

Rhetorical question comprehension by Italian–German bilingual children

Second Language Research

2024, Vol. 40(2) 221–246

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02676583221140861

journals.sagepub.com/home/slr**Miriam Geiss** **Maria F Ferin**

University of Konstanz, Germany

Theo Marinis

University of Konstanz, Germany; University of Reading, UK

Tanja Kupisch

University of Konstanz, Germany; UiT the Arctic University of Norway, Norway

Abstract

This study investigates for the first time the comprehension of rhetorical questions (RhQs) in bilingual children. RhQs are non-canonical questions, as they are not used to request information, but to express the speaker's belief that the answer is already obvious. This special pragmatic meaning often arises by means of specific prosodic and lexical-syntactic cues. Being childhood learners, children have to acquire the concept of rhetoricity, but being bilinguals, they further need to acquire the different cues marking RhQs in their two languages. We tested 85 bilingual children (aged 6–9 years) with Italian as heritage language (HL) and German as majority language (ML) in both of their languages, using a forced-choice comprehension task. Our results show that RhQ comprehension improves with age in both languages. Bilingual children are able to exploit prosodic and syntactic cues to comprehend RhQs in their ML and HL with a slight advantage in the ML. This advantage could be either an effect of the cues used in the experiments in the two languages or of a higher proficiency in the ML. In addition, our results point to a later acquisition of prosodic rhetorical cues, which has implications for bilingual acquisition of non-canonicity in general.

Keywords

bilingualism, cue transfer, prosody, rhetorical questions, syntax

Corresponding author:

Miriam Geiss, Department of Linguistics, University of Konstanz, Universitätsstraße 10, Konstanz, 78457, Germany

Email: miriam.geiss@uni-konstanz.de

I Introduction

Rhetorical questions (RhQs), like (1), are considered to be non-canonical. They differ from canonical, information-seeking questions (ISQs) in not literally asking for information. Instead, the RhQ in (1) implies the speaker's belief that the answer is obvious, as nobody likes paying taxes. The combination of '?' and '!' is used to signal an RhQ reading.

- (1) Who likes paying taxes?!

Understanding RhQs involves the ability to process the linguistic cues that mark rhetoricity in order to interpret the non-literal meaning and understand the speaker's intent. Many languages use prosody to mark rhetoricity, for example by means of specific intonational contours or tone of voice. Prosody serves both linguistic (e.g. differentiating statements and questions) and paralinguistic functions (e.g. expressing emotion), and both are relevant for RhQs.

Why study such a complex phenomenon as RhQs in bilingual children? RhQs are a (interface) phenomenon for which children need to learn to integrate information from different sources, which is particularly challenging when they provide conflicting information, as is the case in RhQs or other pragmatic phenomena (e.g. Ackerman, 1982; Morton and Trehub, 2001). Since prosody-meaning mappings are still developing in primary-school-aged children (Lleó, 2018; Saindon et al., 2016), children at that age are known to rely less on prosodic cues than on literal meaning if there are inconsistencies between the two (Capelli et al., 1990; Glenwright et al., 2014). Interestingly, bilingual children were shown to have advantages over monolingual children in dealing with conflicting information during pragmatic interpretation (Siegal et al., 2009; Yow and Markman, 2011; but see Antoniou et al., 2020; Dupuy et al., 2019). If this is correct, they may also perform well on RhQs. However, previous research on the acquisition of RhQs is very limited for monolingual children and, to our knowledge, no studies have addressed how bilingual children acquire RhQs.

To fill this research gap, the present study investigates RhQ comprehension in Italian–German bilingual children living in Germany. Bilingual children need to acquire not only the concept of RhQs but also the sets of cues that mark rhetoricity in their two languages. If the two languages employ different cues for the same phenomenon, they can potentially influence each other in either direction. This is referred to as cross-linguistic influence and is conditioned by intra- and extra-linguistic factors, both of which will be taken into account here.

As to extra-linguistic factors, bilingual children hardly ever have the same proficiency in their two languages and the amount of input they receive in the two languages is not always the same. This is especially the case if one language is a heritage language (HL) acquired in the family and the other one is the majority language (ML) of the society which eventually becomes the dominant language. Moreover, interpreting RhQs involves many prosodic and lexical-syntactic cues, various (extra-)linguistic and cognitive domains and irony¹ comprehension, as outlined further below, as well as potential language imbalance. Thus, studying RhQs can be informative with respect to which cues are

transferred (lexical-syntactic vs. prosodic), under which conditions (balance vs. imbalance) and in which direction (from stronger to weaker language and/or vice versa).

The aim of this study is thus to investigate whether primary-school-age children are able to comprehend RhQs in their ML (German) and what factors mediate their comprehension. In addition, we compare the children's ability to comprehend RhQs in the ML to their ability to comprehend RhQs in their HL and discuss the acquisition of different sets of cues and their possible interaction. This article is structured as follows: Section II provides an overview of RhQs in German and Italian followed by relevant literature on the acquisition of RhQs and related pragmatic phenomena (Section III). In Sections IV and V, we present our study and the results, which we discuss in Section VI. We conclude with Section VII.

II RhQs in German and Italian

As illustrated in (1), RhQs have the same syntactic surface structure as canonical interrogatives, but they differ from ISQs in their discourse function. Unlike ISQs, RhQs do not require an answer from the addressee (Biezma and Rawlins, 2017) because the answer is assumed to be already obvious, that is, in the common ground shared by the speaker and the addressee (Caponigro and Sprouse, 2007). Pragmatically, RhQs are similar to assertions because they are often used to make a point (Biezma and Rawlins, 2017).

RhQs bear similarity with ironic statements because their comprehension requires the comprehension of both the literal and the intended meaning. Listeners can exploit different cues in order to understand RhQs (see Neitsch, 2019: 46–50). These include world knowledge and context (e.g. *Is the pope catholic?*; Sadock, 1974) and language-specific implementations of prosodic cues (for a cross-linguistic survey on RhQs, see Dehé et al., 2022) and lexical-syntactic cues, such as discourse particles (DiPs).

I RhQs in German

RhQs in German can have the same syntactic surface as ISQs, as illustrated in (2). To disambiguate the two question types, the speaker can use phonological, phonetic, or syntactic cues.

- (2) Wer isst Bananen?/?!
 'Who eats bananas?/?!'

In terms of phonetic and phonological cues, the intonational contour of RhQs containing the *wh*-element *wer* 'who' (*wh*-RhQs) ends in a low tone (Braun et al., 2019). In phonological terms, this is referred to as a low edge tone (L-%). *wh*-ISQs exhibit more variation with respect to edge tones, which can be low (L-%), or exhibiting a shallow rise (L-H%) or high rise (H-^H%).² With respect to the accent associated with the most prominent syllable (i.e. nuclear pitch accent), Braun et al. (2019) report that RhQs are mostly realized with low rising ones (L*+H), and ISQs with high rising ones (L+H*). Phonetically, RhQs have a longer constituent duration than ISQs, especially at the end of

the utterance. Additionally, RhQs are more often realized with a breathy voice quality, also called murmured or whispery voice, on the *wh*-word. Note that, in perception, the type of nuclear pitch accent is the strongest cue to rhetoricity in German, with duration and voice quality being additional secondary cues (Kharaman et al., 2019).

With respect to syntactic cues, rhetoricity can be signaled in German using DiPs, such as *schon* ('already'), as indicated by examples (3a)–(3c) (Bayer and Struckmeier, 2017; Biezma and Rawlins, 2017; Meibauer, 1986).³ Example (3a) presents a syntactically ambiguous question. In most of the semantic literature, there is consensus that *schon* as a DiP is an unambiguous signal for an RhQ (3b). However, it is homophonous with the temporal adverb *schon*; thus, (3b) could be compatible with an ISQ reading. If the particle *schon* is combined with another particle, *denn* ('then') – as in (3c) – the rhetorical reading is reinforced (Meibauer, 1986: 119). On its own, *denn*, which expresses a speaker's concern (Bayer et al., 2016: 593), can be used in both illocution types (Thurmair, 1991) and is therefore considered to be ambiguous.

- (3) a. Wer will heute einkaufen?
 b. Wer will heute schon einkaufen?
 c. Wer will denn heute schon einkaufen?
 'Who will go shopping today?/?!'

(adapted from Bayer and Struckmeier, 2017)

2 RhQs in Italian

String-identical RhQs and ISQs are also possible in Italian, as shown in (4). Similar to German, phonetic and phonological cues serve as disambiguating prosodic cues. According to Sorianello (2018), phonologically, *wh*-RhQs end more often with a low edge tone (L%), as opposed to the high (H%) or rising edge tones (LH%) found for ISQs. Duration and pitch excursion are two additional phonetic cues to rhetoricity: RhQs display a longer duration of the final tonic vowel and the pitch excursion is smaller in RhQs than in ISQs (Sorianello, 2018, 2019). The findings by Ferin (2022) show that the combination of these cues allows adult monolinguals to interpret a question as RhQ or ISQ respectively in a decision task.⁴

- (4) Chi mangia le banane?/?!
 'Who eats bananas?!'

