Language Acquisition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hlac20

Considering the whole paradigm: Preschoolers' comprehension of agreement is not uniformly late

Hannah Forsythe & Cristina Schmitt

To cite this article: Hannah Forsythe & Cristina Schmitt (2021) Considering the whole paradigm: Preschoolers' comprehension of agreement is not uniformly late, Language Acquisition, 28:3, 272-293, DOI: 10.1080/10489223.2021.1887872

To link to this article: https://doi.org/10.1080/10489223.2021.1887872

LANGUAGE ACQUISITION
2021, VOL. 28, NO. 3, 272–293
https://doi.org/10.1080/10489223.2021.1887872

Considering the whole paradigm: Preschoolers' comprehension of agreement is not uniformly late

Hannah Forsythe [a] and Cristina Schmitt [b]

[a]University of California; [b]Michigan State University

ABSTRACT

Many languages encode phi-features via overt morphology, yet children's use of this morphology in comprehension tasks varies widely. Here, we use a picture-selection task to test comprehension of Spanish verbal agreement and clitics, comparing performance across and within each paradigm to examine the effect of two factors: (i) phonological salience, and (ii) semantic (under) specification. Both paradigms encode the same person and number features, but clitics may be easier to comprehend than agreement because they carry more phonological material. Within each paradigm, first- and second-person morphology may be easier to comprehend than third-person because they carry an explicit person feature. We find limited support for phonological salience and stronger support for semantic (under)specification. However, we also find evidence for a third factor affecting interpretation of third-person morphology: discourse prominence. Both adults and children permit third-person agreement and clitics to refer to the speaker and/or addressee if they have been mentioned in the immediately preceding context.

ARTICLE HISTORY
Received 2 July 2019
Accepted 4 February 2021

1. Introduction

Abstract features are part of the core syntax of natural languages, and features like person, number, and gender are commonly realized in the overt morphology of many languages (see Corbett 2006, among others). To acquire this morphology, children must establish a mapping between each morphological exponent (e.g., /s/ in *eat**s***) and one or more underlying features (e.g., third-person singular), which then receive an interpretation (e.g., a single nonspeaker, nonaddressee referent). This article takes up the question of how children acquire this mapping, focusing on the comprehension of Spanish subject-verb agreement and object clitics.

Research on subject-verb agreement and object pronominals presents a very heterogeneous developmental picture, with performance varying widely between languages and between comprehension and production studies. Studies on the acquisition of subject-verb agreement suggest an asymmetry between early production and later comprehension, depending on the language. Children produce adult-like subject-verb agreement particularly early in languages with rich agreement (see Phillips 2010 and references therein), such as Italian and Spanish, where all forms are spontaneously produced by age 2, and errors are extremely rare (Italian: Guasti 1993, Spanish: Clahsen, Aveledo & Roca 2002). But even in the morphologically poor system of English, children spontaneously produce third-singular *-s* in 90% of obligatory context by 2;02–3;10 (Brown 1973) and reliably produce it in elicited production tasks by age 3;05 (Theakston, Lieven & Tomasello 2003). In contrast, the view from comprehension studies is more complicated. In picture-selection studies, children as old as 5 and 6 fail to use number agreement to infer the cardinality of a masked or null subject not only in morphologically poor languages like English (Johnson, de Villiers & Seymore 2005) but also in morphologically

CONTACT Hannah Forsythe ani.forsythe@gmail.com Language Sciences, University of California, 2211 Social Science Plaza B, UCI, Irvine, CA, 92612.

rich languages like Spanish (Pérez-Leroux 2005), Farsi (Rastegar, Shirazi & Sadighi 2012), and Xhosa (Gxilishe et al. 2009). Slightly different versions of the picture-selection paradigm reveal above-chance performance at earlier ages in spoken French (age 2;04–2;08; Legendre et al. 2014:25) and Spanish (3;05–4;02; González-Gómez et al. 2017), indicating that some of the variation in performance between comprehension and production may be because picture-selection tasks are more demanding than spontaneous production. Further evidence of task effects comes from German, where children achieve earlier above-chance performance in preferential looking tasks compared to pointing-and-looking tasks (Brandt-Kobele & Höhle 2010), and in Dutch, where children show a much narrower performance gap between *elicited* production and picture-selection tasks (Verhagen & Blom 2014). On the other hand, children's performance in comparable preferential looking and picture-selection tasks is much more delayed in English relative to spoken French (Legendre et al. 2010a, 2014), with Spanish somewhere in the middle (Legendre et al. 2014), indicating that language differences also contribute to variation in performance. In this article, we compare children's use of different agreement forms within the same language and the same task to shed light on some of the potential reasons for this variable performance across different languages and different tasks.

When it comes to the acquisition of object morphology, results are sparser, and comprehension can't be as easily compared to production because of the so-called null object phase during which children spontaneously omit objects in transitive contexts. Most accounts of object omission assume that children lack the grammatical and/or processing resources to insert the object in its proper place within the derivation (see Mateu 2015 for a brief review; Pérez-Leroux, Pirvulescu & Roberge 2017 for more lengthy theoretical and experimental work). However, these accounts do not tell us one way or the other whether children are able to map the features of an object to an appropriate morphological exponent. Indeed, when children do produce object morphology in elicitation tasks, it tends to be the correct morphology (see Pérez-Leroux et al. 2012 and references therein). Comprehension studies suggest that, as with subject-verb agreement, performance varies across languages. In Xhosa, where both subjects and objects agree with the verb in number and class (and number morphology varies by class), comprehension of object number marking appears to be just as slow to develop as it is for subject number marking (Gxilishe et al. 2009). But in spoken French, subject and object clitics seem to follow separate developmental paths, with early comprehension of subject number marking at about 2½ (Legendre et al. 2010a, 2014) and later comprehension of object number and gender marking between 4 (Pirvulescu & Strik 2014) and 5 (Zesiger et al. 2010).

There are many potential reasons for this wide variation in results, but this article focuses on two factors in particular: (i) the morphophonological properties of individual exponents, and (ii) the semantics of different person and number features. We examine these factors using a picture-selection task in Spanish to compare children's performance across the full range of first-, second-, and third-person agreement and accusative clitics shown in Table 1. This allows us to compare children's comprehension of a wider array of agreement markers and clitics than have previously been compared using a single task.

Table 1. Spanish regular present-tense verbal agreement and accusative object clitics.

	Singular	Plural	Singular	Plural
First person	*-o*	*-mos*	*me*	*nos*
Second person	*-s*	–[1]	*te*	–
Third person	*-Ø*	*-n*	*lo* (masc.) *la* (fem.)	*los* (masc.) *las* (fem.)

[1]This study excludes 2nd person plural agreement and clitics because they are subject to a rather complex pattern of regional variation, with some dialects differentiating between formal and informal register and others not (see Lipski 1994 for details). In Mexico City Spanish, the variety studied here, there is no formality distinction in the 2nd person plural, and this creates syncretism with the 3rd person plural. In subject position, the pronominal subject *ustedes* triggers the same agreement as 3rd person plural (*–n*). In object position, the 2nd person plural accusative clitic is syncretic with 3rd person plural *los* (masculine) or *las* (feminine). To avoid the ambiguity caused by this syncretism, we do not include a 2nd person plural condition.

Table 2. Comprehension asymmetries predicted by the phonological hypothesis and the semantic hypothesis.

	Agreement		Clitics	Semantic hypothesis
	1st plural *–mos*		1st plural *nos*	earlier
	1st singular *–o*		1st singular *me*	
	2nd singular *–s*		2nd singular *te*	
	3rd plural *–n*		3rd plural *los, las*	----------
	3rd singular *–ø*		3rd singular *lo, la*	later
Phonological hypothesis	later	<<	earlier	

First, we examine the effect of phonological salience by comparing performance across agreement and clitics, since clitics encode the same person and number features[2] but are more phonologically salient in the sense that they are comprised of a full syllable. Second, we examine the effect of semantic underspecification by comparing comprehension across first, second, and third persons, since the third person is semantically underspecified for person, in contrast to first and second persons, which explicitly encode reference to the speaker and addressee. Along the way, we compare children's ability to use person agreement in this picture-selection task to their use of number agreement in other Spanish-language picture-selection tasks (e.g., González-Gómez et al. 2017; Legendre et al. 2014; Pérez-Leroux 2005) and analyze the patterns of nontarget responses produced by adults and children.

2. Background and research questions

Children's mixed performance with subject-verb agreement and object pronominals creates a puzzle that researchers have sought to explain by appealing to many factors, including (i) task-based performance factors, (ii) morphophonological factors, and (iii) semantic/pragmatic factors.

As mentioned in the introduction, children's performance varies across tasks. In particular, picture-selection tasks initially seemed to suggest that children comprehend number agreement later than they produce it (Gxilishe et al. 2009; Johnson, de Villiers & Seymore 2005; Pérez-Leroux 2005; Rastegar et al. 2012), while later variations on the picture-selection task (González-Gómez et al. 2017; Legendre et al. 2014; Verhagen & Blom 2014), as well as preferential looking tasks (Brandt-Kobele & Höhle 2010; Legendre et al. 2010a) tended to reveal earlier and/or better performance than the original picture-selection tasks. Currently, the full extent of task effects is unknown, since different task modifications were introduced in different languages. In our study, we will not be able to tease apart cross-linguistic effects from task effects, but what we can do is extend the picture-selection task to the case of person agreement. If the demands of picture selection are the primary driver of children's poor performance in the original picture-selection studies, then children should also fail to use person agreement in this study until around age 5–6.

