# When more is more. L2 agreement improves when listeners can rely on both noun and verbal features

Nicoletta Biondo*[1], Luca Cilibrasi[1,2]

[1]University of Siena, [2]Charles University, Prague

**Abstract**. English verbal agreement has been shown to be a particularly challenging domain in both first and second language acquisition. In this study, we tested the comprehension of sentences with masked and unmasked agreement in 16 Italian children and 36 Italian adults learning English as a second language. In the masked condition, participants were presented with verbs starting with the phoneme /s/, making the plurality of the noun "hidden", and leaving the comprehension of the sentence on the processing of the verb only (i.e., The elephants spill...). In the unmasked condition, the verb started with a different phoneme, which allows for comprehension to be cued by both noun and verbal features (i.e., The elephants drink...). Results show that both children and adults are better at comprehending when both features are available. These findings make a case for a preference for redundant features in L2 agreement processing by learners of English with Italian as L1.

**Keywords:** agreement, English as a second language, bound morphology

## 1 Introduction

Agreement is an important property of languages, and its acquisition is a crucial milestone during development. A large international study directed by Bittner, Dressler & Kilani-Schoch (2011) assessed the acquisition of agreement across a wide variety of languages, showing that the time of acquisition of the agreement properties varies across languages. In particular, one European language emerged as taking longer for its agreement to be mastered than the others: English.

In modern English, present tense agreement only emerges in the third person singular, and all other forms are expressed with a zero-morpheme

* This author has changed affiliations during the submission process. New affiliations are: Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain; University of California Berkeley, Berkeley, USA.

(see "she plays" vs "I/you/we/they play"). Subjects are compulsory and thus, from the semantic point of view, agreement on the third person is redundant (Bock & Eberhard 1993). From the phonetic point of view, the third person of a verb is expressed with 3 different allophones, established by morpho-phonological rules. In short, a voiced version of the phoneme is used if the ending phoneme of the stem is voiced (rule: add /z/), a devoiced version of the phoneme is used if the ending phoneme of the stem is devoiced (rule: add /s/), and finally a vowel is additionally included if the stem ends in a sibilant (rule: add /ɪz/). Phonetically speaking, all these allophones appear as weak and non-salient, making their perception and use challenging in both first and second language acquisition (Song, Sundara & Demuth 2009).

In first language acquisition, children are shown to occasionally omit the third person bound morpheme into their third year during natural sentence production, with a timeline that appears slightly delayed in comparison to other languages (Bittner, Dressler & Kilani-Schoch 2011). In comprehension, this asymmetry is even larger, and errors are still observed into the fourth year. A study by Johnson, de Villiers & Seymour (2005) tested children's ability to comprehend sentences by relying uniquely on verbal morphology. In this ingenious experiment, number marking on the noun was masked by the use of verbs starting with "s", and children could thus only rely on the verb inflection to parse the sentence. For instance, children were presented with the sentence "the ducks swim" and they were then shown two pictures, one with two ducks, and one with only one duck. In order to solve this task, the child could not rely on the plurality of the noun, since that was masked by the beginning of the verb, and she/he needed to carefully parse the verb ending. Their study tested 58 children from age 3 to age 6, and their results showed that only children older than 5 were able to successfully complete the task.

A study by Morales Reyes & Montrul (2020) adopted a similar methodology to investigate agreement processing in L2 learners. In this study, the authors compared the rate of acquisition in English-speaking children learning Spanish (using an adapted version of the task) and Spanish-speaking children learning English (using the method of Johnson, de Villiers & Seymour 2005). The results showed an advantage for the English-speaking children learning Spanish. This finding is consistent with a typological explanation of the acquisition of agreement. In languages with a rich and phonetically salient inflectional morphology, such as Spanish, acquisition can proceed more quickly and efficiently. This variation also explains the patterns observed in monolingual children: Mastering agreement takes longer for children growing up in English-speaking countries than for children growing up in Spanish-speaking countries.

Interestingly, there seems to be a relationship between the timing in the acquisition of specific linguistic phenomena in the L1 and the strategies deployed by the speakers during the acquisition of an L2 (Tsimpli 2014). By looking at a wide variety of linguistic phenomena, Tsimpli (2014) suggests that early grammatical structures (such as the order of verbs and subjects) may be proficiently acquired in L2 independently of the amount of exposure children receive, while late grammatical structures (such as grammatical gender) may be acquired in L2 only in presence of higher amounts of input. Agreement may be considered an early or a late phenomenon depending on the language under investigation. In English, agreement may be described as a late phenomenon (Cilibrasi & Tsimpli 2020). As a consequence, exposure to the L2 is expected to be an important predictor of proficient agreement processing.

# 2 The current study

The main aim of our study was to better explore the comprehension of English verbal morphology in L2 speakers. In particular, we investigated the comprehension of English verbal morphology in young (6–14 y.o.) and older (18–45 y.o.) native speakers of Italian who are acquiring English as L2, thus extending previous work (e.g., Morales Reyes & Montrul 2020) both in terms of ages tested and pairs of languages tested.

In particular, we were interested in (i) testing whether L2 speakers' performance changes as a function of the availability of number features both on the subject NP and the verb (unmasked) or just on the verb (masked); (ii) exploring whether L2 speakers' performance changes as a function of their linguistic background (e.g., L2 age of acquisition, L2 immersion).

Based on previous literature (e.g., Morales Reyes & Montrul 2020; Tsimpli 2014), we expected (i) L2 speakers to be better at comprehending numerosity when number features are "unmasked" (audible both on the subject NP and on the verb), that is higher accuracy and shorter RT in the unmasked condition compared to the masked condition; (ii) L2 accuracy to increase with larger exposure and/or earlier age of acquisition.

These predictions were mainly based on previous literature on child language acquisition but were meant to be extended to L2 adult speakers, especially to speakers who have been exposed to English little (small exposure) and later on in life (late age of acquisition).

