LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Boosting Bilingual Metalinguistic Awareness Under Dual Language Activation: Some Implications for Bilingual Education

Jacopo Torregrossa [a], Sonja Eisenbeiß, [b] and Christiane Bongartz [b]

[a]Goethe University of Frankfurt [b]University of Cologne

CRediT author statement – **Jacopo Torregrossa**: conceptualization; methodology; investigation; data visualization; formal analysis; writing – original draft preparation; writing – review & editing; funding acquisition; resource coordination. **Sonja Eisenbeiß**: conceptualization; methodology; data curation; writing – review & editing. **Christiane Bongartz**: conceptualization; methodology; writing – review & editing; funding acquisition; resource coordination.

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

We would like to thank the children for their enthusiastic participation in our experiments and their parents and teachers for supporting the study. We are also grateful to Maria Andreou, Ioli Baroncini, and Claudia Rizzo for help with the data collection and transcription and to Maximilian Hörl for advice in matters of statistics. The paper benefitted from the constructive comments of our reviewers and our editor as well as from our discussions with other colleagues, in particular Nicole Marx and the members of the department for "Sprache und Lernen" (language and learning) at the Mercator Institute for Literacy and Language Education. The research presented in this paper was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation, Project ID 281511265, SFB 1252 "Prominence in Language") for the project C03 "Reference management in bilingual narratives" at the University of Cologne (Principal Investigators Christiane Bongartz and Jacopo Torregrossa).

Correspondence concerning this article should be addressed to Jacopo Torregrossa, Institute for Romance Languages and Literatures, IG 6.157, Goethe University, Frankfurt, 60629 Frankfurt am Main, Germany. Email: torregrossa@lingua.uni-frankfurt.de

The handling editor for this article was Judit Kormos.

DOI: 10.1111/lang.12552

Abstract: Most studies on bilingual children's metalinguistic awareness assess metalinguistic awareness using monolingual tasks. This may not reflect how a bilingual's languages dynamically interact with each other in creating metalinguistic representations. We tested 33 Greek–Italian bilingual children (8–11 years) for metalinguistic awareness using acceptability-rating tasks in which they had to judge and explain grammatical errors. The tasks were in monolingual and bilingual modes in order to show how far metalinguistic awareness in Italian benefited from the activation of Greek. The participants exhibited better metalinguistic awareness abilities in Italian in the bilingual acceptability-rating task in which Greek was activated. The benefits of the bilingual mode were visible in the judgment and explanation of errors and were modulated by syntactic processing abilities in Italian, length of exposure to Italian, type of structure, and age. The results show that metalinguistic awareness can be shared across languages. We discuss the pedagogical implications of our findings.

Keywords bilingual; metalinguistic awareness; syntactic awareness; syntactic processing; grammaticality judgment; error explanation; biliteracy

Introduction

Over the last 20 years, the topic of bilingual education has received increasing attention from both pedagogical and psycholinguistic points of view (see Bialystok, 2018, for a review). Bilingual education can take different forms, including programs in which two languages are used as mediums of instruction or in which the language serving as the main medium of instruction is different from the language spoken at home or in the community (Baker & Wright, 2021).

Across the literature, there has been an ongoing debate on whether, in bilingual education programs, children's home language(s) should be included in classroom programs or whether the home and school language(s) should be kept separate in order to minimize interference between them (Ballinger et al., 2017; Cummins, 2019). Many studies have suggested that including children's home language(s) has a positive effect on their language and literacy development in several ways, for example, by enhancing children's self-confidence and motivation to learn school subjects (Bartlett & García, 2011; Carbonara & Scibetta, 2022; Creese & Blackledge, 2010; Cummins, 2019; García, 2009) and by boosting learning the school language (Bongartz & Torregrossa, 2020; Cenoz & Gorter, 2017; Luk & Lin, 2015; MacSwan et al., 2017) or a third language in the curriculum (Jessner, 2006). Through the use of their home language(s) in classroom, learners are encouraged to reflect on similarities and differences between the home and the school language(s) at different linguistic levels, for example, at lexical, phonological,

morphological, and syntactic levels (see the study by Kupferberg & Olshtain, 1996, and the pedagogical activities presented in studies by Celic & Seltzer, 2013, and Ballinger et al., 2017). This crosslinguistic reflection contributes to the development of learners' metalinguistic awareness.

In this paper, we aimed to contribute to the debate on whether children's metalinguistic abilities in the school language benefit from the activation of the home language on the basis of a study of Greek–Italian bilingual children ranging in age between 8 and 11 years. At the time of testing, all the participants lived in Greece and spoke Greek at home. They attended a school in which Italian was the main language of instruction. We compared the participants' performance on metalinguistic awareness tasks involving either the use of the school language only or the use of both the school and home language. We designed metalinguistic awareness tasks in which we manipulated the activation of the participants' home language. By doing so, we intended to simulate in a laboratory situation the effect of the use of the participants' home language during classroom activities involving metalinguistic reflections on the school language. Furthermore, our experimental setting allowed us to explore how the effect of the activation of the home language is mediated by variables such as age, language abilities in the school language, length of exposure to the school language, and crosslinguistic differences (or similarities) between children's two languages.

Background Literature

Metalinguistic Awareness: Adopting a Bilingual Lens

Across the literature, the term metalinguistic awareness has been associated with various meanings and used interchangeably with other terms such as language awareness and declarative knowledge (Jessner, 2008). James (1999) distinguished between two ways for learners to develop knowledge about language. The first works from the outside in: Learning about language leads learners to reflect on their own language (consciousness raising in James, 1999). The second involves an inside-out process: Learners have linguistic intuitions and reflect on them based on what they currently know about language.

Our study dealt with learners' syntactic awareness, which refers to the ability to "consciously reflect on, analyze and exert control over syntactic structures, over and above the unconscious processes of understanding and producing sentences" (Simard et al., 2017, p. 440). Simard et al. referred to two different aspects of syntactic knowledge. On the one hand, syntactic processing consists of the "ability to process meaningful language receptively and

productively" (Wu & Ortega, 2013, p. 683). According to Ellis (2009b), this knowledge is implicit, "tacit and intuitive" and "only evident in learners' verbal behavior" (Ellis, 2009b, pp. 11–13). On the other hand, syntactic awareness involves "analyzed knowledge (i.e., structured knowledge of which learners are consciously aware)" and, possibly, verbalization (Ellis, 2009b, p. 38).

Karmiloff-Smith (1992) showed that children seem to be able to engage in metalinguistic reflections from a very young age. In particular, children take their implicit linguistic knowledge as an object of cognitive attention and encode it in a more abstract form through a process of redescription. As children develop, their linguistic representations become progressively more accessible to conscious reflection and verbal report and increasingly associated with explicit knowledge. This progression follows a specific developmental path (D'Souza & Filippi, 2017). Crucially, children are able to redescribe only linguistic structures that they have acquired successfully (see Bialystok, 1988, 2001; Francis, 2012; Sharwood Smith, 2021, for a similar distinction between linguistic and metalinguistic knowledge).

In this study, we analyzed the creation of metalinguistic representations through a bilingual lens. In particular, we investigated whether and under which conditions bilingual children's metalinguistic representations are shared across their two languages. In this sense, metalinguistic awareness can be considered as a component of a common underlying proficiency, which is "the cognitive/academic proficiency that underlies academic performance in both languages" (Cummins, 2000, p. 38; see also Cummins, 1979, 1981). Evidence in favor of this concept of a common underlying proficiency has come from studies of bilingual children's morphological awareness, understood as the ability to manipulate, analyze, and reflect on the smallest units of meaning of language (e.g., Carlisle, 2003; Chung et al., 2019). For example, Ramirez et al. (2010) tested 97 Spanish-speaking children learning English as a second language (L2) in Canada and found correlations between the scores for children's knowledge of derivational suffixes both in English and in Spanish. Likewise, Zhang (2013) showed that Chinese six graders ($N = 204$) learning English as a L2 in China were able to apply their awareness of compound morphology in Chinese to the interpretation of unknown English compound words. Ke et al. (2021) conducted a meta-analysis of correlational coefficients reported in studies on the shared morphological awareness between bilinguals' two languages. They found a small but significant correlation between morphological awareness as assessed both in bilinguals' first language (L1) and in their L2. To the best of our knowledge, bilinguals' syntactic awareness has never been investigated for crosslinguistic sharing.

Measuring Syntactic Awareness at Different Levels of Explicitness

In this study, we assessed bilingual children's syntactic awareness using a grammaticality judgment task combined with an error explanation task. Grammaticality judgment tasks involve a focus on formal aspects of language rather than on meaning (and, hence, on control over syntactic structures, as mentioned in the quote by Simard et al., 2017, above). They may comprise several subtasks such as learners' judging whether a sentence is grammatical or not, identifying possible errors and correcting them, and explaining the reason behind ungrammaticality (e.g., Spada et al., 2015). Notably, these subtasks differ from each other on a continuum of explicitness (Ellis, 2009a; Karmiloff-Smith, 1992). For example, several scholars have claimed that the task of judging sentences as (un-)grammatical taps into implicit knowledge since it is "based on failure of the parser… rather than on conscious reflection about syntax" (Simard et al., 2017, p. 444). However, other studies—mainly in L2 research—have provided solid empirical evidence that grammaticality judgment tasks measure L2 learners' explicit language knowledge. This applies, in particular, to the judgment of ungrammatical (vs. grammatical) sentences (Gutiérrez, 2013) and to grammaticality judgment tasks performed without any time pressure (Ellis, 2005; Loewen, 2009; see Plonsky et al., 2020, for a comprehensive meta-analysis). According to Leow (2015), error identification, correction, and explanation are associated with increasing levels of awareness and depth of processing (see also Leow & Mercer, 2015; Schmidt, 1990, 2001). Error identification requires the lowest level of awareness, that is, awareness at the level of noticing. Both error correction and explanation involve verbalization but differ from each other in the associated level of awareness: Error correction only involves awareness at the level of reporting, whereas error explanation is associated with awareness at the level of understanding. When explaining errors, speakers describe the regularities underlying linguistic data activating their previous knowledge and their knowledge of metalanguage (Ellis, 2005; Leow, 2015; Zhao et al., 2021). On the basis of these considerations, we characterized the identification, correction, and explanation of grammatical errors as three different components of syntactic awareness involving different degrees of explicitness.

