]In our paper, we will use the term (linguistic) "cue" to refer to linguistic marking in general (including prosodic ones), while we reserve the term "marker" to refer to lexico-syntactic cues in particular.

addition, there may be lexical or morphosyntactic cues to rhetoricity. For instance, in English, a RhQ may be accompanied by the expression *the hell*, as in (2a), while Italian makes frequent use of *ma* 'but' at the beginning of a RhQ. Such markers are not unique to RhQs, however. They can also appear in other contexts and not necessarily signal rhetoricity. For example, Italian *ma* can simply be used as the adversative conjunction 'but', and even when used in noncanonical questions it can indicate a speaker's bias without making the question rhetorical. Thus, understanding and producing RhQs also involves the identification of the contexts in which such markers signal rhetoricity.

(2) a. Who the hell likes parking tickets?
b. Ma a chi piacciono le multe?
but to who like-3PL the parking_tickets

When do children start to produce and understand RhQs? There is a dearth of research on the acquisition of RhQs (see Recchia et al. 2010, Geiss et al. 2023, Ferin et al. 2024). However, research on related phenomena allows us to formulate reasonable expectations. RhQs can be seen as part of a larger set of interface phenomena whose use and recognition depend on (para)linguistic cues, such as emotion, sarcasm, and irony recognition, which, in turn, are mediated through the understanding of contextual information (Ackerman 1982), theory of mind (Banasik 2013), and intonation (Capelli et al. 1990, Glenwright et al. 2014). These are all "late acquired". For example, children start to comprehend irony and sarcasm around the age of 6 years (Banasik 2013). Before this age, they tend to interpret utterances literally (Ackerman 1982, Banasik 2013). Glenwright & Pexman (2010) showed some interesting developments between five and nine years of age. In their study, five- to six-year-olds were beginning to understand the non-literal meanings of sarcastic and ironic speakers, but they did not distinguish ironic and sarcastic speakers' intentions. Nine- to ten-year-olds were more accurate at understanding sarcastic and ironic speakers and they distinguished these speakers' intentions, rating sarcastic criticisms as more 'mean' than ironic criticisms. These results show that children can determine the non-literal meanings of sarcasm and irony by six years of age but do not distinguish the pragmatic purposes of these speech acts until later in middle childhood. In the experimental set-up, children were shown prerecorded contextualized remarks, and they had to judge how 'mean' the remarks were, compared to literal remarks. In the case of ironic and sarcastic remarks, this task requires inverting the literal meaning of the remarks. The ability to do so was also reported by Chevallier et al. (2011), Pexman et al. (2011), and Colich et al. (2012) for children aged 8-11 years.

Prosodic cues have a prominent role in disambiguating RhQs from information-seeking questions (ISQs). Infants are sensitive to prosodic cues from a very early age (e.g., Frota et al. 2014, Miller & Eimas 1979), and there is vast evidence that the discrimination of emotional expression based on prosody and/or facial expression emerges during the first year in children's lives (e.g., Walker-Andrews & Grolnick 1983, Fernald 1993, Mumme et al. 1996, Grossmann et al. 2005). Nevertheless, we do not know at what age categories such as irony, criticism, and sarcasm are mapped onto the relevant prosodic forms. Some studies investigated children's ability to infer emotions when the literal meaning and paralanguage do not map. At 4 years, children typically judge a speaker's emotions based on *what* they said rather than *how* they said it (Friend 2000, Morton & Trehub 2001). Morton & Trehub (2001) show that 4-year-olds have a sense of attitudinal prosodic cues, but they need time to learn how to weigh these cues over lexical or situational cues. Between 7 and 9, children show more sensitivity to vocal cues, and at age 10 they can infer the speaker's emotion primarily through prosodic cues (Friend 2000, Morton & Trehub 2001). Bascandziev et al. (2025) investigated preschool children's ability to infer from prosodic cues whether a question is information-seeking or asked with pedagogical intent, finding that they begin doing so from age 5 or 6. Overall, (para)linguistic cues are crucial to emotion, irony, and speaker intent recognition, and we expect them to play a role in the acquisition of RhQs as well.

In summary, the learning task for the acquisition of RhQs involves at least the following components: (i) understanding that a question may be used for different things than

requesting information, (ii) understanding that there may be cues that signal the rhetoricity of a question, and (iii) learning to interpret and use these cues in the given context. We reckon that the acquisition of these components is still ongoing when children have reached primary school age. Our study will therefore focus on primary school children and contribute to answering at what age children master RhQs (and/or their ingredients) based on data from Italian-learning children. Our study will add to the understanding of late acquired phenomena at complex interfaces, which are currently still understudied.

2. Rhetorical questions in Italian

As mentioned in the introduction, RhQs look like questions but they feel like assertions. They can be used for a variety of different functions, including sarcastic or reinforcing comments on a situation (Neitsch 2019, Sorianello 2019). RhQs also vary in terms of whether the intended answer is of opposite polarity (*What has John ever done for Mary?! [Obviously nothing]*) or of the same polarity (*Who always helped Mary when she was in trouble? [Obviously, John]*) (from Han 2002). Herein, we will only be concerned with the former, so called negative-answer RhQs, because we had to restrict the domain of investigation and because more theoretical research has been carried out on this type of RhQs.

In Italian, RhQs can be string identical to ISQs, having the exact same constituents, and the only way to distinguish them is by means of a context or world-knowledge. This is illustrated in (3), which may be a genuine request for information in a context of offering but be interpreted rhetorically when addressed to a group of friends who share the belief that drinking herbal tea is absurd.

(3) Chi beve la tisana?/?!
Who drinks the herbal_tea
'Who drinks herbal tea?/?!'

In addition, there are linguistic cues to RhQs. Based on Ferin (2024a, 2024b), we will illustrate different ways of marking RhQs in Italian.

Most lexico-syntactic markers in Italian RhQs are indirect ones, which means that the rhetorical interpretation is derived through a combination of cues and context, as illustrated in (4). The first cue is the conjunction *ma*, which we had already introduced in (2b). *Ma* is a sentence-initial discourse particle that adds a counter-expectational value to the utterance (Giorgi 2018, Ippolito 2021), roughly signaling to the interlocutor that they are being challenged. *Ma* is compatible with other types of non-canonical questions beyond RhQs, such as surprise questions and extreme ignorance questions (Giorgi 2018, Hinterhölzl & Munaro 2021). The conjunction *e* 'and' can be used in a similar position and also in RhQs, but it is less frequent.

(4) Ma chi mangerebbe mai le banane?
but who eat-COND.3SG ever the bananas
'Who would ever eat bananas?'

Another element that may be found in RhQs is the adverb *mai* 'ever', also present in (4). In its base function, *mai* is the temporal adverb 'ever/never', while in a RhQ it adds the meaning that the answer to the question is impossible to be found (Coniglio 2008, Manzini 2015, Obenauer & Poletto 2000, Hinterhölzl & Munaro 2021). Thus, it is compatible both with a rhetorical reading ('Obviously, no one eats bananas') and with an extreme ignorance reading ('I really can't think of anyone who would eat bananas'). Another typical marker of RhQs, often used in combination with *mai*, also present in (4), is the conditional verb morphology (Obenauer & Poletto 2000, Dikken & Giannakidou 2002), which presents the event as something hypothetical. Finally, another frequently found element in RhQs are aggressive expressions, as, for example, *chi cavolo* corresponding to English 'who the hell'. Aggressive expressions presuppose a negative attitude (Dikken &

Giannakidou 2002) and they intensify the speaker's attitude (Beltrama & Trotzke 2019, Celle et al. 2021). They are not unique to RhQs but can also occur with surprise and surprise-disapproval questions.

(5) Chi cavolo mangia le banane?!
who the_hell eats the bananas
'Who the hell eats bananas?!'

Additional markers, not present in (4) and (5), are illustrated in (6) through (9). All of these go along with specific syntactic configurations. One is the affective reflexive *si*, as in (6). Italian *si* is used in true reflexive contexts or with verbs that are inherently reflexive, but in RhQs the reflexive is used with affective or intensifying value (see Cordin 1995), thus conveying intensification of an emotive attitude, although not as strongly as *wh-the-hell* expressions. This marker has sometimes been described as being regionally biased, because it appears to be used more frequently in Southern varieties of Italian (Ferin 2024b).

(6) Ma chi si mangia i lime?
but who REFL eats the lime?
'But who eats lime?!'

The example in (7) illustrates that RhQs can also be embedded under the verb *volere* 'want' with the wh-word being extracted from the embedded clause. The verb *volere* loses its lexical meaning and becomes a fixed expression indicating that the embedded question is rhetorical (Obenauer 2004). Among the markers investigated here, this is the only unambiguously rhetorical one, which means that it can never have a non-rhetorical interpretation.

(7) Chi vuoi che beva la tisana?!
who want-2SG that drink-SBJV the herbal_tea?
'Who (do you think) drinks herbal tea!'

Other cues that appear together with RhQs are cleft sentences, as in (8), where the clefted element is *chi* 'who'. Cleft constructions are commonly associated with a contrastive or corrective focus and are presuppositional in nature (see Belletti 2012 for Italian); the clefted constituent is focalized, and the subordinate clause is presupposed. RhQs can also appear with right dislocation (RD), as in (9), where a phrase is dislocated to the right edge of the sentence and resumed by a clitic pronoun. The dislocated element in (9) is the object DP *i romanzi* 'the novels', which is resumed by an object clitic, *li*. As previously outlined, RhQs suggest that the answer to the question is already present or inferable from the common ground, and the use of RD is likely used as a pointer to this fact (see Berruto 1986, Crocco 2013 for a similar interpretation). Notice that both cleft questions and RD are very indirect or weak cues to rhetoricity, as they are also compatible with an unbiased information-seeking interpretation.

(8) E chi è che legge i romanzi?!
and who is that reads the novels
'(And) who reads novels?!'