# 3 Sample 1: Italian children with English as L2

## 3.1 Method

### Participants

Sixteen Italian children (8 male, 8 female; age range: 6–14 y.o.) who were acquiring English as L2 were tested. Their age range varied from 6 to 14 years of age (Mean = 9.1; SD = 2.17); in particular, a group of 9 children had between 6 to 9 y.o.a. while a group of 7 children had between 10 and 14 y.o.a. All children had at least one Italian-speaking parent, or both Italian-speaking parents (13 children). All children were raised in Italy, although it was reported that 3 children also lived in a non-English speaking country. Children's age of L2 acquisition varied from 0 to after 6 years of age; in particular, 6 children started to be exposed to the L2 before 4 y.o.a. while 10 children were exposed later on, after 4 y.o.a. All parents declared that their child did not have hearing disorders, and they also provided their consent for the participation of their children in this study, in line with the Helsinki declaration. For further information and a qualitative analysis on the children see Appendix B.

### Design and materials

The design was the same as proposed by Morales Reyes & Montrul (2020) for L1 Spanish L2 English (sequential bilingual) children, in turn inspired by Johnson et al.'s (2005) study on monolingual English children. In particular, the design consisted of two conditions. In the *masked* condition (10 items), number features could only be extracted from the verb. The plural /s/ on the noun was masked by verbs beginning with an /s/ consonant cluster, thus leading to a co-articulation of both /s/ in rapid speech (e.g., *The elephants spill…*). In the *unmasked* condition (10 items), number features could also be extracted from the noun, since no verb began with an /s/ (e.g., *The boys write…*). Singular sentences (e.g., *The elephant spills…*, *The boy writes…*) were also tested. A schema of the design is provided in Table 1.

The experiment also included 3 filler items and 3 practice items with present progressive verb forms (e.g., *The girl is taking a nap*). It should be noted that in Morales Reyes and Montrul's study only 6 unmasked sentences were used, while we decided to add 4 sentences to the unmasked condition to have an equal number of trials (10) in the two experimental conditions. The masked sentences were exactly the same as adopted by

Johnson, de Villiers & Seymour (2005) and by Morales Reyes & Montrul (2020). See Johnson et al.'s appendix for a complete list.

The sentences were read by a female English native speaker, who was instructed to read the sentences at a natural speech rate and to not pause during the reading of the subject and the verb, as in natural-connected speech (Morales Reyes & Montrul 2020). All sentences were prerecorded and equalized in terms of volume, background noise, and duration.

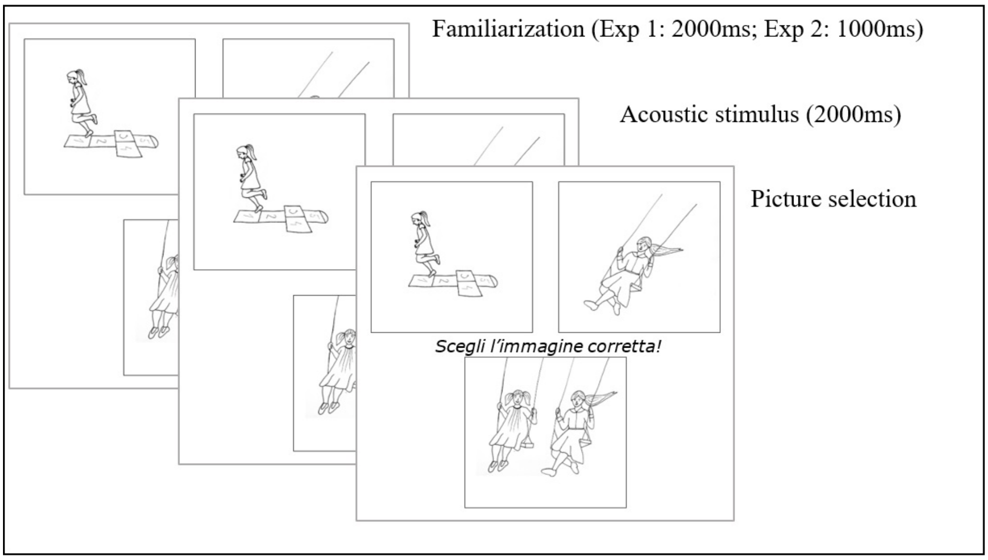
Each sentence was paired with 3 pictures, as shown in Figure 1. One picture represented a single subject performing the action described by the sentence (correct), one picture represented the same action performed by two subjects (distractor), and one picture represented the same single subject who was performing a different action (filler). The position (up left, up right, down center) of the correct/distractor/filler pictures was counterbalanced across items to prevent biased behavior.

The stimuli were then divided into 2 lists so that each participant saw only one version of each item.

Table 1: Example of the experimental material tested

| **Masking** | **Number** | **Sentence** |
|---|---|---|
| Masked | sg | The ape swings in the tree. |
| Masked | pl | The apes swing in the tree. |
| Unmasked | sg | The horse gallops across the field. |
| Unmasked | pl | The horses gallop across the field. |

Figure 1: Experimental routine



### Procedure

The task was administered through Ibex Farm, a javascript-based platform developed for hosting psycholinguistic experiments on the web (Drummond 2013). The experiment was thus presented on a web page that could be accessed both through a notebook and through a tablet (smartphones were excluded because of the limited size of the screen).

At the beginning of the experiment, participants' parents were asked to provide their consent about their children's participation in the study and to fill out a questionnaire on their (and their children's) linguistic background. The questionnaire included questions on participants' age, gender, education level, self-reported L2 proficiency, L2 age of acquisition, L2 daily use, context of L2 use, level of motivation in L2 acquisition, reported hearing/reading disorders. Afterwards, participants could read the instructions and start a practice session with 3 trials. After practice, the real experiment began.