Variables Mediating the Development of Bilingual Metalinguistic Awareness

The literature review in the previous section has suggested that some aspects concerning the creation of metalinguistic representations across languages deserve further investigation. The studies reported in our literature review

were based on correlational analyses between two measures of metalinguistic awareness, one taken at Time 1 in the L1 and the other at Time 2 in the L2 (cf. Ke et al., 2021). Crucially, this has made it difficult to determine whether the L1 contributes to the development of metalinguistic representations in the L2 or whether metalinguistic representations develop in parallel, separately for each language (cf. Zhang et al., 2010).

Furthermore, in our literature review, we identified three tasks that tap into different levels of explicitness of syntactic awareness, that is, error identification, correction, and explanation. It is still unclear whether the interaction of a bilingual's two languages in creating metalinguistic representations involves all of these levels of explicitness.

As a last methodological consideration, it should be pointed out that two variables seem to mediate the crosslinguistic sharing of metalinguistic abilities among bilingual children: the typological distance of the two languages involved and language proficiency. For the role of typological distance, crosslinguistic sharing of metalinguistic representations related to a certain linguistic phenomenon is more likely to occur if the phenomenon at stake is productive in both languages (Koda, 2005, 2008). For example, Zhang (2013) showed that children's compound awareness in Chinese contributed to their compound awareness in English much more than their derivational awareness in Chinese did to their derivational awareness in English. While compounding is highly productive in both languages, derivational processes in Chinese are, in general, less productive than in English. Other studies have shown that crosslinguistic sharing of morphological awareness is likely to occur from a more transparent to a less transparent language (Chung et al., 2018; Hayashi & Murphy, 2013; Ramirez et al., 2010). From a morphosyntactic point of view, a language is more transparent than another if it allows for a greater number of one-to-one mappings between morphosyntactic forms and linguistic functions (Liceras et al., 2012; Torregrossa & Bongartz, 2018). For example, inflectional morphology is more transparent in Spanish than in English, since, for instance, person and number are more consistently marked in Spanish than in English. The notion of transparency is particularly relevant for the language constellation of Greek and Italian in this study, especially in the domain of case-marking.

Another variable that may modulate the crosslinguistic sharing of metalinguistic awareness is learners' language proficiency or their language abilities in general. Bialystok has shown that at (relatively) lower levels of language proficiency, bilinguals tend to perform worse than monolinguals on tasks of syntactic awareness like the one used in our study (Altman et al., 2018; Barac & Bialystok, 2011; Bialystok, 2001; Friesen & Bialystok, 2012). Focusing on

the sharing of metalinguistic awareness across languages, some studies have reported that the language in which bilingual children are more proficient tends to boost their metalinguistic representations in the other language (Chung et al., 2018, 2019; Zhang et al., 2010, 2014). Crucially, this means that whenever bilinguals' proficiency in one language is not able to support the creation of metalinguistic representations, they can rely on the corresponding metalinguistic representations in their other language.

To our knowledge, no study has investigated how far length of exposure to the L2 affects the sharing of metalinguistic awareness across languages. Children with less L2 exposure might not have had enough time to develop metalinguistic representations in their L2. As a result, while performing tasks tapping into their metalinguistic awareness in their L2, they might rely on the metalinguistic representations developed in their L1 more heavily than would children with greater L2 exposure similarly to what has been previously discussed with reference to proficiency (e.g., see Blom & Paradis, 2016; Paradis, 2011, on the effects of language exposure variables on bilingual children's acquisition outcomes). Our study would be the first to explore how far the effect of the activation of the L1 on children's metalinguistic awareness in the L2 is mediated by length of L2 exposure.

The Present Study

In our study, we investigated Greek–Italian bilingual children who were exposed to the language of school (Italian) between birth (if they came from bilingual families) and the age of 72 months (upon entrance in school) and were attending an Italian immersion primary school in Athens, Greece, at the time of testing. The study's aim was to investigate how far the activation of bilingual children's home language (Greek) affects their syntactic awareness in the school language (Italian). Specifically, we examined whether the effects of the activation of Greek are visible at different levels of explicitness of syntactic awareness in Italian, that is, judgment and explanation of grammatical errors. Furthermore, we intended to identify which variables modulate the effects of the activation of Greek on the participants' syntactic awareness in Italian, with our focus being on the role of syntactic processing abilities in Italian, length of exposure to Italian, and type of ungrammatical syntactic structure (considering typological differences between Greek and Italian). We formulated our research questions considering the abovementioned two levels of explicitness (judgment and explanation). With respect to the judgment of ungrammatical sentences in Italian, we asked the following research questions:

1. Does children's ability to judge ungrammatical sentences in Italian benefit from the activation of Greek?
2. How do the (possible) benefits of the activation of Greek on the judgment of ungrammatical sentences in Italian relate to children's syntactic processing abilities in Italian, length of exposure to Italian, and type of syntactic structure?

With respect to the explanation of ungrammatical sentences in Italian, we asked the following research questions:

3. Does children's ability to explain the errors contained in ungrammatical sentences in Italian benefit from the activation of Greek?
4. To what extent are the (possible) benefits of the activation of Greek on the explanation of grammatical errors in Italian modulated by children's syntactic processing ability in Italian, length of exposure to Italian, type of syntactic structure, and age?

We included the role of age in the explanation of grammatical errors since children's ability to verbalize metalinguistic knowledge is a more elaborate ability than is the judgment of ungrammatical sentences and seems to increase with age.

Method

Participants

We tested 33 Greek–Italian bilingual children (16 females) ranging in age from 7 years 11 months to 11 years 7 months (M_{age} = 9 years 4 months, SD = 12 months). We conducted the study in a school where Italian was the main medium of instruction with a total number of 24 hours of instruction per week. All subjects were taught in Italian, and Italian was also offered as a language subject. Greek was taught as an additional language for 5 hours per week. Crucially, the main language of the school, Italian, was different from the language spoken in the society, Greek. Before the study, the parents provided written informed consent and all the participants were told that they did not have to take part in the activity if they did not want to do so. The participants were all born and raised in Greece except for one who had arrived in Greece at the age of 1 year. The sample included 19 simultaneous bilinguals (i.e., exposed to both Italian and Greek from birth) and 12 sequential bilinguals. Among the sequential bilinguals, six were first exposed to Italian at age 3 years (upon entrance in kindergarten). For the remaining six participants, exposure to Italian began at age 6 years when they had entered elementary school.[1] Both the parents and

the teachers reported that none of the participants had previously identified speech, hearing, or visual impairments.

Research Instruments

Background Questionnaires

We administered a background questionnaire to the parents in order to identify patterns of language and literacy exposure outside school, with particular attention to the participants' exposure to Greek (the home language) and length of exposure to Italian (the school language; see Bongartz & Torregrossa, 2020; Caloi & Torregrossa, 2021; Mattheoudakis et al., 2016, for the design of the questionnaire). For the analysis presented in this paper, we used only the information related to the participants' length of exposure to Italian (cf. the methodological discussion in our literature review). We had no specific hypothesis related to the effect of the quantity of language exposure to Greek or Italian on bilingual participants' metalinguistic abilities in Italian. In Appendix S1 in the Supporting Information online, we have described the structure and analysis of the questionnaire information. In Appendix S2 in the Supporting Information online, we have reported the results of the analysis of the questionnaire. The participants' language dominance profile emerging from the analysis was useful for our interpretation of some of the results of the study.

Sentence Repetition Task

We used a sentence repetition task (SRT) in Italian as a measure of the participants' syntactic processing abilities, that is, of their implicit linguistic knowledge. Research on L2 acquisition and child bilingualism has shown that SRTs involve sentence processing in both comprehension and production (Andreou et al., 2021; Klem et al., 2015; Marinis & Armon-Lotem, 2015; Spada et al., 2015). For comprehension, participants have to decode and interpret the stimulus sentences. For production, participants have to reconstruct the meaning of the sentences and reproduce them, which involves lexical retrieval, grammatical encoding, and phonological realization by participants.

The SRT that we used for this study consisted of 25 sentences targeting 28 structures. In Appendix S3 in the Supporting Information online, we have described the design of the SRT and the procedure for its administration and reported a complete list of all sentences and target structures of the SRT. We based the choice of these structures on existing SRTs (e.g., Marinis & Armon-Lotem, 2015). We have made our SRT available through a public OSF profile (https://osf.io/kfqy4) and in Torregrossa et al. (2022) through IRIS (https://www.iris-database.org).

Metalinguistic Awareness Tasks

We designed three metalinguistic awareness tasks, in order to assess the participants' ability to judge and explain grammatical errors in three different conditions:

- when the task included only single sentences in Italian, which could be either grammatical or ungrammatical—a monolingual Italian one-sentence task (coded as MAT1 in our R analyses) serving as a baseline task;
- when the task included both single and pairs of sentences in Italian—a monolingual two-sentence task (coded as MAT2 in our R analyses); among the pairs of sentences, we juxtaposed grammatical Italian sentences and their ungrammatical Italian counterparts;
- when the task included both single sentences in Italian and pairs of sentences in Italian and Greek—a bilingual two-sentence task (coded as MAT3 in our R analyses); the pairs of sentences involved the activation of Greek by our pairing each grammatical or ungrammatical Italian sentence with its grammatical or ungrammatical Greek counterpart.

The use of the monolingual Italian two-sentence task allowed us to explore whether the participants' syntactic awareness benefited from the presence of two sentences, one grammatical and one ungrammatical, independently of the activation of Greek. By contrast, the bilingual two-sentence task involved both two sentences and dual language activation.

In each of these three tasks, the participants had to listen to sentences (some grammatical and some not) that described pictures (see Figures 1 and 2). Then, they had to rate their grammatical acceptability based on a 5-point Likert scale. Finally, the participants had to explain the reason of their ratings for all sentences that they did not rate with the highest level of the Likert score. As we discussed in our literature review, rating and explaining tap into different degrees of explicitness of syntactic awareness.