A second factor that may help explain the cross-linguistic differences observed in children's use of agreement markers is their morphophonological properties. Legendre et al. (2014) administered

[2]In the third person, clitics additionally encode gender.

similar video-selection tasks across spoken French, Spanish, and English and found early comprehension of French singular and plural preverbal agreement markers[3] (2;05), slightly later comprehension of Spanish plural (2;05–3;11) but not singular verbal agreement, and no evidence for comprehension of either singular or plural verbal agreement in English (ages tested: 2;00–3;10). The authors propose that these cross-linguistic differences can be attributed to differences in the perceptual salience of agreement morphology, complemented by the concept of "cue reliability," drawn from early research on morphosyntax such as Brown (1973), Slobin (1973), and MacWhinney, Bates & Kliegl (1984). In spoken French, preverbal agreement markers *ils* 'they' and *il* 'he' are phonetically distinguishable in prevocalic position, where a process known as liaison causes the normally silent final segment of plural *ils* to surface as [z] in the onset of the following syllable (e.g., *ils embrassent* /il.**z**eã.bʁas/ or /i.**z**eã.bʁas/ 'he hugs,' where it contrasts with the final segment of the singular *il* (e.g., *il embrasse* /i.**l**eã.bʁas/ 'they hug'; see Legendre et al. 2014:23). This contrast is perceptually salient for two reasons. First, since [z] is a strident it is more acoustically salient than the segment [n] that distinguishes Spanish plural agreement (e.g., *agarra**n*** '[they] grab' from singular agreement (e.g., *agarra_* '[s/he] grabs.' Second, unlike the strident allomorphs [s] and [z], which distinguish English singular agreement (e.g., *grab**s***) from plural agreement (e.g., *grab_*), French liaison consistently places the segment [z] in the onset of the syllable—an environment that has been argued to enhance consonant perceptibility (Benki 2003; Redford & Diehl 1999). In addition to being perceptually salient, the segment [z] can be considered a reliable cue because it is consistently linked to a single meaning. In spoken French, not only does [z] mark plural on third-person verbs, it also marks plural on determiners (e.g., *les amis* /le**z**ami/ 'the friends' and is thus a reliable cue to plurality across verbal and nominal domains. In Spanish, word-final [n] is associated with plurality when it appears on verbs, but on nouns it has no meaning, making it a somewhat weaker cue to plurality. And in English, word-final /s/, realized as [s] or [z] or [əz], is associated to singularity when appearing on a verb but to the exact opposite meaning when found on nouns, making this segment a downright unreliable cue to singularity. Thus, a combination of phonological salience and cue reliability may explain why we observe the earliest above-chance performance in spoken French.

Phonological effects may also explain some within-language comprehension asymmetries as well. For instance, Childers et al. (2001) report that even children under 3 perform above chance in a picture-selection task when it comes to interpreting the number features of auxiliaries *is* and *are*, which are comprised of a full syllable. Also, Pérez-Leroux (2005) observes that in both Spanish and English, children perform more poorly in the condition with a null exponent—namely, the third-singular form in Spanish (e.g., *agarra_* '[s/he] grabs') and the third-plural form in English (e.g., *grab_*).[4] Phonological salience even appe'are Xing.' ;ars to condition comprehension of nominal plural -s in English: In a preferential looking task, children looked longer at a plural picture in response to nouns inflected with the plural allomorph [s], which has the longest duration, but they did not do so for nouns with other plural allomorphs (Davies, Rattanasone & Demuth 2016).

In this article, we will probe how children's comprehension is conditioned by the property of phonological salience, as defined and operationalized in Polišenska (2010). *Phonological salience* is defined as "how easy it is to detect the root and the affix in a given word form" and is determined by a combination of the morpheme's content and its environment: Inflectional morphemes should be acquired earlier when they have more phonological material, when they are word-initial (or at least, not word-internal), when they are stressed, and when they are not subject to reductive processes like shortening. If phonological salience is a factor in children's use of agreement morphology in comprehension, then we would expect children acquiring Spanish to achieve better performance with

[3]These authors assume that preverbal subject clitics have the status of agreement markers. Although not all authors agree with this analysis (e.g., de Cat 2002), there seems to be evidence for it applying to modern spoken French (see Culbertson 2010; Legendre et al. 2010b). For ease of exposition we will describe results as if these forms are agreement markers, as Legendre et al. (2014) assume to be the case.

[4]Note that Childers et al. (2001) find the opposite asymmetry in Chilean Spanish for comprehension of present progressives *está Xndo* versus *están Xndo* 'is Xing' vs. 'are Xing.'

accusative clitics than with agreement markers because, except for the first-person plural, clitics have more phonological material than their corresponding agreement markers (see Table 1).

The last factor we will address in this article is the semantics of the person and number features themselves. Literature on the acquisition of person suggests that morphemes instantiating first- and second-person features should be simpler to interpret than their third-person counterparts because third person is considered to be formally unmarked (e.g., Benveniste 1971 and subsequent work). Whereas first- and second-person features denote the speaker and addressee respectively,[5] the third person imposes no restrictions on the potential referent. As long as other features like gender are satisfied, third-person expressions may refer to anyone, even the speaker or addressee, such as in example (1), where the third-person possessive *her* picks out the same referents as the first-person plural pronouns *us* and *we*.

(1) Every one of us loves **her** mother. That's why we are here planning this mother-daughter brunch.

The developmental implications of this formal difference are most clearly spelled out by Legendre & Smolensky (2012), using data originally reported in Legendre et al. (2011) for spoken French. In this task, the child and two experimenters fish for animal pictures out of a basket, with each participant fishing for a different kind of animal. Children's comprehension of first-, second-, and third-person preverbal agreement markers was tested by asking what animal each participant was fishing for, as in (2).

(2)

(a) Experimenter 1:
Qu'est-ce que ***tu*** *attrapes?*
'What are you catching?'

(b) Experimenter 2:
Qu'est-ce que ***j'*** *attrape?*
'What am I catching?'

(c) Experimenter 2:
Qu'est-ce qu' ***elle*** *attrape?*
'What is she catching?'

Children aged 2;06 reliably produced target responses in the first- and second-person conditions (78% and 84% respectively) but not in the third-person condition (19%), even though their production as reported in parental MCDI questionnaires revealed no such person asymmetry, and even though elicited production studies show children ages 2;00–2;07 producing plenty of third-person preverbal markers (Jakubowicz & Rigaut 1997). This suggests that even in spoken French, where comprehension of preverbal number agreement matures early on, children still comprehend the third person *elle* 'she' later than they produce it and, crucially, later than they comprehend first- and second-person preverbal markers *je* 'I' and *tu* 'you (sg).'

Adopting an OT framework, Legendre & Smolensky (2012) argue that the reason children comprehend third-person markers later than they produce them and later than they comprehend first- and second-person markers is because the third person is semantically underspecified. In a nutshell, third-person forms carry no person features, and this causes ambiguity for the listener without causing optionality for the speaker. On the comprehension side, when children hear a third-person expression like *elle*, they must decide which interpretation is the most optimal. Since this form

[5]More precisely, these features introduce presuppositions that restrict the denotation of whatever referring expression they accompany, returning a value if their requirements are satisfied and returning nothing if they are violated (e.g., Sauerland 2003; Charnavel 2019, and many others).

carries no person features, it is a grammatically licit choice for *any* feminine referent, including the speaker and addressee if she is female. Adults resolve this ambiguity by performing a kind of pragmatic inference: They assume that the speaker *would* have used a form with person features if they *could* have, and the fact that the speaker did *not* therefore implies that first- and second-person features are ungrammatical in this situation. In other words, *elle* must refer to a nonspeaker, nonaddressee referent because otherwise the speaker would have used a more informative form like *je* or *tu*. This inference is referred to as an "Implicated Presupposition" (see Sauerland 2003), and the assumption that speakers will always use the most informative form possible is referred to as "Maximize Presupposition" (see Heim 1991). Legendre and Smolensky argue that children do not initially assume this maxim and therefore fail to calculate Implicated Presuppositions. On the production side, however, there is no need to make this kind of pragmatic calculation. When children intend to refer to a nonspeaker, nonaddressee referent, they simply cannot use forms like *je* and *tu* because these forms carry person features that explicitly clash with this interpretation. The only grammatical choice is to use a form without person features, such as *elle*. Thus, while the comprehension of *elle* and other third-person expressions requires children to choose between different possible interpretations, production does not require a similar choice between different possible forms, once they know which features go with each form.

Further support for the claim that children do not calculate Implicated Presuppositions comes from the specific kind of errors that children committed in Legendre et al.'s (2011) fishing task: Children allowed the third person to refer to the addressee (i.e., themselves, 5/12 errors) or to the speaker (7/12 errors), but they never allowed the first person *je* or second person *tu* to refer to the nonspeaker, nonaddressee experimenter—all errors consisted of mixing up the speaker and the hearer (3/3 errors). Additionally, there is evidence from other languages that the third person is generally more difficult for children to comprehend, relative to the first and second persons. In English, Brener (1983) found that preschoolers comprehend third-person pronouns in overheard speech later than first- and second-person pronouns in the same context. In both Chilean and Mexican Spanish, children completing an act-out task (Miller & Schmitt 2014, Expt. 3) achieve greater accuracy with the second-person agreement marker /-s/ compared to the third-person plural /-n/ and the third-person singular null exponent. This is especially remarkable given that syllable-final /s/ in Chilean Spanish is variably aspirated ([h]) or even deleted (see Miller 2007 and references therein). This suggests that *even in a language with /s/ weakening*, children comprehend the second-person singular /-s/ agreement marker earlier than either the singular or plural third-person agreement markers.