(9) Chi li legge, i romanzi?!
who CL reads the novels
'Who reads novels?!'

The overview shows that in RhQ marking there is no one-to-one correspondence between form and interpretation. The various elements that form a RhQ may come from other linguistic means, where each element adds a pragmatic nuance, but all jointly leading to a rhetorical interpretation. What is

more, all markers we have introduced are optional, which makes their acquisition even more challenging.

In addition to optional syntactic markers, RhQs often differ prosodically from ISQs through a variety of phonological and phonetic cues (Ferin et al. forthcoming, Sorianello 2018). More specifically, RhQs are typically longer in duration than ISQs, and they can be characterized by a smaller pitch excursion. A difference is found also in the final part of the utterance, which more often consists of a high or rising terminus in wh-ISQs (H% or LH%)[2] and is almost exclusively low in wh-RhQs (L%) (Ferin et al. forthcoming, Sorianello 2018). The nuclear pitch accent (NPA) offers a more complex picture: while differences are found in the position and shape of the NPA, the contrast was not categorical. In particular, Ferin et al. (forthcoming) found that an early falling accent (H + L*) on the verb was present in both types of questions, but is more frequent in ISQs; a later-aligned fall (H* + L) was also shared, but it is more frequent in RhQs; finally, a nuclear rise (L + H*) is strongly associated with RhQs but is also possible in ISQs. Ferin (2024a) showed that Italian speakers of diverse regional backgrounds are able to categorize questions as information-seeking or rhetorical based solely on a combination of prosodic cues: RhQs were characterized by longer duration and low final contour, while ISQs had shorter duration and a rising final contour.

Our contribution addresses the question when children acquire the relevant lexico-syntactic and prosodic cues to RhQs and how to combine them, and when they learn to interpret and use them in the given pragmatic contexts.

3. Research questions and hypotheses

Our paper addresses the following questions:

Research Question 1 (RQ1): Can primary school age children (aged 6-10 years) produce and understand RhQ marking in Italian?

Research Question 2 (RQ2): Do production and comprehension of cues develop during this period?

Research Question 3 (RQ3): Does the ability to produce RhQ markers imply the ability to comprehend them?

Given the absence of previous research on this topic, it is unclear whether children will produce any kind of marking, and, in case they do, of what type it will be. Since cues are optional, there are three alternative scenarios. The first is that children have not acquired the relevant markers and thus do not use them in production. The second is that children have knowledge of (part of) the markers, but do not use them in production, or do not use them consistently in an adult-like manner. The third possibility is that children know the markers and use them at adult-like rates. In terms of comprehension, we expect that at least some, but not all children can discriminate RhQs and ISQs based on their prosodic form alone. Given that in natural language RhQs are often signaled by a combination of prosodic and lexical-syntactic cues, we expect that the combination of the two types of cues will be more effective than prosody alone, and thus more easily interpretable for children.

With respect to the second research question concerning development, we expect that children may have started acquiring the relevant cues at age 6, with an improvement throughout primary school. This expectation is based on the observation that children start developing non-literal language and

[2]For the annotation of intonation, we follow the conventions of the Tones and Break Indices system (ToBI) for Italian (Gili Fivela et al. 2015). An H represents a high tonal target and an L a low tonal target; the symbol % represents the terminus of the prosodic unit. Thus, the label 'LH%' indicates a rise from a low to a high F0 frequency at the end of the question. The symbol * represents the nuclear pitch accent. For example, the label H*+L indicates an F0 movement from high to low anchored to the metrically strongest syllable in the utterance.

complex interface phenomena at around age 6. However, a high rate of individual variation is expected, as is typical for complex interface phenomena.

As to the relation between comprehension and production (RQ3), it is natural to conjecture that children who are able to produce a certain cue (e.g., *ma*) will also comprehend its function. However, the reverse does not hold: a child might know the pragmatic function of *ma* without producing it, since it is optional in RhQs. Therefore, we expect that, if children produce *ma*, they will also be able to understand it, but not necessarily vice versa. Comprehension is further supported by prosody, but prosodic and lexical-syntactic cues may not be related: children may be able to produce and comprehend RhQs based on only a subset of the cues available in their language. For example, they may be fully aware of prosodic cues, yet without knowing any of the lexico-syntactic ones, or vice versa.

4. Methods

Data was collected online during the Covid-19 pandemic, via Zoom© videocalls with the participants. Children were tested with a variety of tasks, tapping into RhQs.[3] The production task investigated children's ability to mark RhQs, focusing on optional lexico-syntactic cues.[4] The comprehension task aimed at investigating whether children were able to discriminate between different types of questions based on lexico-syntactic cues and/or prosodic cues. The comprehension task was preceded by a perception (discrimination) task, to ensure that children were sensitive to the prosodic difference between the test items used in comprehension.

The tasks were presented as games within the fictional setting of a detective school to keep the children motivated and engaged throughout the data collection process. The experimenter shared a PowerPoint presentation through the screen share function in Zoom, including the computer audio. The production task was done on a different (preceding) day than the perception and comprehension task.

Adult controls were tested on the same tasks as children, but without the detective setting. The procedure for the production task remained the same and the task was carried out during a Zoom call with the experimenter, while the perception and the comprehension tasks were completed autonomously by the adult participants in an online questionnaire on the web platform SoSci Survey (www.soscisurvey.com, Leiner 2019), as explained in the following section. All materials summarized in the following are available in the OSF repository: https://osf.io/8kurf/.

4.1. Participants

Eighty-four monolingual[5] Italian children aged 6-10 years participated in the study. We controlled for their origin in terms of North vs. Centre/South (Centre/South were collapsed because they are both south of the major isogloss La Spezia-Rimini dividing Italy). 42 Italian monolingual adults served as controls. Table 1 reports information on age, gender and origin of each participant group. Participants were recruited through personal contacts and social media. Informed consent was obtained from participants themselves (adults) or from the parents (children). After the last session, adults were compensated with a 10€ Amazon voucher and children with an 8€ bookstore chain voucher.

[3]The full battery of tasks included also some potentially related skills (irony comprehension, theory of mind, language proficiency), which were also tested in connection with bilingual acquisition (see Ferin et al. 2025 for further details; the additional tasks are not addressed here).

[4]The data from the production task can also be used for a prosodic analysis of the children's productions. This cannot be done within the scope of this paper, however.

[5]When referring to the children as 'monolingual Italian' we do not consider potential knowledge of other Italo-romance varieties and exposure to L2 English at primary school.

Table 1. Participant information.

	N	Age (M, SD)	Gender (F/M)	Origin (N/C, S)
6-year-olds	24	6.5 (.3)	14/10	14/10
7-year-olds	21	7.6 (.2)	9/12	11/10
8-year-olds	20	8.4 (.3)	15/5	14/6
9-year-olds	19	9.5 (.3)	7/12	7/12
Adults	42	26.0 (4)	26/16	10/32

4.2. Structure of the experimental items

The structure of the experimental items was the same in all three experiments with wh-questions of the structure *Who* + *verb* + *DP*, as in 'Who eats melon?'. All questions contained the wh-word *chi* 'who', the syntactic subject, followed by one out of four transitive verbs and one out of six nouns, preceded by a definite article (the resulting DP was always the direct object). The nouns included were all trisyllabic with stress on the penultimate syllable (e.g., *vioLIno* 'violin'), and they were all part of child vocabulary. The nouns were selected so that they did not favor or disfavor an RhQ interpretation. The form of the prosodic and syntactic conditions was chosen after careful pretesting with Italian monolingual adults (Ferin 2024a). Table 2 summarizes the conditions.

RhQs were characterized by a longer duration ($M = 1.6$ s, $SD = 0.1$), a smaller pitch excursion ($M = 7.9$ semitones [st], $SD = .5$), and a low boundary tone (L%). ISQs, by contrast, had a rising boundary tone (LH%), a greater pitch excursion ($M = 11.5$ st, $SD = 0.9$) and a shorter duration ($M = 1.3$ s, $SD = 0.1$).

In the perception and comprehension task, sentences were also manipulated for lexico-syntactic cues, with three conditions. In the *neutral* condition, the wh-question was presented without any modification (10a); in the *ambiguous* condition, the object was right-dislocated (10b). We refer to the latter condition as "ambiguous" because RD can also occur with ISQs. The rhetorical condition cued rhetoricity more strongly by the combination of *ma* 'but' and RD (10c). Neutral and ambiguous sentences were crossed with both prosodic conditions. Rhetorical sentences were only presented with RhQ prosody, as previously described.

(10) a. Chi vuole il budino? (neutral: no cue)
who wants the pudding
b. Chi lo vuole, il budino? (ambiguous: RD)
who CL wants the pudding
c. Ma chi lo vuole, il budino? (rhetorical: *ma*, RD)
but who CL wants the pudding
'Who wants pudding?/?!'

The sentences were recorded by the first author, a female native speaker of Italian. Each stimulus was recorded twice. The files were resynthesized using Praat (Boersma & Weenink 2015), adapting their duration and smoothing the pitch contour. The duration of each pair of questions was artificially averaged, eliminating potential small differences in duration between the two sentences of the pair. Additionally, the files were stylized with a resolution of 2.0 st (i.e., the F0 track was simplified to represent only the main turning points in the contour). The filler items (polar RhQs and ISQs of the form *Qualcuno mangia i piselli?/?!* 'Does anyone eat green peas?/?!') were created in the same way.

Table 2. Overview of prosodic and syntactic cues used in the perception and comprehension task.

	ISQs	RhQs
Prosodic cues	• Shorter duration • Greater pitch excursion • Rising boundary tone (LH%)	• Longer duration • Smaller pitch excursion • Low boundary tone (L%)
Syntactic cues	• Neutral (no cue) • Ambiguous (RD)	• Neutral (no cue) • Ambiguous (RD) • Rhetorical (*ma* + RD)

Abbreviations: ISQs = information-seeking questions; RhQs = rhetorical questions; RD = right dislocation; *ma* = Italian 'but'.