For a child-friendly version of a Likert scale, we followed Ambridge et al. (2008) and represented the different degrees of grammatical acceptability by means of five faces with expressions that ranged from a frown to a smile: two levels of frowning faces, two levels of smiling faces, and a neutral face in the middle (see Figures 1 and 2). We additionally marked the divide between grammatical acceptability and unacceptability with color: The smiley faces were green, the frowning faces red, and the face in the middle was half green and half red. Beside emphasizing the distinction between acceptable and unacceptable sentences, the use of colors permitted the elicitation of binary judgments in case the participants, especially the youngest ones, could not provide graded

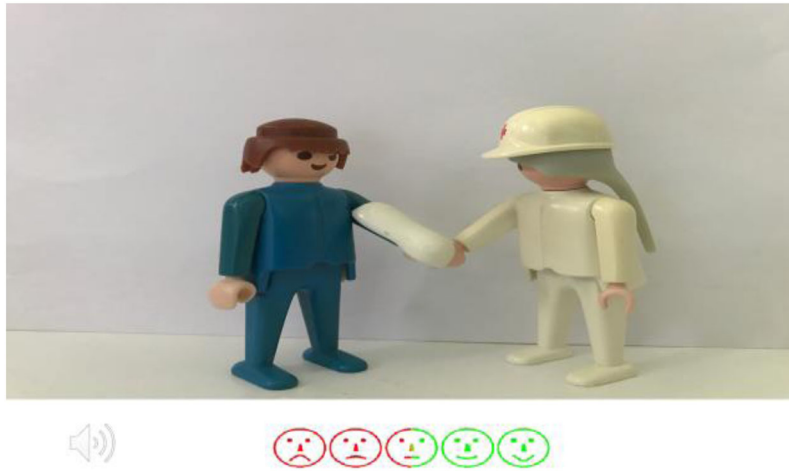

Figure 1 Picture included in the monolingual one-sentence task. The corresponding sentence is *Guarda il paziente! L'infermiera lo* $_{CL.ACC.MASC.SG.}$* *cura la ferita* "Look at the patient! The nurse heals him* the wound (treats his wound)." The sentence contains a clitic (*lo*) which is incorrectly case-marked (%err: *gli* $_{CL.DAT.MASC.SG.}$).

Figure 2 Picture included in the bilingual two-sentence task. The corresponding sentences are: *Oi agrotes* $_{PL.}$ *mazevoun* $_{PRES.3PL.}$ *louloudia* "The farmers collect flowers" (in Greek, correct) and *I contadini* $_{PL.}$ *raccoglie* $_{PRES.3SG.}$ *i fiori* "The farmers collects the flowers" (in Italian). The Italian sentence involves incorrect subject–verb agreement (%err: *raccolgono* $_{PRES.3PL.}$).

judgments (Ambridge et al., 2008, p. 105). However, we did not need to rely on this option in our study: All the participants were able to use the full range of possible answers.

The Target Structures

In each task, the ungrammaticality of the sentences involved violations of subject–verb agreement or incorrect case, gender, or number markings on clitics (see Appendix S4 in the Supporting Information online). The choice of these structures was mainly motivated by the observation that nominal and verbal inflectional morphology are particularly vulnerable in bilingual language acquisition (see Benmamoun et al., 2013, for a general review; Torregrossa & Bongartz, 2018, on clitic production by Italian heritage speakers). Furthermore, these structures differ from each other in the extent to which they are similar across Greek and Italian. The subject–verb agreement and the grammatical number system are similar in both languages: Verbal morphology marks three persons, and verbal and nominal morphology both use two numbers, singular and plural. However, Greek and Italian differ in how they mark case. Greek consistently marks case on pronouns, determiners, and full nouns, distinguishing between nominative, genitive, accusative, and vocative. In contrast, Italian marks case only on accusative and dative clitics and nominative, accusative, and dative personal pronouns (Torregrossa et al., 2020).

For gender marking, Greek has a three-way system, masculine, feminine, and neuter, whereas Italian has a two-way system, masculine and feminine. In both languages, gender is marked on noun endings, attributive and predicative adjectives, determiners, and pronouns. Furthermore, word suffixes can be used to predict gender in the majority of cases (see Egger et al., 2018, for Greek; Chini, 1995, for Italian). Gender marking on nouns and adjectives is conflated with case and number marking in Greek, whereas gender marking is conflated only with number marking in Italian. This makes the gender system less transparent in Greek compared to Italian.[2]

Design of the Three Metalinguistic Awareness Tasks

Figure 1 and Figure 2 illustrate experimental items from the monolingual Italian one-sentence task and the bilingual two-sentence task, respectively. All tasks are available via the OSF (https://osf.io/kfqy4) and in Torregrossa et al. (2022). In each task, eight single sentences occurred, that is, two for each target structure (subject–verb agreement and case, gender, and number-marking on clitics), one correct and one incorrect. The Italian one-sentence task contained eight additional sentences (i.e., two for

each target structure, one correct and one incorrect) and, hence, 16 sentences in total. By contrast, the Italian two-sentence task involved two pairs of sentences for each target structure, where one sentence was correct and one sentence incorrect. Thus, the task contained eight pairs of sentences (16 sentences) in addition to the abovementioned eight single sentences, that is, 24 sentences in total. Finally, the bilingual two-sentence task featured two pairs of sentences for each target structure, one correct in Greek and one incorrect in Italian or one incorrect in Greek and one correct in Italian. Thus, the task contained eight pairs of sentences (16 sentences), of which eight were in Italian (four correct and four incorrect) and eight in Greek (four correct and four incorrect), in addition to the eight single sentences, that is, 24 sentences in total. It should be noted that for each type of structure, both the monolingual Italian one-sentence task and the bilingual two-sentence task involved four sentences in Italian (two correct and two incorrect), while the monolingual Italian two-sentence task contained eight sentences in Italian (four correct and four incorrect). Therefore, the number of items per type of structure was low. We took this into consideration in the statistical analysis. However, we decided not to increase the number of items in order not to make the task too long and challenging for the participants. In each task, we randomized the order of the trials (corresponding to the different types of structures) by using a computer-generated sequence of numbers. We did not randomize the stimuli across subjects.

Procedure

We tested the children in three different sessions, each session one week apart. Each session was dedicated to a different metalinguistic awareness task. We varied the order of administration of the three metalinguistic tasks across the participants, considering all possible orderings (six in total). We always administered the sentence repetition task after the monolingual Italian one-sentence task because this task was the shortest one among the three metalinguistic awareness tasks.

Data Analysis

Analysis of the Sentence Repetition Task

After transcribing the participants' answers, we coded them according to whether the participants were able to repeat the target structure or not. We gave 1 point if a participant's repetition of the target structure was accurate, independent of the accuracy of the whole sentence. In contrast, we gave 0 points if a participant made an error in the target structure or substituted another

structure for the target structure, for instance, by using an active verb instead of a passive verb (cf. Marinis & Armon-Lotem, 2015, for this methodology). SRTs can be analyzed in many different ways, for example, on the basis of verbatim repetitions, the grammaticality of the sentence as a whole, or the omission of function and content words. We decided to consider only the accuracy of the reproduction of the target structure since some studies have suggested that it reflects children's syntactic processing abilities more accurately than do the other measures of analysis because it is less affected by vocabulary knowledge (Hamann & Abed Ibrahim, 2017). The first author and a student assistant, both native speakers of Italian, independently coded the participants' answers. The interrater reliability expressed as a percentage agreement was 97%. The item reliability of the SRT, measured using Cronbach's alpha based on the scores of the participants in this study ($N = 33$), was .67, which can be considered as a fair to moderate one (Brown, 2014).

Analysis of the Metalinguistic Awareness Tasks

Since the main focus of the study was on children's ability to judge ungrammatical sentences and explain their errors, our statistical model included only the incorrect sentences in the three metalinguistic awareness tasks. For the analysis of the Likert scores, we converted the five-faces Likert scale into a Likert scale from 1 (*saddest face*) to 5 (*happiest face*) and used these scores as the outcome variable in a linear mixed-effects model. Before doing this, however, we separately conducted a Cronbach's item reliability analysis of the three metalinguistic awareness tasks. This analysis revealed that the reliabilities of the monolingual Italian one-sentence task and the monolingual Italian two-sentence task were at an acceptable level, with alphas of .84 and .81, respectively, whereas the reliability of the bilingual two-sentence task was moderate, with an alpha of .77.

For the analysis of the participants' explanations, after transcribing their answers, we gave each explanation a score ranging from 0 to 2. We gave 0 points whenever the participants did not judge the sentence as incorrect, were unable to give an explanation (i.e., "I don't know", "I am not sure"), or the explanation was wrong as in Example 1a or irrelevant as in Example 1b.

Example 1

a. Target sentence: *Guarda il cameriere! La cliente la* ACC.FEM.SG. *chiama*.
"Look at the waiter! The female customer calls her."

Child 06: "La chiama" è sbagliato perché non è due, è uno. ["Calls her" is wrong, because it is not two, it is one.]

b. Target sentence: *Guarda il paziente! L'infermiera lo* ACC.MASC.SG. *cura la ferita*. "Look at the patient! The nurse him heals the wound."
Child 09: "La dottoressa cura la ferita" non è giusto. Si dice "la dottoressa cura la mano." ["The female doctor heals the wound" is not right. You have to say "The female doctor heals the hand."]

We gave 1 point if participants corrected a sentence (e.g., *LO chiama*, *not LA chiama* "calls *him*, not calls *her*," after the target structure in Example 1a). The participants scored 2 points if they referred explicitly to relevant form–function relations such as the incorrect mapping between a third person singular and a plural subject as in Example 2a, the impossibility of using an accusative masculine clitic to refer to a female discourse referent as in Example 2b, or the use of accusative clitics for direct objects as in Example 2c.

Example 2

a. Target sentence: *Gli infermieri cura* PRES.3SG. *il malato*. "The nurses heals the sick."
Child 13: Dice "cura," deve essere "curano" perché sono tanti, plurale. [She says "heals," it must be "heal" because they are many, plural.]

b. Target sentence: *Guarda la studentessa! Lo studente lo* ACC.MASC.SG. *tocca*. "Look at the female student! The male student hits him."
Child 14: È una femmina e deve essere "LA tocca." [It is a female and it should be "hits HER."]

c. Target sentence: *Guarda il gelataio! La cuoca lo* ACC.MASC.SG. *prepara la cena*. "Look at the ice-cream man! The cook prepares him the dinner."
Child 20: ha detto "LO prepara," perché sennò prepara l'uomo e non la cena. [She said "prepares HIM," because otherwise, (she) prepares the man and not the dinner.]

As Examples 2a, 2b, and 2c show, some participants provided both a correction and an explanation. In these cases, the occurrence of the explanation was sufficient for ascribing 2 points to the item. The distinction between 1 point and 2 points in the explanation score was meant to reflect the distinction between correcting and explaining, which corresponded to two different degrees of refinement of verbalization of syntactic awareness, with explanation being associated with a higher degree of awareness than correction since it involved

knowledge of metalanguage (cf. our literature review). The first author and a student assistant independently coded the participants' explanation scores. A Cronbach's interrater reliability analysis revealed substantial interrater agreement, with an alpha amounting to .91.