The paradigm we adopted was a sentence-picture matching task. Participants were asked to listen to the sentence and to select the picture that matched the sentence. A schema of the experimental routine is provided in Figure 1. Participants had a 2 second familiarization period to look at the three pictures before hearing the experimental sentence. At the end of the sentence presentation, the sentence "Choose the right picture!" appeared in the center of the screen and participants could choose one of the pictures (no picture could be chosen before the end of the sentence acoustic presentation). Participants could choose the right picture by clicking on it (with a finger if they were using a tablet, or with a mouse if they were using a laptop[1]).

## 3.2 Data analysis

Both accuracy and reaction times were collected.

Accuracy had value equal to 1 when the participant chose the right picture, while the value 0 was recorded when the participant either chose the wrong picture or the distractor[2]. Accuracy data were used as dependent variables of logit mixed-effect models (Jaeger 2008) performed through the R package *lme4* (Bates et al. 2014).

Reaction times were recorded from the moment in which the message "Choose the right picture" appeared in the center of the screen (i.e., at the end of the acoustic stimulus presentation) to the moment of the picture

1. No difference was found in the accuracy and reaction times data provided through a notebook and a tablet ($M_{acc}$ laptop = 0.56, $M_{acc}$ tablet = 0.51, $t_{(12)} = 0.93$, $p = 0.37$; $M_{logRT}$ laptop = 8.05, $M_{logRT}$ tablet = 8.07, $t_{(14)} = -0.03$, $p = 0.97$).
2. The percentage of selected correct/distractor/filler pictures for each subject is reported in Appendix A.

selection. Reaction time data were used as the dependent variable of linear mixed-effect models. Before entering the analysis, outliers were removed by using a fixed threshold (20000 msec) that affected 5% of the data. In order to normalize the residuals, reaction time data were log-transformed before entering the data analysis.

The advantage of using mixed-effect models, compared to traditional non-parametric/parametric tests (e.g., Chi-square, t-tests, ANOVA) adopted in previous studies, is the possibility of accounting for by-subject and by-item variability through the definition of a random effect structure. The random effect structure was selected by using a parsimonious approach (Bates et al. 2015), that is by performing a Principal Component Analysis so as to reduce the complexity of the random effect structure and to identify the most parsimonious model properly supported by the data. The best-fitting model both for accuracy and reaction time data was the one containing only random intercepts for subject and item grouping factors.

The fixed-effect factors of the models were selected *a priori* based on the research questions and were all sum-contrast coded (Schad et al. 2020). In particular, we considered *masking* (unmasked = 1, masked = -1) and *number* (singular = 1, plural = -1) and their interaction as fixed-effect factors to address the first research question, i.e., to verify whether the presence of one (masked) or two number cues (unmasked) on subject NPs and verbs affected accuracy and reaction times. In order to address the second research question, we added *age* (6-to-9 = 1, 10-to-14 = -1) and *L2 age of acquisition* (before4 = 1, after4 = -1) and their interaction with *masking* in the model. We expected age and L2 age of acquisition to have inverse relation with accuracy and RTs. In particular, we expected accuracy to increase with older age and decrease with later L2 age of acquisition. We also expected reaction times to decrease with older age and increase with later L2 age of acquisition. The interaction with *masking* allowed us to see whether the presence of one (masked) or two number cues (unmasked) differentially affected the participants based on age and/or L2 age of acquisition.

## 3.3 Results

Table 2 reports the output of the analyses.

The analysis of the accuracy data showed that all children better comprehended the sentences when the number features were "unmasked", that is acoustically perceivable both on the subject NP and on the verb, as shown by the significant effect of *masking*. We also found an interaction between *masking* and *number* showing that in the masked sentences accuracy was lower in the singular condition compared to the plural con-

Table 2: Output of the best-fitting model for accuracy and reaction time data, followed by the output of the models built to solve the interactions

| | **Reaction time data** | | | | **Accuracy data** | | | |
|---|---|---|---|---|---|---|---|---|
| | *Estimate* | *SE* | *z* | *p* | *Estimate* | *SE* | *t* | *p* |
| Mask | .03 | .15 | .29 | .77 | .63 | .17 | 3.73 | **.0002** |
| Numb | -.02 | .09 | -.4 | .69 | -.18 | .13 | -1.41 | .16 |
| Mask x Numb | .1 | .06 | 1.72 | .09 | .47 | .13 | 3.62 | **.0003** |
| Age | .43 | .17 | 2.47 | **.03** | -.19 | .19 | -1.01 | .31 |
| L2aoa | -.2 | .18 | -1.16 | .26 | -.18 | .18 | 0.79 | .43 |
| Age x Mask | -.07 | .08 | -.88 | .38 | -.78 | .19 | -4.2 | **.00003** |
| L2aoa x Mask | .03 | .08 | -.36 | .72 | .39 | .18 | 2.12 | **.03** |
| ***Mask x Numb*** | | | | | | | | |
| Masked (sg vs pl) | | | | | -.65 | .27 | -2.41 | **.02** |
| Unmasked (sg vs pl) | | | | | .28 | .19 | 1.46 | .15 |
| ***Age x Mask*** | | | | | | | | |
| Old (mask vs unmask) | | | | | .96 | .2 | .49 | **.000001** |
| Young (mask vs unmask) | | | | | .03 | .16 | .16 | .87 |
| ***L2aoa x Mask*** | | | | | | | | |
| Late (mask vs unmask) | | | | | .54 | .15 | 3.56 | **.0004** |
| Early (mask vs unmask) | | | | | .23 | .22 | 1.07 | .28 |

dition while no difference was found between singular and plural condition in the unmasked condition. These two findings are represented visually in Figure 2.

The interaction between *age* groups and *masking* showed that older children were sensitive to the audibility of number features, and in particular they were more accurate when the number features were unmasked, that is audible on both the subject NP and the verb. Conversely, younger children did not show the same sensitivity. Finally, the interaction between *L2 age of acquisition* and *masking* showed that, in particular, the children who started acquiring English later, after 4 years of age, were the ones that were especially sensitive to the presence of un/masked number features on the subject NP, while the children who acquired English earlier on in life did not show the same sensitivity. Age of onset and age effects are presented visually in Figure 3.