RhQs can also be marked by additional lexical and morphosyntactic cues. In an elicited production study Ferin (2022) found that the initial adversative particle *ma* 'but', as in (5a), is used very frequently in spontaneous speech to mark RhQs that express disagreement. This particle is used with counter-expectational value (i.e. something in the previous context does not conform to the speaker's beliefs) and as such can also occur in other types of non-canonical questions expressing a conflict with the previous context. The particle *ma* was often found in combination with clitic right dislocation (CLRD), a syntactic structure marking a familiar topic (5b). Its frequent occurrence in RhQs is presumably due to the fact that the answer is given, thus establishing a link with the common

ground. However, like German *denn*, CLRD on its own is also compatible with ISQs, constituting an ambiguous cue. Another cue for rhetoricity is the adverbial particle *mai* ‘ever’ (Coniglio, 2008; Obenauer and Poletto, 2000), which is also possible in combination with a conditional verb, as in (5c). However, as Ferin (2022) shows, this cue seems to belong to a more formal register.

- (5) a. Ma chi mangia le banane?
 but who eats the bananas
 b. Chi le mangia le banane?
 who CL eats the bananas
 c. Chi mangerebbe mai le banane?
 who would eat ever the bananas
 ‘Who eats bananas?!?’

In sum, both languages exploit specific prosodic cues for rhetoricity. With respect to the syntactic cues, Italian *ma* and German *schon* can be considered strong correlates of rhetoricity, especially if they combine with another syntactic cue (Italian: CLRD, German: *denn*), which on their own would be ambiguous. However, while (*denn*) *schon* appears to be a direct marker of rhetoricity in German, Italian resorts to cues that trigger a rhetorical reading in a more indirect way.

III The acquisition of RhQs

Research on the acquisition of RhQs is scarce. To our knowledge, the only study investigating RhQs in children is the study by Recchia et al. (2010). Therefore, the following literature review also includes studies on related pragmatic phenomena. Whenever possible, reference to studies on bilingual children is made.

According to Recchia et al. (2010), 4-year-old monolingual children can comprehend, at least to some extent, RhQs in a naturalistic (family) environment. This ability improves with age: 6-year-old children in Recchia et al.’s (2010) study were better than 4-year-olds. Studies on phenomena comparable to RhQs, such as irony, show similar results (e.g. Banasik, 2013; Giustolisi et al., 2017). Children around the age of 6 years can comprehend ironic statements (for monolingual Polish children, see Banasik, 2013; for bilingual Polish–English children, see Banasik and Podsiadło, 2016), while younger (monolingual) children tend to interpret them literally (Ackerman, 1982; Banasik, 2013). The studies also show that irony comprehension is facilitated by contextual information (Ackerman, 1982), advanced Theory of Mind development (Banasik, 2013), and prosody (Capelli et al., 1990; Glenwright et al., 2014).⁵

As outlined above, RhQs involve the ability to infer the speaker’s intent, which goes beyond the literal meaning of the question and is often signaled by prosodic cues. This poses a challenge for language acquisition as the different sources (i.e. literal meaning and prosodic cues) can contradict each other (e.g. Ackerman, 1982; Morton and Trehub, 2001). Interestingly, studies suggest that bilinguals are particularly good at integrating cues from different sources (Siegal et al., 2009; Yow and Markman, 2011). For example, bilingual children aged 4 to 6 years were better in detecting violations of the Gricean

conversational maxims than monolingual children (Siegal et al., 2009). However, more recent studies on children's comprehension of pragmatic meaning show no difference between monolingual and bilingual children (e.g. Antoniou et al., 2020; Dupuy et al., 2019).

RhQs also involve the ability to decode prosodic and paralinguistic cues. Although children's ability to interpret such cues in RhQs have not been investigated so far, studies on emotion recognition might tell us what to expect. These studies show that, by the age of 4 years, children rely primarily on what the speaker said to make their judgements, although they have acquired the paralinguistic cues associated with happiness and anger or sadness (Friend, 2000; Morton and Trehub, 2001).⁶ When presented with contextual cues, children between 5 and 7 years rely on the situation to make their judgement (Aguert et al., 2013). Between 7 and 9 years, children begin to pay more attention to the prosodic cues, and by the age of 10 years, children can make their judgement solely on the basis of the prosodic cue (Friend, 2000; Morton and Trehub, 2001). More recent studies suggest that the ability to infer emotions through prosodic cues is not fully mastered by age 13 years (Aguert et al., 2013; Chronaki et al., 2015). Studies of bilingual children suggest that they acquire the ability to weigh prosodic over lexical cues faster than monolingual children, but still not in an adult-like manner (for a comparison, see, for example, Yow and Markman, 2011). According to Champoux-Larsson and Dylman (2019), this bilingual advantage is a prosodic bias: bilingual children rely on prosody when asked to identify the semantics of the word. In other words, bilingual children have more difficulties in ignoring the irrelevant discrepant prosodic cue.

With respect to the acquisition of prosodic cues more generally, studies have shown that infants are already sensitive to the acoustic parameters associated with questions and statements (for yes/no questions in European Portuguese, see, for example, Frota et al., 2014; for declarative questions in English, see Soderstrom et al., 2011). However, early sensitivity does not entail that infants acquired the concepts of questions and statements and their corresponding prosodic form. The ability to reliably map prosodic forms onto the respective category (i.e. statement or question) is still developing around the age of 6 years (see, for example, Saindon et al., 2016). Bilingual children showed a delay of several months compared to monolingual children in producing nuclear pitch accents in a target-like shape (i.e. differences in alignment and scaling were found; for an overview, see Lleó, 2018). At the same time, transfer from the ML to the HL can occur, resulting in non-target-like intonational contours in the HL (Lléo et al., 2004). The observed differences in production might be an indication that bilingual children also comprehend prosody differently than monolinguals, but this has not been investigated. As a consequence of their yet unstable prosody–meaning mappings, children may rely less on prosody than adults, especially when the literal meaning and (para)linguistic cues are discrepant, as is the case for RhQs.

In summary, previous research has shown that (para)linguistic cues play an important role in differentiating questions from statements, and in the recognition of emotion and irony, suggesting that this might also be the case for RhQ comprehension, where prosody can be used to disambiguate RhQs from ISQs (see Section II.1). To date, no study has investigated at what age bilingual children are able to comprehend RhQs in their ML and

which of the relevant linguistic cues discussed in the literature they use to make their judgements. Therefore, we address the following research questions:

- Research question 1: How do prosodic and syntactic cues affect bilingual children's comprehension of RhQs in their ML?
- Research question 2: To what extent is RhQ comprehension mediated by irony comprehension, language dominance, and age?
- Research question 3: Is there a difference in RhQ comprehension between the ML and the HL?

IV Method

In this study, we examine comprehension of RhQs by Italian–German bilingual children.⁷ All bilingual children were tested in both languages on different days using two main experiments (a Perception Task and a Comprehension Task). The aim of these tasks is to investigate the effect of prosodic and syntactic cues on RhQ comprehension (linguistic factors) as well as of language dominance, age and irony comprehension.

I Participants

Eighty-four Italian–German bilingual children took part in this study, divided into four age groups (6-, 7-, 8-, and 9-year-olds). All children were exposed to Italian from birth. Forty-seven had one Italian- and one German-speaking parent (2L1), while 37 had two Italian-speaking parents and acquired German sequentially between the ages of 1 and 7 years (eL2).⁸ At the time of testing, all children were living in Germany. An overview of participant profiles is provided in Table 1.

Before taking part in the study, parents signed a consent form, approved by the ethics committee of the University of Konstanz, and completed a questionnaire about the child's language history. Information from the questionnaire was used to determine relative language dominance. A language score was calculated for both languages. The scores are comprised of four macro areas, with a maximum of 5 points each: informal quantity (i.e. time spent with people in informal contexts, e.g. *How much time does your child spend with the grandparents?*), formal quantity (i.e. time spent with people in formal contexts, e.g. *How many hours does your child receive formal instruction in school?*), informal quality (i.e. quality of input and language use in informal settings, e.g. *With how many people (outside of the family) does your child speak German and Italian?*), and formal quality (i.e. quality of input and language use in formal settings, e.g. *Is your child attending an Italian language class?*). For more details, see Furlani (2021).

Relative language dominance, shown in Table 2, was calculated by subtracting the Italian score from the German score. Positive scores indicate German dominance, negative scores indicate Italian dominance. Scores close to zero indicate balanced bilingualism. On an individual level, most children received a positive score, indicating that they are German-dominant. Herein, we treat dominance as a continuous variable, that is, in each group (German-dominant, Italian-dominant), a greater dominance indicates a greater difference between German and Italian exposure and language use.

Table 1. Participant profiles.

Group	Total N	Gender	Age (years)	N of 2L1	N of eL2	eL2 AoO German
6-year-olds	23	7m; 16f	6.5 (<i>SD</i> = .36, range = 6.1–6.9)	14	9	2.5 (<i>SD</i> = 1.27, range = 1–5)
7-year-olds	21	11m; 10f	7.5 (<i>SD</i> = .29, range = 7.07–7.9)	14	7	3 (<i>SD</i> = .8, range = 1.5–4)
8-year-olds	18	6m; 12f	8.7 (<i>SD</i> = .19, range = 8.2–8.9)	6	12	3 (<i>SD</i> = 1.46, range = 1.5–7)
9-year-olds	22	15m; 7f	9.5 (<i>SD</i> = .32, range = 9–10)	13	9	3 (<i>SD</i> = 1.7, range = 1.3–5.5)

Notes. m = male. f = female. 2L1 = simultaneous bilingual. eL2 = early sequential bilingual. AoO = age of onset.