Statistical Analyses

For the statistical analysis of the Likert scores, we used R (R Core Team, 2015) and lme4 (Bates et al., 2015) to fit a series of linear mixed-effects models, each related to one (or more) of the predictors of interest in our investigation. For each model, we considered the Likert scores (from 1 to 5) for ungrammatical sentences as the dependent variable. In the first model, we used the type of task (Italian one-sentence task, Italian two-sentence task, and bilingual two-sentence task) as a fixed effect, choosing the Italian one-sentence task as the reference level. There is some debate as to whether it is advisable to run linear mixed-effects models with Likert scales as outcome variables, given that the outcome is ordinal and not continuous (e.g., Norman, 2010). Therefore, we also modeled the data related to the first model by using an ordinal logistic regression. In the second model, the predictor was the interaction of type of task with the participants' syntactic processing abilities, whereas in the third model, the predictor was the interaction of type of task with length of exposure to Italian. In the fourth model, we used the interaction of type of task with type of structure as a predictor, choosing the Italian one-sentence task and subject–verb agreement as reference levels. However, we must note that for each structure, the number of ungrammatical items was low across all tasks, especially in the Italian one-sentence task and the bilingual two-sentence task (which included two ungrammatical structures of each type). Therefore, the results of the analysis related to the fourth model should be taken with caution. For all four models, we fit the models with random slopes for participants using a maximal random effects structure and random intercepts for items unless convergence issues arose. We chose an alpha level of .05 to determine statistical significance. We used partial eta squared for effect sizes. In order to interpret the effect sizes, we referred to Cohen's (1988) benchmarks, with a partial eta squared of .0099 indicating a small effect, a partial eta squared of .0588 indicating a medium effect, and a partial eta squared of .1379 indicating a large effect (see also Richardson, 2011). In order to identify relevant pairwise contrasts whenever needed, we used the function emmeans in the emmeans R package (Lenth, 2020).

For the analysis of the explanation scores, we used nnet (Venables & Ripley, 2002) to perform a multinomial logistic regression for the

explanation task with its possible values of 0, 1, and 2 as our outcome variable and type of task (Italian one-sentence task, Italian two-sentence task, and bilingual two-sentence task), type of structure (subject–verb agreement, clitic-gender, clitic-number, and clitic-case), the participants' syntactic processing abilities in Italian, length of exposure to Italian, and age as predictor variables. We mean-centered the variables participants' syntactic processing abilities in Italian, length of exposure to Italian, and age in order to make the intercept more interpretable. In multinomial logistic regression models, the estimates refer to log odds ratios. We chose the score 0 as the reference level of the outcome variable.[3] The main idea in carrying out a multinomial logistic regression is to model the probability of a categorical outcome variable (with three or more levels, as in our case) as a linear combination of the predictor variables.[4] For this model, we calculated marginal R^2 for effect size.

Results

Background Questionnaires and Sentence Repetition Task

Table S2.1 in Appendix S2 in the Supporting Information online provides a general overview of the participants' profile in terms of dominance of exposure in Greek or Italian. Overall, the participants appeared to be Greek-dominant in their home language history as well as in their current language use. By contrast, they were Italian-dominant in their literacy practices outside school. Table S2.1 in Appendix S2 in the Supporting Information online also reports the descriptive statistics related to the participants' length of exposure to Italian and the scores that they obtained in the SRT. For their SRT scores, the participants exhibited a relatively high degree of accuracy, with the worst result corresponding to a score of 18 (out of 28), indicating correct repetition of almost 65% of syntactic structures. Table S2.2 in Appendix S2 in the Supporting Information online reports the correlation matrix between the participants' age, length of exposure to Italian, and SRT scores that we used to determine potential multicollinearity among the predictors of the models that we used for the analysis.

Metalinguistic Awareness Tasks

Likert Scores

The violin plot in Figure 3 shows the distribution of the Likert scores assigned to ungrammatical sentences across the three tasks. The mean for the Likert scores was 1.66 ($SD = 1.21$) for the Italian one-sentence task, 1.51 ($SD = 1.00$) for the Italian two-sentence task, and 1.41 ($SD = 0.94$) for the bilingual two-sentence task.

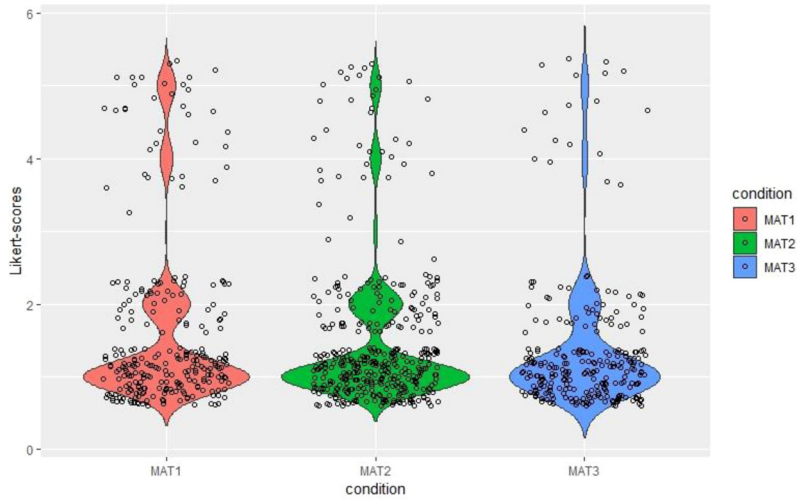

Figure 3 Violin plot of the Likert scores assigned to ungrammatical sentences across the three tasks (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task) with individual data points with jitter on top of it. MAT1 = monolingual Italian one-sentence task; MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

Table 1 Parameters of the linear mixed-effects analysis concerning the Likert scores assigned to ungrammatical sentences for type of task (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task)

Fixed effects	*b*	*SE*	95% CI	*df*	*t*	*p*
Intercept	1.64	0.16	[1.33, 1.95]	34.89	10.06	<.001
Type of task [MAT2]	−0.12	0.11	[−0.33, 0.08]	37.94	−1.15	.260
Type of task [MAT3]	−0.23	0.11	[−0.45, −0.02]	39.86	−2.09	.043

Note. MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

Research Question 1 concerned whether the activation of Greek in the bilingual two-sentence task enhanced the participants' ability to judge ungrammatical sentences in Italian. Table 1 reports the statistical results for each level of the fixed effect type of task (the Italian two-sentence task and the bilingual two-sentence task) compared to the reference level (the Italian one-sentence task).[5] There was a significant effect of the bilingual two-sentence task. The negative estimate indicated that in the bilingual two-sentence task,

Table 2 Parameters of the linear mixed-effects analysis concerning the Likert scores assigned to ungrammatical sentences for type of task (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task) and sentence repetition task scores

Fixed effects	*b*	*SE*	95% CI	*df*	*t*	*p*
Intercept	1.62	0.16	[1.32, 1.93]	32.33	10.10	<.001
Type of task [MAT2]	−0.17	0.11	[−0.38, 0.04]	31.51	−1.54	.130
Type of task [MAT3]	−0.27	0.11	[−0.48, −0.06]	38.33	−2.47	.020
Sentence repetition task	−0.28	0.14	[−0.56, −0.01]	7.14	−1.96	.090
Type of task [MAT2] × SRT	0.18	0.11	[−0.02, 0.39]	34.75	1.72	.090
Type of task [MAT3] × SRT	0.32	0.10	[0.13, 0.52]	34.12	3.14	.003

Note. MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task; SRT = sentence repetition task.

the participants tended to assign lower scores to ungrammatical sentences than in the Italian one-sentence task. We found no effect of the Italian two-sentence task. The partial eta squared of the effect of type of task was .09, which is a medium effect. The same pattern of results emerged when we modeled the data using ordinal logistic regression (see our discussion below and the results presented in Appendix S5 in the Supporting Information online).

Research Question 2 concerned how the benefits of the activation of Greek in the judgment of ungrammatical sentences in Italian related to the participants' syntactic processing abilities in Italian. Specifically, we were interested in understanding whether the benefits of the activation of Greek in the bilingual two-sentence task varied with the participants' syntactic processing abilities as measured by the SRT.[6] The results of the linear mixed-effects analysis reported in Table 2 revealed that there was a significant lower-order effect of the bilingual two-sentence task, indicating that, in the bilingual task, the participants tended to assign lower scores (negative estimate) to ungrammatical sentences than they did in the Italian one-sentence task (when the SRT was at its mean value). The significant interaction of the bilingual two-sentence task with the SRT suggested that the difference between the Likert scores assigned to ungrammatical sentences in the Italian one-sentence task and in the bilingual two-sentence task tended to be more pronounced among the participants with lower syntactic processing abilities (see Figure 4). In other words, the activation of Greek in the bilingual two-sentence task seemed to enhance the participants' ability to judge ungrammatical sentences in Italian, especially among the participants with more reduced syntactic processing abilities in Italian. The partial eta squared of both the effect of type of task and the effect of the SRT was .13,

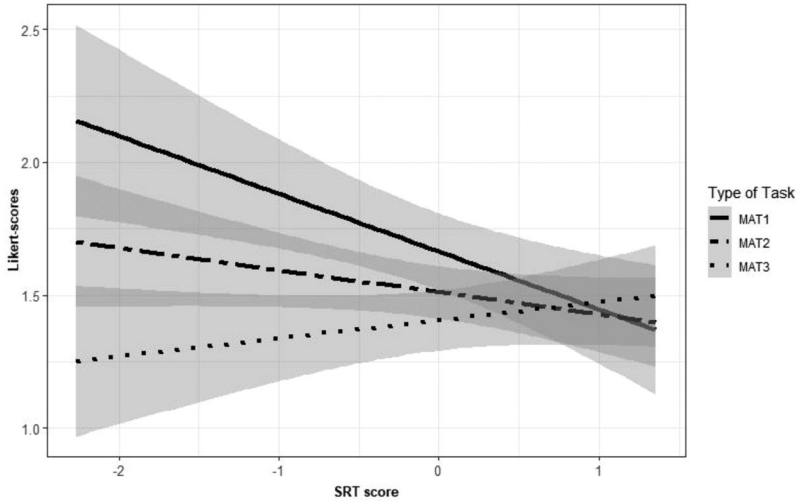

Figure 4 Linear regression lines related to the effect of syntactic proficiency (mean-centered sentence repetition task score in the x-axis) on the Likert scores associated with ungrammatical sentences across the three tasks (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task). Shaded areas indicate the 95% confidence regions. The figure was created by using the function ggplot of the ggplot2 package (Wickham, 2016). MAT1 = monolingual Italian one-sentence task; MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

which corresponds to a medium effect. The partial eta squared of the interaction between type of task and the SRT scores was .21, which is a large effect.