4.3 Statistical analyses

The statistical analyses were conducted in R (R Core Team 2023). We fitted linear mixed effects regression models or logistic mixed effects regression models with the functions *lmer* and *glmer* in the R package *lmerTest* (Kuznetsova et al. 2017). For each model, we first fitted a maximal model and then removed any non-significant interactions (first) and fixed effects (second) following a stepwise selection procedure: we first removed one non-significant interaction or fixed effect and then compared the AIC of the simpler model to the more complex model, to ensure that the simpler model was a better fit for the data.[6] Anovas and *p*-values were obtained with the *Anova* function in the *car* package. Post-hoc pairwise comparisons with Tukey correction of *p*-values were obtained with the *emmeans* and *emtrends* functions in the package *emmeans* (Lenth 2023). Categorical independent variables (IVs) were coded as factors and received treatment coding. Continuous IVs were scaled to improve model fit. Participant and Item were added as random effects; random slopes were included whenever possible. The full model selection procedure for each model and a complete report of their effects can be found in the Online Supplementary Materials at https://osf.io/8kurf/.

5. Production task

5.1 Procedure

To elicit RhQs and ISQs, we used four fictitious characters, two curious and friendly ones (Olaf and Rapunzel) and two grumpy and unkind ones (Grumpy and Drizella). The former were used to elicit ISQ and the latter to elicit RhQs. Each character was introduced with a context and illustrations in PowerPoint that highlighted the characters' attitudes. The two friendly characters asked curious questions (ISQs) to their friends; the grumpy characters used RhQs to reply unkindly to another character. With each character, the child heard an example of a typical ISQ or RhQ (model question) with an intonation characteristic of the respective question type. Each sentence was recorded by a different native speaker of Italian. Table 3 provides a summary of the four contexts and model questions. The model questions for RhQs were either presented as 'unmodified', presenting the minimal question form of wh + verb + object, or with some additional marking: the sentence-initial particle *ma* and CLRD.[7]

After the introduction of the character and the model question, the experimenter asked the participant to impersonate the character, asking questions like the one they had just heard. The experimenter showed the image of an object, the target of the question (e.g., a banana), using a prompt of the type: 'Now Rapunzel would like to know who eats bananas, so she asks . . . '. The participant was thus prompted to produce a question of the form 'Who eats bananas?'. Eight questions per character were elicited. The first two served as practice items, which left 12 target ISQs and RhQs, respectively. Participants were free to add particles or other syntactic features to the base form of the question (wh + verb + object). The order of the characters was counterbalanced across children.

The participants' productions were transcribed and tagged for any optional markers, that is, any lexical addition or syntactic modification to the base form of the question. For example, the question in (11) would have received a tag for 'ma' and one for 'conditional'. Questions were modified by no cue, one cue or more than one cue.

(11) Ma chi mangerebbe le banane?
but who eat-COND.3SG the bananas
'Who would eat bananas?!'

[6] If more than one interaction or predictor was available for removal at a certain stage of the selection procedure, we used the function *drop1* to determine which removal best improved the model.

[7] Since we had no specific predictions on the age of acquisition of the cues under analysis, we speculated that if the children would not use any additional marking spontaneously, they would at least correctly interpret and re-use the cues of the model question.

Table 3. Character introductions and model questions in the production task.

Character	Summary of the context	Model question
Olaf	Olaf is a curious snowman. One day he finds some objects lying around. For example, a book of fairy tales. He asks his friends who reads fairytales …	Chi legge le favole? who reads the fairytales 'Who reads fairytales?'
Rapunzel	Rapunzel is very friendly. She wants to find out what her new friends like, to get to know them better. For example, she asks …	Chi **li** mette, gli stivali? who CL put-on the boots 'Who wears boots?'
Grumpy	Grumpy is always unfriendly with Snow White and replies unkindly to her. For example, Snow White asks if he eats zucchini, but he replies …	Chi mangia le zucchine?! who eats the zucchini 'Who eats zucchini?!'
Drizella	Drizella is always unfriendly with Cinderella and replies unkindly to her. For example, Cinderella asks if she eats melon, but Drizella replies …	**Ma** chi **lo** mangia, il melone? !but who CL eats the melon 'Who eats melon?!'

5.2. Results

Practice items and non-target sentences (e.g., polar questions instead of wh-questions) were removed from the dataset. The final dataset contained a total of 2,009 target questions for children and 1,001 for adults.

We first looked at the rate of modification for each group of participants, calculating the average number of cues modifying questions. We fitted a linear mixed-effects regression model with Number of cues (i.e., how many cues were present in a question) as dependent variable; Condition (2 levels: RhQ vs ISQ) and AgeGroup (5 levels: 6 vs. 7 vs. 8 vs. 9 vs. adults) were added as independent variables, as well as an interaction term between the two predictors. Participant and Item were included as random intercepts. The model yielded a significant main effect of Condition (χ^2 (1) = 34.58, $p < .001$), revealing that all groups (6-, 7-, 8-, 9-year-olds, and adults) used more optional marking in RhQs (M = 0.88) than in ISQs (M = 0.17), as shown in Figure 1.

The interaction between AgeGroup and Condition was also significant (χ^2 (1) = 113.04, $p < .001$), while AgeGroup was not significant on its own (χ^2 (1) = 4.18, $p = .382$). The interaction, as observed in Figure 1, indicates that, while there was no predicted difference between groups in ISQs, there was an age effect within RhQs. In particular, 6- and 8-year-olds used fewer additional cues than 9-year-olds[8] (6 vs. 9: $\beta = -0.52$, $p < .001$; 8 vs. 9: $\beta = -0.56$, $p < .001$) and adults (6 vs. Adults: $\beta = -0.58$, $p < .001$; 8 vs. Adults: $\beta = -0.62$, $p < .001$). 7-year-olds were in between: they used fewer cues than adults, but the trend did not reach significance ($\beta = -0.3$, $SE = 0.11$, $t = -2.62$, $p = 0.06$). All other comparisons were not significant. In particular, 9-year-olds were not significantly different from adults.

In a second step, we analyzed the types of cues produced within each group. In Figure 2, each cue is associated with a percentage, indicating how many questions in percentage were modified by that cue. For both children and adults, the adversative particle *ma* was the most frequent. Adults and 9-year-olds produced it with the highest frequency (49% each), while the rate was lower for the younger groups (6: 29%; 7: 33%; 8: 20%). The next cue in terms of frequency was RD, which also presents similar percentages for adults (35%) and 9-year-olds (32%) and lower frequencies for 6- (19%), 7- (22%), and 8-year-olds (12%). Cleft questions were also comparatively frequent in RhQs (7-11%), except amongst 6-year-olds (1%), but this structure also occurred in ISQs at only slightly lower rates (5-7%); thus, differently from the other cues, it did not clearly differentiate RhQs from ISQs. The remaining cues occurred with a lower frequency in all groups (1-9%), indicating that only a few participants used them in RhQs. These were the sentence-initial particle *e* 'and', the adverbial particle

[8]We only report beta values and *p*-values here; the full statistics and additional information on the models can be found in the Online Supplementary Materials available from OSF (see Footnote 4).

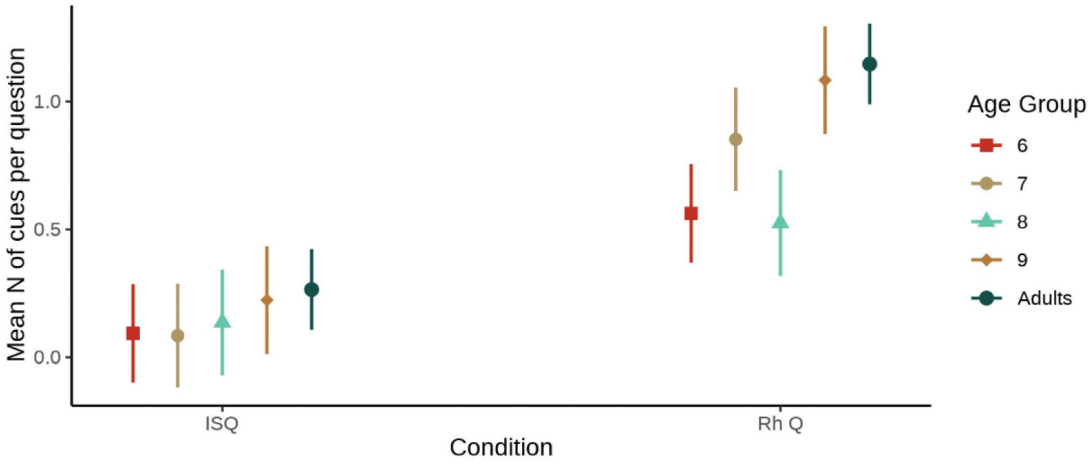

Figure 1. Model estimates of the mean number of cues modifying information seeking questions (ISQs) and rhetorical questions (RhQs) in the production task, by age group.