Table 2. Overview of dominance and irony measures.

Group	Mean dominance score	Mean irony score	Parental irony score
<i>6-year-olds:</i>			
Mean	5	.88	5.24
SD	4.09	.16	1.54
Range	–8.52–11.84	.57–1	1–7
<i>7-year-olds:</i>			
Mean	5.9	.91	5.62
SD	3.95	.15	.49
Range	.09–12.84	.5–1	2–7
<i>8-year-olds:</i>			
Mean	2.7	.99	5.14
SD	5.81	.04	1.64
Range	–8.81–12.05	.86–1	1–7
<i>9-year-olds:</i>			
Mean	5.4	.97	5.3
SD	5.42	.08	1.81
Range	–7.64–12.47	.71–1	1–7

Given the similarity between irony and RhQs, we used different measures to assess children's ability to understand and use irony. This will allow to explore whether the two phenomena are driven by the same underlying factors. Therefore, all children completed an irony comprehension task based on Giustolisi et al. (2017). We selected seven out of 10 stories and translated them from Italian into German. Each story ends in a remark. In three of the stories, the ironic remark was expressing a criticism (target condition). The remaining four stories served as control conditions and contained literal remarks. In half of the remarks a compliment was expressed, in the other half a criticism (for details, see Giustolisi et al., 2017). The children completed the task in their preferred language as

Table 3. Overview of test items in the Perception and Comprehension Task.

Sentence	Translation
Wer mag (DiP) Bananen?	'Who likes (DiP) bananas?'
Wer spielt (DiP) Gitarre?	'Who plays (DiP) the guitar?'
Wer isst (DiP) Melonen?	'Who eats (DiP) melons?'
Wer will (DiP) Orangen?	'Who wants (DiP) oranges?'
Wer will (DiP) Sandalen?	'Who wants (DiP) sandals?'
Wer mag (DiP) Vanille?	'Who likes (DiP) vanilla?'

reported by the parents.⁹ Overall, accuracy was very high (94%) showing that the children were able to understand irony. Note that although there was higher variability in the 6-year-olds, their mean accuracy was above 87%, while the other groups reached over 90% (7-year-olds: 92%, 8-year-olds: 99%, 9-year-olds: 97%). As an additional measure, the questionnaire included a question on general irony use across languages in the family and on the child's general ability to comprehend and use irony as assessed by the parent, on a scale from 1 (minimum) to 7 (maximum). The results of this measure are reported in Table 2.

2 Materials and procedure

Two experiments were used to assess the children's acquisition of RhQs: a Perception Task (Experiment 1) and a Comprehension Task (Experiment 2). Children were tested in both tasks in both languages on different days. The order of the languages was counter-balanced.

a Experimental items. The experiments in both languages use string-identical *wh*-questions of the structure *Who eats bananas?* All questions contain the *wh*-word *wer*_{GERMAN}/*chi*_{ITALIAN} 'who', one out of four verbs, and one out of six nouns. In both languages, the nouns were trisyllabic with stress on the penultimate syllable (e.g. *BaNAnen* 'bananas'). The nouns used were expected to be known by children of the age tested and they do not (dis)favour an RhQ interpretation (for an overview of test items, see Table 3; for a list of all items, see Appendix 1 in supplemental material¹⁰).

The prosodic cues used in the German experiments are based on Kharaman et al. (2019). ISQs were produced with an early-peak accent (H+!H*) and normal (i.e. modal) voice quality, RhQs with a late-peak accent (L*+H) and a breathy voice quality on the *wh*-word. Recordings were manipulated in Praat (Boersma and Weenink, 2015) ensuring an overall shorter sentence duration for ISQs (-10% of mean duration) and longer sentence duration for RhQs (+10% of mean duration); see Figure 1.

Additionally, the questions were manipulated syntactically. Following Meibauer (1986: 199), the combination of the two DiPs *denn* and *schon* is used as a strong cue to rhetoricity (rhetorical condition). This condition was only used with RhQ prosody. The ambiguous condition contained the DiP *denn*, which can appear in both question types, ISQs and RhQs. The third syntactic condition does not contain any DiP (neutral

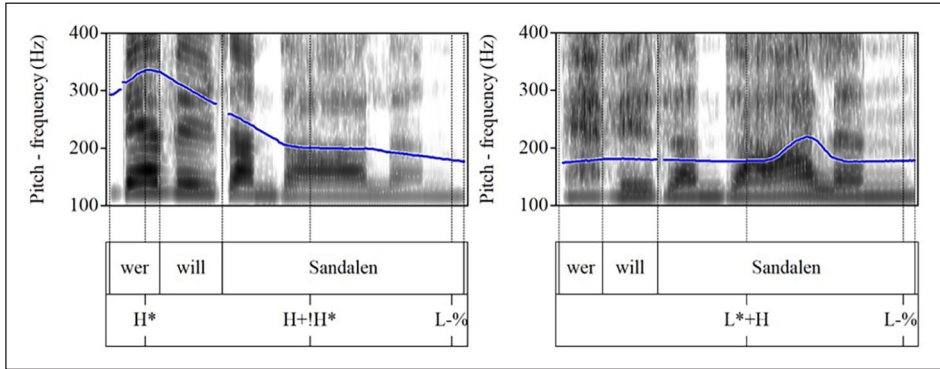


Figure 1. Example of a string-identical sentence pair: *Wer will Sandalen?!?! 'Who wants sandals?!?!'*. The ISQ (left) has an early-peak accent (H+!H*) and the RhQ (right) has a late-peak accent (L*+H).

Table 4. Overview of prosodic and syntactic cues used in the German experiments.

	ISQs	RhQs
Prosodic cue	H+!H* Modal voice quality Shorter duration	L*+H Breathy voice quality Longer duration
Syntactic cue	Neutral (no DiP) Ambiguous (<i>denn</i>)	Neutral (no DiP) Ambiguous (<i>denn</i>) Rhetorical (<i>denn schon</i>)

Note. ISQs = information-seeking questions. RhQs = rhetorical questions.

condition). This condition is treated as neutral as it does not contain any syntactic cue that might trigger a certain question interpretation. The filler items (yes/no ISQs and RhQs, e.g. *Does anyone like limes?!?!)* were created in the same way (for an overview, see Table 4). All questions were recorded by the first author, a female native speaker of German, in a sound-attenuated booth at the University of Konstanz. The stimuli were judged as natural and correctly identified as either RhQ or ISQ by two groups of monolingual German adults in two pilot studies.

The stimuli in the Italian Perception and Comprehension Tasks were designed in parallel to the German ones. Following Soriano (2018, 2019) and Ferin (2022) the prosodic and syntactic cues presented in Table 5 were selected. The items were recorded by the second author, a female native speaker of Italian. A study with monolingual Italian adults ensured the effectiveness of the cues. For a more detailed description of the Italian stimuli, see Ferin and Geiss (2022).

b Experiment 1: Perception Task. The aim of this task was to find out if the children are able to perceive the prosodic cues used in Experiment 2. Therefore, children listened to

Table 5. Overview of prosodic and syntactic cues used in the Italian experiments.

	Information-seeking questions (ISQs)	Rhetorical questions (RhQs)
Prosodic cue	Shorter duration	Longer duration
	Greater pitch excursion	Smaller pitch excursion
	Low edge tone	Rising edge tone
Syntactic cue	Neutral (no syntactic marker)	Neutral (no syntactic marker)
	Ambiguous (CLRD)	Ambiguous (CLRD)
		Rhetorical (<i>ma</i> + CLRD)

Note. CLRD = clitic right dislocation. ISQs = information-seeking questions. RhQs = rhetorical questions.

Table 6. Number of experimental and filler items used in Experiment 1 in each language.

	Same		Different	
	ISQ–ISQ	RhQ–RhQ	RhQ–ISQ	ISQ–RhQ
Neutral	4	4	2	2
Ambiguous	4	4	2	2
Rhetorical		4		
Filler			6	6
Total	20		20	

Note. ISQs = information-seeking questions. RhQs = rhetorical questions.

two string-identical questions and they had to decide whether the pairs sounded *gleich* ‘same’ or *unterschiedlich* ‘different’. The two options were presented as two cards containing two notes symbolizing the two options as well as *gleich* and *unterschiedlich* below the notes (in Italian: *uguale* and *diverso*, respectively).

Each task consisted of 20 experimental items of the category *same*, that is, sentence pairs with the same syntactic structure and the same prosody. Another eight experimental items were used in the category *different*, where the sentence pairs had the same syntactic structure but a different prosodic contour. In this category, half of the pairs started with an RhQ followed by an ISQ, while the other half had the opposite order. Additionally, 12 filler items (yes/no ISQs and RhQs) were used in the category *different* to balance the number of same and different pairs. This resulted in 40 items (Table 6). All sentence pairs were concatenated with an inter-stimulus interval of 1,000 ms and presented in a pseudo-randomized order.