Research Question 3 concerned whether the benefits of the activation of Greek in the bilingual two-sentence task varied based on the participants' length of exposure to Italian.[7] The results of the linear mixed-effects analysis reported in Table 3 did not show any significant effect. The partial eta squared of the effect of type of task was .11 (a medium effect), the partial eta squared related to the effect of length of exposure to Italian was .01 (a small effect). However, the partial eta squared of the interaction of type of task and length of exposure was .32 (a large effect).

Finally, Research Question 4 asked whether the benefits of the activation of Greek in the bilingual two-sentence task could be observed with all syntactic structure or just some of them. The model and the results are reported in Table S6.1 of Appendix S6 in the Supporting Information online. The linear mixed-effects analysis showed a significant lower-order effect of type of

Table 3 Parameters of the linear mixed-effects analysis concerning the Likert scores assigned to ungrammatical sentences for type of task (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task) and length of exposure to Italian

Fixed effects	*b*	*SE*	95% CI	*df*	*t*	*p*
Intercept	1.69	0.18	[1.35, 2.03]	28.00	9.45	<.001
Type of task [MAT2]	−0.12	0.12	[−0.35, 0.10]	29.17	−1.03	.310
Type of task [MAT3]	−0.25	0.12	[−0.48, 0.01]	30.88	−1.98	.057
Length of exposure (Italian)	−0.04	0.17	[−0.37, 0.29]	7.41	−0.24	.820
Type of task [MAT2] × Length of exposure (Italian)	−0.10	0.11	[−0.32, 0.10]	11.42	−0.95	.360
Type of task [MAT3] × Length of exposure (Italian)	0.08	0.12	[−0.16, 0.30]	9.62	0.61	.560

Note. MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

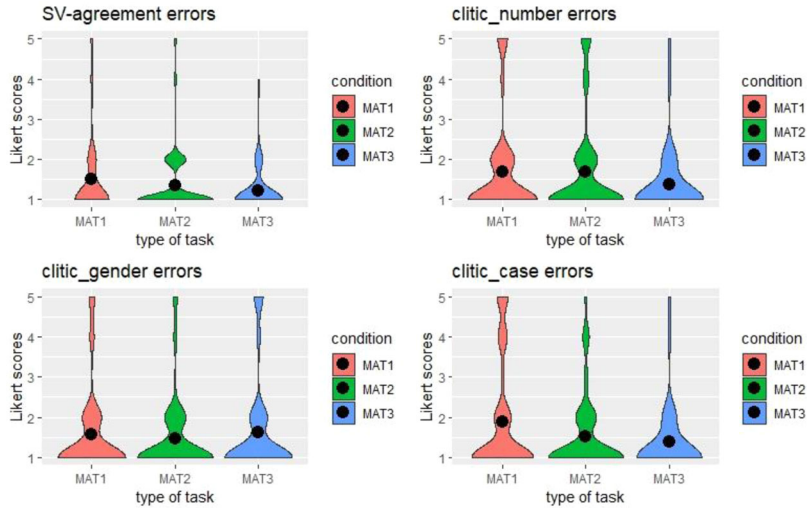

Figure 5 Violin plots of the distribution of Likert scores assigned to ungrammatical sentences across the three tasks (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task) and the four types of structures (subject–verb agreement, number, gender, and case marking on clitics). The black dots indicate the mean value for the Likert scores. SV = subject–verb; MAT1 = monolingual Italian one-sentence task; MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

structure (incorrect case-marking) only, indicating that in the Italian one-sentence task, the participants tended to assign higher scores (positive estimate) to ungrammatical sentences exhibiting incorrect case marking on clitics than to ungrammatical sentences exhibiting incorrect subject–verb agreement (the reference level of the corresponding variable). The emmeans analysis revealed that the Likert scores that the participants assigned to sentences exhibiting incorrect number marking on clitics tended to be higher in the Italian two-sentence task than in the bilingual two-sentence task, and the ones that they assigned to sentences exhibiting incorrect case marking on clitics tended to be higher in the Italian one-sentence task than in the Italian two-sentence task and the bilingual two-sentence task (see Table S6.2 in Appendix S6 in the Supporting Information online, where we have reported all pairwise comparisons). Figure 5 also illustrates these contrasts.

For the model reported in Table S6.1 in Appendix S6 in the Supporting Information online, the partial eta squared of the effect of type of task was

.13, the partial eta squared of the effect of type of structure was .20. and the partial eta squared of the effect of the interaction of type of task with type of structure was .02. These are considered to be medium, large, and small effects, respectively.

Explanation Scores

The aim of the multinomial logistic regression analysis for the participants' explanation scores was to investigate how far the activation of Greek in the bilingual two-sentence task enhanced the participants' ability to explain the errors contained in the ungrammatical sentences in Italian, considering the role of the participants' syntactic processing abilities, length of exposure to Italian, age, and type of syntactic structure. As in the case of the analysis of the Likert scores, we considered the results of the analysis related to the effect of the type of structure as not conclusive. Therefore, we decided to report the analysis related to all predictors in Table S7.1 in Appendix S7 in the Supporting Information online, whereas the analysis reported below does not account for the role of type of structure.

Table 4 reports the results of our multinomial logistic regression analysis for each predictor. Overall, the prediction capacity of the model results reported in Table 4 was 73% and its adjusted R^2 (calculated by using the function r2_mcfadden of the R package performance; Lüdecke et al., 2021) was .16. The log odds of providing an explanation scored as 1 (vs. explanation scored as 0) and explanation scored as 2 (vs. explanation scored as 0) significantly increased by 0.60 and 0.71, respectively, when we compared the Italian one-sentence task to the bilingual two-sentence task. In contrast, the coefficients associated with the Italian two-sentence task were not significant in either cases. Figure 6 plots the predicted probabilities for each category (0, 1, and 2) across the three types of tasks (the Italian one-sentence task, the Italian two-sentence task, and the bilingual two-sentence task). With the bilingual two-sentence task, there was a clear decrease in the probability of observing an explanation score of 0, and as a consequence, a relative increase in the probability of observing a score of 1 or 2.

Furthermore, Table 4 shows that a one-unit increase in the variable SRT was associated with a significant increase in the log odds of providing an explanation scored as 1 (vs. explanation scored as 0) and an explanation scored as 2 by 0.78 and 0.61, respectively. There was also a significant increase in the log odds of providing an explanation scored as 2 (vs. explanation scored as 0) by 0.70 if a participant's age increased by one unit. Finally, we observed a significant decrease in the log odds of providing an explanation scored as 1 (vs.

Table 4 Parameters of the multinomial logistic analysis concerning the explanation scores associated with ungrammatical sentences for type of task (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task), participants' age, score in the sentence repetition task, and length of exposure to Italian

Predictors	Explanation[a]	Log *OR*	95% CI	*SE*	*p*[b]
Intercept	1	1.58	[1.21, 1.95]	0.19	<.001
	2	−0.20	[−0.69, 0.29]	0.25	.430
Type of task [MAT2]	1	−0.13	[−0.59, 0.33]	0.23	.570
	2	0.09	[−0.50, 0.69]	0.30	.760
Type of task [MAT3]	1	0.60	[0.04, 1.16]	0.29	.036
	2	0.71	[0.01, 1.41]	0.36	.046
Sentence repetition task	1	0.78	[0.57, 0.99]	0.11	<.001
	2	0.61	[0.33, 0.89]	0.14	<.001
Length of exposure (Italian)	1	−0.32	[−0.56, −0.09]	0.12	.007
	2	−0.20	[−0.50, 0.11]	0.16	.200
Age	1	0.01	[−0.24, 0.26]	0.13	.940
	2	0.70	[0.39, 1.01]	0.16	<.001

Note. The reference category for comparing MAT2 and MAT3 was the monolingual Italian one-sentence task. MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task; *OR* = odds ratio. [a]A value of 1 indicates that an explanation scored as 1 was compared to a baseline explanation scored as 0; a value of 2 indicates that an explanation scored as 2 was compared to a baseline explanation scored as 0. [b]Calculated using the Wald test.

explanation scored as 0) by 0.32 if a participant's length of exposure to Italian increased by one unit.

Discussion

Differences Between Monolingual and Bilingual Modes

The first result emerging from our study was that the participants tended to rate ungrammatical sentences with higher Likert scores in the monolingual one-sentence syntactic awareness task compared to the bilingual two-sentence task. In other words, the activation of Greek in the bilingual two-sentence task boosted the participants' sensitivity to the ungrammaticality of sentences in Italian. The positive effect of the bilingual task was also visible in the participants' ability to explain grammar errors. Crucially, we did not find the same effects when we compared the rating of ungrammatical sentences in the Italian two-sentence task with the rating in the Italian one-sentence task.

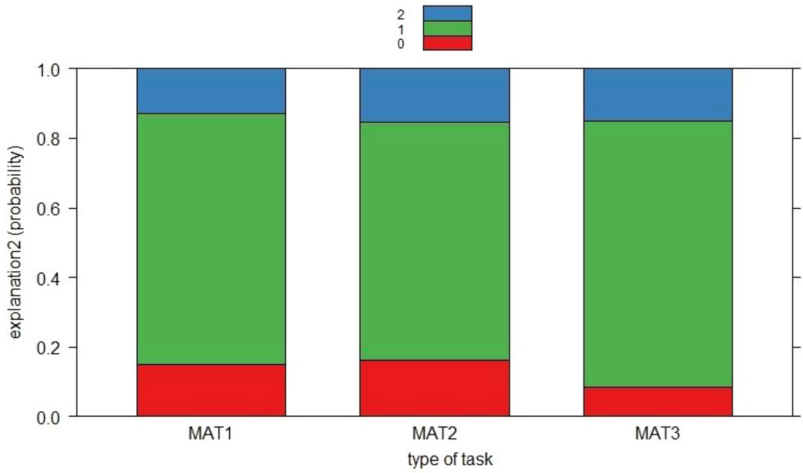

Figure 6 Predicted probability plot for each outcome category of the explanation score (0, 1, 2) across the three types of tasks (Italian one-sentence task, Italian two-sentence task, bilingual two-sentence task). The plot was created using the effects package (Fox & Hong, 2009) based on the lattice library (Sarkar, 2008). MAT1 = monolingual Italian one-sentence task; MAT2 = monolingual Italian two-sentence task; MAT3 = bilingual Greek–Italian two-sentence task.