The analysis of the reaction time data only showed that older children were faster than younger children, as shown by the main effect of *age*. This is a straightforward result easily explainable by the development of domain-general skills, and is not represented in a figure.

Figure 2: Mean accuracy across conditions. Bars represent standard errors. The figure displays a main effect of masking (unmasked sentences are overall more accurate) and an interaction between number and masking (only in the masked condition there is a significant difference between singular and plural).

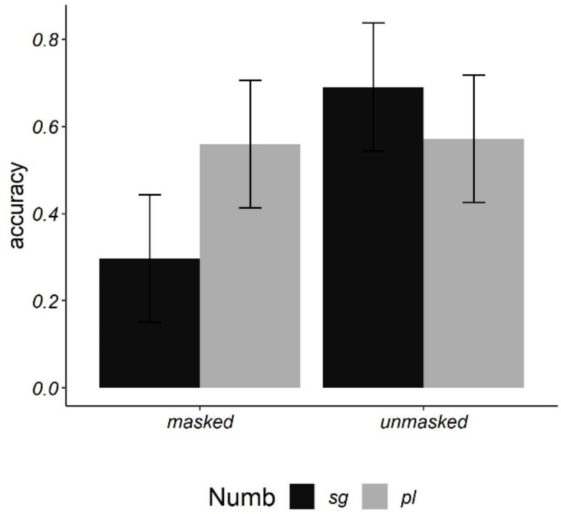
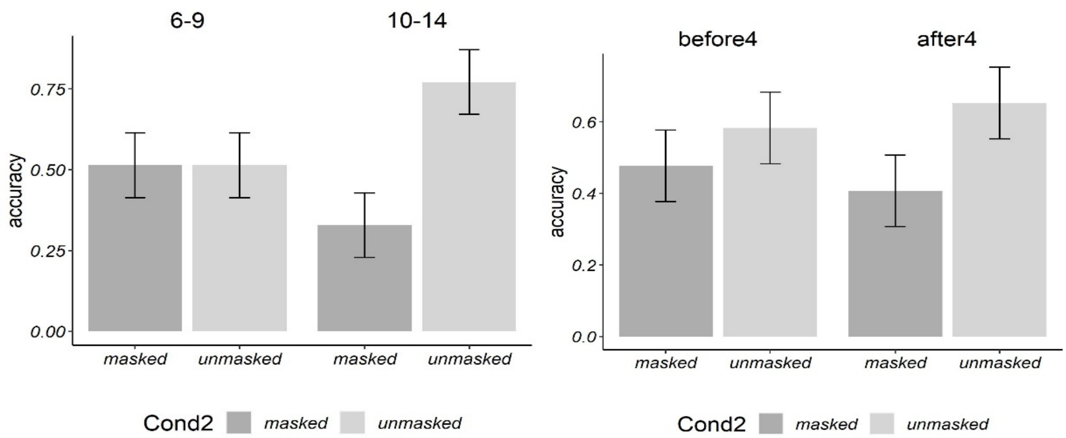



Figure 3: Mean accuracy divided by group. Bars represent standard errors. The graph shows that the significant effect of masking is present only in older children and only in children that were exposed to English after the age of 4. Notice that due to ongoing changes in the Italian educational system, younger children tend to be those that were exposed to English before.



# 4 Sample 2. Italian adults with English as L2

## 4.1 Method

### Participants

Forty L1 Italian L2 English speakers (17 female, 22 male, 1 other; age range 18–42 y.o.a., Mean = 25, SD = 5.9) were tested. They all grew up in Italy, and 14 of them lived abroad at least once (11 in a non-English speaking country; 3 in an English-speaking country for no more than 12 months). Nineteen speakers declared that they started to acquire the English language before 6 y.o.a. while the other 20 speakers declared that they acquired English after 6 y.o.a.

Given that no study has ever tested the paradigm presented in this study with adult speakers, we also collected data from 40 native English speakers (27 female, 13 male; age range 18–43 y.o.a., Mean = 29, SD = 5.9) as a control. Only 2 participants, 1 in each group, declared that they have/had hearing disorders and were thus excluded from the data analysis. All participants gave their consent to participate in the study, in line with the Helsinki Declaration.

### Materials

We adopted the same design and material used in Experiment 1.

### Procedure

This study followed the procedure described in Experiment 1. Only three parameters were changed. First, the consent and the linguistic questionnaire were directly filled by the (adult) participants. Second, the linguistic questionnaire also included a question about the daily use of L2 (from 0 to 100), a factor that has been known to affect adult L2 acquisition (Carhill, Suárez-Orozco & Páez 2008, De Bruin, Carreiras & Duñabeitia 2017). Third, the familiarization period for the pictures was reduced to 1000 ms. All participants performed the task via laptop.

## 4.2 Data analysis

Data from 40 native speakers and 36 L2 speakers[3] were included in the final analysis. We adopted the same type of analysis described in Experiment 1. However, we adapted the fixed-effect factors to the research questions. In particular, we ran two different analyses to address research question (i) and research question (ii). We could not perform a unique analysis, as in Experiment 1, because in this study we also had a control group of English native speakers who did not have measures of L2 age of acquisition or daily L2 use. Therefore, in the first analysis, we compared the two groups: The model was built to include the two factors manipulated in the experimental material, i.e., *masking* and *number*, in addition to the *group* factor (native = 1, L2 = -1). A three-way interaction among these factors was also included to test whether the two groups of speakers behaved differently in the four conditions. The best-fitting model for accuracy data was *ACC ∼ 1 + Mask*Group*Numb + (1 + Mask:Numb || subj) + (1 + Group*Numb | item),* while the best-fitting model for reaction time data was *logRT ∼ 1 + Mask*Group*Numb + (1 + Mask || subj) + (1 + Group || item).*

In the second analysis, we focused on the group of L2 speakers, thus adding to the interaction between *condition* and *number*, the factor *L2 age of acquisition* (before 6 y.o.a. = 1, after 6 y.o.a. = -1), the (centered) numeric factor *daily L2 use* and their interaction with *masking* (unmasked = 1, masked = -1).