The Perception Task was carried out online through Zoom. The children were tested individually and they were asked to be in a quiet room and to wear headphones, if possible. During the test, the children saw the PowerPoint slide with the two cards via screen sharing, while the experimenter played the sound files and noted down the answers on SoSci Survey (Leiner, 2019). The task started with a training session, consisting of six practice items that were similar to the experimental items. The children had to identify at least two sentence pairs of each category correctly in order to proceed to the actual task.

Table 7. Number of experimental and filler items used in Experiment 2 in each language.

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

Note. ISQs = information-seeking questions. RhQs = rhetorical questions.

The training session could be repeated up to three times. If the children did not pass the training, the task was interrupted.

c Experiment 2: Comprehension Task. The aim of this task was to find out whether children can identify ISQs and RhQs and which role prosodic and syntactic cues play in the interpretation. The Comprehension Task was a forced-choice experiment. To explain the difference between ISQs and RhQs types, the children were (re)introduced to two Disney characters, depicted on cards. The children knew the two characters from a previous production task. Rapunzel was introduced as a friendly, curious character and Drizella as Cinderella's grumpy unfriendly sister. The two characters stressed the different attitudes expressed by the two question types: inquisitiveness for ISQs (Rapunzel-type questions) and ironic criticism for RhQs (Drizella-type questions) (for full instructions, see Appendix 2 in supplemental material). Additionally, the character description included an example sentence. The children's task was to decide which character uttered the question.

This task consisted of 30 experimental items and 14 yes/no questions used as filler items (Table 7). The items were presented in pseudo-randomized order without context. The Comprehension Task took place after Experiment 1. Again, the experimenter used SoSci Survey to play the sound files and to note down the given answers while the children saw another PowerPoint slide with the characters. The task started with a training session with four practice items.

3 Statistical analyses

For the statistical analyses, we used generalized linear mixed-effects regression models (*glmer*) included in the R package *lme4* (Bates et al., 2015). The ANOVA and summary functions of the *lmerTest* package (Kuznetsova et al., 2017) were used to obtain *p*-values. To investigate interactions, we ran post-hoc tests based on pair-wise comparisons using the *emmeans* package (Lenth, 2020). For Experiments 1 and 2, we fitted logistic mixed-effects regression models with 'correct' (i.e. correctly identified sentence pair/question type) as binary categorical dependent variable. Different independent variables (IVs) were used in the models. 'Age group' (6-, 7-, 8-, 9-year-olds) was used in the between-group analyses. In Experiment 1, 'condition' (same, different) was used as an additional IV. In Experiment 2, the IVs 'prosodic cue' (ISQ, RhQ) and 'syntactic cue' (neutral,

ambiguous, rhetorical) were used to determine the linguistic effects. To investigate the effects of language dominance and irony comprehension, the IVs ‘dominance’ (continuous variable), ‘irony score’ (raw score obtained through Irony Comprehension Task), and ‘parental use of irony’ (information taken from the questionnaire) were added. The IV ‘language’ (German, Italian) was used to compare the children’s performance in Experiment 2 in their two languages using a within-group design. Participant and item were added as random effects. Random slopes for participant and item were included if they improved the fit of the model.¹¹ We followed stepwise model comparison based on the Akaike Information Criterion (AIC). The first model always contained all the factors and relevant interactions described above. Subsequently, we removed non-significant interactions and main effects.¹² Appendix 3 in supplemental material provides the model specification of the final model, and Appendix 4 in supplemental material presents the complete effects of the final models.

V Results

In this section, we briefly present the results of Experiment 1, which was the prerequisite for Experiment 2. Then, we address the potential effects of prosody and syntax (research question 1) and other potentially mediating factors (research question 2). Finally, we compare the children’s performance in German with their performance in Italian (research question 3).¹³ Each section begins with the descriptive statistics followed by the statistical analyses.

1 Perception Task

The results of the German Perception Task, summarized in Figure 2, show the mean participant accuracy for each condition and age group. Six children (two 6-year-olds, three 7-year-olds, one 8-year-old) did not pass the training session and were removed from all subsequent analyses.

Overall, the mean accuracy is very high in this task (94.8%). There are slight differences between the groups (6-year-olds: 91%, SD=12%; 7-year-olds: 95%, SD=6%; 8-year-olds: 98%, SD=6%; 9-year-olds: 95%, SD=7%). The effect of age group did not reach significance ($X^2=6.94$, $df=3$, $p=.07$). No significant effect of condition is found either ($X^2=1.64$, $df=1$, $p=.2$).

One 6-year-old child scored below chance (50%) in the *different* condition and was therefore excluded from the Comprehension Task analyses. To account for individual variation in the Perception Task and the fact that the ability to discriminate string-identical sentence pairs with a different prosody is an important prerequisite for the Comprehension Task, the children’s performance in the category *different* is used as a fixed effect in the following analyses.

2 Comprehension Task

The results of the German Comprehension Task are summarized in Figure 3, showing the response accuracy of each participant averaged over group, prosodic cue, and syntactic

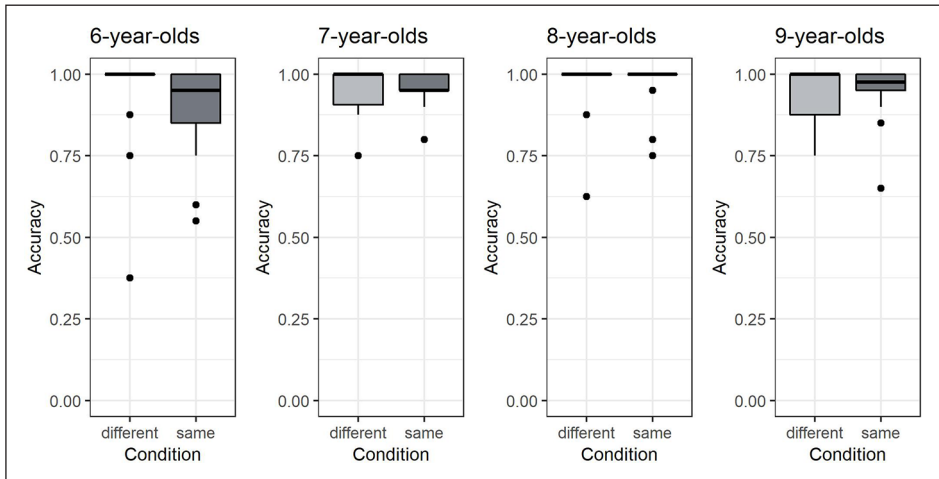


Figure 2. Accuracy, averaged over participant, per age group and condition (different vs. same).

cues. For all groups, mean accuracy is higher for ISQs than for RhQs, and there is considerably more variation in the RhQ condition (see Table 8). In the ISQ condition, the neutral condition tends to have a higher accuracy than the ambiguous condition, except for the 7-year-olds where the pattern is the opposite. In the RhQ condition, sentences with the rhetorical condition *denn schon* have the highest accuracy in all age groups, followed by the ambiguous condition *denn*. The comparison of the different age groups suggests that accuracy improves with age, except for the 9-year-olds who seem to perform worse than the 8-year-olds.

To address research question 1, we ran two mixed effects models. The first model includes only the neutral and ambiguous condition crossed with ISQ and RhQ prosody (two-level comparison). We find a significant effect of prosodic cue ($X^2=12.85$, $df=1$, $p<.001$). This analysis confirms that children, irrespective of age, have significantly more difficulties in identifying RhQs compared to ISQs ($\beta=1.81$, $SE=.51$, $z=3.58$, $p<.001$). In addition, we find an effect of age group ($X^2=8.74$, $df=3$, $p<.05$). A post-hoc test shows that 7-year-olds have significantly more problems identifying the two question types than 8-year-olds ($\beta=-1.53$, $SE=.59$, $z=-2.6$, $p<.05$). All other comparisons are not significant. As expected, the performance in the Comprehension Task is affected by the performance in the Perception Task (*different* condition), that is, a higher accuracy in the Perception Task is related to a higher accuracy in the Comprehension Task ($\beta=5.39$, $SE=2.22$, $z=2.43$, $p<.05$).