If semantic underspecification delays children's use of person features in comprehension, what about their use of number features? The plural has been argued to be semantically underspecified for number in the same way that the third person is underspecified for person (Sauerland 2003), which would predict slower acquisition of plural relative to singular agreement. In fact, Legendre et al. (2011) argue in favor of this interpretation based on additional results from their fishing task, which were not mentioned by Legendre & Smolensky (2012). The fishing task also included a plural condition testing comprehension of first-person plural *on* (meaning either 'we' or 'one'), second-person plural *vous* ('you' plural or second person singular formal), and third-person plural *elles* ('them' feminine) in a separate block. Children performed significantly worse in the plural block, where they frequently produced singular responses, relative to the singular block, where they never produced plural responses. Legendre et al. (2011) interpret this as evidence that children struggle to calculate the Implicated Presupposition associated to the plural in the same way as they struggle with the third person.

However, evidence for the number-based comprehension asymmetry is a little more mixed than it is for the person-based asymmetry, and our study will not be able to resolve the question either. First of all, the fishing task itself may have biased children against plural responses—the child had to name two different animal names instead of just one. Second, results from number agreement studies do not consistently show that children comprehend singulars before plurals. Picture-selection tasks sometimes report better performance in the plural condition (Spanish: González-Gómez et al. 2017;

Legendre et al. 2014; Pérez-Leroux 2005; Dutch: Verhagen & Blom 2014) and other times in the singular (Farsi: Rastegar, Shirazi & Sadighi 2012; Chilean Spanish and American English auxiliaries (*está/están X*, 'is/are Xing'; Childers et al. 2001). Preferential looking studies report no singular/plural asymmetry for agreement (spoken French: Legendre et al. 2010a, 2014; German: Brandt-Kobele & Höhle 2010) or nominals (English: Davies, Rattanasone & Demuth 2016). Our study uses a picture-selection task similar to the early versions of this paradigm (e.g., Pérez-Leroux 2005), and like those tasks it allows subjects to pick a picture including two individuals in response to singular verbs (for example, *baila* '[s/he] dances' is compatible with a picture of two individual people dancing because if two are dancing then one is also dancing). This is not an ideal setup to compare performance across singular and plural conditions; therefore, we will set aside the issue of number asymmetries and concentrate instead on person asymmetries.

In sum, studies point to a variety of performance-based and linguistic factors that may interfere with children's ability to map agreement and pronominal morphology to their features. As yet, no study has directly compared any of these factors to each other, and no study except for Legendre et al. (2011) has even tested the entire agreement or clitic paradigm in a single task. In this study we will begin by pitting morphophonological and semantic factors against each other, using the full paradigm of subject-verb agreement and accusative clitics in Mexican Spanish.

3. Hypotheses and predictions

The first hypothesis we test is that phonological salience conditions children's comprehension of agreement in Mexican Spanish. Generally speaking, for any two morphemes with the same features, the one with more phonological material, more acoustically salient material, and a more perceptually enhancing phonological environment will be used earlier in comprehension tasks. Since Spanish accusative clitics instantiate the same person and number features as agreement, and since both agreement markers and accusative clitics can appear in a prosodically dependent, postverbal position as in (3), we can test this hypothesis by comparing children's comprehension across the two paradigms.

(3) (a) *Muéstrame la foto en donde baila**mos/o/s/n/ø**.*
Show-me the photo in where dance-1P/1S/2S/3P/3S
'Show me the photo where we/I/you/they/(s)he dance(s).'

(b) *Muéstrame la foto en donde Nemo está tapándo**nos/me/te/las/la**.*
Show-me the photo in where Nemo is covering-1P/1S/2S/3P/3S
'Show me the photo where Nemo is covering us/me/you/them/her.'

Every agreement marker except the first-person plural has less phonological material than the corresponding accusative clitic (which does not combine with the verbal stem directly). We therefore predict significantly poorer and/or later comprehension of each agreement marker relative to its clitic counterpart (except for first-person plural).

The second hypothesis we test is that the semantic difference between third person, on the one hand, and first and second person, on the other, condition children's comprehension of agreement. This hypothesis predicts that, for any two morphemes with the same number features and/or gender features, the one that encodes first or second person will be easier to comprehend than the one that encodes third person. We test this hypothesis by comparing third-person to first- and second-person forms with the same number (and gender) features, within each paradigm. We predict poorer and/or later comprehension of each third-person clitic or agreement marker relative to its first-person and second-person counterparts. These predictions are summarized in Table 2.

4. Methods

4.1. *Subjects*

We tested 46 native Spanish-speaking children ages 2;03–6;07 from a day care facility in Mexico City, Mexico; four were excluded from the final analysis (see the following). Adult subjects included 25 adults (12 women) recruited from among the teachers and administrators at the day care or from the Latin American community of Michigan State University. All adults were born and raised in Mexico with Spanish as their first language and were younger than 40 years old, to ensure the felicitous use of the informal second-person pronoun *tú* during the task. Two adults were excluded because they used the formal second-person pronoun *usted* when addressing the experimenter. The primary experimenter in Mexico was a female teacher from the school, and the primary experimenter in the United States was a female native Spanish-speaking undergraduate at Michigan State University. Adult subjects in Mexico were not compensated, while those in the United States received US$15 for their participation.

4.2. *Experimental stimuli and fillers*

Subjects' interpretation of present-tense agreement and object clitics was tested using a picture-selection task consisting of 30 test items (15 agreement, 15 clitics), as in (4)–(5). Subjects were presented with an array of five photos depicting (i) the subject, (ii) the experimenter, (iii) the subject and experimenter together, (iv) an unrelated teacher, and (v) two unrelated teachers together. Everyone in the photos was depicted performing the same action so that the target photo could only be identified by interpreting the agreement or clitic morpheme used in the prompt. Only feminine clitics were tested, and all experimenters were female.

(4) *Muéstrame la foto en donde salta**mos/o/s/n/Ø***
show-me the photo in which jump-**1P/1S/2S/3P/3**
'Show me the photo where we/I/you/they/she is/are jumping.'

(5) *Muéstrame la foto en donde Nemo está besándo-**nos/me/te/las/la***
show-me the photo in which Nemo is kissing-CL.**1P/1S/2S/3P.fem/3S.fem**
'Show me the photo where where Nemo is kissing us/me/you/them/her.'

Fourteen out of 30 experimental trials were preceded by fillers like those in (6)–(7). Subjects were presented with an array in which each of the five person(s) performed a different action, so that the target photo could be identified by interpreting the lexical content of the verb. Fourteen out of 30 experimental trials were preceded by distractors like (8), asking the subject which of two cartoon characters had more of various objects and substances. Two out of 30 experimental trials appeared at the beginning of the block and were not preceded by anything.

(6) *Muéstrame la foto en donde hay alguien saltando/bailando/ . . .*
show-me the photo in which there-is someone jumping/dancing/ . . .
'Show me the photo where someone is jumping/dancing/etc.'

(7) *Muéstrame la foto en donde Nemo está besando/peinando/ . . . a alguien.*
show-me the photo in which Nemo is kissing/combing . . . A someone.
'Show me the photo where Nemos is kissing/combing/etc. someone.'

(8) *¿Cuál pato tiene más jabón/jabones?*
Which duck has more soap/soaps?
'Which duck has more soap/soaps?'

For the agreement condition, the following intransitive (unaccusative or unergative) or detransitivized verbs were used: *saltar* 'jump,' *aplaudir* 'clap,' *dormir* 'sleep,' *dibujar* 'draw,' and *bailar* 'dance.' In the clitic condition, photos showed a puppet named Nemo performing the following actions on each of the five person(s): *besar* 'kiss,' *peinar* 'comb,' *lavar* 'wash,' *tapar* 'cover,' and *tocar* 'touch.'

Fillers used the same 10 actions, plus an additional four items in which the target action was sitting (*sentada*) and lying down (*acostada*).

4.3. *Design and procedure*

Agreement and clitics were presented in separate blocks, with the agreement block first. The goal was to make the task less taxing by (i) allowing children to concentrate on subject properties and object properties one at a time, and (ii) allowing them to start with those forms that encode fewer features (agreement does not carry gender).[6] Subjects were randomly assigned to one of two different versions of the task, each with a different random ordering of test items. Fillers and distractors followed the experimental items in alternating order. Fillers were presented in pseudorandom order such that each filler depicted a different action from the immediately preceding or following test item, and this ordering was reshuffled after every other subject to mitigate the possibility that any particular order would bias the interpretation of the test items.

Photos not including the subject were taken beforehand and preinserted into the arrays in random order. Photos including the subject were taken no more than one week before testing and then inserted into the arrays in random order.

Testing was preceded by a familiarization phase in which the subject was asked to identify each of the actors by name (him/herself, the primary experimenter, and the other two adult women), and any errors were corrected. Next, the primary experimenter introduced the task and obtained consent through the following:

(9) Vamos a ver algunas fotos de personas haciendo varias cosas y tú me vas a señalar la foto que yo te diga, ¿te parece? Pero sólo me vas a señalar una foto nada más, ¿bien?

'We're going to see some photos of people doing different things and you're going to point out the one I tell you, sound good? But you can only pick one photo, okay?'

Halfway through each block and between blocks, there was a short break in which children were given a sticker. Any child who refused or repeatedly displayed unwillingness to participate in any part of the test was excused (see the following). The entire procedure, including the photo shoot and testing, lasted approximately 30 minutes.