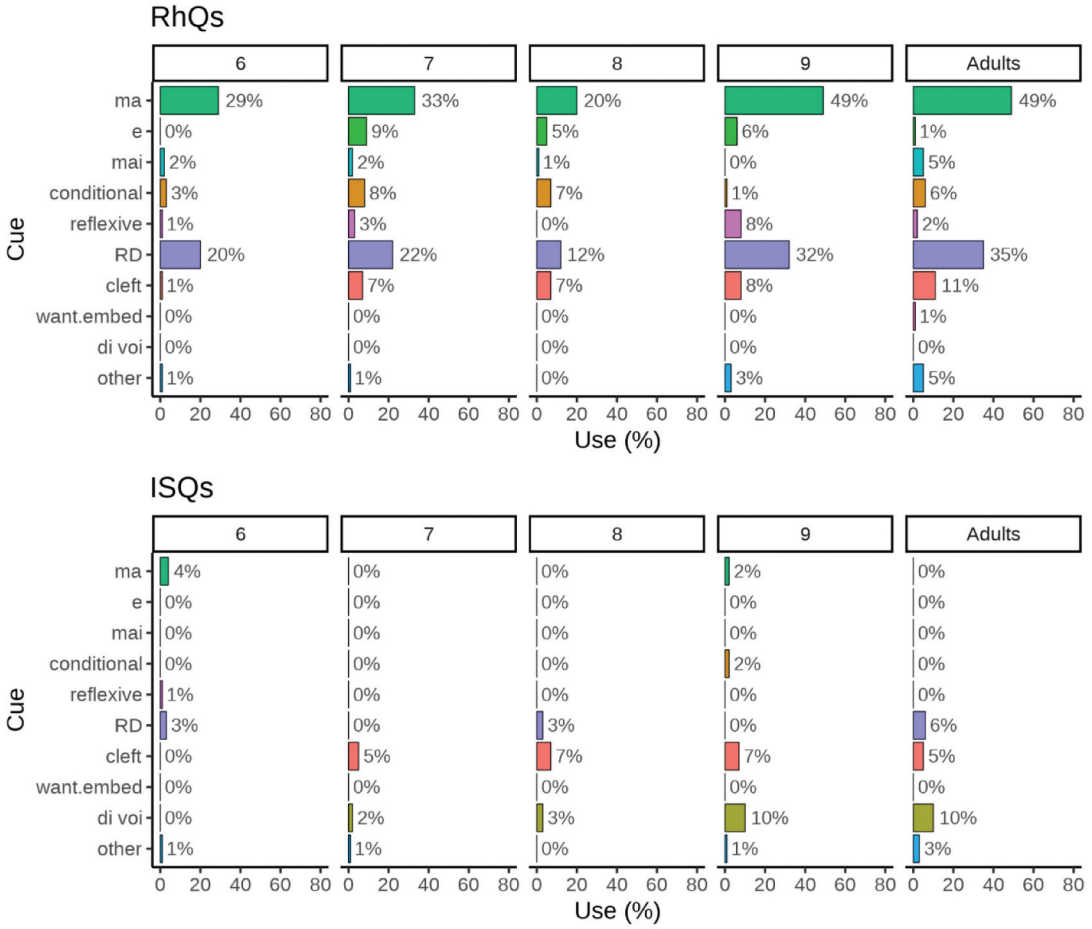

Figure 2. Types of cues used to modify information-seeking questions (ISQs) and rhetorical questions (RhQs) in the production task, by age group. The percentage indicates how many questions in proportion were modified by that cue; a question may be modified by more than one cue.

mai 'ever', conditional morphology on the verb and the reflexive clitic *si* with affective value. Only adults used 'want'-embedding (in fact, only one adult participant made use of this cue). All cues except for cleft questions were either absent in ISQs, or present in negligible percentages compared to RhQs (Figure 2). ISQs sometimes presented a wh-constituent modified by *di voi* 'of you': instead of asking 'Who eats bananas?', participants asked 'Who of you eats bananas?'. This happened particularly in adults and 9-year-olds (10% each) (see Dehé et al. 2022 for a similar finding in German ISQs).

5.2.1. Effect of the model question

The results shown in Figure 2 and previously described include all RhQs in the dataset. However, recall that model question (iv), which participants heard before producing one block of RhQs, was modified by *ma* and RD (Table 3). This may have influenced the results in two ways: directly, by prompting the use of these two cues and causing their high frequency; indirectly, by reminding participants that RhQs can be overtly marked by optional markers, thus boosting the use of lexico-syntactic markers overall (as opposed to focusing on prosodic marking alone). To control for the potential priming effect of the model question, we examined only the first block of RhQs for each participant. Since the order of characters was counterbalanced, some participants heard the unmodified question first (Grumpy), and some heard the modified question first (Drizella). The results for this subset of data are shown in Figure 3. All groups used fewer additional cues in the unprompted condition than in the prompted condition ($\beta = -0.82$, $SE = 0.08$, $t = -9.85$, $p < .001$), but additional marking was not entirely absent in the unprompted condition.

The frequencies of the various cues also differed across age groups. In the prompted condition, 6-year-olds used only *ma* and RD (62% and 46%, respectively) but no other markers; in the unprompted condition, *ma* was not used at all, and RD was at 3%, while they spontaneously produced other cues, such as *mai*, conditional, reflexive and clefts. 8-year-olds used a variety of cues when prompted, while in the unprompted condition they produced hardly any additional marking. 7-year-olds used almost all cues in both conditions, but there was a sharp difference in the use of *ma*, present in 70% of the RhQs in the prompted condition and only in 9% in the unprompted condition. 9-year-olds, similarly to 6-year-olds, used *ma* and RD frequently in the prompted condition, with the additional spontaneous use of *e* and reflexive. In the unprompted condition a greater variety of cues was found (all but *e*). *Ma* is still the most frequent (33%), though far less frequent than when prompted (74%). Finally, adults used all cues in both conditions (except for *e*). *Ma* was the most frequent in both conditions; RD was the second most frequent only in the prompted condition, while in the unprompted condition clefts were more prominent.

6. Perception and comprehension tasks

6.1 Perception task: Procedure

The aim of the perception task was to confirm that children were sensitive to the prosodic cues subsequently used in the comprehension experiment. Children listened to two string-identical sentence pairs, and they had to decide whether the sentences sounded *uguale* 'same' or *diverso* 'different' (Figure 1).

The task consisted of 28 sentence pairs, twenty in the *same* category (string-identical, same prosody), and eight in the *different* category (string-identical, different prosody). In the category *different*, half of the pairs started with an RhQ followed by an ISQ, the other half had the opposite order. To balance the number of same and different pairs, 12 polar questions were included in the category *different*. This resulted in a total of 40 items. The two sentences of each pair were presented with a 1,000 ms silence in between and the order of items was pseudo-randomized. The children listened to a sentence pair, played by the experimenter on their computer and audio-shared through Zoom, and had to decide whether the two sentences sounded same or different. The experimenter covertly recorded the answer by clicking on an online survey page implemented on SoSci survey

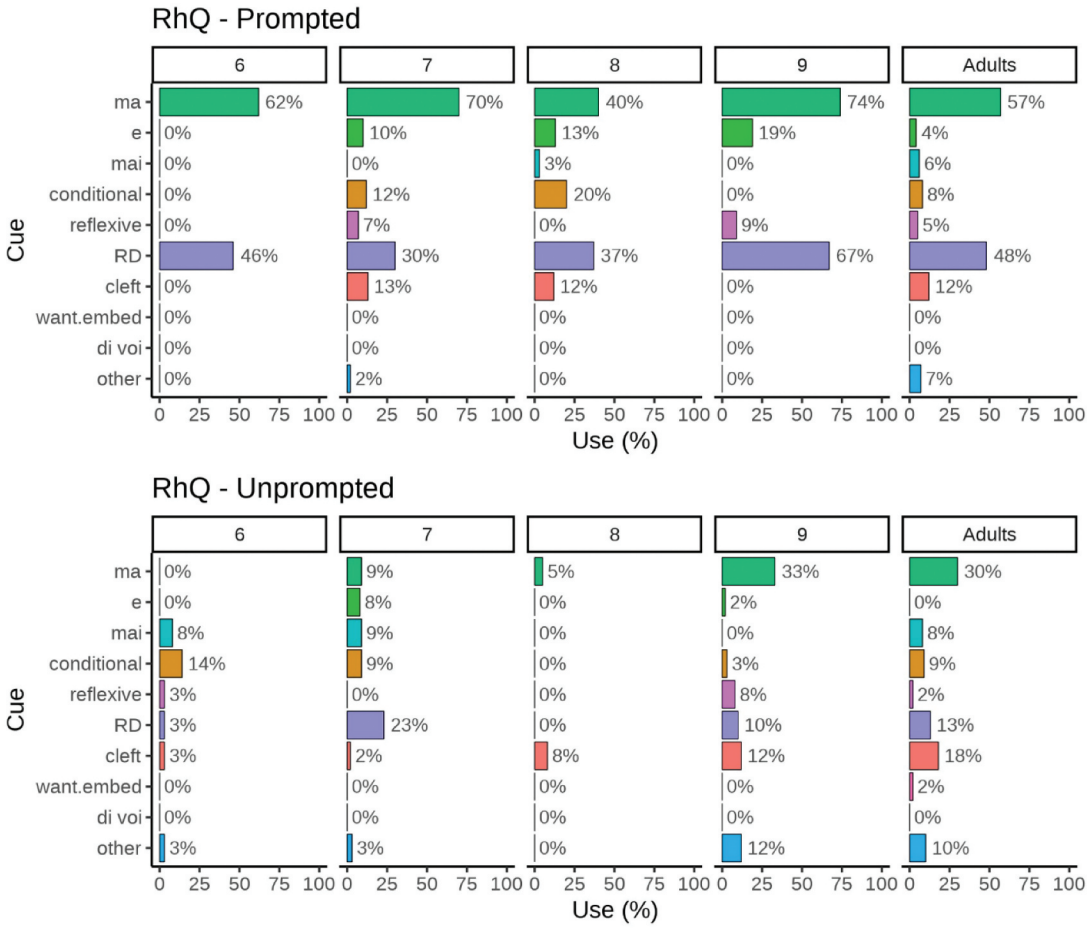

Figure 3. Types of cues used to modify prompted and unprompted rhetorical questions (RhQs) in the production task, by age group. The percentage indicates how many questions in proportion were modified by that cue; a question may be modified by more than one cue.

(Leiner 2019). The children had to pass a training session (6 items) before the actual task, making sure they paid attention to the prosodic difference.

The procedure for adults was partially different: they completed an online questionnaire on their own, implemented on SoSci Survey. Like the children, they listened to a sentence pair and saw the two cards depicted in Figure 4 on the screen; they had to decide whether the sentences were *same* or *different* by clicking on the corresponding card.