The second model compares the accuracy of all three levels (neutral, ambiguous, rhetorical) in the RhQ condition, leaving aside the ISQ condition (three-level comparison). We find a significant interaction between syntactic cue and age group ($X^2=19.63$, $df=6$, $p<.01$). A post-hoc test¹⁴ reveals a significant difference between the neutral and the rhetorical condition for all age groups; 6-, 7- and 9-year-olds perform significantly

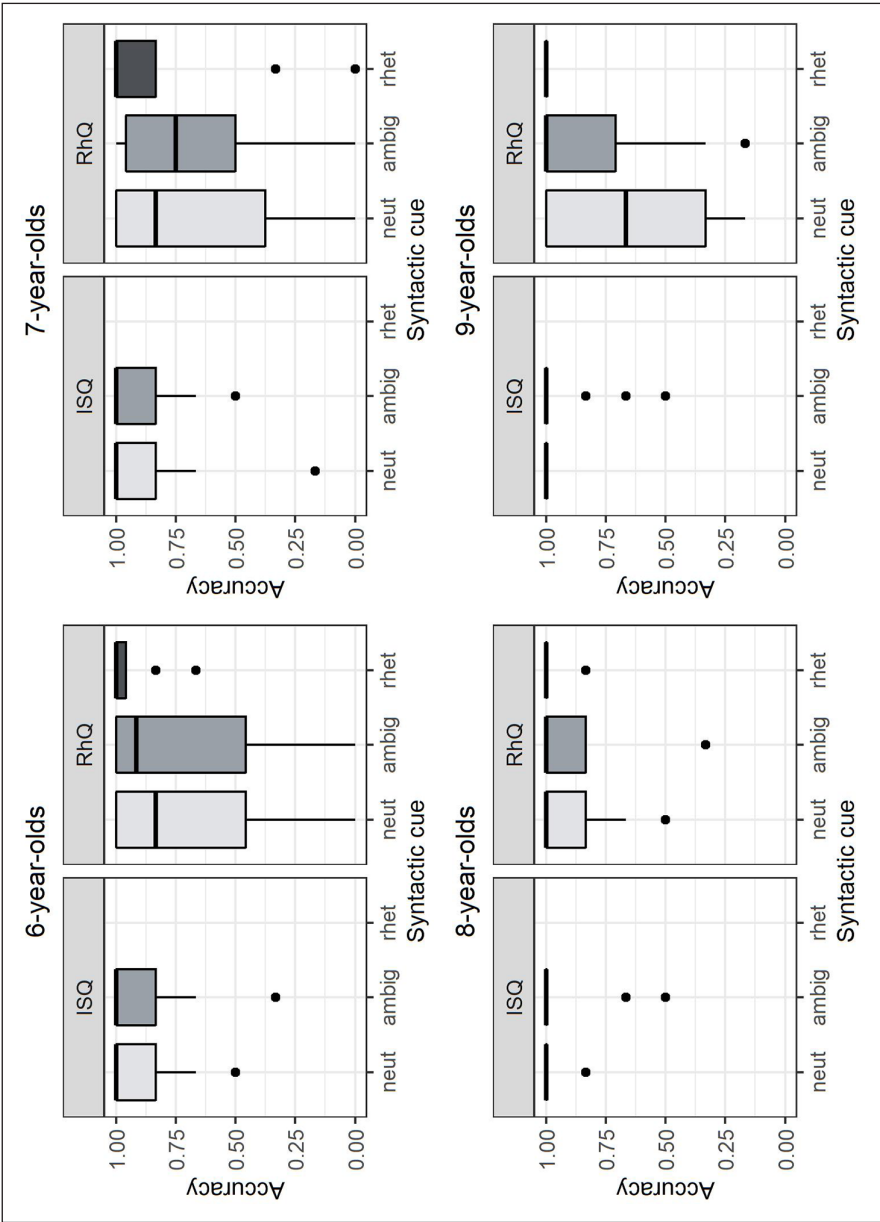


Figure 3. Response accuracy in the German Comprehension Task per age group, divided by prosodic cue (ISQ vs. RhQ) and syntactic cue (neutral, ambiguous, rhetorical).

better in the rhetorical compared to the ambiguous condition; 9-year-olds perform significantly better in the ambiguous compared to the neutral condition. All other comparisons are not significant.

Regarding research question 2, we find no effect of irony comprehension (two-level: irony score: $X^2=.11$ $df=1$, $p=.73$; parent use of irony: $X^2=.06$, $df=1$, $p=.81$; three-level: irony score: $X^2=.17$ $df=1$, $p=.68$; parent use of irony: $X^2=.51$ $df=1$, $p=.48$)¹⁵ and of dominance (two-level: $X^2=.39$, $df=1$, $p=.53$, three-level: $X^2=2.03$, $df=1$, $p=.15$) in either analysis.

3 Comparison between German and Italian

As a next step, we compare the children's performance in the HL with their performance in the ML (research question 3; for the ISQ condition, see Figure 4; for the RhQ condition, see Figure 5). One child was excluded from this analysis due to their poor performance in the Italian Perception Task.

As indicated by the mean accuracy per age group, children of all ages are better at identifying questions in German than Italian, irrespective of the question type except for the 8-year-olds who are slightly worse at identifying ISQs in German than in Italian (see Table 8).

The statistical analysis (two-level comparison) shows a significant three-way interaction between language, age group, and prosodic cue ($X^2=12.92$, $df=3$, $p<.01$). In the ISQ condition, there is no significant difference between German and Italian irrespective of age.¹⁶ In the RhQ condition, a post-hoc test shows that children of all age groups are significantly better in identifying RhQs in German than in Italian (6-year-olds: $\beta=1.45$, $SE=.29$, $z=5.06$, $p<.0001$; 7-year-olds: $\beta=.84$, $SE=.3$, $z=2.76$, $p<.01$; 8-year-olds: $\beta=2.37$, $SE=.38$, $z=6.23$, $p<.0001$; 9-year-olds: $\beta=.86$, $SE=.28$, $z=3.08$, $p<.01$). There was no significant effect of syntactic cue (neutral, ambiguous; $X^2=2.71$, $df=1$, $p=.1$).

In the three-level comparison (RhQ prosody only), the three-way interaction between language, age group, and syntactic cue is not significant ($X^2=7.91$, $df=6$, $p=.24$). However, we find a significant interaction between language and age group ($X^2=11.61$, $df=3$, $p<.01$), between age group and syntactic cue ($X^2=25.57$, $df=6$, $p<.01$), and between language and syntactic cue ($X^2=6.36$, $df=2$, $p<.05$). The post-hoc test reveals that the 6- and 8-year-old children were better in German than in Italian for all syntactic cues.¹⁷ The 7-year-olds are better in identifying questions with the neutral condition (RhQ prosody only; $\beta=.87$, $SE=.44$, $z=1.96$, $p<.05$) in German than in Italian. For the other conditions, the difference between the languages is not significant (ambiguous: $\beta=.7$, $SE=.44$, $z=1.57$, $p=.12$; rhet.: $\beta=.16$, $SE=.5$, $z=.32$, $p=.75$). The 9-year-olds show no significant difference between the neutral ($\beta=.5$, $SE=.42$, $z=1.21$, $p=.23$) and the rhetorical condition ($\beta=1.72$, $SE=1.16$, $z=1.49$, $p=.14$), but they are better at identifying the ambiguous condition in German than in Italian ($\beta=1.39$, $SE=.45$, $z=3.1$, $p<.01$).

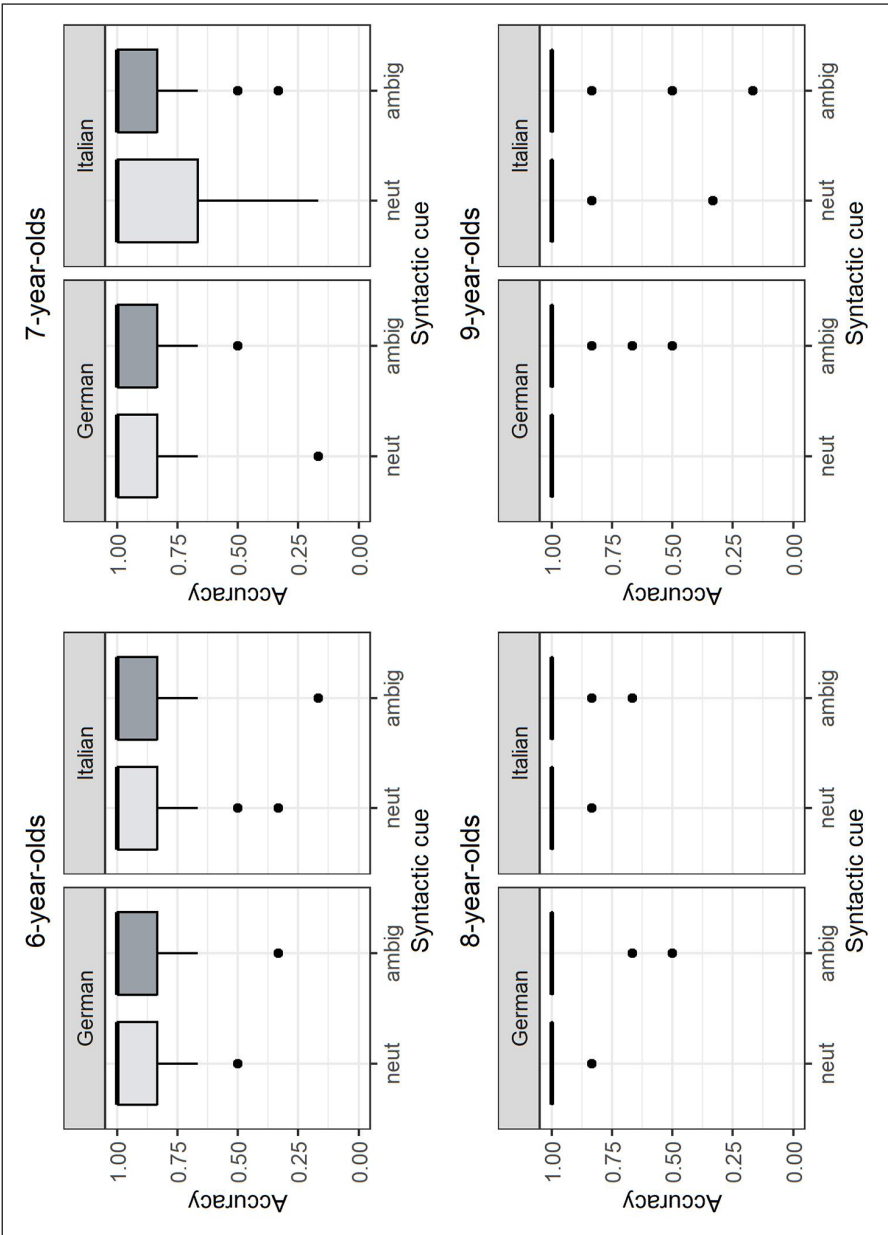


Figure 4. Mean response accuracy in ISQ condition of German and Italian Comprehension Task divided by age group and syntactic cue (neutral, ambiguous).