4.4. *Coding and exclusions*

Responses were recorded on a sheet of paper by the secondary experimenter and then transferred to a spreadsheet for coding and analysis in R (R Core Team 2013). Any photo containing the target referent was counted as a target response, even if it also included another referent as well. This is because a photo with additional referents is still technically compatible with the prompt. The listener can only exclude the additional referent by appealing to the uniqueness presupposition of the singular definite *la foto* 'the photo,' which implies that only one photo in the display uniquely meets the speaker's requirements, i.e., the photo with the target referent and no additional referents. Since children find this presupposition difficult to generate, particularly when the definite also happens to be singular (Munn, Miller & Schmitt 2006), we thought this was too high a standard to impose. Note that this coding decision makes singular conditions inherently easier, since plural responses could be accepted in singular conditions (as long as they include the target referent), but singular responses

[6] A mixed-effects logistic regression model failed to find a significant effect of item order on children's performance within either the agreement block ($\beta = 0.01$, $SE = 0.01$, $z = 0.49$, $p = .62$) or the clitic block ($\beta = -0.01$, $SE = 0.01$, $z = -0.96$, $p = .34$). We therefore assume that any differences between blocks are due to differences between agreement and clitics rather than to the order in which the blocks were presented.

could not be accepted in plural conditions. Therefore the following analyses only compare and contrast forms with the same number feature.

Three children were excused before completing the task, and one child was excluded due to an extremely low score on filler questions (50% correct of all intelligible responses), leaving a total of 42 child subjects. Two adults were excluded for failure to address the primary experimenter using the informal second person *tú*, leaving a total of 23 adult subjects.

5. Results

Adult and child responses are reported in Tables 3 and 4 (highlighted cells represent target answers). To test whether comprehension is conditioned by (i) the contrast between less phonologically salient agreement markers and more salient clitics, and/or (ii) the contrast between semantically

Table 3. Adult responses ($N = 23$; target responses in shaded cells).

	Agreement conditions					Clitic conditions				
Picture choice	**1Sg**	**1Pl**	**2Sg**	**3Sg**	**3Pl**	**1Sg**	**1Pl**	**2Sg**	**3Sg**	**3Pl**
investigator	56	0	1	14	1	67	0	0	19	1
adult & investigator	7	66	3	5	26	1	68	6	0	14
adult	2	0	64	17	0	1	0	63	5	0
other female-Sg	4	0	0	30	3	0	0	0	44	0
other female-Pl	0	3	1	3	38	0	1	0	1	54
other	0	0	0	0	1	0	0	0	0	0
no answer	0	0	0	0	0	0	0	0	0	0
proportion target	0.91	0.96	0.97	0.48	0.55	0.99	0.99	1.00	0.65	0.78
(SD)	(0.22)	(0.15)	(0.13)	(0.37)	(0.27)	(0.06)	(0.06)	(0)	(0.35)	(0.31)

Table 4. Child responses ($N = 42$; target responses in shaded cells).

	Agreement conditions					Clitic conditions				
Picture choice	**1Sg**	**1Pl**	**2Sg**	**3Sg**	**3Pl**	**1Sg**	**1Pl**	**2Sg**	**3Sg**	**3Pl**
investigator	57	9	2	15	12	88	11	5	29	6
child & investigator	54	86	31	33	44	23	90	39	9	40
child	6	14	88	39	34	4	14	77	28	22
other female-Sg	3	1	2	20	13	3	4	0	48	8
other female-Pl	5	15	3	17	21	8	6	5	11	49
other	1	0	0	0	2	0	0	0	1	1
no answer	0	1	0	2	0	0	1	0	0	0
proportion target	0.88	0.69	0.94	0.30	0.17	0.88	0.72	0.92	0.47	0.39

Table 5. Estimated effects (*SE*) of binary logistic regression models predicting the probability of a target response as a function of form (reference: agreement, contrast: clitic) and person (reference: third, contrast: first, second). Singular and plural items analyzed separately.

	Adults		Children	
	Singulars	Plurals	Singulars	Plurals
Intercept	−0.1 (0.4)	0.28 (0.39)	−1.05 (0.33)**	−1.77 (0.32)***
Form: clitic	0.96 (0.41)*	1.26 (0.52)*	0.98 (0.42)*	1.27 (0.37)***
Person: first	3.1 (0.58)***	3.33 (0.72)***	3.35 (0.41)***	2.59 (0.39)***
Person: second	4.39 (0.85)***	–	4.09 (0.53)***	–
First person x clitic	1.08 (1.2), *n.s.*	−0.24 (1.28), *n.s.*	*−1.02 (0.54), p = .059*	−1.04 (0.48)*
Second person x clitic	17.49 (256), *n.s.*	–	*−1.25 (0.67), p= .063*	–
Random effects structure:	(1|subj/item)	(1|subj) + (1|item)	(1|subj) + (1|item)	(1|subj) + (1|item)

underspecified third-person and semantically specified first- and second-person forms, we ran a binary logistic regression, modeling the probability of a target response as a function of form (reference: agreement, contrast: clitic) and person (reference: third person, contrast: first person, second person), which we ran separately for singular and plural conditions. Each model included random intercepts and/or slopes for subjects and items, which were removed one by one if the model failed to converge. Fixed effects and final random effects are reported in Table 5.

For adults and for children, there was a significant effect of form and a significant effect of person in both the singular and plural conditions. Both adults and children produced more target answers in the clitic condition than they did in the agreement condition, consistent with the phonological hypothesis, and they produced more target answers in first- and second-person conditions than they did in the third-person condition, consistent with the semantic hypothesis. Interestingly, however, the estimated coefficients for first and second persons (β = 2.59–4.39) were larger than those for clitics (β = 0.96–1.27), suggesting that person has a stronger effect than form. Looking at the group-level accuracy in Tables 3 and 4 we see why: The vast majority of all participants' nontarget responses, in both the agreement and clitic conditions, occurred in the third person. We analyze these nontarget responses in section 5.2.

For children, there was also a significant interaction between clitics and first-person plural, as well as trends toward a significant interaction between clitics and first-person singular and between clitics and second-person singular. All three of these interactions are negative, suggesting that even though children may have produced more target answers in the clitic condition compared to the agreement condition overall, this contrast may be weaker or perhaps nonsignificant in first- and second-person conditions. We therefore followed up with a series of five logistic regressions comparing each individual agreement marker to the corresponding clitic with the same person and number features (e.g., first-person singular *-o* vs. first-person singular *me*). Each model included random intercepts and/or slopes for participant.[7] As expected, there was a significant effect of form within the third-person singular (β = 0.91, SE = 0.30, z = 3.09, $p < .01$) and third-person plural (β = 1.50, SE = 0.36, z = 4.16, $p < .001$) with more target answers in response to the clitics than to their corresponding agreement markers. However, there was no evidence of a form effect within the first-person singular (β = −0.00, SE = 0.40, z = 0.00, p = 1.00), second-person singular (β = −0.43, SE = 0.54, z = −0.79, p = .43), or first-person plural (β = 0.18, SE = 0.29, z = 0.60, p = .55). In other words, the overall effect of form is driven by the third person.

In sum, performance in this task seems to be conditioned by both phonological and semantic differences, but there is more consistent evidence for the semantic effect. Consistent with the semantic hypothesis, children perform better in first- and second-person conditions relative to third-person conditions. Meanwhile, we only find partial support for the phonological hypothesis. Children perform better in the clitic condition than in the agreement condition—but this effect is only significant within the third person, not the first or second persons. The next two sections take a closer look at children's comprehension of first- and second-person agreement and the nature of children and adults' nontarget responses in the third person.

[7]Random effects for item were not included, since agreement and clitics were tested using different items.

5.1. *Developmental analysis of first- and second-person agreement*

As a group, children in our sample appear to comprehend person agreement, insofar as their raw accuracy scores in first-person singular, first-person plural, and second-person singular agreement conditions are all well above chance (88%–94% accuracy versus 40% chance in first- and second-person singular; 69%–72% accuracy versus 20% chance in first-person plural). These raw scores are also much higher than those reported in picture-selection studies on number agreement (e.g., González-Gómez et al. 2017; Legendre et al. 2014; Pérez-Leroux 2005), where raw accuracy scores fall between 50% and 68% even though chance accuracy is higher in those tasks (50%). However, since this study spans a very wide age range, it is important to determine how much these overall high accuracy rates change with age. To determine the effect of age on accuracy, we ran one logistic regression with a fixed effect of age in years (centered around the mean) in each of the first-person singular agreement, second-person singular agreement, and first-person plural agreement conditions. None of these models revealed a significant effect of age, suggesting that children's high level of accuracy in these conditions is not driven solely by the oldest children.

Next, we divided children into groups by their age year (2;03–2;11, N= 7; 3;00–3;09, N= 10; 04;01–-4;11, N= 13; 5;00–5;11, N= 7; 6;00–6;07, N= 5) and compared their performance to chance. As a first-pass analysis, we compared each group's proportion of target answers in each condition to chance using single-sample t-tests. However, since proportions are not normally distributed (thus violating the t-test's assumption of normality), we also ran a second analysis using logistic regression. For each age group and condition, we built a model with no fixed effects and a random intercept for participant. The intercept of these models can be interpreted as an estimate of each age group's log odds of a target response in each condition, and this can be compared to the log odds of getting a target response simply by chance (first- and second-person singular conditions: chance accuracy = 0.4, chance log odds = $\log\left(\frac{.4}{1-.4}\right) = -0.405$; first-person plural condition: chance accuracy = 0.2, chance log odds = $\log\left(\frac{.2}{1-.2}\right) = -1.386$). See the appendix for full details of both analyses. The t-test analysis showed significantly greater than chance performance by all age groups in all three conditions (all $M > 0.57$, all $t > 2.74$, all $p < .05$), and the regression analysis showed greater than chance performance by all ages in all conditions except 5-year-olds in the first-person plural condition. Thus, even in a fairly demanding picture-selection task, we find evidence that children as young as 2;03–2;11 comprehend first- and second-person agreement markers —much younger than even the earliest above-chance performance with Spanish number agreement reported in a picture-selection task (3;05–4; González-Gómez et al. 2017).