6.2. Perception task: Results

Two children who failed the training section (both six-year-olds) were removed from all further analyses. Figure 5 shows that accuracy in this task was overall very high. Adults performed almost at ceiling in both conditions ($M = 97\%$, $SD = 10$), while children had lower accuracy in the *different* ($M = 86\%$, $SD = 18$) than in the *same* condition ($M = 95\%$, $SD = 10$), although accuracy was still very high. Accuracy improved with age, with the 6-year-olds having the lowest (86%) and the 9-year-olds the highest (97%).

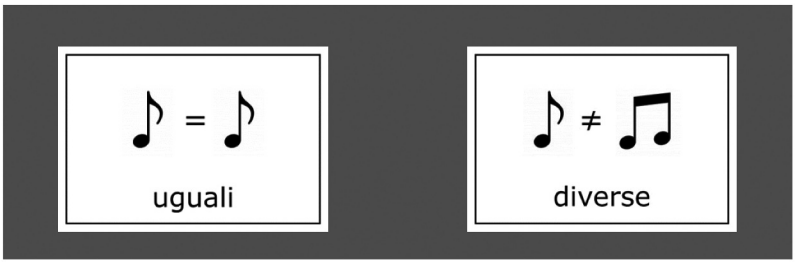

Figure 4. Screen setup for perception task.

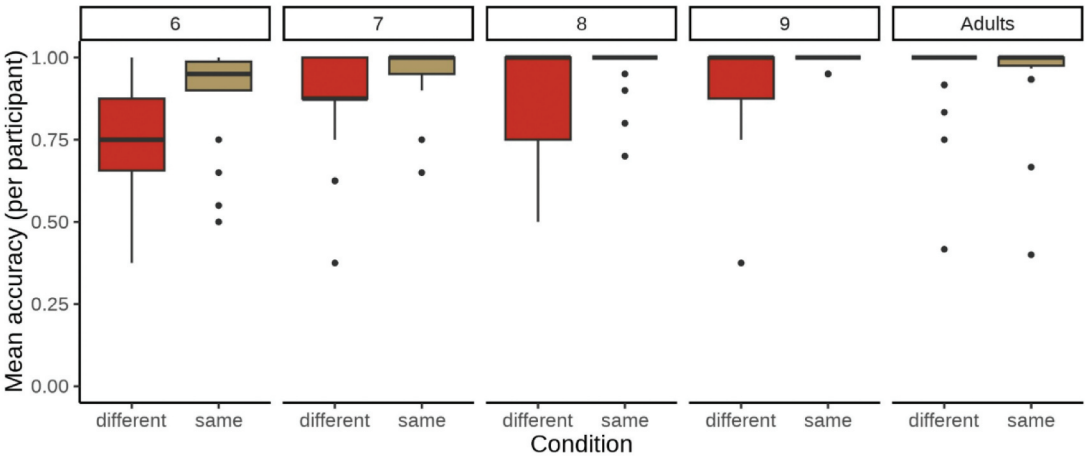

Figure 5. Boxplot showing the mean accuracy per participant in the perception task, divided by condition and age group.

To test these results statistically, we fitted a logistic mixed effects regression model with Accuracy as binary dependent variable (1 = correctly identified same or different pair). The model included Condition (2 levels: same vs. different) and AgeGroup (5 levels: 6 vs. 7 vs. 8 vs. 9 vs. adults) as independent variables, and an interaction term between the two. Participant and Item were included as random intercepts, and Condition was added as random slope to Participant. The interaction between AgeGroup and Condition was significant (χ^2 (4) = 10.02, $p = 0.04$). There was also a significant fixed effect of Condition (χ^2 (1) = 8.15, $p = 0.004$) and AgeGroup (χ^2 (4) = 28.65, $p < .001$). The analysis of contrasts with *emmeans* confirmed that all child groups were significantly more accurate in identifying the *same* than the *different* contrast pairs (6: $\beta = -1.51$, $p = 0.004$; 7: $\beta = -2.08$, $p = 0.004$; 8: $\beta = -2.41$, $p = 0.003$; 9: $\beta = -2.77$, $p = 0.004$), while adults were equally accurate in both conditions ($\beta = -0.2$, $p = 0.787$).

Although most children were highly accurate in the perception task, there was some variability within the groups, in particular for the condition *different* and within 6-year-olds. Since the ability to discriminate sentence types based on prosody was considered crucial for the performance in the comprehension task, we calculated the accuracy for condition *different* for each child, to use it as a predictor in the analysis of the comprehension task (henceforth, 'Perception-different score'). One adult scored low in both conditions and was removed from the analysis of the comprehension task[9]

[9]Completion metrics from the questionnaire revealed that this participant completed it unreasonably fast compared to the other participants.

Table 4. Number of items per condition used in the comprehension task.

	ISQ	RhQ
Neutral	6	6
Ambiguous	6	6
Rhetorical		6
Filler	10	4
Total	22	22

Abbrevations: ISQ = information-seeking question; RhQ = rhetorical question.

6.3. Comprehension task: Procedure

Comprehension was tested by means of a forced-choice task, following the perception task. The children saw a PowerPoint slide with two cards, each with one character that was familiar from the production task. As previously explained, Rapunzel was introduced as curious and friendly, asking questions to gather information about her friends (ISQs). Genoveffa 'Drizella', Cinderella's stepsister, was characterized as grumpy and unfriendly, replying to Cinderella's suggestions with unfriendly remarks (RhQs), such as "Who eats bananas?!". The RhQ remark was explicitly associated with a negative answer and its obviousness (e.g., "Who eats bananas?!" = "No one eats bananas!"). The comprehension task consisted of 30 experimental items and 14 fillers (Table 4). The fillers were used to balance the number of ISQ and RhQ test items. The questions were presented in a pseudo-randomized order without context. The experimenter played a question and asked the child who might have said that question, whether Rapunzel or Drizella, and covertly recorded the child's answer on their computer.

Adults completed the task as an online questionnaire, following the perception task. Adults were explicitly instructed on the difference between rhetorical and information seeking questions, which were associated with one emoji each. After hearing the question, they had to decide whether the question was rhetorical or information-seeking by clicking on the corresponding image.

6.4. Comprehension task: Results

Figure 6 summarizes the results of the comprehension task. Each datapoint indicates the accuracy of one participant, divided by group, age group, prosodic cue, and syntactic cue. Adults were at ceiling in both ISQs ($M = 98\%$, $SD = 7$) and RhQs ($M = 95\%$, $SD = 15$). In children, instead, RhQs ($M = 72\%$, $SD = 29$) were less accurate and had a higher inter-individual variation than ISQs ($M = 92\%$, $SD = 16$). Neutral and ambiguous sentences appear to have the same distribution, both within ISQs and within RhQs, while in the RhQ condition the 'rhetorical' sentences (with *ma* and RD) have the highest accuracy in all the child groups. There is, however, a difference between age groups: The 6-year-olds display substantial variance, with accuracy ranging from 0 to 1 (i.e., some children were 0% accurate, some were 100% accurate, others in between); accuracy in this condition improves almost to ceiling in the older age groups. For the neutral and the ambiguous cue, there is less improvement, and individual variation remains very high.

The inferential analysis was conducted in two separate steps. In a first model, we analysed the fully crossed design, that is, only neutral and ambiguous cue (IV: Syntactic cue, 2 levels) crossed with ISQ and RhQ prosody (IV: Prosodic cue, 2 levels). We fitted a logistic mixed effects regression model with Accuracy as binary dependent variable (1 = correctly identified RhQ or ISQ). The maximal model included Prosodic cue (2 levels: RhQ vs. ISQ), Syntactic cue (2 levels: neutral vs. ambiguous) and AgeGroup (5 levels: 6 vs. 7 vs. 8 vs. 9 vs. adults) as IVs, and an interaction term between the three predictors. The Perception-different score was also used as a fixed effect. Participant and Item were included as random intercepts, and Prosodic cue was added as random slope to Participant.

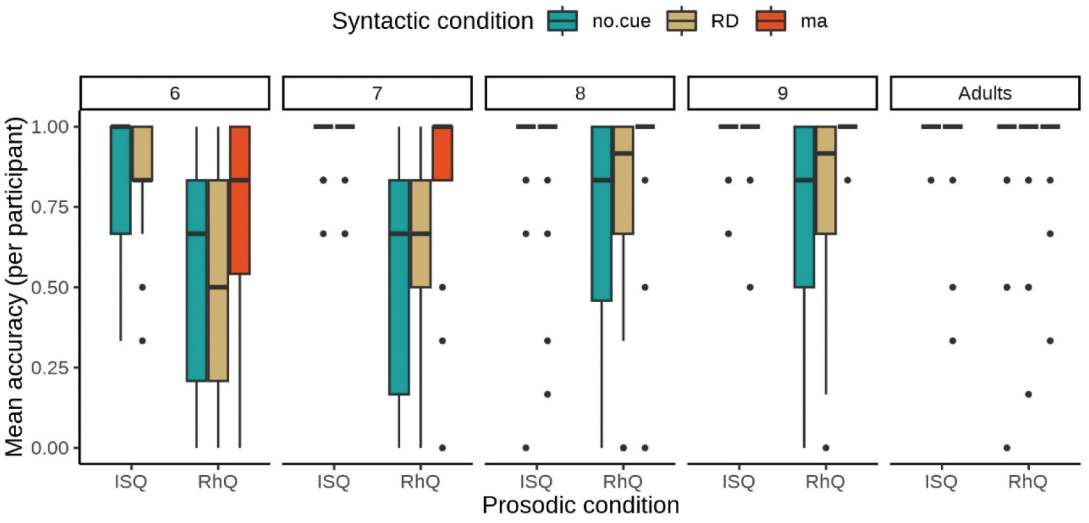

Figure 6. Boxplot showing the mean accuracy per participant in the comprehension task, divided by age group, prosodic condition and syntactic condition.