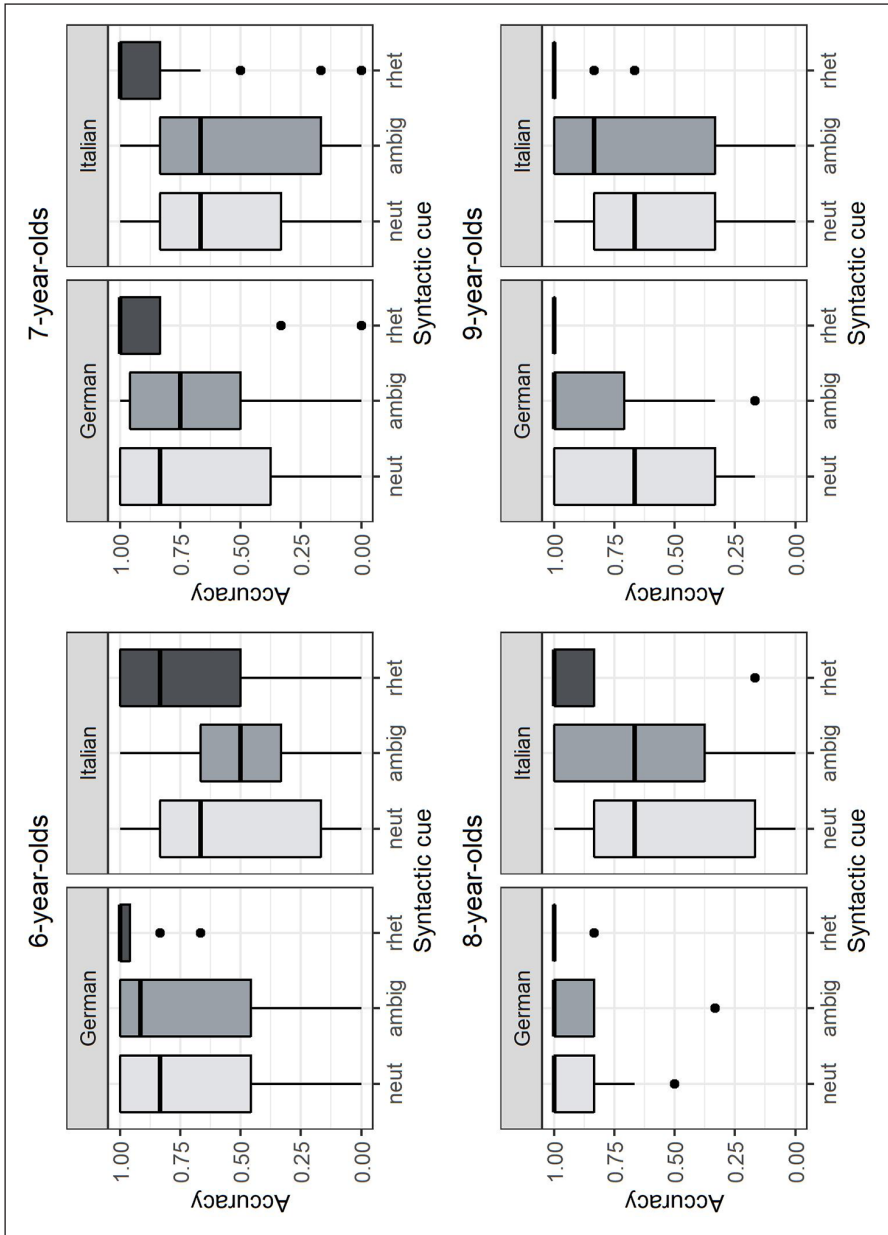


Figure 5. Mean response accuracy in RhQ condition of German and Italian Comprehension Task divided by age group and syntactic cue (neutral, ambiguous, rhetorical).

Table 8. Mean accuracy in percent (SD) in Comprehension Task by question type (ISQ vs. RhQ), language, and age group.

	German task		Italian task	
	ISQs	RhQs	ISQs	RhQs
6-year-olds	91 (14)	79 (23)	87 (18)	57 (28)
7-year-olds	90 (16)	71 (31)	84 (22)	66 (29)
8-year-olds	95 (10)	93 (10)	97 (6)	71 (27)
9-year-olds	97 (8)	83 (17)	93 (17)	73 (25)

Note. ISQs = information-seeking questions. RhQs = rhetorical questions.

VI Discussion

We examined the comprehension of RhQs in German by Italian–German bilingual children of four different age groups (6-, 7-, 8-, and 9-year-olds) by means of two experiments. In this section, we summarize our findings, discuss potential explanatory factors, and compare the children’s performance in the Comprehension Task in their two languages. For a more detailed discussion of the Italian data, see Ferin and Geiss (2022).

I Understanding RhQs in the ML

Overall, our results indicated that bilingual children of all age groups were able to identify RhQs in German, which suggests that they acquired the concept of RhQs (for RhQ comprehension in context for 4- and 6-year-olds, see Recchia et al., 2010). However, the fact that they were significantly better in identifying ISQs over RhQs shows that they have not yet acquired all relevant cues.¹⁸ This is in line with previous studies on question–statement differentiation (e.g. Saindon et al., 2016) and emotion recognition (Friend, 2000; Morton and Trehub, 2001), which reported improvement with age.

The first two research questions aimed at identifying which linguistic factors (prosodic and syntactic cues) provide stronger cues for RhQ comprehension, and whether RhQ comprehension is mediated by irony comprehension, language dominance, and age. In order to answer this question, the results of Experiment 1 (Perception Task) need to be taken into consideration. The aim of the Perception Task was to ensure that the children are sensitive to the prosodic cues used in Experiment 2 (Comprehension Task) and, more specifically, that they can tell apart the two prosodic contours (category *different*). Our results show, that, irrespective of age, the children were able to differentiate the two prosodic forms. However, since children varied in their performance, we included accuracy in this task as an additional predictor for the results of the Comprehension Task; see Section VI.2.

2 Factors influencing RhQ comprehension in the ML

First, as outlined in Section II, prosody is an important cue to differentiate ISQs from RhQs. Our results show that children were significantly better at identifying sentences

with ISQ prosody (above 90%) than sentences with RhQ prosody (71% to 83%). This comes as no surprise, since ISQs are canonical questions, which are very frequent in everyday speech and early acquired (for monolingual and bilingual children, see, for example, Lléo and Rakow, 2011). In addition, higher accuracy in the Perception Task went along with higher accuracy in the Comprehension Task (two-level comparison). This was expected because only children who can perceive the differences in the prosodic cues can also exploit them in comprehension.

Second, syntactic cues can serve as (additional) cues to rhetoricity. Our results show that *denn schon* (in combination with RhQ prosody) was the strongest cue, followed by the ambiguous condition *denn*, and the neutral condition (RhQ prosody only) with the lowest accuracy. The high accuracy for the rhetorical condition is in line with previous literature arguing that questions with *denn schon* are unambiguously rhetorical (Meibauer, 1986). The lowest accuracy for the neutral condition shows that RhQ prosody alone is not always strong enough to trigger an RhQ interpretation. Nevertheless, children did not ignore the prosodic cue (mean accuracy above 65%). If that had been the case, we would have expected more ISQ answers in this category. Instead, there was considerable individual variation in the 6- and 7-year-olds (and to a lesser extent also in the 8- and 9-year-olds), for the neutral and the ambiguous condition. That is, some children can consistently interpret the two conditions while others fail to do so. This suggests that not all children have acquired these cues and/or that they have difficulties basing their judgements on prosody. The latter would be in line with previous studies showing that the ability to weigh prosody over content is still developing in children between 6 and 9 years (Aguert et al., 2013; Chronaki et al., 2015; Friend, 2000; Morton and Trehub, 2001).

Third, the ability to understand RhQs can be mediated by other (non-)linguistic factors. As our data show, age plays an important role (two- and three-level comparison). While the 6- and 7-year-olds had more difficulties understanding RhQs when there was no rhetorical cue, 8-year-olds have mastered these cues. One possible explanation for this is that with increasing age, the children's mental abilities, such as Theory of Mind, become more advanced, which might in turn affect RhQ comprehension (for irony comprehension, see Banasik, 2013). However, to date, the relationship between RhQs and Theory of Mind development has not been investigated.