5.2. *Error analysis of third-person agreement and clitics*

To further explore children's comprehension of third-person agreement and clitics, we must first understand why adults themselves produced so many nontarget answers in this condition. As shown in Table 3, the vast majority of adults' nontarget responses involved a violation of the person feature (agreement: 62 out of 67 total errors; clitic: 38 out of 39 errors). This was also true for children's errors: 177 out of 190 total errors in the agreement condition and 134 out of 143 total errors in the clitic condition involved a person violation. Why would participants allow a third-person clitic or agreement marker to refer to a picture of themselves or the experimenter?

One possibility is that syncretism caused participants to interpret third-person forms as second-person forms. For example, third-singular forms like *baila* 'dance-3S' and *besarla* 'kiss-3S' may have been interpreted as second-person formal 'you (singular formal) dance' and 'kiss you (singular feminine formal),' and third-person plural forms like *bailan* 'dance-3P' and *besarlas* 'kiss-3P' may have been interpreted as second-person plurals 'you (plural) dance' and 'kiss you (plural feminine).' However, syncretism does not actually explain the majority of participants' nontarget responses. First of all, it fails to explain why adults allowed third-person singulars to refer to the speaker *even more* often than to themselves (33 "investigator" responses versus 22 "adult" responses) and why children did so almost as often (44 "investigator" responses versus 67 "child" responses). As for the third-

person plural condition, the display did not even contain a picture of the addressee and a third person, which would be the expected response if third-person plural forms were interpreted as 'you (plural).' Rather, participants frequently chose the picture of the addressee and the *speaker*—the referent corresponding to 'us.' Syncretism can explain at most 22 out of the adults' 106 total nontarget responses and 67 of children's 333 total nontarget responses, failing to explain why they seem to freely allow third-person forms to refer to the speaker, as well as the addressee.

We therefore consider a second possibility: that participants are affected by the preceding discourse. As noted in the introduction, third-person pronouns do not *literally* exclude reference to the speaker or addressee—listeners simply infer this based on the assumption that speakers avoid using third person in this way (i.e., they assume that speakers adhere to the maxim "Maximize Presupposition"). However, this inference can be relaxed, such as when speakers refer to themselves in the third person. What may be driving participants to relax the pragmatic inference in this case is the content of the preceding filler. In particular, for experimental items preceded by fillers, such as (10a), participants may interpret the experimental item as a continuation of the topic established in the filler—and that topic may well be the speaker and/or the addressee. In contrast to distractors, such as (10b), fillers like (10a) depict the same set of people as experimental items, so participants may interpret both the filler and the subsequent experimental item as parts of the same discourse segment. If the experimental item uses an agreement marker or clitic that just so happens to be compatible in number and/or gender with the previously selected referent, participants may be tempted to select the same referent again rather than the target referent. This is illustrated in (11), where the referent selected in the preceding filler trial (the speaker) is compatible with the number of the third person singular null pronoun in the experimental trial and may potentially compete with the target referent (the other woman) to be the antecedent.

(10) *(a) Muéstrame la foto en donde hay alguien saltando*
show-me the photo in which there-is someone jumping.
Display: speaker, addressee, other wom(e)n performing different actions
(b) *¿Cuál pato tiene más jabón/jabones?*
Which duck has more soap/soaps?
Display: two cartoon ducks

(11) *(a) Muéstrame la foto en donde hay alguien saltando*
show-me the photo in which there-is someone jumping.
Selected referent: **speaker** jumping
(b) *Muéstrame la foto en donde ø baila.*
show-me the photo in which *pro* dance-3S
Target referent: **other woman** dancing
Repeat referent: **speaker** dancing

There are two reasons that participants may link third-person agreement and clitics to a referent from the preceding filler rather than to the target referent. First, when pronouns appear in a multisegment discourse, they typically maintain reference to a previously introduced referent instead of introducing new referents, since this increases coherence between segments (Grosz, Joshi & Weinstein 1995). Second, not only do pronouns prefer discourse-old referents, they prefer discourse-prominent referents in particular. While many factors contribute to prominence, including order of mention (Arnold et al. 2000), agentivity (Pyykkönen et al. 2010), and syntactic position (Walker, Joshi & Prince 1998), the factors that are most relevant here are parallelism (Chambers & Smyth 1998; Smyth 1994) and cognitive status (Gundel, Hedberg & Zacharski 1993). The referring expression *alguien* 'someone' that introduces a new referent in each filler has the same semantic role as the referring expression in the following experimental item: In the agreement block both are agents—see (10)–(11)—and in the clitic block both are patients—see (12)–(13). Pronouns are argued to favor antecedents in parallel semantic roles, even if they are not in parallel syntactic positions (Kehler 2002). Additionally, the referent selected by the participant in the preceding filler has a high cognitive status, as illustrated in (12)–(13). At the point when

experimental item (13) is uttered, the experimenter and participant have been recently introduced by the speaker and have just been pointed at by the participant. Assuming that mentioning a referent activates it in short-term memory and that pointing places it in the current focus of attention, this would give the experimenter and participant the highest cognitive status, "in focus" (Gundel, Hedberg & Zacharski 1993).[8] In contrast, the target referent (the other woman) has been neither explicitly talked about nor pointed at yet, giving it a lower cognitive status in comparison.

(12) *Muéstrame la foto en donde Nemo está besando a alguien.*
show-me the photo in which Nemo is kissing A someone.
Target photo: Nemo kissing **experimenter and participant**

(13) *Muéstrame la foto en donde Nemo está tapándolas.*
show-me the photo in which Nemo is covering-3P.
Target photo: Nemo covering other woman
Repeat photo: Nemo covering **experimenter and participant**

If participants are influenced by these discourse considerations, we would expect them to repeat their response from the preceding filler trial whenever that response is compatible in number (and gender, if applicable) with the agreement or clitic form being tested. In the rest of this section we test this hypothesis, first for adults and then for children.

Figure 1 shows the proportion of responses that match the response provided during the immediately preceding filler trial.[9] The proportion of matching responses in first- and second-person

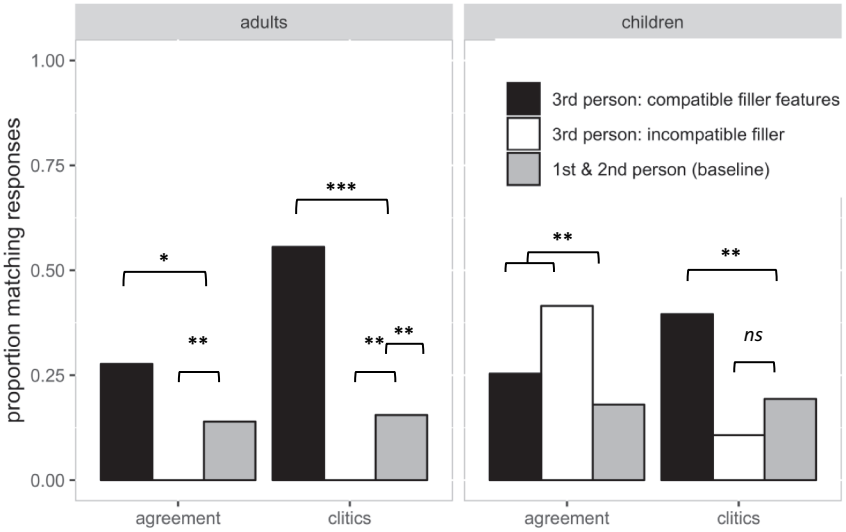

Figure 1. Proportion of responses matching the immediately preceding filler trial response. (i) Black bars represent third-person experimental items in which the preceding filler response was compatible with the number and/or gender features of the clitic or agreement marker in the experimental item. (ii) White bars represent third-person items in which the preceding filler response had incompatible features. (iii) Gray bars represent first- and second-person items, which serve as a baseline.

[8]It is our assumption that the referent has the highest status "in focus" as opposed to the next-highest status "activated." Gundel, Hedberg and Zacharski (1993) claim that mentioning a referent places it in short-term memory (elevating it to "activated" status), but they do not explicitly discuss pointing as a way to bring it into the current focus of attention (further elevating it to "in focus"). Nevertheless the important distinction is that this referent has a cognitive status that is equal to or higher than that of the target referent—which crucially has *not* yet been mentioned at the point when the clitic or agreement marker is uttered.

[9]These data are taken from the 14 experimental trials that were immediately preceded by a filler and excludes the 14 trials preceded by a distractor and the two trials appearing at the beginning of each block (see section 4.2.)

conditions (gray) serves as a baseline. Because first- and second-person forms do not need to rely on the preceding discourse to select a referent, such repeat responses should be due to sheer coincidence. Indeed, one-sample *t*-tests revealed that the proportion of responses matching the preceding filler response in first- and second-person conditions was not significantly different from 20% in either block, for either adults or children (all $t < 1.8$, all $p > .08$). Therefore, we use participants' proportion of matching responses in first- and second-person conditions as a baseline against which to compare their proportion of matching responses in the third person.

For the third-person condition, responses are subdivided depending on whether or not the picture chosen during the immediately preceding filler trial was compatible in number and/or gender with the third-person agreement marker or clitic being tested (black: compatible in number and/or gender; white: incompatible in number and/or gender). For the agreement block, only number features are relevant because agreement markers do not carry gender; however, for the clitic block both number and gender are relevant. For example, the response to filler item (12) (the photo of the experimenter and the participant together) is compatible with the clitic *las* in experimental item (13) only if the participant herself is female. (Recall that all experimenters were female, so only the gender of the participant him/herself could influence gender compatibility.) As can be seen from Figure 1, adults never repeated their response from the preceding filler trial if that response clashed with either the number or gender features of the third-person agreement or clitic being tested (white bars have height zero).