In the course of the stepwise model selection, all interactions were removed, as well as the main effect of Syntactic cue, as they were not significant. The final model included a significant main effect of Prosodic cue (χ^2 (1) = 20.77, $p < .001$), of AgeGroup (χ^2 (4) = 27.68, $p < .001$) and of the Perception-different score (χ^2 (1) = 25.39, $p < .001$). The effect of Prosodic cue indicates that, while accuracy in the ISQ condition (i.e., with ISQ prosody) is predicted to be almost at ceiling, it is lower in the RhQ condition ($\beta = -3.16$, $p < .001$). The effect of AgeGroup consists of all child groups being less accurate than adults (with p always lower than .01), while no difference was observed among the four child groups (with p always above .5). The full comparisons are found in Table 4.5 in the Online Supplementary Materials. Finally, the effect of the Perception-different score indicates that children who were less accurate in the perception task were also more likely to be less accurate in the comprehension task (i.e., there was a positive effect: $\beta = 6.66$, $p < .001$).

A second model analyzed the accuracy in all three levels of the syntactic condition, excluding the ISQ condition. The maximal model, with Accuracy as binary dependent variable, included Syntactic cue (3 levels: no vs RD vs *ma*) and AgeGroup (5 levels: 6 vs. 7 vs. 8 vs. 9 vs. adults) as independent variables and their interaction, as well as the Perception-different score as independent variable. Participant and Item were included as random intercepts. To test whether the rate of production of lexico-syntactic cues was related to the ability to exploit linguistic cues in comprehension, we also added 'Production score' (i.e., the mean number of cues per RhQ used by each participant) as a fixed effect, in interaction with Syntactic cue.

Both interactions were significant (Syntactic cue * AgeGroup: χ^2 (8) = 24.41, $p = 0.002$; Syntactic cue * Production score: χ^2 (2) = 3.91, $p = 0.001$). AgeGroup (χ^2 (4) = 23.26, $p < .001$) was also significant as a main effect, while Syntactic cue was not (χ^2 (2) = 1.69, $p = 0.4$). The effect of the Perception-different score was also significant (χ^2 (1) = 19.9, $p < .001$). The first interaction, which is visualized in Figure 7 (left) can be described as follows: while adults were equally accurate in all three conditions (i.e., there is no significant difference between conditions for adults), in all children groups the rhetorical condition was significantly more accurate than the ambiguous and the neutral condition, while there was no difference between the latter two. Additionally, the interaction is visible in the comparisons between groups. All child groups were significantly less accurate than adults in the neutral and ambiguous condition, while in the rhetorical condition there was only a difference

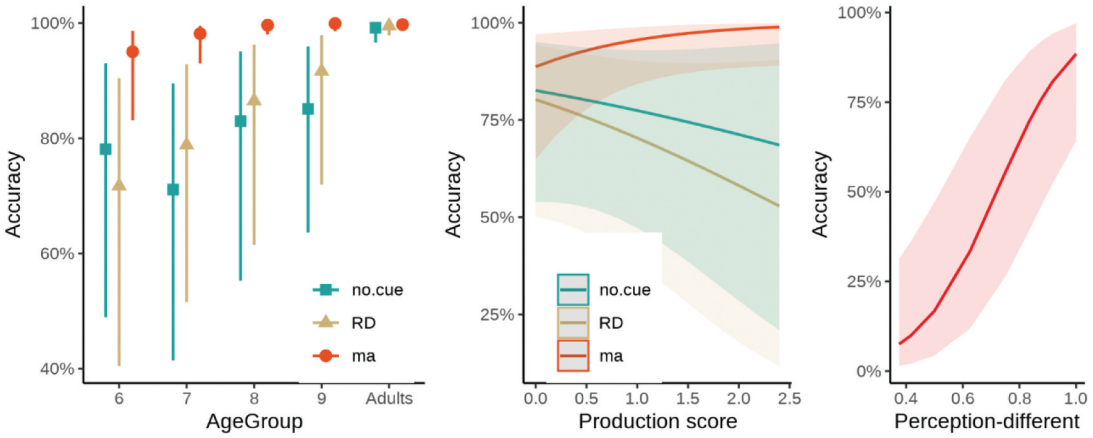

Figure 7. Model estimated accuracy in the comprehension task, predicted by the interaction between age group and condition (left), by the interaction between condition and Production score (centre) and by the Perception-different score (right).

between 6- and 9-year-olds and 6-year-olds and adults, with the former being less accurate. All comparisons are reported in the Online Supplementary Materials, Tables 4.8 and 4.9.

The interaction between Syntactic cue and Production score is shown in Figure 7 (centre). A post-hoc analysis with the *emtrends* function reveals that the effect of mean rate of markers in production varied across the three syntactic conditions of the comprehension task, because the slopes for the ambiguous and the neutral condition are significantly different than the slope for the rhetorical condition (neutral–rhetorical: $\beta = -1.34$, $p = .004$; ambiguous–rhetorical: $\beta = -1.55$, $p < .001$), while not differing from each other ($\beta = 0.21$, $p = .704$). As also visible in Figure 7, this means that children who used more lexico-syntactic markers in production were also more likely to have higher accuracy in the rhetorical condition specifically.

Finally, the effect of the Perception-different score was positive: children who were less accurate in the perception task were also likely to be less accurate in the comprehension task (i.e., there was a positive effect: $\beta = 7.27$, $p < .001$).

7. Discussion

We investigated whether children at primary school age produce and understand RhQ marking in Italian (RQ1), whether production and comprehension are still developing during this period (RQ2), and whether the ability to produce RhQs implies the ability to comprehend them (RQ3). By "marking" we referred to prosodic cues, which we investigated in comprehension, as well as specific lexico-syntactic marking, which we investigated in both comprehension and production.

7.1. Production and comprehension of RhQs

As there was no previous research on this phenomenon, we had no specific expectations except that the development of RhQs would be ongoing during children's primary school years, and that production skills would imply comprehension skills. The results of both the production and the comprehension tasks showed that some children in this age range are able to discriminate canonical and non-canonical questions on the basis of their linguistic form, or that they are in the process of acquiring this ability.

Children at all ages used lexico-syntactic cues to mark RhQs, and they did so significantly more than in ISQs, which indicates sensitivity to rhetorical marking. As to the types of cues, children and adults most frequently produced the adversative particle *ma;* adults and 9-year-olds did so with the

highest frequency (about 50%). Right dislocation was also very frequent, around 30% for adults and 32% for 9-year-olds and lower for 6-8-year-olds. Cleft questions were also comparatively frequent in RhQs, except amongst 6-year-olds, but this cue did not clearly differentiate RhQs from ISQs. The remaining cues occurred with low frequency in all groups, including sentence-initial particle *e* 'and', the adverbial particle *mai* 'ever', conditional verb morphology and the reflexive clitic *si* with affective value. All cues except cleft questions were either absent in ISQs, or present to a very low degree. The comparison of cue production in sentences with and without a modified model question further showed that only 9-year-olds were on par with adults when not prompted (i.e., when the production was completely spontaneous).

When discussing whether or not children have acquired RhQ markers, we need to take into account that these markers are optional. Thus, if children do not use a certain cue, this does not automatically mean that they have not acquired it, as they may simply avoid or chose not to use it. Nevertheless, the trends observed at the group level do reveal some developmental patterns, as we will discuss below. Altogether, the production task proved that at least some children throughout the investigated age period were able to appropriately use some lexical and morphosyntactic devices with a specialized pragmatic function in a contextually appropriate fashion.

In comprehension, children were generally capable of correctly identifying ISQs from RhQs, especially when the difference was prominently marked: they correctly identified ISQs as such, and they did identify RhQs when a rhetorical contour was combined with *ma* and RD (with some exceptions amongst the 6-year-olds). When they could only rely on prosody to identify RhQs (i.e., in the neutral and ambiguous condition), by contrast, their accuracy varied considerably: some children were able to recognize a rhetorical prosody (i.e., children with a high accuracy in these conditions), while other children consistently identified these questions as ISQs, and some children replied at chance.

The ability to exploit prosody alone in the interpretation of RhQs thus appears to be the greatest difference between children and adults. Through a perception task, we controlled for the possibility that children could not use prosody to identify RhQs simply because they could not perceive a difference in the contours. This, however, was only marginally borne out: while children who were less accurate in the perception task also tended to be less accurate in the comprehension task, the performance in the former was overall very good, indicating that most children were able to perceive the difference. We therefore assume that the children were simply more likely to focus their attention to the overt lexical cue (*ma* in the comprehension task) rather than the prosodic cues, failing to map the difference in prosodic form – although they could perceive it—to the communicative intention of the "rhetorical" characters. This is rendered more likely by the fact that none of the prosodic cues used in RhQs are markers of rhetoricity per se, but are rather various types of indirect cues that help induce a rhetorical reading of the sentence (see Dehé et al. 2022, Ferin 2024a). While the combination of cues used in the comprehension task was sufficient for adults to draw the rhetorical inference, it was not strong enough for some of the children, in line with research on other phenomena such as irony and emotion recognition, as discussed in Section 7.2. Additionally, a higher reliance on *ma* may have been partially boosted by the use of this cue in the production task. However, the two tasks were conducted on different days; therefore, we do not expect that this potential boosting effect be a decisive factor in the interpretation of these results.

7.2. Age development

Given previous research on similar and related phenomena, such as comprehension of sarcasm and irony, we expected that children would start developing non-literal language and complex interface phenomena as well as the associated linguistic cues around age 6, with an improvement throughout primary school. As expected, there were age effects, although the development was not perfectly linear from younger to older children: 6- and 8-year-olds used fewer additional cues than 9-year-olds and

7-year-olds were in between[10] Additionally, the presence of a marker (*ma*) in the model question was crucial for the younger groups, who used *ma* itself much more infrequently and who modified less overall. In comprehending RhQs, where all children were particularly good when a combination of lexico-syntactic and prosodic cues was present, the 6-year-olds varied substantially, while older children were at ceiling. For the neutral and the ambiguous cue, there was less improvement with age, and all child groups were less accurate than adults.