## 4.3 Results

Table 3 reports the output of the analyses.

The first analysis showed that English native speakers were more accurate and faster than the group of L2 speakers, as shown by the main effect of *group*. Moreover, the main effect of *masking* (with no interaction with the *group* factor) showed that both groups behave similarly, namely they were both more accurate in the unmasked condition, in which both

3. An anonymous reviewer pointed out that the participants who experienced different types of immersive contexts could not be grouped together. Immersion in a L2-speaking country can positively affect L2 comprehension. However, to our knowledge, there is no evidence showing a direct relation between immersion in a non-L2-speaking/foreign country and L2 comprehension abilities. In order to reduce variability in our sample without underpowering our study, we excluded only the 3 participants who spent time in a L2-speaking country. Experience abroad ("yes" for the 11 participants who lived in a non-English-speaking country) was used as a covariate during the analysis of the L2 speaker data. This factor never reached significance (accuracy data: Estimate = -.16, SE = .22, z = -.72, p = .47; reaction times: Estimate = .09, SE = .07, 1 = 1.14, p = .26).

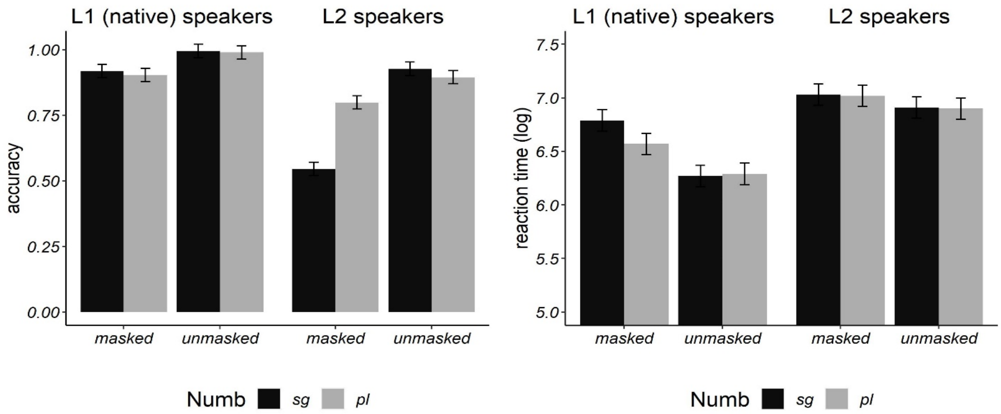
the subject NP and the verb clearly expressed number features, compared to the masked condition in which only the verb showed audible number cues. The three-way interaction was not significant, despite the numerical difference between the singular and the plural condition in the L2 group. Figure 4 illustrates the mean accuracy values and (log) reaction times in the two groups of adult speakers.

The second analysis which was run to investigate the role of L2 speaker-related factors such as L2 age of acquisition and L2 daily use replicated the effect of *masking* found in the first analysis. It also showed a significant *masking* x *number* interaction that did not appear to be significant in the first analysis. Post-hoc models built to resolve the interaction showed that in the masked condition the L2 speakers were less accurate when there was the /s/ on the verb (singular) than when there was not (plural), while no difference was found between singular and plural condition in the unmasked sentences. No other significant effect was found.

Table 3: Output of the best-fitting model for accuracy and reaction time data, followed by the output of the models solving the significant interactions.

| | **Reaction time data** | | | | **Accuracy data** | | | |
|---|---|---|---|---|---|---|---|---|
| | *Estimate* | *SE* | *z* | *p* | *Estimate* | *SE* | *t* | *p* |
| *First analysis: Native vs L2* | | | | | | | | |
| Mask | 1.2 | .26 | 4.6 | **.000005** | -.13 | .07 | -1.87 | .08 |
| Group | 1.5 | .33 | 4.7 | **.000003** | -.24 | .06 | -4.32 | **.0001** |
| Numb | -.4 | .32 | -1.19 | .24 | .03 | .03 | .91 | .36 |
| Mask x Group | .18 | .25 | .73 | .47 | -.07 | .04 | -1.74 | .09 |
| Mask x Numb | .36 | .29 | 1.26 | .21 | -.03 | .03 | -1.21 | .23 |
| Group x Numb | -.26 | .3 | -.87 | .38 | .03 | .03 | -1.32 | .19 |
| Mask x Group x Numb | -.23 | .27 | -.88 | .38 | -.03 | .03 | -0.97 | .33 |
| *Second analysis: Only L2* | | | | | | | | |
| Mask | .96 | .19 | 5.2 | **.0000002** | -.06 | .09 | -.73 | .47 |
| Numb | -.19 | .17 | -1.1 | .27 | -.01 | .04 | -.38 | .71 |
| L2 aoa | -.03 | .21 | -.15 | .88 | -.05 | .07 | .78 | .44 |
| L2 use | .02 | .01 | 1.64 | .1 | -.001 | .003 | -.3 | .78 |
| Mask x Numb | .51 | .18 | 2.91 | **.004** | -0.3 | .04 | -.11 | .92 |
| Mask x L2 aoa | .11 | .13 | .85 | .39 | -.03 | .04 | -.7 | .48 |
| Mask x L2 use | .01 | .01 | 1.54 | .12 | -.003 | .002 | -1.72 | .09 |
| ***Mask x Numb*** | | | | | | | | |
| Masked (sg vs pl) | -.82 | .33 | -2.52 | **.01** | | | | |
| Unmasked (sg vs pl) | .3 | .21 | 1.41 | .16 | | | | |

Figure 4: Mean accuracy and log reaction times. Bars represent standard errors. Both accuracy and reaction times are characterized by a main effect of group: native speakers are more accurate and faster. The accuracy analysis also presents a significant effect of masking: accuracy in the unmasked sentences is overall higher (in both groups).