Finally, increasing age and language experience result in more opportunities to learn the cues relevant to rhetoricity. Relatedly, the children in our study receive not only input in the ML but also in the HL which might affect RhQ comprehension. Our results show that dominance, calculated based on a questionnaire, did not affect RhQ comprehension in the ML, but in the HL (for the results of the HL, see Ferin and Geiss, 2022; for similar findings on the role of dominance, see Yip and Matthews, 2007). In addition, given the similarities between RhQs and irony, we suspected that irony comprehension and RhQ comprehension would be correlated. However, we did not find an effect. Since accuracy in the Irony Comprehension Task was overall very high, in line with previous literature showing that 6-year-olds can already understand irony (Banasik and Podsiadło, 2016), it is possible that the Irony Task was not sensitive enough to reveal differences between the children. Alternatively, the ability to understand irony may develop earlier than the ability to understand RhQs. Based on our results, we cannot tease these two possibilities apart. Further investigation of the relationship between the acquisition of irony and RhQs is needed.

3 Comparing ML and HL

Research question 3 was concerned with a potential difference in RhQ comprehension between the ML and the HL. The results of the German and Italian Comprehension Task showed that bilingual children can identify ISQs equally well in their two languages based on prosody, while all groups were better at identifying RhQs in their ML than in their HL. A potential explanation is that the selected cues in the two languages, despite their semantic similarities, are not acquired at the same time. This offers the possibility to transfer prosodic and syntactic cues from one language to the other (for a discussion of cue transfer in multilingual settings, see Westergaard, 2021), which will be discussed in the following paragraphs.

First, the RhQs in this study were produced with distinct RhQ prosodies (i.e. prosodic cues) in both languages. It is important to keep in mind that the selected cues in Italian and German are not identical (i.e. they cannot be translated 1:1). With respect to prosody, both languages use duration as a cue to rhetoricity (Dehé et al., 2022). However, in German the nuclear pitch accent is crucial in differentiating RhQs and ISQs (Kharaman et al., 2019), while in Italian edge tones and pitch excursion are used (Ferin, 2022; Soriano, 2018, 2019). The data from the neutral condition (RhQ prosody only) shows that, in both languages, some children have difficulties interpreting questions as rhetorical and default to the canonical question interpretation (ISQ) instead. This is in line with previous research (Aguert et al., 2013; Chronaki et al., 2015; Friend, 2000; Morton and Trehub, 2001) and points to a later or slower development of prosody in its function of conveying additional pragmatic meaning (e.g. irony or rhetoricity). The question whether there has been prosodic cue transfer cannot be answered.

Second, we used two lexical-syntactic cues as additional markers for rhetoricity. In German two DiPs (*denn, denn schon*) were used, while in Italian CLRD and the particle *ma* in combination with CLRD were used. Similar to CLRD, the particle *denn* can be considered an information structure device, as it makes reference to a previous topic which is part of the common ground (Bayer and Obenauer, 2011). In addition, the use of *denn* signals that the speaker is concerned about the answer, which is not the case for CLRD, making *denn* a potentially stronger cue to rhetoricity than CLRD. In its rhetorical use, *schon* translates as ‘against expectations’ (Viesel and Freitag, 2019: 1), being similar to the counter-expectational adversative Italian particle *ma*.

When comparing the effects of the syntactic cues within RhQ prosody, in German both the ambiguous and the rhetorical condition had a facilitative effect in RhQ comprehension (at least for some age groups); in Italian only the rhetorical condition showed a facilitative effect. In other words, the particle *denn* was exploited as an additional cue by itself, unlike CLRD. In particular, 6-year-olds showed a lot of variation for all conditions in Italian, including the rhetorical condition (Figure 5), which is more reliably used at later ages. Conversely, *denn schon* was used as a rhetorical marker by the majority of children in German, including the 6-year-olds. This discrepancy points to a later development of RhQ comprehension in Italian than in German. Given the similarity of cues and the fact that in Italian the rhetorical condition is the condition that worked best across all groups, it is possible that the children transferred this cue to interpret RhQs from German to Italian. Alternatively, it is possible that children are good in

interpreting questions in Italian in the rhetorical condition because *ma* is a strong marker for non-canonical questions, which is frequently used in colloquial speech (Ferin, 2022), while the other cues need more time to be acquired.

The fact that children are better in the ML than in the HL could also be related to a slower acquisition process of the relevant cues of the HL compared to the ML, possibly due to a more limited amount of language experience. Monolingual baseline data are needed to determine whether the differences between the languages are imputable to purely linguistic factors, cue transfer, or bilingualism in general.

VII Conclusions

To date, research on how monolingual and bilingual children acquire RhQs is scarce. The present study addressed for the first time how bilingual primary school children comprehend RhQs in their ML (German), to what extent RhQ comprehension is acquired differently in the ML and the HL (Italian), and which linguistic cues facilitate RhQ comprehension. Our results show that bilingual children can identify RhQs in the ML and that RhQ comprehension is facilitated by the DiPs *denn* and *denn schon* accompanied by RhQ prosody. In the HL, a facilitative effect was found only for the particle *ma* and RhQ prosody. This might be the result of cue transfer or of the strength of the particle *ma* in signaling non-canonicity. In both languages RhQ comprehension improved with age, although in German the children already have a higher accuracy at the age of 6 years. This could be either an effect of the selected cues or of a higher proficiency in the ML, the dominant language for most of the children. In addition, dominance affected only the performance in Italian, while irony comprehension had no effect in either language. Taken together, these findings suggest that RhQs in bilingual children are acquired hand-in-hand in the two languages with a slight advantage in the ML, and their acquisition is not completed by age 6 years. Irrespective of age and language, some children have difficulties interpreting questions as rhetorical and interpreted them in the canonical way as ISQs. This has important implications for the acquisition of prosody, which is used to convey additional pragmatic meaning, such as rhetoricity.

Acknowledgements

We thank Sebastiano Arona, Alexandra Besler, Noemi Furlani, Carolin Hennerich, Chiara Ochsenreiter, and Mariapaola Piccione for help with the data collection and Massimiliano Canzi for help with the data analysis.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) to the project KU 2439/5-1 ‘Non-Canonical Questions in Early and Late Bilingual Language Acquisition’ within the research unit FOR2111 ‘Questions at the Interfaces’.

ORCID iD

Miriam Geiss  <https://orcid.org/0000-0002-1724-9748>

Notes

1. Herein, we do not differentiate between irony and sarcasm and use irony as an umbrella term for both phenomena.
2. Transcription follows Gili Fivela et al. (2015) for Italian and Grice and Baumann (2002) for German.
3. Herein, DiPs are considered part of syntax and will thus be considered syntactic cues.
4. The prosodic cues reported in Soriano (2018, 2019) were attested for the variety of Italian spoken in Bari, while the experiment in Ferin (2022) included speakers from all over Italy, proving that, at least in comprehension, these cues are also used in other varieties of Italian.
5. Both studies report a facilitative effect of prosody on irony comprehension. Capelli et al. (1990) recorded their stimuli with an ironic tone of voice, while Glenwright et al. (2014) described the prosodic form of their stimuli on the basis of F0 means and SDs.
6. In Friend (2000), children had to judge utterances as happy or angry. Utterances were recorded by a native speaker of English. The phonetic and phonological properties of the stimuli are not reported. The study by Morton and Trehub (2001) used happy paralinguistic consisting of a higher pitch level, greater pitch and a faster speaking rate, as opposed to sad paralinguistic which was realized with a lower and attenuated pitch and a slower speaking rate.
7. This study is part of a larger project on the acquisition of RhQs conducted at the University of Konstanz. For more details on the project, see <https://typo.uni-konstanz.de/questionsInterfaces/index.php/project-p10> (accessed November 2022).
8. Herein, 'sequential bilinguals' refers to children who acquire one language from birth and a second one sometime during early childhood.
9. The aim of this task was to test children's ability to understand irony in general. We chose to administer it in the child's preferred language to reduce the possibility of underperformance due to language ability rather than a failure to understand irony.
10. The appendix is provided either online (https://www.researchgate.net/publication/364694845_APPENDIX_Rhetorical_question_comprehension_by_Italian-German_bilingual_children) or by the first author.
11. Using syntactic cue as random slope was not possible due to convergence issues.
12. When a factor was not significant, it was removed from the model. In this case, we report the ANOVA output for that factor in the last model that included it before it was removed.
13. The full results of the Italian Perception and Comprehension Task are reported in Ferin and Geiss (2022).
14. For full presentation of results, see Appendix 4.
15. To be able to test for the effect of irony comprehension in the two- and three-level comparison, the random intercept item could not be included in the model due to convergence issues. The results of the final model were very similar to the model with item as random intercept.
16. For full presentation of results, see Appendix 4.
17. For full presentation of results, see Appendix 4.
18. We tested monolingual German adults and German-dominant heritage speakers of Italian with the same stimuli. For both groups, mean accuracy was above 95% for both question types.