Overall, there was a clear developmental trend: the 9-year-olds performed on par with adults in all measures except for those conditions in the comprehension task in which they had only prosodic cues to rely on. Thus, RhQs can be considered as a "very late acquired" phenomenon à la Tsimpli (2014), that is, a complex phenomenon at the interface between language, discourse and cognition that is highly dependent on language external factors, such as the amount of input. We cannot entirely exclude that specific cognitive abilities required for RhQ comprehension are still in the process of maturating in some of the children. However, the children also took part in an irony task (reported in Ferin 2024a), where they performed at ceiling, in line with previous research showing that 6-year-olds can already understand irony (Banasik & Podsiadło 2016). Complex interface phenomena of this kind also pose an interesting case for studies that explore the effects of shifting input conditions in early bilingual children, who face the pressure of reduced input and of cross-linguistic influence from the other language. Thus, the acquisition of RhQ marking constitutes a promising field of enquiry for this population (see Ferin et al. 2025).

More generally, the results are in line with the literature on interface phenomena whose use and recognition depend on (para)linguistic cues, such as emotion, sarcasm and irony recognition. Although infants are sensitive to prosodic cues and can discriminate emotional expression based on prosody and facial expression from a very early age (e.g., Walker-Andrews & Grolnick 1983, Fernald 1993, Mumme et al. 1996, Grossmann et al. 2005), it appears to take a few more years until they can evaluate based on contextual information and (para)linguistic cues whether lexical and grammatical information should be interpreted literally or not (Morton & Trehub 2001), and until they make use of a larger range of optional linguistic cues to signal non-literal meaning themselves. The interpretation and production of rhetorical questions is mediated through the understanding of contextual information and of the communicative intention of the speaker. In this respect, understanding RhQs is similar to understanding ironic and sarcastic language, an ability that is still in the process of developing after the age of 6 years (Glenwright & Pexman 2010, Chevallier et al. 2011, Pexman et al. 2011, Colich et al. 2012, Banasik 2013).

7.3. Are comprehension and production related?

Since comprehension tends to precede production in language acquisition, we expected children who produced certain cues (e.g., *ma*) to also comprehend their function, but not necessarily the reverse. We further expected comprehension to be supported by prosody. However, since prosody, lexicon and syntax evolve separately, it was also natural to expect that not all cues would be acquired at the same time.

We found that children who used more lexico-syntactic markers in production were also more likely to have higher accuracy in the rhetorical condition specifically.

In other words, a higher production of cues was tied to children being more likely to focus on *ma* as the discriminating cue in the comprehension task. Given that *ma* was the most frequent cue in

[10]The results for 8-year-olds cannot be easily explained with the information at our disposal. SES information (such as education and profession of mother and father, time spent reading) were collected through the background questionnaire, but there were no detectable differences between the groups. An extremely tentative explanation may come from the fact that children were tested during and after the restrictions imposed by the COVID 19 pandemic. If 8 years is the turning point at which children start approaching adult-like levels, the reduced interactions with the larger community may have reduced also their exposure to the relevant interactional type of input. However, there is no evidence to support such a conclusion, and the observed pattern may simply be due to a sampling issue, determined by the cross-sectional nature of the design and the optionality of the phenomenon.

production, it also had the largest effect on the production rate (i.e., high production rates are largely due to a high rate of *ma* production). This suggests that comprehension and production do go hand in hand. There was no relation, instead, between use of lexico-syntactic markers and prosodic cues in comprehension, as indicated by the lack of an effect in the neutral and ambiguous condition.

7.4. Do children's markings signal rhetoricity in sensu-strictu?

We have shown that children can use optional pragmatic markers to differentiate canonical and non-canonical questions in production and in comprehension, and that the acquisition of this ability is still underway. In other words, we have provided evidence that they are able to use markers of bias, of special attitude or epistemic stance, that can then be used to draw the "rhetorical inference" (see Section 2). One weakness in our setup was, however, that we have not tested whether children actually produce and recognize these questions as properly rhetorical, because the cues in our comprehension task are not resolutive in this sense. In other words, we have not been able to provide conclusive evidence that RhQs have been acquired as a pragmatic category. The only unambiguously rhetorical cue is *want*-embedding (Ferin 2024b), but this was only produced by one adult and none of the children. All the other cues were ambiguous, either with canonical questions (RD, clefts, affective reflexive, *e*, conditional) or with surprise-disapproval and extreme ignorance questions (*ma, mai*). Thus, children may be using the cues to express the individual semantic/pragmatic contribution of the single cue, but without drawing the rhetorical inference.

In comprehension, we used "rhetorical prosody" and the cue *ma*, which is not unambiguously rhetorical. In prosody, the combination of a low boundary tone, longer duration of the stressed vowels and a more compressed final boundary tone seems to be a solid combination of cues, and some children as well as all adults were able to use this combination to discriminate between question types. However, each cue on their own is not a cue to rhetoricity. Thus, we do not know whether children paid attention to the combination of cues (thus likely deriving the rhetorical interpretation) or to only one cue (e.g., a prosodic marker of attitude) to map the question onto the character. While a fully conclusive answer to the question whether children have acquired "RhQs as a pragmatic category" will require a separate study, we have some indirect evidence that allows us to draw some preliminary conclusions.

A first indication that children have indeed acquired RhQs as a pragmatic category is that some children combined several cues in production, as for example in (12), which features *e* 'and', affective reflexive *si* and a conditional verb. Such combinations of cues strongly increase the plausibility of a rhetorical reading, because they are not appropriate in information-seeking contexts. Again, this does not imply that all children can produce RhQs, but it is crucial that some children do, which means that the acquisition process is ongoing.

(12) E chi si mangerebbe il budino?! [7-year-old]
and who REFL eat-COND.3SG the pudding
'Who would eat the pudding?!'

Another indication that at least some children have acquired the rhetorical interpretation is the use of *mai* "never". Use of *mai* itself does not tell us directly that questions are interpreted rhetorically per se, but when *mai* is combined with conditional mood it becomes a very strong non-canonical questions marker, either rhetorical or indicating extreme ignorance. Either way, a question modified by *mai* is not information-seeking because the speaker is stating the impossibility of finding an answer. Moreover, while *mai* is grammatical in two positions, the wh-position preceding the inflected verb (*Chi mai mangerebbe* ...) and the adverbial position (*Chi mangerebbe mai* ...), in the former it is necessarily interpreted as a discourse particle triggering a non-canonical reading. Interestingly, there was one 6-year-old child in our dataset

who used *mai* in wh-position, signaling that they are using it a discourse particle with this specific pragmatic function, as illustrated in (13).

(13) Chi mai mangerebbe una banana?! [6-year-old]
who ever eat-COND.3SG a banana
'Who would ever eat a banana?!'

Our study leaves open the question whether all children would accept *mai* in the wh-position, which could be explored in future studies, employing comprehension experiments that target this specific cue or sentence repetition tasks with contextualized sentences biased to either the rhetorical or the information-seeking reading.

8. Conclusion and future directions

In conclusion, we have shown that primary school aged children are in the process of acquiring RhQs and their linguistic correlates. Children aged 6 to 9 were able to exploit lexico-syntactic cues in production, and lexico-syntactic and prosodic cues in comprehension. We found a developmental effect, such that 9-year-old children were the most similar to adults, while younger children displayed more variability at the group level. Future research should focus on the acquisition of RhQ prosody also in production, and on its (potential) relation with the production of lexico-syntactic marker. Further studies could further explore the acquisition of RhQs in languages with different markers to see how the types and combinations of markers modulates acquisition speed. Finally, future research could take into account the role of gestures, to see whether they are acquired contemporaneously with the other set of (para)linguistic markers, and whether they are used in a compensatory or reinforcing way.

Acknowledgments

The project was done in collaboration with Theodoros Marinis and Miriam Geiß. We thank Sebastiano Arona, Alexandra Besler, Sara Cavaglià, Federica De Cristofaro, Noemi Furlani, Mariapaola Piccione, and Laura Steck for help with data collection and analysis.

Disclosure statement

No potential conflicts of interest are reported by the authors(s).

Funding

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project KU 2439/5-1 "Non-Canonical Questions in Early and Late Bilingual Language Acquisition" of the FOR2111 "Questions at the Interfaces."

ORCID

Maria Francesca Ferin http://orcid.org/0009-0000-1203-3967
Tanja Kupisch http://orcid.org/0000-0002-2653-2692

Conflict of interest

The authors report there are no competing interests to declare.

Data availability statement

The data supporting the findings presented in this paper, the analysis scripts and the Online Supplementary Materials are available at the following OSF repository: https://osf.io/8kurf/

References

Ackerman, Brian P. 1982. Contextual integration and utterance interpretation: The ability of children and adults to interpret sarcastic utterances. *Child Development* 53(4). 1075–1083.

Banasik, Natalia. 2013. Non-literal speech comprehension in preschool children - An example from a study on verbal irony. *Psychology of Language and Communication* 17(3). 309–324. http://dx.doi.org/10.2478/plc-2013-0020.

Banasik, Natalia & Kornelia Podsiadło. 2016. Comprehension of ironic utterances by bilingual children. *Psychology of Language and Communication* 20(3). 316–335.

Banuazizi, Atissa & Cassandre Creswell. 1999. Is that a real question? Final rises, final falls, and discourse function in yes-no question intonation. *Chicago Linguistic Society* 35. 1–14.

Bascandziev, Igor, Patrick Shafto, & Elizabeth Bonawitz. 2025. Prosodic cues support inferences about the question's pedagogical intent. *Open Mind* 9. 340–363. https://doi.org/10.1162/opmi_a_00192

Belletti, Adriana. 2012. Revisiting the CP of clefts. In Günther Grewendorf and Thomas Ede Zimmermann (eds.), *Discourse and grammar. From sentence types to lexical categories*, 91–114. Boston/Berlin: De Gruyter Mouton.