# 5 Discussion

This study investigates the comprehension of masked and unmasked agreement in Italian children and adults learning English as a second language. The main finding we obtained from these experiments is that both children and adults are significantly more accurate in the unmasked condition. In terms of parsing, the main difference between comprehension in the masked vs unmasked condition consists in the number of cues available. In the masked condition, exemplified by sentences such as "the ducks swim", the correct comprehension of the sentence must rely on the correct parsing of the verbal morphology. The noun is ambiguous (as it is not possible to understand if it is plural or singular due to the fact that it ends with the same phoneme of the verb beginning). In the unmasked condition, exemplified by sentences such as "the ducks run", parsing can be completed relying on two cues: The noun morphology and the verb morphology. Our results demonstrate that Italian second language learners of English are significantly more accurate when both cues are available.

This study complements similar studies conducted on other populations. The first experiment with this methodology was conducted on English monolingual children of varying ages (Johnson, de Villiers & Seymour 2005). The study of Johnson, de Villiers & Seymour (2005) showed that monolingual children are inaccurate in the masked condition until about the age of 5. When presented with "the ducks swim", English speaking 4-year-olds will randomly choose 1 or 2 swimming ducks in the picture matching task. On the contrary, at the same age they are already accurate with the unmasked condition. After the age of 5, they become accurate with the masked condition as well, reaching ceiling performance.

Morales Reyes & Montrul (2020) extended this line of research to children learning English as a second language (with Spanish as L1). In their experiment, a group of Spanish natives learning English was compared, with an adapted task, to a group of English native children learning Spanish as L2. The study showed that, given similar learning conditions, English children learning Spanish were significantly more accurate. This asymmetry suggests that the difficulties found in children learning English are related to its typology. As explained by Bittner, Dressler & Kilani-Schoch (2011), in English the third person singular morpheme, the only person that does not use a zero morpheme at the present, is semantically redundant (as subjects are compulsory) and phonetically weak. This makes its learning particularly challenging.

Our results complete the work of Johnson, de Villiers & Seymour (2005) and Morales Reyes & Montrul (2020) and show that both children and adults learning English as L2 (and having Italian as L1) are significantly more accurate when both noun and verbal morphology are available. These findings suggest that, in early stages of learning, a reliance on a larger number of morphological cues might be the preferred strategy. This statement would hold in both monolingual children (see the 4-year-olds English monolinguals in Johnson, de Villiers & Seymour 2005) and in L2 learners (whether they are children or adults), since patterns in these three groups overlap. Importantly, an overall main effect of masking was also observed in the adult control group, suggesting that the presence of more cues has a facilitatory effect for native speakers as well. This may seem in contradiction with the results of Johnson, de Villiers & Seymour (2005), who observed a reduction of this asymmetry in older children. However, what we observe here might be an instance of U-shaped development (Siegler 2004; Pauls, Macha & Petermann 2013): Children may be initially very reliant on cues and prefer the unmasked condition, then become less reliant and display a reduced asymmetry, and finally, as adults, when presented with a task of this kind, they may again resort to the analysis of cues to improve their performance.

Interestingly, when our children's sample was divided according to age, it appears that the effect of masking was only present in the older children. One possibility for such a finding lies exactly in the pattern we just observed for adults. Older children may be able to consciously resort to the analysis of cues and thus may be more sensitive to the effect of masking. Alternatively, such result may be connected to a related variable: Age of onset. When dividing children according to their ages of L2 onset (instead of demographic age), only children with a late age of onset displayed this asymmetry. Italian is morphologically rich, and number features are morphologically expressed on determiners, nouns, and verbs. Italian children who were exposed to the L2 later could rely more on L1 morphological properties, that is they are more sensitive to the lack/presence of number features compared to the children that are exposed to the L2 earlier in life, and thus had more time to handle the two (L1 and L2) different morphological systems.

A subsequent post-hoc investigation of these two variables (age and age of onset) revealed that there was a relationship between them: A later age of onset was more common in the older children. Given this relationship, the lack of a condition mismatch in the early age of onset children is to be expected, since they are predicted to have a more native-like performance in this kind of task, and in monolingual English children a lack of mismatch is the most common pattern after the age of 5 (Johnson, de Villiers & Seymour 2005). Consistently with this explanation, the late age of onset children are expected to display a masking effect, as they are expected to perform similarly to the second language learners (looking for as many cues as possible).

Both children and adult second language learners additionally displayed an interaction between number and masking, which was also found in previous investigations (Morales Reyes & Montrul 2020). In particular, the masking effect was significantly larger in singular sentences than in plural sentences (for children see Figure 2; for adults see L2 speakers in Figure 4). While "the duck runs" was comprehended better than "the duck swims", there was no difference between "the ducks run" and "the ducks swim". Our interpretation of this finding is that the masked noun may be automatically interpreted as plural. In other words, when a listener hears the /s/, "plurality" is activated, both when the noun phrase is truly plural (e.g., "the ducks run/swim) and when the noun is singular and the verb start with /s/ (e.g., "the duck swims"). Under this scenario, the only odd sentence (i.e., the one that requires additional reanalysis) is the "the duck swims". Further research is necessary to understand if this pattern is consistent and replicated, and, if so, to understand if our interpretation of such result is correct. Particularly, we want to stress that the sample of Italian children tested in this study is too small to draw certain conclusions, and, as such, this may be seen as an exploratory study. Further research with a larger and more controlled sample may help ascertain the validity of our findings.

In summary, this study showed that, as a general pattern, second language learners (whether children or adults) and also native speakers are more accurate in interpreting agreement when both noun and verbal cues are available. Additional results indicate that the effect of masking is particularly evident when the sentence is singular, and also that for children this effect is larger in individuals with a later age of onset and in the older subjects.

# Acknowledgements

We would like to thank Kathleen Furi for recording the stimuli, and all the children and adults who participated in the test. This project was funded by the Primus Research Program of Charles University, grant 19/Hum/14. The online adult data collection was funded by the Department of Social, Political and Cognitive Science of the University of Siena (DISPOC young researchers research funds 2021). We are thankful for their support.

# Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Appendix A

The type of picture selected during the sentence-picture matching task are represented in Figure A1, for the group of young L2 speakers, and in Figure A2 for the group of adult L2 speakers.

“Correct” stands for correct picture (e.g., one girl swinging). “Distractor” identifies the picture representing the same action but incorrect numerosity of the subject (e.g., two girls). “Filler” identifies the picture representing a different action (e.g., one girl playing hopschotch).

Figure A1: Percentage of correct/distractor/filler picture selection for each young L2 speaker

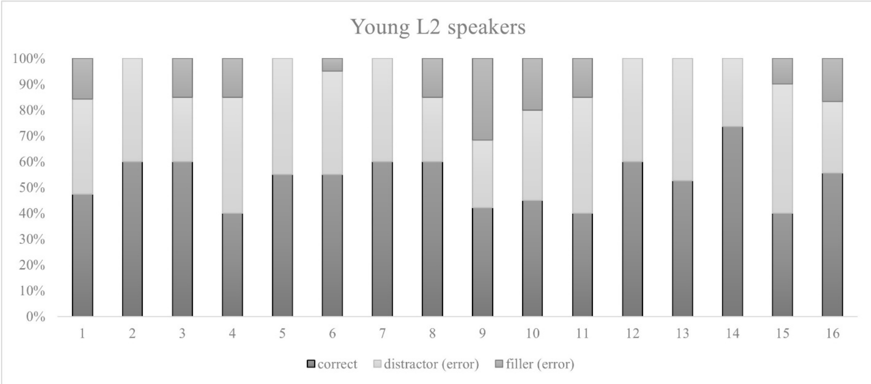
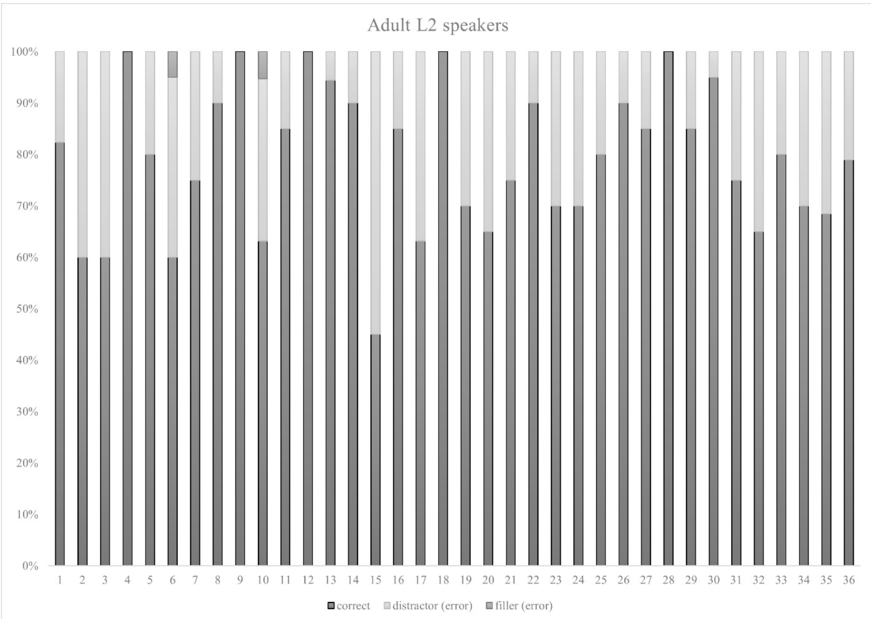


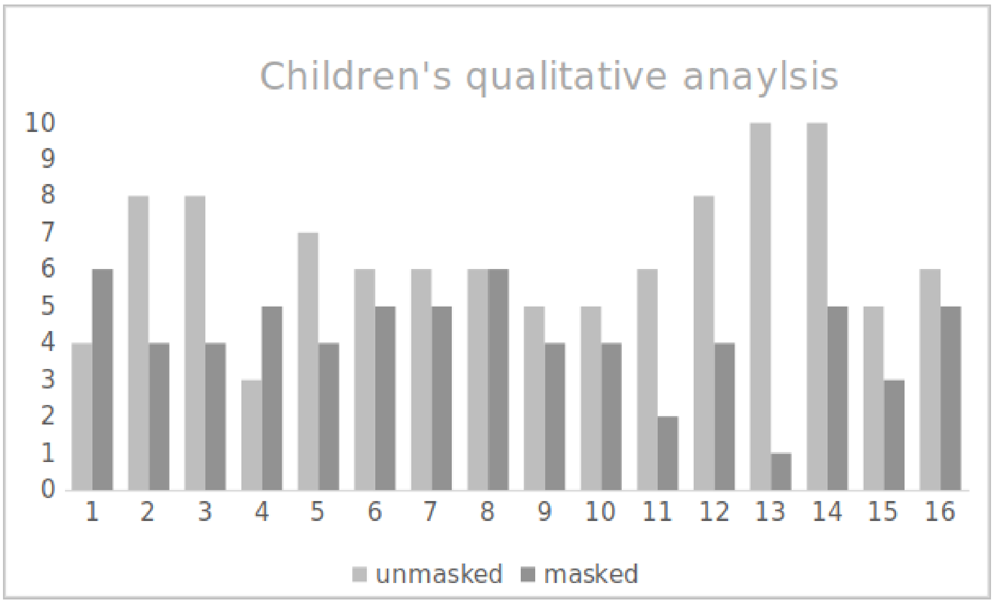

Figure A2: Percentage of correct/distractor/filler picture selection for each adult L2 speaker




# Appendix B

Figure B1 displays a comparison in overall accuracy for each child participating in this study. As it may be observed, most participants are shown to be more accurate in the control condition, meaning that most children are more accurate when agreement information is available both from the subject and from the verb. This is a reflection of the main effect of masking we observed in the statistical analysis. A few

Figure B1: Children's individual accuracy (max score per condition = 10)




subjects however escape this general pattern, and we will describe their metadata in more detail to understand whether there are additional reasons for that. The subjects to escape the general pattern are 1, 4 and 8. First, it is interesting to observe that there is a very substantial overlap in the answers regarding English use among these three participants:

All 3 children report use of English with teachers, use of Italian with classmates, relatives, friends, and family friends. All report no effort of the parents for them to use English, and no use of English in TV, in audiobooks or in books. The first difference observed regards homework, where only child 8 is reported as having difficulties. All children are not reported having private English classes outside of school.