Therefore, our study suggests that the participants' sensitivity to ungrammaticality benefits from the activation of Greek more than from the mere occurrence of two sentences in the task. Both the Italian two-sentence task and the bilingual two-sentence task were supposed to enhance the participants' sensitivity to form–function relationships. However, they did this in different ways. While the Italian two-sentence task activated the grammatical alternative to an ungrammatical sentence, the bilingual two-sentence task encouraged the participants to reflect, in a more or less explicit way, upon differences and similarities between their two languages. Therefore, in the bilingual task, the comparison between grammatical and ungrammatical sentences operated at a more abstract level, involving metalinguistic representations that are possibly shared between the two languages: The participants seemed to be able to connect the metalinguistic representation of a linguistic structure in Italian to the corresponding metalinguistic representation in Greek. This crosslinguistic sharing of metalinguistic awareness was visible at all levels of awareness and depth of processing considered in our study, that is, error identification, correction, and explanation. Following Leow (2015), we argued that error correction and

explanation involve more explicit knowledge than error identification due to verbalization. Furthermore, error explanation requires a deeper level of processing since it involves the use of metalanguage.

Overall, our study suggests that metalinguistic representations can be shared across a bilingual's two languages. A similar conclusion has been reached in previous studies based on correlational analyses between measures of metalinguistic awareness in children's L1 and L2. Our study adds to the previous literature by establishing a causal (rather than correlational) inference between metalinguistic abilities in a bilingual's two languages. The comparison between tasks in monolingual and bilingual mode was crucial for us to draw this causal inference.

Effects of Syntactic Processing Abilities

Another finding that emerged from our study is that the positive effects related to the activation of Greek on the participants' syntactic awareness in Italian were mediated by their syntactic processing abilities in Italian. The participants with lower syntactic processing abilities exhibited the greatest improvement in their syntactic awareness when Greek was activated compared to when they performed the task in a monolingual mode (the Italian one-sentence and two-sentence tasks). The pattern observed in Figure 4 suggests that when performing the task in a bilingual mode, the participants with lower syntactic processing abilities caught up with their peers with higher processing abilities.

The participants' syntactic processing abilities in Italian affected the extent to which they were accurate in their correction and explanation of sentence ungrammaticality (see Table 4; see also Phillips & Ehrenhofer, 2015; Pearl, 2022, for a more general discussion on the impact of processing abilities on children's language acquisition). This result is in line with Karmiloff-Smith's (1992) observation that children's metalinguistic representations rely on successful acquisition of the corresponding structures (see also Altman et al., 2018; Bialystok, 2001; Duncan et al., 2009). However, we observed no lower-order effect of participants' syntactic processing abilities in the analysis related to how the interaction of the activation of Greek with the participants' syntactic processing abilities in Italian affected their judgment of ungrammatical sentences (see Table 2). The observation that we found an effect of syntactic processing abilities only in association with the more explicit part of the task (i.e., explanation) suggests that more elaborate metalinguistic awareness abilities, that is, abilities requiring the highest degree of awareness, were more sensitive to the participants' syntactic processing abilities than less elaborate ones, that is, abilities involving lower degrees of awareness.

Effects of Age and Length of Exposure

The participants' explanation abilities were affected by their age, too. In particular, age played a significant role in the transition from correcting structures to explaining the reason behind the ungrammaticality of structures. This suggests that the creation of the most elaborate metalinguistic representations requires children to attain a certain degree of cognitive maturity related to increasing age (D'Souza & Filippi, 2017; Karmiloff-Smith, 1992).

As for the impact of length of exposure to the school language (Italian) on the participants' metalinguistic abilities, we expected to find similar results as the ones emerging from the analysis of the participants' syntactic processing abilities: The emergent bilinguals who participated in our study (i.e., with a more reduced exposure to the home language and allegedly lower syntactic processing abilities in it) should have benefitted the most from the activation of Greek. Contrary to our hypothesis, we did not find any effect of the interaction of the activation of Greek with the participants' length of exposure to Italian on their ability to judge ungrammatical sentences in Italian (see Table 3). More surprisingly, we observed that the participants with more reduced language exposure to Italian provided more corrections to ungrammatical sentences compared to the participants with greater exposure. Our study did not allow us to draw any conclusion about why this was the case. At a very speculative level, we propose that the mechanisms underlying the creation of metalinguistic representations may be different for earlier-onset learners compared to later-onset learners. While earlier-onset learners rely on their syntactic processing abilities for creating metalinguistic representations (through a process of redescription as defined in our literature review), later-onset learners may rely more directly on explicit declarative knowledge of grammatical rules "to bypass the increasingly inefficient implicit mechanisms" (DeKeyser, 2000, p. 518). This would explain why we observed the advantage of later-onset bilinguals in association with error correction only, which reflects a more explicit knowledge than error identification.

Effects of Syntactic Structure

We observed that the positive effect of the activation of Greek on the participants' rating of ungrammatical sentences was modulated by the type of syntactic structure. It should be mentioned again that this result should be interpreted with caution given the low number of ungrammatical sentences for each structure. However, we noticed that the participants tended to perform better in the bilingual metalinguistic task compared to the monolingual metalinguistic ones (the Italian one-sentence and two-sentence tasks) when considering structures

with respect to which Greek and Italian pattern similarly to each other (see the judgment of number-marking errors in the bilingual two-sentence task opposed to the Italian two-sentence task) or Greek is more transparent than Italian (see the judgment of case-marking errors in the bilingual as opposed to the Italian one-sentence task). We also observed better performance in the bilingual metalinguistic task in association with subject–verb agreement errors in Figure 5, but it was not confirmed statistically (see Appendix S6 in the Supporting Information online). These results are in line with what has been found in previous literature (e.g., Chung et al., 2018; Hayashi & Murphy, 2013; Ramirez et al., 2010; Zhang et al., 2010). By contrast, the gender system in Greek is not more transparent than the Italian gender system, which may be due to the observed syncretisms in the Greek clitic system (see Note 2) or to the fact that inflectional endings on nominals conflate more information in Greek compared to Italian (as explained in our Method section). This may be the reason why we found no effect of the bilingual two-sentence task on the participants' metalinguistic awareness related to incorrect gender marking on clitics.

It should be noted that the participants' rating of sentences featuring case-marking errors on clitics also benefited from the Italian two-sentence task even if to a lesser extent compared to the bilingual two-sentence task. This may be because the children in this study exhibited more difficulties with structures featuring case-marking errors on clitics compared to the other structures. For instance, these structures were associated with the highest Likert scores in the Italian one-sentence task (see Figure 5 and Table S6.1 in the Supporting Information online). Furthermore, we found a decrease in the likelihood of observing the highest explanation score (a score of 2) with errors related to case marking on clitics (see Table S7.1 in Appendix S7 in the Supporting Information online). These results suggest that metalinguistic awareness related to case marking in Italian was particularly vulnerable among the children who participated in this study. This is in line with previous studies showing that bilingual children may show some delay in the mastery of case morphology (e.g., Schulz & Grimm, 2019). There has been no study on bilingual children's acquisition of case marking on clitics in Italian. However, some studies on Italian L2 acquisition reveal that the distinction between accusative and dative clitics is particularly difficult for learners to acquire (Santoro, 2007). Therefore, the general difficulty associated with case marking on clitics may be the reason why the participants' metalinguistic abilities related to this structure were boosted by both the explicit mention of the correct forms (in the Italian two-sentence task) and the activation of Greek (in the bilingual two-sentence task).

Implications

The results emerging from our study have clear pedagogical implications. Across the literature, several studies have shown that bilingual pedagogies have a positive effect on children's literacy skills. In particular, children's learning the school language seems to be improved by schools' including the children's home languages in classroom activities (see our literature review). In the context considered in this paper, this would correspond to the use of Greek during literacy activities in Italian given that the participants mostly spoke Greek outside school (see Appendix S2 in the Supporting Information online). The direct observation of classroom routines was not the focus of our study. However, we intended to show whether children's metalinguistic reflections benefit from the activation of their home language in a context outside the classroom, that is, in a laboratory situation. Our results suggest that this is the case and are, thus, consistent with previous studies having a stronger focus on education practices. Therefore, this study draws on bilingual education practices and, at the same time, validates them, by showing that these practices are effective in other contexts beyond the classroom. The psycholinguistic approach adopted in this contribution has also allowed us to understand under which conditions the activation of children's home language is most effective. The findings related to the interaction of the bilingual two-sentence task with the participants' syntactic processing abilities suggest that integrating children's home language during classroom activities may be particularly beneficial for bilinguals with more reduced abilities in the school language. Notably, these children are not necessarily the ones with a later onset to the school language (see the above discussion on the effects of length of exposure and the correlation matrix in Table S2.2 in Appendix S2 in the Supporting Information online).

Limitations and Future Directions

This study was the first one to show the positive effects of the activation of children's home language(s) in their syntactic awareness in the school language(s). However, it should be considered as exploratory. First, the sample of participants in the study was relatively small. This was mainly related to the difficulty of gathering a set of representative speakers of the target population. Furthermore, we made sure that the tasks were not too long (and boring) for the children by including few (grammatical and ungrammatical) items for each kind of structure. We hope that we will be able to replicate the results of this study using a larger sample of participants, considering, for instance, similar immersion language learning contexts as the one at issue here (e.g., French immersion schools in Canada or Spanish immersion schools in the United States).

We have made our research instruments available if other researchers are interested in conducting replication studies.

Our second remark concerns how far the findings of this study are generalizable to other groups of bilingual children. First, the results shown in this study might be related to the bilingual profile of the children who were relatively balanced in their exposure to Greek and Italian. Furthermore, the results might be related to the language combination: Greek and Italian are very similar to each other with respect to the morphosyntactic phenomena chosen for the metalinguistic awareness task. Even in the domain of gender marking (which is the one in which the two languages exhibit the greatest number of dissimilarities), Greek and Italian are more similar to each other than to other languages displaying no gender marking at all. Therefore, it remains to be seen whether the same results would be found for language pairs with a larger typological distance. Researchers should observe the same results if our hypothesis holds true that the activation of children's home language enhances crosslinguistic metalinguistic reflections related to both similarities and differences between the school and the home language.