If adults allow third-person null subjects and clitics to refer to the topic of the preceding filler and relax the inference that third person excludes the speaker and addressee, then we would expect them to repeat the immediately preceding filler response more often in the third person than they do at baseline (i.e., more often than occurs by chance in first- and second-person conditions). However, we do not expect adults to relax the number and gender requirements of third-person agreement and clitics, and we therefore only expect above-baseline matching rates when the preceding filler response is compatible with the number and/or gender of the agreement marker or clitic being tested (black bars). When the preceding filler response has incompatible number and/or gender features (white bars) we expect below-baseline matching. To test these expectations, we used one-sided chi-squared tests of proportion to compare the frequency of matching responses in filler-compatible cases (black) to baseline (gray) and in filler-incompatible cases (white) to baseline (gray). As expected, adults produced significantly more matching responses in third-person conditions with number and gender-compatible preceding responses, relative to baseline, in both the agreement block, ($M1 = 0.28$, $M2 = 0.14$, $\chi(1) = 2.77$, $p = .048$), and the clitic block, ($M1 = 0.56$, $M2 = 0.15$, $\chi(1) = 18.83$, $p < .001$), and they produced significantly fewer matching responses in third-person conditions with *in*compatible preceding responses, relative to baseline, (agreement block: $M1 = 0$, $M2 = 0.14$, $\chi(1) > 6.9$, $p < .01$; clitic block: $M1 = 0$, $M2 = 0.14$, $\chi(1) > 6.9$, $p < .01$).

Next, we ran the same two analyses on child responses to test whether they too relax the inference associated to third person in order to allow third-person null subjects and clitics to refer to the topic of the preceding filler (without relaxing number and gender requirements). In the clitic block, children showed mostly the same pattern as adults: When the response from the preceding filler was compatible in number and gender with the third-person form being tested (black), then children chose this photo significantly more often relative to baseline, ($M1 = 0.40$, $M2 = 0.19$, $\chi(1) = 6.44$, $p < .01$); however, when the preceding filler response was incompatible with those number and/or gender features (white), they did not repeat this response significantly less often relative to baseline, ($M1 = 0.11$, $M2 = 0.19$, $\chi(1) = 1.58$, $p = .10$). In other words, it appears that children allow third-person clitics to refer to the topic of the preceding filler just like adults do, but they are not as good as adults at blocking this response when that topic has the wrong number and/or gender features.

In the agreement block, children differed more starkly from adults: For number-compatible cases (black), children failed to produce significantly more matching responses relative to

baseline, ($M1$ = 0.25, $M2$ = 0.18, $\chi(1)$ = 1.13, p = .14), and for number-incompatible cases (white) they actually produced *more* instead of *fewer* matching responses relative to baseline, ($M1$ = 0.42, $M2$ = 0.18, $\chi(1)$ = 10.57, p < .01, two-sided). This would appear to suggest that children actually prefer to link third-person agreement to the topic of the preceding filler when it has the wrong number properties as opposed to the right ones. However, matching rates were not significantly higher in number-incompatible cases (white) than in number-compatible cases (black), ($M1$ = 0.25, $M2$ = 0.42, $\chi(1)$ = 2.82, p = .09, two-sided). This leaves us with the alternative possibility that children simply consider the preceding filler response *regardless* of number compatibility. We therefore collapsed across compatible and incompatible cases and found that, indeed, children were more likely to repeat the preceding filler response in third-person agreement conditions *overall*, relative to the first- and second-person baseline conditions, ($M1$ = 0.33, $M2$ = 0.18, $\chi(1)$ = 6.82, p < .01).

In sum, adults tend to interpret a third-person clitic or agreement marker by looking to the most recently selected photo compatible with its number and/or gender features. Children also look to the most recently selected photo for a referent, but they are more willing than adults to accept number and/or gender violations, especially when it comes to interpreting agreement, where they appear to disregard number altogether.

6. Discussion[10]

This study used a picture-selection task to test children's comprehension of agreement and accusative clitics in Spanish, comparing their accuracy across and within each paradigm to determine whether children's performance is conditioned by phonological salience and/or semantic underspecification. We found partial support for the role of phonological salience and more complete support for the role of semantic underspecification. Consistent with the phonological hypothesis, children were more accurate in their responses to third-person singular and plural clitics *la* and *las* compared to the less phonologically salient third-person singular and plural agreement markers /-ø/ and /-*n*/ respectively; however, we failed to find reliably higher accuracy for the first- and second-person singular clitics *me* and *te* compared to their less salient agreement counterparts /-*o*/ and /-*s*/. Consistent with the semantic hypothesis, children were more accurate in their responses to first- and second-person clitics and agreement markers compared to third-person clitics and agreement markers of the same number. Children frequently allowed third-person agreement and clitics to refer to the speaker or addressee but generally did not allow first- and second-person agreement and clitics to refer to a nonspeaker, nonaddressee referent. This replicates other person-based asymmetries reported for English (Brener 1983) and French (Legendre et al. 2011) and is consistent with the claims by Legendre et al. (2011) and Legendre & Smolensky (2012) that children fail to calculate Implicated Presuppositions.

However, thanks to the inclusion of adult participants, our study offers additional insight into why children might want third-person clitics and null subjects to refer to the speaker and addressee. We found that not only children but also adults committed a substantial number of person errors in the third person, allowing reference to the speaker or addressee almost half the time in the agreement block (62 of 138 responses) and almost one-third of the time in the clitic block (38 out of 138 responses). In those cases where the experimental item and preceding filler could potentially be interpreted as segments of the same discourse, we found that adults and children were affected by the prominence of the previously selected referent. Specifically, we found that adults tended to choose the same referent that they had chosen in the preceding filler, allowing them to maintain reference to an "in focus" referent in a parallel semantic and syntactic role, instead of referring to the less prominent target referent. Children showed the same tendency as adults, except that they permitted reference to the discourse-prominent referent

[10]We thank two anonymous reviewers for thoughtful comments leading to a substantial reframing of these results.

even when it had the wrong number properties. In fact, in the agreement condition the referent's number properties didn't seem to make any difference at all.

Although we find no fault with Legendre et al. (2011) account of Implicated Presuppositions and children's difficulty calculating them, our results suggest that this is only part of the story. As Legendre et al. themselves point out, tasks like the fishing game and the present study present a special challenge because they require the listener to compute over multiple alternatives. Unlike other picture-selection tasks where participants use a single feature (e.g., number) to choose between two referents (e.g., one duck, a pair of ducks), participants in this task must simultaneously compare the properties of five potential referents to the person *and* number features of the form being tested. What our study shows is that listeners also take discourse prominence into account when comparing these alternatives. That is, listeners choose referents based on the person, number, *and* discourse prominence of potential referents. Adults in our study showed that they were willing to relax the person requirement to satisfy the prominence requirement, while children showed a willingness to sacrifice both person and number for the sake of prominence.

We cannot say for sure what caused adults to relax the person requirement in this particular case. One possibility is that the sheer number of alternatives provided (five photos) placed a heavy demand on processing, causing adults and children alike to skip calculating the Implicated Presupposition associated with the third person. Another possibility is that the pragmatics of the task made the Implicated Presupposition truly optional. It is certainly possible to felicitously refer to the speaker or the addressee in the third person, as illustrated in (14) and (15), and participants may have interpreted the items in this study as one such case. Regardless of which reason led participants to cancel the Implicated Presupposition, this study shows that canceling it allowed participants to rely on discourse prominence instead—and that children are every bit as sensitive to discourse prominence as adults, if not more so.

(14) Jake: Who ate my yogurt?
Terri: I confess! You know how Terri loves yogurt!

(15) Owner to his dog Cheddar: Who's my fluffy boy? Cheddar is my fluffy boy!

Children's sensitivity to discourse prominence also offers a second potential explanation for why children were more accurate in the third-person clitic condition compared to the less phonologically salient third-person agreement condition. Across languages, there is a universal tendency for phonologically reduced referring expressions to favor more prominent referents than phonologically robust referring expressions (Ariel 1988, Gundel, Hedberg & Zacharski 1993). If reduced expressions like agreement are more strongly biased toward discourse-prominent referents than more robust expressions like clitics, this would explain why children were particularly willing to choose a nontarget, discourse-prominent referent in the agreement condition, thus leading to more person and number errors. (And it would simultaneously explain why adults committed more person errors in the agreement block compared to the clitic block.) In other words, we interpret both the person-based asymmetry and the phonological asymmetry through the lens of children's sensitivity to discourse. Children recognize that third-person referring expressions favor discourse-prominent antecedents, and so they relax the requirement that third person exclude the speaker and addressee, thereby giving rise to the person asymmetry. Furthermore, they recognize that this preference is even stronger for third-person agreement than it is for third-person clitics, and so they also relax the need for number-compatible referents, giving rise to the agreement-clitic asymmetry.