References

- Ackerman BP (1982) Contextual integration and utterance interpretation: The ability of children and adults to interpret sarcastic utterances. *Child Development* 53: 1075–83.

- Aguert M, Laval V, Lacroix A, Gil S, and Le Bigot L (2013) Inferring emotions from speech prosody: Not so easy at age five. *PLoS One* 8: e83657.
- Antoniou K, Veenstra A, Kissine M, and Katsos N (2020) How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings? *Bilingualism: Language and Cognition* 23: 186–203.
- Banasik N (2013) Non-literal speech comprehension in preschool children: An example from a study on verbal irony. *Psychology of Language and Communication* 17: 309–24.
- Banasik N and Podsiadło K (2016) Comprehension of ironic utterances by bilingual children. *Psychology of Language and Communication* 20: 316–35.
- Bates D, Maechler M, Bolker BM, and Walkers SC (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.
- Bayer J and Obenauer H-G (2011) Discourse particles, clause structure, and question types. *The Linguistic Review* 28: 449–91.
- Bayer J and Struckmeier V (2017) The status quo of research on discourse particles in syntax and semantics. In: Bayer J and Struckmeier V (eds) *Discourse particles: Formal approaches to their syntax and semantics*. Linguistische Arbeiten 564. Berlin: de Gruyter, pp. 1–14.
- Bayer J, Häußler J, and Bader M (2016) A new diagnostic for cyclic wh-movement: Discourse particles in German questions. *Linguistic Inquiry* 47: 591–629.
- Biezma M and Rawlins K (2017) Rhetorical questions: Severing asking from questioning. *Proceedings of Semantics and Linguistic Theory (SALT)* 27: 302–22.
- Boersma P and Weenink D (2015) *Praat: Doing phonetics by computer: Version 6.1.05* [computer program]. Available at: <http://www.praat.org> (accessed November 2022).
- Braun B, Dehé N, Neitsch J, Wochner D, and Zahner K (2019) The prosody of rhetorical and information-seeking questions in German. *Language and Speech* 62: 779–807.
- Capelli CA, Nakagawa N, and Madden CM (1990) How children understand sarcasm: The role of context and intonation. *Child Development* 61: 1824–41.
- Caponigro I and Sprouse J (2007) Rhetorical questions as questions. In: Puig-Waldmüller E (ed.) *Sinn und Bedeutung 11*. Barcelona: Universitat Pompeu Fabra, pp. 121–33.
- Champoux-Larsson M-F and Dylman AS (2019) A prosodic bias, not an advantage, in bilinguals' interpretation of emotional prosody. *Bilingualism: Language and Cognition* 22: 416–24.
- Chronaki G, Hadwin JA, Garner M, Maurage P, and Sonuga-Barke EJS (2015) The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology* 33: 218–36.
- Coniglio M (2008) Modal particles in Italian. *Working Papers in Linguistics* 18: 91–129.
- Dehé N, Braun B, Einfeldt M, Wochner D, and Zahner K (2022) The prosody of rhetorical questions: A cross-linguistic view. *Linguistische Berichte* 269: 3–42.
- Dupuy L, Stateva P, Andreetta S, et al. (2019) Pragmatic abilities in bilinguals: The case of scalar implicatures. *Linguistic Approaches to Bilingualism* 9: 314–40.
- Ferin MF (2022) Rhetorical questions in Italian varieties. Paper presented at the Annual Conference of the Centre for Excellence in Estonian Studies, Subjectivity and Intersubjectivity in Language and Culture, Tartu, Estonia.
- Ferin MF and Geiss M (2022) Investigating rhetorical question comprehension in Italian by Italian–German bilingual children. *Lingue e Linguaggio* 21: 29–51.
- Friend M (2000) Developmental changes in sensitivity to vocal paralinguistic. *Developmental Science* 3: 148–62.
- Frota S, Butler J, and Vigário M (2014) Infants' perception of intonation: Is it a statement or a question? *Infancy* 19: 194–213.
- Furlani N (2021) Language abilities of Italian–German bilingual children. Unpublished master's thesis, University of Konstanz, Konstanz, Germany.

- Gili Fivela B, Avesani C, Barone M, et al. (2015) Intonational phonology of the regional varieties of Italian. In: Frota S and Prieto P (eds) *Intonation in Romance*. Oxford: Oxford University Press, pp. 140–97.
- Giustolisi B, Panzeri F, and Bogani A (2017) Irony comprehension from 4 to 9: The role of language and theory of mind. Poster presented at Generative Approaches to Language Acquisition 13 (GALA-13), Palma de Mallorca, Spain.
- Glenwright M, Parackel JM, Cheung KR, and Nilsen ES (2014) Intonation influences how children and adults interpret sarcasm. *Journal of Child Language* 41: 472–84.
- Grice M and Baumann S (2002) Deutsche intonation und GTöBl. *Linguistische Berichte* 191: 267–98.
- Kharaman M, Xu M, Eulitz C, and Braun B (2019) The processing of prosodic cues to rhetorical question interpretation: Psycholinguistic and neurolinguistics evidence. In: *Proceedings of Interspeech*, pp. 1218–22.
- Kuznetsova A, Brockhoff PB, and Christensen RH (2017) lmer Test package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26.
- Leiner DJ (2019) *SoSci Survey: Version 3.2.32*. Available at: <https://www.sosicisurvey.de> (accessed November 2022).
- Lenth R (2020) *emmeans: Estimated Marinal means, aka Least-Squares Means: R package version 1.4.7*. Available at: <https://CRAN.R-project.org/package=emmeans> (accessed November 2022).
- Lleó C (2018) Bilingual children’s prosodic development. In: Prieto i Vives P and Esteve-Gibert N (eds) *The development of prosody in first language acquisition*. Amsterdam / Philadelphia, PA: John Benjamins, pp. 317–42.
- Lleó C and Rakow M (2011) Intonation targets of yes/no questions by Spanish and German monolingual and bilingual children. In: Rinke E and Kupisch T (eds) *The development of grammar: Language acquisition and diachronic change: In Honour of Jürgen M. Meisel*. Amsterdam: John Benjamins, pp. 263–86.
- Lleó C, Rakow M, and Kehoe MM (2004) Acquisition of language-specific pitch accent by Spanish and German monolingual and bilingual children. *Laboratory Approaches to Spanish Phonology* 7: 3–27.
- Meibauer J (1986) *Rhetorische Fragen [Rhetorical questions]*. Tübingen: Niemeyer.
- Morton JB and Trehub SE (2001) Children’s understanding of emotion in speech. *Child Development* 72: 834–43.
- Neitsch J (2019) Who cares about context and attitude? Prosodic variation in the production and perception of rhetorical questions in German. Unpublished dissertation, Universität Konstanz, Konstanz, Germany.
- Obenauer H-G and Poletto C (2000) Rhetorical wh-phrases in the left periphery of the sentence. *Rivista di Grammatica Generativa* 26: 121–51.
- Recchia HE, Howe N, Ross HS, and Alexander S (2010) Children’s understanding and production of verbal irony in family conversations. *British Journal of Developmental Psychology* 28: 255–74.
- Sadock JM (1974) *Toward a linguistic theory of speech acts*. New York: Academic Press.
- Saindon MR, Trehub SE, Schellenberg EG, and Van Lieshout P (2016) Children’s identification of questions from rising terminal pitch. *Journal of Child Language* 43: 1174–91.
- Siegal M, Iozzi L, and Surian L (2009) Bilingualism and conversational understanding in young children. *Cognition* 110: 115–22.
- Soderstrom M, Ko E-S, and Nevzorova U (2011) It’s a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development* 34: 107–10.

- Sorianello P (2018) Tra prosodia e pragmatica: Il caso delle domande retoriche [Between prosody and pragmatics: The case of rhetorical questions]. *Studi e Saggi Linguistici* 56: 39–71.
- Sorianello P (2019) ‘A che serve saperlo?’ Funzioni pragmatiche e variazioni intonative della domanda retorica [‘What’s the use of knowing this?’: Pragmatic functions and intonational variations of rhetorical questions]. In: Nuzzo E and I Vedder (eds) *Studi AltLA 9: Lingua in contesto: La prospettiva pragmatica* [Language in context: The pragmatic perspective]. Milan: Officinaventuno, pp. 89–108.
- Thurmair M (1991) Zum Gebrauch der Modalpartikel ‘denn’ in Fragesätzen: Eine korpusbasierte Untersuchung [On the use of the modal particle ‘denn’ in interrogative clauses: A corpus-based study]. In: Klein E, Pouradier Duteil F, and Wagner KH (eds) *Betriebslinguistik und Linguistikbetrieb*. Tübingen: Niemeyer, pp. 377–88.
- Viesel Y and Freitag C (2019) Wer kann denn schon ja sagen? [Who can say yes?] *Zeitschrift für Sprachwissenschaft* 38: 1–56.
- Westergaard M (2021) Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research* 37: 379–407.
- Yip V and Matthews S (2007) Relative clauses in Cantonese–English bilingual children: Typological challenges and processing motivations. *Studies in Second Language Acquisition* 29: 277–300.
- Yow WQ and Markman EM (2011) Bilingualism and children’s use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition* 14: 562–69.