Finally, we referred to the pedagogical implications of our results. However, we have provided no empirical evidence related to possible didactic implementations of our hypotheses and their effects on children's learning outcomes. It would be interesting to investigate how children's ability to rely on the activation of their home language(s) while performing syntactic awareness tasks related to the school language(s) varies based on the way in which children are taught grammar. This kind of investigation would require triangulation among classroom observations and experiments to assess children's metalinguistic abilities such as the ones used in our study. Intervention studies fostering children's crosslinguistic reflections would also allow substantial progress in this direction.

Conclusion

In our study, we investigated how far the activation of children's home language (Greek) enhanced their syntactic awareness abilities as assessed in the school language (Italian). Furthermore, we analyzed to what extent the benefits of activating the home language Greek were modulated by the participating children's syntactic processing abilities in and length of exposure to their school language Italian, the type of syntactic structure, and the level of explicitness of the metalinguistic awareness task.

The main finding of the study is that the activation of Greek enhanced the participants' metalinguistic abilities in Italian. This was the case

particularly for the participants with lower syntactic processing abilities in Italian. These findings are consistent with previous studies showing that metalinguistic representations can be shared across a bilingual's two languages (e.g., Ke et al., 2021). However, our findings provide an original contribution to this issue since they establish a causal rather than correlational relation between the metalinguistic abilities in each language. Likewise, the result related to the modulating effect of the participants' syntactic processing abilities suggests that the activation of children's home language may have a scaffolding effect on their reflections on the grammar of the school language. This is consistent with research on education showing that emergent bilinguals' literacy skills benefit from including their home language(s) during classroom activities (e.g., Cummins, 2019).

The results related to the effect of length of exposure to Italian on the participants' metalinguistic abilities point to the need to conduct larger-scale studies. Likewise, we were not able to provide any conclusive evidence on whether the positive effects of activating Greek were most visible in association with syntactic structures for which Greek is more transparent than Italian since the overall effect size of the interaction between type of task and type of structure was small. More research is also needed to investigate how far the observed positive effects of activating Greek on the participants' metalinguistic abilities in Italian is related to the typological proximity between the two languages.

As a final remark, we would like to reflect on the fact that previous studies have not always found a bilingual advantage in metalinguistic awareness (Bialystok, 2001, and references in our literature review). Our investigation suggests that this may be due to the fact that the administration of traditional monolingual tasks did not account for the way in which metalinguistic representations are formed in the bilingual mind (see our discussion in the literature review). Traditional assessments may not have been able to provide the right conditions for the emergence of any bilingual advantage. In this study, the use of a bilingual metalinguistic awareness task appeared to be a more ecological way to elicit metalinguistic judgments from bilingual children than traditional forms of assessment have been.

Final revised version accepted 9 September 2022

Open Research Badges

This article has earned an Open Materials badge for making publicly available the components of the research methods needed to reproduce the reported

procedure. All materials that the authors have used and have the right to share are available at https://osf.io/kfqy4 and http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Notes

1 Two learners' questionnaires did not contain the relevant information related to the age of onset of Italian. It is interesting to note that all the parents of the children exposed to Italian at the age of 6 years ($n = 6$) declared that, when their child entered school, Italian was introduced at home, too, with one or both parents speaking Italian to the child regularly.

2 For gender-marking, the clitic system in Greek exhibits paradigmatic syncretism (see Appendix S8 in the Supporting Information online for the full paradigm of accusative and genitive/dative third person clitics in the two languages). As one of the reviewers suggested, the occurrence of syncretisms within the clitic system of each language may also affect the possibility of crosslinguistic sharing of syntactic awareness in this domain. Furthermore, it should be noted that in general, the use of clitics is acquired earlier in Greek than in Italian (Grohmann et al., 2012; Theodorou & Grohmann, 2015; see also the Discussion section on the effect of timing of acquisition on children's syntactic awareness abilities).

3 The resulting R model was: m ← multinom (explanation2 ~ type of task + type of structure + SRT + age + length of exposure, data = MAT), in which "explanation2" results from the choice of the score 0 of our outcome variable as the baseline. For our analysis, we followed the procedure described in the introduction to multinomial logistic regression written by the Statistical Consulting Group at the University of California at Los Angeles (https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression; accessed July 20, 2020).

4 In principle, we could have modeled our data by using an ordinal logistic regression given that the values corresponding to the explanation score could be ordered in ascending order of accuracy ($0 < 1 < 2$). However, the Brant test for the full model (as implemented in Schlegel & Steenbergen's, 2020, study) indicated that our data violated the proportional odds assumption (which is one of the main assumptions of ordinal logistic regressions), $\chi^2(8) = 77.64$, $p < .001$.

5 The resulting R model was: m1 ← lmer (likert ~ 1 + type of task + (1+type of task|participant) + (1|item), data = MAT_data, control = lmerControl(calc.derivs = FALSE)).

6 The resulting R model was: m2 ← lmer (likert ~ 1 + type of task * SRT + (1+type of task*SRT|participant) + (1|item), data = MAT_data, control = lmerControl(calc.derivs = FALSE)).

7 The resulting R model was: m3 ← lmer (likert ~ 1 + type of task * length of exposure + (1+type of task*length of exposure|participant) + (1|item), data = MAT_data, control = lmerControl(calc.derivs = FALSE)).

References

Altman, C., Goldstein, T., & Armon-Lotem, S. (2018). Vocabulary, metalinguistic awareness and language dominance among bilingual preschool children. *Frontiers in Psychology*, *9*, Article 1953. https://doi.org/10.3389/fpsyg.2018.01953

Ambridge, B., Pine, J., Rowland, C., & Young, C. R. (2008). The effects of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*(1), 87–129. https://doi.org/10.1016/j.cognition.2006.12.015

Andreou, M., Torregrossa, J., & Bongartz, C. (2021). Sentence Repetition Task as a measure of language dominance. In D. Dionne, & L.-A. Vidal Covas (Eds.), *BUCLD 45: Proceedings of the 45th Boston University Conference on Language Development* (pp. 14–25). Cascadilla Press. http://www.lingref.com/bucld/45/BUCLD45-02.pdf

Baker, C., & Wright, W. E. (2021). *Foundations of bilingual education and bilingualism*. Multilingual Matters. https://doi.org/10.21832/BAKER9899

Ballinger, S., Lyster, R., Sterzuk, A., & Genesee, F. (2017). Context-appropriate cross-linguistic pedagogy. Considering the role of language status in immersion education. *Journal of Immersion and Content-Based Language Education*, *5*(1), 30–57. https://doi.org/10.1075/jicb.5.1.02bal

Barac, R., & Bialystok, E. (2011). Cognitive development of bilingual children. *Language Teaching*, *44*(1), 36–54. https://doi.org/10.1017/S0261444810000339

Bartlett, L., & García, O. (2011). *Additive schooling in subtractive times: Bilingual education and Dominican immigrant youth in the Heights*. Vanderbilt University Press. https://doi.org/10.2307/j.ctv16b78cp

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3–4), 129–181. https://doi.org/10.1515/tl-2013-0009

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*, *24*(4), 560–567. https://doi.org/10.1037/0012-1649.24.4.560

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Bialystok, E. (2018). Bilingual education for young children: Review of the effects and consequences. *International Journal of Bilingual Education and Bilingualism*, *21*(6), 666–679. https://doi.org/10.1080/13670050.2016.1203859

Blom, E., & Paradis, J. (2016). Introduction: Special issue on age effects in child language acquisition. *Journal of Child Language*, *43*(3), 473–478. https://doi.org/10.1017/S030500091600012X

Bongartz, C., & Torregrossa, J. (2020). The effects of balanced biliteracy on Greek–German bilingual children's secondary discourse ability. *Journal of Bilingual Education and Bilingualism*, *23*(8), 948–963. https://doi.org/10.1080/13670050.2017.1355888

Brown, J. D. (2014). Classical theory reliability. In A. J. Kunnan (Ed.), *The companion to language assessment*. Wiley Blackwell. https://doi.org/10.1002/9781118411360.wbcla054

Caloi, I., & Torregrossa, J. (2021). Home and school language practices and their effects on heritage language acquisition: A view from heritage Italians in Germany. *Languages*, *6*(1), Article 50. https://doi.org/10.3390/languages6010050

Carbonara, V., & Scibetta, A. (2022). Integrating translanguaging pedagogy into Italian primary schools: Implications for language practices and children's empowerment. *International Journal of Bilingual Education and Bilingualism*, *25*(3), 1049–1069. https://doi.org/10.1080/13670050.2020.1742648

Carlisle, J. F. (2003). Morphology matters in learning to read. A commentary. *Reading Psychology*, *24*, 291–322. https://doi.org/10.1080/02702710390227369

Celic, C., & Seltzer, K. (2013). *Translanguaging: A CUNY-NYSIEB Guide for Educators*. CUNY NYSIEB. https://www.cuny-nysieb.org/wp-content/uploads/2016/04/Translanguaging-Guide-March-2013.pdf

Cenoz, J., & Gorter, D. (2017). Translanguaging as a pedagogical tool in multilingual education. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism. Encyclopedia of language and education* (pp. 309–321). Springer. https://doi.org/10.1007/978-3-319-02240-6_20

Chini, M. (1995). *Genere grammaticale e acquisizione. Aspetti della morfologia nominale in italiano L2* [Grammatical gender and acquisition. Aspects of nominal morphology in Italian L2]. Franco Angeli.

Chung, S. C., Chen, X., & Deacon, S. H. (2018). The relationship between orthographic processing and spelling in Grade 1 French immersion children. *Journal of Research in Reading*, *41*(2), 290–311. https://doi.org/10.1111/1467-9817.12104

Chung, S. C., Chen, X., & Geva, E. (2019). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*, *50*, 149–161. https://doi.org/10.1016/j.jneuroling.2018.01.003

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal*, *94*(1), 103–115. https://doi.org/10.1111/j.1540-4781.2009.00986.x

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3–49). National Dissemination and Assessment Center. https://doi.org/10.13140/2.1.1334.9449

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the cross-fire*. Multilingual Matters. https://doi.org/10.21832/9781853596773

Cummins, J. (2019). The emergence of translanguaging pedagogy: A dialogue between theory and practice. *Journal of Multilingual Education Research*, *9*(13), 19–36. https://research.library.fordham.edu/jmer/vol9/iss1/18

D'Souza, D., & Filippi, R. (2017). Progressive modularization: Reframing our understanding of typical and atypical language development. *First Language*, *37*(5), 518–529. https://doi.org/10.1177/0142723717720038

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/S0272263100004022

Duncan, L. G., Casalis, S., & Colé, P. (2009). Early metalinguistic awareness of derivational morphology: Observations from a comparison of English and French. *Applied Psycholinguistics*, *30*(3), 405–440. https://doi.org/10.1017/S0142716409090213

Egger, E., Hulk, A., & Tsimpli, I. M. (2018). Crosslinguistic influence in the discovery of gender: The case of Greek-Dutch bilingual children. *Bilingualism Language and Cognition*, *21*(4), 694–709. https://doi.org/10.1017/S1366728917000207

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Multilingual Matters.

Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.). *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). Multilingual Matters.

Francis, N. (2012). *Bilingual competence and bilingual proficiency in child development*. The MIT Press. https://doi.org/10.7551/mitpress/9780262016391.001.0001

Friesen, D. C., & Bialystok, E. (2012). Metalinguistic ability in bilingual children: The role of executive control. *Rivista di Psicolinguistica Applicata*, *12*(3), 47–56. http://digital.casalini.it/2614897

Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, *32*(1), 1–24. https://doi.org/10.18637/jss.v032.i01

García, O. (2009). *Bilingual education in the 21st century. A global perspective*. Wiley.

Grohmann, K. K., Theodorou, E., Pavlou, N., Leivada, E., Papadopoulou, E., & Martinez-Ferreiro, S. (2012). The development of clitic placement in Cypriot Greek and the Romance connection. In P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 128–152). Cambridge Scholars Publishing.

Gutiérrez, X. (2013). The construct validity of grammaticality judgement tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449. https://doi.org/10.1017/S0272263113000041

Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, *2*, Article 16. https://doi.org/10.3389/fcomm.2017.00016

Hayashi, Y., & Murphy, V. A. (2013). On the nature of morphological awareness in Japanese-English bilingual children. A cross-linguistic perspective. *Bilingualism: Language and Cognition*, *16*(1), 49–67. https://doi.org/10.1017/S1366728912000181

James, C. (1999). Language awareness: Implications for the language curriculum. *Language, Culture and Curriculum*, *12*(1), 94–115. https://doi.org/10.1080/07908319908666571

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh University Press. https://doi.org/10.1515/9780748626540

Jessner, U. (2008). Language awareness in multilinguals. In J. Cenoz, & N. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 6. Knowledge about language*, (pp. 357–369). Springer.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press. https://doi.org/10.7551/mitpress/1579.001.0001

Ke, S., Miller, R. T., Zhang, D., & Koda, K. (2021). Crosslinguistic sharing of morphological awareness in biliteracy development: A systematic review and meta-analysis of correlation coefficients. *Language Learning*, *71*(1), 8–54. https://doi.org/10.1111/lang.12429

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.-E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills

rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. https://doi.org/10.1111/desc.12202

Koda, K. (2005). Learning to read across writing systems: Transfer, metalinguistic awareness and second language reading development. In V. J. Cook (Ed.). *Second language writing systems* (pp. 311–334). Multilingual Matters. https://doi.org/10.21832/9781853597954-014

Koda, K. (2008). Impacts of prior literacy experience on learning to read in a second language. In K. Koda & A.M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development* (pp. 68–96). Routledge.

Kupferberg, I., & Olshtain, E. (1996). Explicit contrastive instruction facilitates the acquisition of difficult L2 forms. *Language Awareness*, *5*(3–4), 149–165. https://doi.org/10.1080/09658416.1996.9959904

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.4.5. https://CRAN.R-project.org/package=emmeans

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge. https://doi.org/10.4324/9781315887074

Leow, R. P., & Mercer, J. D. (2015). Depth of processing in L2 learning: theory, research and pedagogy. *Journal of Spanish Language Teaching*, *2*(1), 69–82. https://doi.org/10.1080/23247797.2015.1026644

Liceras, J. M., Fernández Fuertes, R., & Alba de la Fuente, A. (2012). Overt subjects and copula omission in the Spanish and the English grammar of English-Spanish bilinguals: On the locus and directionality of interlinguistic influence. *First Language*, *32*(1–2), 88–115. https://doi.org/10.1177/0142723711403980

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Multilingual Matters. https://doi.org/10.21832/9781847691767-006

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison, and testing of statistical models. *Journal of Open Source Software*, *6*(60), Article 3139. https://doi.org/10.21105/joss.03139

Luk, J., & Lin, A. (2015). Voices without words: Doing critical literate talk in English as a second language. *TESOL Quarterly*, *49*(1), 67–91. https://doi.org/10.1002/tesq.161

MacSwan, J., Thompson, M., Rolstad, K., McAlister, K., & Lobo, G. (2017). Three theories of the effects of language education programs: An empirical evaluation of bilingual and English-only policies. *Annual Review of Applied Linguistics*, *37*, 218–240. https://doi.org/10.1017/S0267190517000137

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.) *Methods for assessing multilingual children:*

Disentangling bilingualism from language impairment (pp. 95–124). Multilingual Matters. https://doi.org/10.21832/9781783093137

Mattheoudakis, M., Chatzidaki, A., Maligkoudi, C., & Agathopoulou, E. (2016). Family and school language input: Their role in bilingual children's vocabulary development. *Journal of Applied Linguistics*, *31*, 49–69. https://doi.org/10.26262/jal.v0i31.8291

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*, 625–632. https://doi.org/10.1007/s10459-010-9222-y

Paradis, J. (2011). Individual differences in child English second language acquisition. Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par

Pearl, L. (2022). Modeling syntactic acquisition. In J. Sprouse (Ed.), *The Oxford handbook of experimental syntax* (Chapter 7). Oxford University Press.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, *5*(4), 409–453. https://doi.org/10.1075/lab.5.4.01phi

Plonsky, L., Mardsen, E., Crowther, D., Gass, S.M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgement tasks in second language research. *Second Language Research*, *36*(4), 583–621. https://doi.org/10.1177/0267658319828413

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Ramirez, G., Chen, X., Geva, E., & Kiefer, H. (2010). Morphological awareness in Spanish-speaking English language learners: Within and cross-language effects on word reading. *Reading and Writing: An Interdisciplinary Journal*, *23*, 337–358. https://doi.org/10.1007/s11145-009-9203-9

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Santoro, M. (2007). Second language acquisition of Italian accusative and dative clitics. *Second Language Research*, *23*(1), 37–50. https://doi.org/10.1177/0267658307071603

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. Springer. http://lmdvr.r-forge.r-project.org

Schlegel, B., & Steenbergen, M. (2020). Brant: Test for parallel regression assumption (R package, Version 0.3-0) [Computer software]. https://cran.r-project.org/web/packages/brant/brant.pdf

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Schulz, P., & Grimm, A. (2019). The age factor revisited: Timing in acquisition interacts with age of onset in bilingual acquisition. *Frontiers in Psychology*, *9*, Article 2732. https://doi.org/10.3389/fpsyg.2018.02732

Sharwood Smith, M. (2021). The cognitive status of metalinguistic knowledge in speakers of one or more languages. *Bilingualism: Language and Cognition*, *24*(1), 185–196. https://doi.org/10.1017/S1366728920000371

Simard, D., Labelle, M., & Bergeron, A. (2017). Measuring metasyntactic abilities: On a classification of metasyntactic tasks. *Journal of Psycholinguistics Research*, *46*, 433–456. https://doi.org/10.1007/s10936-016-9445-z

Spada, N., Shiu, J. L-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparison with other validation studies. *Language Learning*, *65*(3), 723–751. https://doi.org/10.1111/lang.12129

Theodorou, E., & Grohmann, K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua*, *161*, 144–158. https://doi.org/10.1016/j.lingua.2014.11.011

Torregrossa, J., & Bongartz, C. (2018). Teasing apart the effects of dominance, transfer and processing in reference production by German–Italian bilingual adolescents. *Languages*, *3*(3), Article 36. https://doi.org/10.3390/languages3030036

Torregrossa, J., Andreou, M., & Bongartz, C. (2020). Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa: A Journal of General Linguistics*, *5*(1), Article 95. https://doi.org/10.5334/gjgl.1011

Torregrossa, J., Eisenbeiß, S., & Bongartz, C. (2022). *Materials. Datasets from "Boosting bilingual metalinguistic awareness under dual language activation: Some implications for bilingual education"* [Collection: Language test and questionnaire]. IRIS Database, University of York, UK. https://doi.org/10.48316/24q2-cr23

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://www.stats.ox.ac.uk/pub/MASS4

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://ggplot2.tidyverse.org

Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, *46*(4), 680–704. https://doi.org/10.1111/flan.12063

Zhang, D. (2013). Linguistic distance effect on cross-linguistic transfer of morphological awareness. *Applied Psycholinguistics*, *34*(5), 917–942. https://doi.org/10.1017/S0142716412000070

Zhang, D., Anderson, R. C., Li, H., Dong, Q., Wu, X., & Zhang, Y. (2010). Cross-language transfer of insight into the structure of compound words. *Reading and Writing*, *23*(3), 311–336. https://doi.org/10.1007/s11145-009-9205-7

Zhang, D., Koda, K., & Sun, X. (2014). Morphological awareness in biliteracy acquisition: A study of young Chinese EFL readers. *International Journal of Bilingualism*, *18*(6), 570–585. https://doi.org/10.1177/1367006912450953

Zhao, C., Kormos, J., Rebuschat, P., & Suzuki, S. (2021). The role of modality and awareness in language learning. *Applied Psycholinguistics*, *42*(3), 703–737. https://doi.org/10.1017/S0142716421000023

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Description and Analysis of the Questionnaire.

Appendix S2. Results of the Analysis of the Questionnaire Information.

Appendix S3. Description of the Sentence Repetition Task.

Appendix S4. Items of the Three Metalinguistic Awareness Tasks and Procedure for the Administration.

Appendix S5. Cumulative Link Mixed Model Analysis of the Effect of Type of Task on the Likert Scores Associated with Ungrammatical Sentences.

Appendix S6. Linear Mixed-Effects Analysis for the Likert-Scores Associated with Ungrammatical Sentences for Type of Task and Type of Structure.

Appendix S7. Parameters of the Multinomial Logistic Analysis Including Explanation Score as Outcome Variable and Type of Syntactic Structure as Predictor.

Appendix S8. Paradigm of Accusative and Genitive/Dative Third Person Object Clitic Pronouns in Greek and Italian.