If this interpretation is correct, it would certainly not be the first time that discourse considerations have influenced children's use of grammatical information in an experimental task. Children have widely been observed to violate Principle B in their interpretations of object pronouns (Chien & Wexler 1990; Thornton & Wexler 1999; see Elbourne 2005 for review), yet when pragmatic conditions are altered such that the grammatically illicit referent is either less discourse-prominent or no longer

part of the QUD, performance becomes adult-like (Conroy et al. 2009; Spenader, Smits & Hendriks 2009).[11]

Before closing, we wish to point out two ways in which children's performance in first- and second-person conditions also sheds light on previous work; in particular, the role of task-based performance factors and the formal syntactic status of agreement. Previous work has pointed out that picture selection imposes greater processing demands than spontaneous production (Brandt-Kobele & Höhle 2010; González-Gómez et al. 2017; Legendre et al. 2014; Verhagen & Blom 2014), potentially exaggerating children's difficulty with number agreement in some comprehension tasks. In this study, children demonstrated early comprehension of first- and second-person agreement, even in an unusually demanding version of the picture selection task—where the choice is between not two but five different pictures. Although this study involves person agreement rather than number agreement, it nevertheless suggests that children's failure to use number agreement in various picture-selection tasks should not be entirely blamed on the task itself. As this study demonstrates, success in a picture-selection task is possible for children as young as 2;03–2;11.

Other work argues that children's difficulty with number agreement is because agreement carries uninterpretable features (de Villiers & Gxilishe 2008; Johnson, de Villiers & Seymore 2005). If one adopts this assumption for Spanish agreement, it could potentially explain children's mixed success with number agreement (González-Gómez et al. 2017; Legendre et al. 2014; Pérez-Leroux 2005). However, this fails to capture our results, which show very early success with person agreement; in fact, children are every bit as successful with first- and second-person agreement as they are with first- and second-person clitics. This implies that either Spanish agreement instantiates *interpretable* person and number features just like clitics (along the lines of Alexiadou & Anagnostopoulou 1998) or that the interpretability of agreement features makes no difference to children's performance.

Summing up, our results show that children's interpretation of Spanish agreement and clitics is conditioned by the semantic (under)specification of their features and somewhat less consistently by the phonological salience of these morphemes. The surprising finding, however, is that a large number of children's errors can be linked to their remarkably adult-like sensitivity to discourse prominence. Future studies on children's comprehension of person and number morphology should therefore take into account the discourse properties of the elements instantiating these features. What may seem like errors on the surface could in fact be driven by surprisingly adult-like awareness of the surrounding discourse.

Acknowledgment

We would like to thank the members of the Michigan State University Language Acquisition Lab, in particular Alan Munn, Jessica Gamache, Ni La Le, and Camila Alfonso. Many thanks to our adult participants at MSU and to the teachers, parents, and children at Servicios Educativos del Desarrollo Infantil, Mexico City, D.F., as well as Beti López Juárez and Patricia de la Fuente Zuno. Thanks to the audience at BUCLD39 for insightful feedback.

Disclosure statement

No potential conflict of interest was reported by the authors.

[11] Although both this study and these Principle B studies both suggest that discourse-pragmatic considerations can override more strictly linguistic restrictions, it is important to acknowledge the limits of this suggestion. There is still a lot we don't know about how children (and adults) rank different discourse and syntactic properties when choosing referents or how this ranking differs across languages. What we know so far from a variety of processing studies is that the choice of an antecedent for a pronoun depends on (i) properties of the potential antecedents, (ii) properties of the pronominal element itself (null pronoun vs. overt pronouns vs. demonstrative, etc.), and (iii) their respective syntactic positions, among other considerations (see Miltsakaki 2002; de la Fuente 2015; Forsythe 2018 for reviews of this literature).

Funding

This research was funded in part by a Tinker Field Research Grant from the MSU Center for Latin American and Caribbean Studies to Hannah Forsythe, a NSF Postdoctoral research grant [#SPRF1810159] to Hannah Forsythe, and a NSF Grant [#BCS-1656133] to Cristina Schmitt.

ORCID

Hannah Forsythe http://orcid.org/0000-0003-4495-9814

References

Alexiadou, A., and E. Anagnostopoulou. 1998. "Parametrizing Agr: Word order, V-movement and EPP-checking." *Natural Language & Linguistic Theory* 16 (3): 491–539. doi:10.1023/A:1006090432389

Ariel, M. 1988. "Referring and accessibility." *Journal of Linguistics* 24 (1): 65–87. doi:10.1017/S0022226700011567

Arnold, J. E., J. G. Eisenband, S. Brown-Schmidt, and J. C. Trueswell. 2000. "The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking." *Cognition* 76 (1): B13–B26. doi:10.1016/S0010-0277(00)00073-1

Benki, J.R. 2003. "Analysis of English nonsense syllable recognition in noise." *Phonetica* 60 (2): 129–57. doi:10.1159/000071450

Benveniste, E. 1971. *Problems in general linguistics*. vol. 8, University of Miami Press.

Brandt-Kobele, O.-C., and B. Höhle. 2010. "What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition." *Lingua* 120 (8): 1910–25. doi:10.1016/j.lingua.2010.02.008

Brener, R. 1983. "Learning the deictic meaning of personal pronouns." *Journal of Psycholinguistics* 12 (3): 235–62.

Brown, R. 1973. *A first language: The early stages*. Oxford, England: Harvard University Press.

Chambers, C.G. and R. Smyth. 1998. "Structural parallelism and discoure coherence: A test of Centering Theory." *Journal of Memory and Language* 39 (4): 593–608. doi:10.1006/jmla.1998.2575

Charnavel, I. 2019. "Presupposition failure and intended pronominal reference: Person is not so different from gender after all." *Linguistic Inquiry* 50 (2): 372–87. doi:10.1162/ling_a_00307

Chien, Y.-C., and K. Wexler. 1990. "Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics." *Language Acquisition* 1 (2): 225–95. doi:10.1207/s15327817la0103_2

Childers, J., A. M. Fernandez, C. Echols, and M. Tomasello. 2001. "Experimental investigations of children's understanding and use of verb morphology: Spanish-and English-speaking 2 1/2-and 3-year-old children." In M. Almgen, A., Bareña, MJ Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on language acquisition. Proceedings of the International Association for the Study of Child Language*, 104–27.

Clahsen, H., F. Aveledo, and I. Roca. 2002. "The development of regular and irregular verb inflections in Spanish child language." *Journal of Child Language* 29 (3): 599–622. doi:10.1017/S0305000902005172

Conroy, A., E. Takahashi, J. Lidz, and C. Phillips. 2009. "Equal treatment for all antecedents: How children succeed with Principle B." *Linguistic Inquiry* 40 (3): 446–86. doi:10.1162/ling.2009.40.3.446

Corbett, G. 2006. *Agreement*. vol. 109. Cambridge: Cambridge University Press.

Culbertson, J. 2010. "Convergent evidence for categorial change in French: From subject clitic to agreement marker." *Language* 86 (1): 85–132. doi:10.1353/lan.0.0183

Davies, B., N. X. Rattanasone, and K. Demuth. 2016. "Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects." *Language Learning and Development* 13 (1): 38–53. doi:10.1080/15475441.2016.1219257

de Cat, C. 2002. "French Dislocation." University of York PhD Thesis.

de la Fuente, I. 2015. "Putting pronoun resolution in context: The role of syntax, semantics, and pragmatics in pronoun interpretation." Université Paris Diderot PhD thesis.

de Villiers, J. G., and S. Gxilishe. 2008. "The acquisition of number agreement in English and Xhosa." *Merging Features: Computation, Interpretation, and Acquisition*, 104–123. Oxford, UK: Oxford University Press.

Elbourne, P. 2005. "On the acquisition of Principle B." *Linguistic Inquiry* 36 (3): 333–65. doi:10.1162/0024389054396908

Forsythe, H. 2018. "Discourse and grammatical cues in the acquisition of Spanish pronouns." Michigan State University PhD Thesis.

González-Gómez, N., L. Hsin, I. Barriere, T. Nazzi, and G. Legendre. 2017. "Agarra, agarran: Evidence of early comprehension of subject-verb agreement in Spanish." *Journal of Experimental Child Psychology* 160:33–49. doi:10.1016/j.jecp.2017.02.010.

Grosz, B.J., A.K. Joshi, and S. Weinstein. 1995. "Towards a computational theory of discourse interpretation." *Computational Linguistics* 21:203–25.

Guasti, M. T. 1993. "Verb syntax in Italian child grammar: Finite and nonfinite verbs." *Language Acquisition* 3 (1): 1–40. doi:10.1207/s15327817la0301_1

Gundel, J., N. Hedberg, and R. Zacharski. 1993. "Cognitive status and the form of referring expressions in discourse." *Language* 69 (2): 274–307. doi:10.2307/416535

Gxilishe, S., M.R. Smouse, T. Xhalisa, and J. de Villiers 2009. "Children's insensitivity to information from the target of agreement: The case of Xhosa." In J. Crawford, K. Otaki and M. Takahashi (eds.), Proceedings of the third Conference on Generative Approaches to Language Acquisition North America (GALANA 2008), 46–53. Somerville, MA: Cascadilla Proceedings Project.

Heim, I. 1991. "Artikel und definitheit." In A. von Stechow and D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgen Aussischen Forschung*, 487–535. Berlin: de Gruyter.

Jakubowicz, C., and C. Rigaut. 1997. "L'acquisition des clitiques nominatifs en français." In A. Zribi- Hertz (ed.), *Les pronoms: Morphologie, syntaxe, et typologie*, 57–99. Paris: Presses Universitaires de Vincennes.

Johnson, V., J. de Villiers, and H. Seymore. 2005. "Agreement without understanding? The case of third person singular/s/." *First Language* 25 (3): 317–30. doi:10.1177/0142723705053120

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Legendre, G., I. Barriere, L. Goyet, and T. Nazzi. 2010a. "Comprehension of infrequent subject-verb agreement forms: Evidence from French-learning children." *Child Development* 81 (6): 1859–75. doi:10.1111/j.1467-8624.2010.01515.x

Legendre, G., I. Barrière, L. Goyet, and T. Nazzi. 2011. "On the acquisition of implicated presuppositions: Evidence from French personal pronouns." In M. Pirvulescu, M. C. Cuervo, A. T. Pérez-Leroux, J. Steele and N. Strik (eds.), Selected papers from the 4th Conference on Generative Approaches to Language Acquisition North America (GALANA 2010), 150–62, Somerville, MA: Cascadilla Proceedings Project.