All parents are highly educated, having at least a bachelor's degree. All parents obtained their degrees in Italy, but the parents of child 4 and child 8 are not of Italian origin: one parent of child 8 was born in an English-speaking country, but moved to Italy shortly after. One parent of child 4 was born in a non-English speaking foreign country, but moved to Italy at the age of 2.

Interestingly, 2 of the children displaying unusual behaviour are born in December (4 and 8). The years of birth of children 1, 4 and 8 are: 2012, 2011, 2014. Child 8, born in December 2014, is the youngest child of this sample. All 3 children are male. They attend rather different grades (1 = 2, 4 = 4, and 8 = 1), due to their different ages. Child 8 is the only child in the sample to have lived abroad, and specifically in a non-anglophone country. This is also reflected in another question, where child 8 is the only one in the sample to report that most of the parents of his friends are not from Italy or from an anglophone country.

Very interesting answers come from the use of the computer. First, when it comes to the use of Italian and English all three children have answers that fit well with the rest of the sample. Children 1 and 8 use the computer in Italian very often, and in child 4 uses it sometimes. Children 4 and 8 use the computer in English sometimes, and child 1 uses it rarely.

What is striking, however, is the use in other languages: children 4 and 8 are the only ones to report use of a third language in the entire sample (in both cases "sometimes").

The evaluation of the English proficiency of these three children indicate that they may be low on the spectrum of this sample: For the question on their comprehension of oral English content, parents report "not at all", "a little" and "not at all" for 1, 4 and 8 respectively.

For comprehension of TV in English, similarly, they report "not at all", "a little" and "a little" for 1, 4 and 8 respectively. For their English production, the values indicated are "a little", "a little" and "not at all" for 1, 4 and 8 respectively. In terms of comprehension in reading, the scores are "a little", "a little" and "not at all", while for writing they are "a little", "a little" and "not at all". Overall, it appears that these three children are not comfortable in their use of English. In terms of onset, there is complete variability among them, so onset does not seem to be the reason these three children stand on their own: child 1 was exposed to English after 4, child 4 after 6, and child 8 between 1 and 3. All three parents reported that English learning is important for them, like most parents in the sample, so also this motivational aspect does not seem to play a role.

Finally, questions assessing language difficulties indicate that none of these children have hearing problems or experienced issues learning to read and write. However, the parent of child number 8 reports that the child experienced difficulties learning to pronounce Italian words (one of only two cases in the whole sample) and difficulties in learning to pronounce English words (one of only three cases in the whole sample).

In conclusion, this qualitative analysis indicates that the entire sample is behaving very consistently with the exception of three participants. Of these three participants, two happen to be those that experience the use of a third language in everyday life, and all three seem to be on the lower part of the spectrum when it comes to their English skills. Their ages vary, but it may be worth noticing that one of them is the oldest child in the sample. These factors, combined, may explain why their behaviour differs from the behaviour of the other children in the sample.

## References

Bates, D., R. Kliegl, S. Vasishth & H. Baayen. 2015. Parsimonious mixed models. *arXiv*, Note = 04967.

Bates, D., M. Mächler, B. Bolker & S. Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv*, Note = 5823.

Bittner, D., W. U. Dressler & M. Kilani-Schoch. 2011. *Development of verb inflection in first language acquisition: a cross-linguistic perspective*. Vol. 21. Walter de Gruyter. 21.

Bock, K. & K. M. Eberhard. 1993. Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes* 8(1). 57–99.

Carhill, A., C. Suárez-Orozco & M. Páez. 2008. Explaining English language proficiency among adolescent immigrant students. *American Educational Research Journal* 45(4). 1155–1179.

Cilibrasi, L. & I. Tsimpli. 2020. Sensitivity to morphophonological cues in monolingual and bilingual children: evidence from a nonword task. In E. Babatsouli & M. J. Ball (eds.), *An anthology of bilingual child phonology*. Bristol: Multilingual Matters.

De Bruin, A., M. Carreiras & J. A. Duñabeitia. 2017. The best dataset of language proficiency. *Frontiers in psychology* 8. 522.

Drummond, A. 2013. Ibex farm. http://spellout.net/ibexfarm.

Jaeger, T. F. 2008. Categorical data analysis: away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language* 59(4). 434–446.

Johnson, V. E., J. G. de Villiers & H. N. Seymour. 2005. Agreement without understanding? The case of third person singular /s/. *First Language* 25(3). 317–330.

Morales Reyes, A. & S. Montrul. 2020. Variational learning in child L2 acquisition? Comprehension of verbal agreement in Spanish and English. *Language Learning and Development*. 1–20.

Pauls, F., T. Macha & F. Petermann. 2013. U-shaped development: an old but unsolved problem. *Frontiers in Psychology* 4. 301.

Schad, D. J., S. Vasishth, S. Hohenstein & R. Kliegl. 2020. How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *Journal of Memory and Language* 110. 104038.

Siegler, R. S. 2004. U-shaped interest in u-shaped development-and what it means. *Journal of Cognition and Development* 5(1). 1–10.

Song, J. Y., M. Sundara & K. Demuth. 2009. Phonological constraints on children's production of English third person singular -s. *Journal of Speech, Language, and Hearing Research* 52. 623–642.

Tsimpli, I. M. 2014. Early, late or very late?: timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* 4(3). 283–313.