Beltrama, Andrea & Andreas Trotzke. 2019. Conveying emphasis for intensity: Lexical and syntactic strategies. *Language and Linguistics Compass* 13(7). 1–13. https://doi.org/10.1111/lnc3.12343

Berruto, Gaetano. 1986. La dislocazione a destra in italiano [Right dislocation in Italian]. In *Tema-Rema in Italiano*, edited by Harro Stammerjohann, 55–69. Tübingen: Gunter Narr Verlag.

Biezma, María & Kyle Rawlins. 2017. Rhetorical questions: Severing questioning from asking. In Dan Burgdorf, Jacob Collard, Sireeman Maspong, & Brynhildur Stefánsdóttir (eds.), *Proceedings of Semantic and Linguistic Theory (SALT), University of Maryland, College Park* 27, 302–322. Linguistic Society of America.

Boersma, Paul & David Weenink. 2015. Praat: Doing phonetics by computer (Version 6.1.05) [Computer software]. Retrieved from http://Praat.Org/.

Capelli, Carol A., Noreen Nakagawa, & Cary M. Madden. 1990. How children understand sarcasm: The role of context and intonation. *Child Development* 61(6). 1824–1841.

Caponigro, Ivano & Jon Sprouse. 2007. 'Rhetorical questions as questions'. In Estela Puig-Waldmüller (ed.), *Proceedings of Sinn Und Bedeutung 11*, 121–133. Barcelona: Universitat Pompeu Fabra.

Celle, Agnes, Anne Jugnet, & Laure Lansari. 2021. Expressive questions in English and French: What the hell versus mais qu'est-ce que. In Andreas Trotzke and Xavier Villalba (eds.), *Expressive meaning across linguistic levels and frameworks*, 138–166. Oxford: Oxford University Press.

Chevallier, Coralie, Ira Noveck, Francesca Happé, & Deirdre Wilson (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia* 49(3). 507–517.

Colich, Natalie, Audrey-Ting Wang, Jeffrey Rudie, Leanna Hernandez, Susan Bookheimer, & Mirella Dapretto. 2012. Atypical neural processing of ironic and sincere remarks in children and adolescents with autism spectrum disorders. *Metaphor & Symbol* 27(1). 70–92.

Coniglio, Marco. 2008. Modal particles in Italian. *Working Papers in Linguistics* 18. 191–129.

Cordin, Patrizia. 1995. I pronomi riflessivi [Reflexive pronouns]. In Lorenzo Renzi, Giampaolo Salvi & Anna Cardinaletti (eds.), *Grande Grammatica Italiana Di Consultazione*, 606–617. Bologna: Il Mulino.

Crocco, Claudia. 2013. Is Italian clitic right dislocation grammaticalised? A prosodic analysis of yes/no questions and statements. *Lingua* 133. 30–52. https://doi.org/10.1016/j.lingua.2013.03.003

Dayal, Veneeta. 2016. *Questions*. Oxford: Oxford University Press.

Dehé, Nicole, Daniela Wochner, & Marieke Einfeldt. 2022. The interaction of discourse markers and prosody in rhetorical questions in German. *Journal of Linguistics* 60(1). 1–25. https://doi.org/10.1017/S0022226722000299.

Dikken, Marcel den, & Anastasia Giannakidou. 2002. From hell to polarity: "Aggressively non-D-Linked" wh-phrases as polarity items. *Linguistic Inquiry* 33(1). 31–61. https://doi.org/10.1162/002438902317382170.

Ferin, Maria F. 2024a. Rhetorical questions in Italian between theory and acquisition. Konstanz: Universität Konstanz dissertation.

Ferin, Maria F. 2024b. Rhetorical questions in colloquial Italian. *Italian Journal of Linguistics* 36(1). 1–38. https://doi.org/10.26346/1120-2726-218.

Ferin, Maria F., Theodoros Marinis, & Tanja Kupisch. 2025. The acquisition of rhetorical questions in bilingual children with Italian as a Heritage Language. *Bilingualism: Language and Cognition* 28(1). 1–14. https://doi.org/10.1017/S1366728923000974.

Ferin, Maria F., Simona Sbranna, & Aviad Albert. forthcoming. Prosodic patterns of Italian rhetorical questions. In Nicole Dehé, Regine Eckardt, and George Walkden (eds.), *The Oxford handbook of non-canonical questions*. Oxford: Oxford University Press.

Fernald, Anne. 1993. Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development* 64(3), 657–674.
Friend, Margaret. 2000. Developmental changes in sensitivity to vocal paralanguage. *Developmental Science* 3(2). 148–162.
Frota, Sónia, Joseph Butler, & Marina Vigário. 2014. Infants' perception of intonation: Is it a statement or a question? *Infancy* 19(2). 194–213.
Geiss, Miriam, Maria F. Ferin, Theo Marinis, & Tanja Kupisch. 2023. Rhetorical question comprehension by Italian–German bilingual children. *Second Language Research* 40(2). 02676583221140861. https://doi.org/10.1177/02676583221140861.
Gili Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Crocco, Mariapaola D'Imperio, Rosa Giordano, Giovanna Marotta, Michelina Savino, & Patrizia Sorianello. 2015. Intonational phonology of the regional varieties of Italian. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 140–197. Oxford: Oxford University Press. http://hdl.handle.net/1854/LU-5743090.
Giorgi, Alessandra. 2018. *Ma Non Era Rosso?* (But Wasn't It Red?): On Counter-Expectational Questions in Italian. *Romance Languages and Linguistic Theory* 14. 69–84. https://doi.org/10.1515/ip-2019-0017.
Glenwright, Melanie, Jayanthi M. Parackel, Kristene RJ Cheung, & Elizabeth S. Nilsen. 2014. Intonation influences how children and adults interpret sarcasm. *Journal of Child Language* 41(2). 472–484.
Glenwright, Melanie & Penny M. Pexman. 2010. Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language* 37(2). 429–451.
Grossmann, Tobias, Tricia Striano, & Friederici, Angela. 2005. Infants' electric brain responses to emotional prosody. *NeuroReport: For Rapid Communication of Neuroscience Research* 16, 1825–1828.
Han, Chung-Hye. 2002. Interpreting interrogatives as rhetorical questions. *Lingua* 112. 201–229.
Hinterhölzl, Roland & Nicola Munaro. 2021. On the illocutionary force of exclamatives and non-canonical questions in German and Italian. In Andreas Trotzke and Xavier Villalba (eds.), *Expressive meaning across linguistic levels and frameworks*, 43–65. Oxford: Oxford University Press.
Ippolito, Michela. 2021. The contribution of gestures to the semantics of non-canonical questions. *Journal of Semantics* 38(3). 363–392. https://doi.org/10.1093/jos/ffab007.
Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. Lmer test package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(3). 1–26.
Leiner, Dominik Johannes. 2019. SoSci Survey (Version 3.2.32). https://www.soscisurvey.de .
Lenth, Russel. 2023. Emmeans: Estimated marginal means, aka least-squares means (Version 1.8.7). R package. https://CRAN.R-project.org/package=emmeans.
Manzini, Maria Rita. 2015. Italian adverbs and discourse particles. Between recategorisation and ambiguity. In Josef Bayer, Roland Hinterhölzl, and Andreas Trotzke (eds.), *Discourse-oriented syntax*, 93–120. Amsterdam: John Benjamins.
Miller, Joanne & Peter Eimas. 1979. Organization in infant speech perception. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *33*(4). 353–367.
Morton, Bruce J. & Sandra E. Trehub. 2001. Children's understanding of emotion in speech. *Child Development* 72(3). 834–843.
Mumme, Donna, Anne Fernald, Anne, & Carla Herrera. 1996. Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development* 67(6). 3219–3237. https://doi.org/10.2307/1131775.
Neitsch, Jana. 2019. *Who cares about context and attitude? Prosodic variation in the production and perception of rhetorical questions in German*. Konstanz: Universität Konstanz dissertation.
Obenauer, Hans-Georg. 2004. Nonstandard wh-questions and alternative checkers in Pagotto. *Syntax and Semantics of the Left Periphery, Interface Explorations* 9. 343–384.
Obenauer, Hans-Georg & Cecilia Poletto. 2000. Rhetorical wh-phrases in the left periphery of the sentence. *Venice Working Papers in Linguistics* 26. 121–153.
Pexman, Penny, Kristin Rostad, Carly McMorris, Emma Climie, Jaqueline Stowkowy, & Melanie Glenwright. 2011. Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of Autism Developmental Disorders* 41(8). 1097–1112.
R Core Team. 2023. R: *A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. https://www.R-project.org/.
Recchia, Holly E., Nina Howe, Hildy S. Ross, & Stephanie Alexander. 2010. Children's understanding and production of verbal irony in family conversations. *British Journal of Developmental Psychology* 28(2). 255–274. https://doi.org/10.1348/026151008X401903.
Sorianello, Patrizia. 2018. Tra prosodia e pragmatica: il caso delle domande retoriche [Between prosody and pragmatics: the case of rhetorical questions]. *Studi e Saggi Linguistici LVI* (2). 39–71.
Sorianello, Patrizia. 2019. "A Che Serve Saperlo?" Funzioni Pragmatiche e Variazioni Intonative Della Domanda Retorica ['What's the point of knowing that?' Pragmatic functions and intational variation of rhetorical questions. In E. Nuzzo and I. Vedder (eds.), *Studi AItLA 9: Lingua in Contesto. La Prospettiva Pragmatica*, 89–108. Milan: Officinaventuno.
Tsimpli, Ianthi. 2014. Eary, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313. https://doi.org/10.1075/lab.4.3.01tsi
Walker-Andres, Arlene, & Wendy Grolnick. 1983. Discrimination of vocal expressions by young infants. *Infant Behavior and Development* 6(4). 491–498.