Legendre, G., J. Culbertson, E. Zaroukian, L. Hsin, I. Barrière, and T. Nazzi. 2014. "Is children's comprehension of subject-verb agreement universally late? Comparative evidence from French, English, and Spanish." *Lingua*. doi:10.1016/j.lingua.2013.05.004.

Legendre, G., J. Culbertson, I. Barriere, T. Nazzi, and L. Goyet. 2010b. "Experimental and empirical evidence for the status and acquisition of subject clitics and agreement marking in adult and child spoken French." In V. Torrens, L. Escobar, A. Gavarró and J. Gutiérrez (eds.), *Movement and Clitics: Adult and Child Grammar*, 333–60. Newcastle, UK: Cambridge Scholars Publishing, Newcastle.

Legendre, G., and P. Smolensky. 2012. "On the asymmetrical difficulty of acquiring person reference in French: Production versus comprehension." *Journal of Logic, Language and Information* 21 (1): 7–30. doi:10.1007/s10849-011-9150-0

Lipski, J. M. 1994. *Latin American Spanish*. London: Longman.

MacWhinney, B., E. Bates, and R. Kliegl. 1984. "Cue validity and sentence interpretation in English, German, and Italian." *Journal of Verbal Learning and Verbal Behavior* 23 (2): 127–50. doi:10.1016/S0022-5371(84)90093-8

Mateu, V. E. 2015. "Object clitic omission in child Spanish: Evaluating representational and processing accounts." *Language Acquisition* 22 (3): 240–84. doi:10.1080/10489223.2014.962232

Miller, K. 2007. "Variable input and the acquisition of plurality in two varieties of Spanish." Michigan State University PhD thesis.

Miller, K., and C. Schmitt. 2014. "Spanish-speaking children's use of verbal inflection in comprehension." *Lingua* 144:40–57. doi: 10.1016/j.lingua.2013.06.008.

Miltsakaki, E. 2002. "Toward an aposynthesis of topic continuity and intrasentential anaphora." *Computational Linguistics* 28 (3): 319–55. doi:10.1162/089120102760276009

Munn, A., K. Miller, and C. Schmitt 2006. "Maximality and plurality in children's interpretation of definites." In D. Bamman, T. Magnitskaia and C. Zaller (eds.), *Proceedings of BULCD 30*, 377–87. Boston, MA: Cascadilla Press.

Pérez-Leroux, A. T. 2005. "Number problems in children." In C. Gurski (ed.), *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*. London, Canada: University of Western Ontario.

Pérez-Leroux, A. T., A. P. Castilla-Earls, and J. Brunner. 2012. "General and specific effects of lexicon in grammar: Determiner and object pronoun omissions in child Spanish." *Journal of Speech, Language, and Hearing Research* 55 (2): 313–27. doi:10.1044/1092-4388(2011/10-0004)

Pérez-Leroux, A.T., M. Pirvulescu, and Y. Roberge. 2017. *Direct objects and language acquisition*. Cambridge, UK: Cambridge University Press.

Phillips, C. 2010. "Syntax at age two: Cross-Linguistic differences 1." *Language Acquisition* 17 (1–2): 70–120. doi:10.1080/10489221003621167

Pirvulescu, M. and N. Strik. 2014. "The acquisition of object clitic features in French: A comprehension study." *Lingua* 144: 58–71.

Polišenska, D. 2010. "Dutch children's Acquisition of verbal and adjectival inflection." University of Amsterdam, Center for Language and Communication PhD thesis.

Pyykkönen, P., D. Matthews, and J. Järvikivi. 2010. "Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution." *Language and Cognitive Processes* 25 (1): 115–129.

R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rastegar, Z., H. Shirazi, and F. Sadighi. 2012. "An amazing conundrum in children's comprehension and production of verb inflection." *World Applied Sciences Journal* 18 (8): 1095–1101.
Redford, M. A., and R. L. Diehl. 1999. "The relative perceptual distinctiveness of initial and final consonants in CVC syllables." *The Journal of the Acoustical Society of America* 106 (3):1555–65. doi:10.1121/1.427152
Sauerland, U. 2003. "On the semantic markedness of phi-features." In D. Harbour, D. Adger and S. Bejar (eds.), *Phi-Theory: Phi-Features Across Modules and Interfaces*, 57–82. New York, NY.
Slobin, D. 1973. "Cognitive prerequisites for the development of grammar." *Studies of Child Language Development* 1:75–208.
Smyth, R. 1994. "Grammatical determinants of ambiguous pronoun resolution." *Journal of Psycholinguistic Research* 23 (3): 197–229. doi:10.1007/BF02139085
Spenader, J., E.-J. Smits, and P. Hendriks. 2009. "Coherent discourse solves the pronoun interpretation problem." *Journal of Child Language* 36 (1): 23–52. doi:10.1017/S0305000908008854
Theakston, A. L., E. V. M. Lieven, and M. Tomasello. 2003. "The role of the input in the acquisition of third person singular verbs in English." *Journal of Speech, Language, and Hearing Research* 46 (4): 863–77. doi:10.1044/1092-4388(2003/067)
Thornton, R., and K. Wexler. 1999. *Principle B, VP ellipsis, and interpretation in child grammar*. Cambridge, MA: MIT Press.
Verhagen, J., and E. Blom. 2014. "Asymmetries in the acquisition of subject–verb agreement in Dutch: Evidence from comprehension and production." *First Language* 34 (4): 315–35. doi:10.1177/0142723714544412
Walker, M.A., A.K. Joshi, and E.F. Prince. 1998. *Centering in discourse*. Oxford: Clarendon Press.
Zesiger, P., L. Chillier Zesiger, M. Arabatzi, L. Baranzini, S. Cronel-Ohayon, J. Franck, U. H. Frauenfelder, C. Hamann, and L. Rizzi. 2010. "The acquisition of pronouns by French children: A parallel study of production and comprehension." *Applied Psycholinguistics* 31 (4): 571–603.

Appendix: Accuracy by age year in first- and second-person agreement conditions

Table A1 shows the proportion of target answers provided in each condition by children in the following age groups: 2;03–2;11 (N= 7), 3;00–3;09 (N= 10), 4;01–4;11 (N= 13), 5;00–5;11 (N= 7), 6;00–6;07 (N= 5). These rates are compared to chance using two different analyses: (i) single-sample t-tests comparing proportion target to chance accuracy in each condition, and (ii) intercept-only binary logistic regression models with random intercepts for participant: Children's performance is considered to be significantly different from chance if the 95% confidence interval for the model's intercept (β_0) is greater than the log odds of chance accuracy in that condition ($\log\left(\frac{chance}{1-chance}\right)$). For 4- and 6-year-olds in the second-person singular condition, the t-test and/or the 95% confidence interval of the regression model is not reported, due to a lack of variation in responses (accuracy at or near ceiling).

Explanatory note

This experiment was previously published in the supplemental proceedings of the 39th Boston University Conference on Language Development. The current version makes significant updates to the statistical analysis, the framing of the

Table A1. Proportion of target answers in first-person singular, first-person plural, and second-person singular agreement conditions with comparisons against chance.

Condition	Chance accuracy; log odds of chance	Age year	Target (%)	t-test: $t(df)$, p value	Logistic regression: β_0 (95% CI)
first-person singular	40%; −0.405	2	81.0	$t(6) = 6.08$, $p < .001$*	1.45 (0.45, 1.98)*
		3	86.7	$t(9) = 6.33$, $p < .001$*	2.39 (0.99, 11.3)*
		4	89.7	$t(12) = 8.53$, $p < .001$*	3.23 (1.38, 16.99)*
		5	90.5	$t(6) = 8.21$, $p < .001$*	2.25 (1.01, 8.32)*
		6	93.3	$t(4) = 8$, $p < .01$*	2.64 (1.04, 46.43)*
first-person plural	20%; −1.386	2	57.1	$t(6) = 3.9$, $p < .01$*	0.29 (−0.6, 1.62)*
		3	66.7	$t(9) = 5.42$, $p < .001$*	0.69 (−0.06, 1.73)*
		4	74.4	$t(12) = 5.81$, $p < .001$*	1.64 (0.34, 5.99)*
		5	61.9	$t(6) = 2.74$, $p < .05$*	0.83 (−1.4, 8.52) *ns*
		6	80.0	$t(4) = 4.5$, $p < .05$*	1.75 (−0.03, 14.47)*
second-person singular	40%; −0.405	2	90.5	$t(6) = 8.21$, $p < .001$*	2.25 (1.01, 8.32)*
		3	93.3	$t(9) = 12$, $p < .001$*	2.64 (1.44, 14.69)*
		4	97.4	$t(12) = 22.4$, $p < .001$*	3.64 (*NA*)
		5	90.5	$t(6) = 8.21$, $p < .001$*	2.25 (1.01, 8.32)*
		6	100.0	*NA*	*NA*

research problem, and the interpretation of the results. This manuscript represents a dramatic (and, we are confident, improved) reworking of the original presentation.

A note regarding experimental stimuli

We do not include an appendix of experimental and filler stimuli for two reasons: (i) The linguistic stimuli are fully described in the methods section. (ii) The visual stimuli consisted of photos of the research assistants and participants performing various actions, which have been destroyed to protect participants' privacy.