Generics as default? Comparing the acquisition of universals and generics in Spanish

Elena Castroviejo, Dimitra Lazaridou-Chatzigoga, Marta Ponciano & Agustín Vicente*†

Abstract. This paper reports an experiment that investigates interpretive distinctions between two different expressions of generalization in Spanish. In particular, our aim was to find out when the distinction between generic statements (GS) such as *Tigers have stripes* and universal quantified statements (UQS) such as *All tigers have stripes* was acquired in Spanish-speaking children of two different age groups (4/5-year-olds and 8/9-year-olds), and then compare these results with adults. The starting point of this research was the semantic distinction between GS and UQS in that the former admits exceptions, unlike the latter. On the other hand, cognitive psychologists have observed a Generic overgeneralization effect (GOG) consisting in allowing for UQS to be felicitous in the face of exceptions, thus proposing that this "error" stems from people misinterpreting UQS as GS and from GS being defaults (simpler, more easily learned and processed) instead of involving quasi-universal quantification, which was the learned view from semantics. In the current paper we aimed to test the "Generics as Default" (GAD) hypothesis by comparing GS and UQS in three different age ranges. Our data show that, overall, participants accept GS more often than they reject UQS. Moreover, we also confirm a hypothesized interaction between age and NP type (GS vs UQS). Further, we present several data points that are not predicted by the GAD, including an observed decline in the accuracy of GS in the older group of children as well as in adults with respect to younger children, and that children fail at rejecting generics that adults reject.

Keywords. genericity; quantification; generics as default; acquisition; Spanish

1. Introduction. The expression of generalization is pervasive in everyday language. Across languages and within the same language, different mechanisms are used to this effect. The goal of the present paper is to compare the acquisition and interpretation of two such strategies in Spanish. Specifically, generic statements (henceforth GS), (1-a),(1-b), and universally quantified statements (henceforth UQS), (2-a),(2-b).

*We are extremely grateful to the teachers and staff at Angel Ganivet Ikastetxea, Vitoria-Gasteiz, for organizing everything for us and for their help and patience. We extend our gratitude to all the families and students from Angel Ganivet who agreed to collaborate with us as well as to the adult controls for participating in the study. We are also indebted to A. Elordieta, V. Escandell, M.J. Ezeizabarrena, O. Fernández-Soriano, N. Katsos, J. Mateu, J. Ormazabal, M. Santesteban, E. Vallduví, B. Vicente, and the audience of ELM-1 for their insightful remarks, and to M. Santesteban for his assistance with the statistical analysis of a previous version of this work. We are of course responsible for any remaining mistakes. This research has been partially supported by projects VASTRUD (PGC2018-096870-B-I00) and PROLE (PGC2018-093464-B-I00), funded by the Ministry of Science and Innovation (MCI) / Spanish Research Agency (AEI) and the European Regional Development Fund (FEDER, EU), the IT1396-19 Research Group (Basque Government), and GIU18/221 (University of the Basque Country, UPV/EHU).
Authors: Elena Castroviejo, UPV/EHU (elena.castroviejo@ehu.eus), Dimitra Lazaridou-Chatzigoga, University of Cambridge (dl518@cam.ac.uk), Marta Ponciano, UPV/EHU (marta.ponciano@ehu.eus) & Agustín Vicente, Ikerbasque and UPV/EHU (agustin.vicente@ehu.eus).
†The order of authors is alphabetical.

(1) a. Cats have whiskers. ENGLISH
b. Los gatos tienen bigotes. SPANISH
DET.PL cats have.3PL whiskers
'Cats have whiskers.'

(2) a. All cats have whiskers. ENGLISH
b. Todos los gatos tienen bigotes. SPANISH
all DET.PL cats have.3PL whiskers
'All cats have whiskers.'

One key distinction between GS and UQS is that the former, but not the latter, are tolerant to exceptions. Following the work of Lazaridou-Chatzigoga & Stockall (2013) and Lazaridou-Chatzigoga et al. (2019), in this paper, we want to stress the importance of exceptions as a means to identify differences in the interpretation of GS and UQS both in children and adults. Moreover, we also make a case for the need of experimental studies in this domain that direct our attention to languages other than English and which pay close attention to linguistic differences in the expression of generalizations. In particular, here we will analyze the results of a language acquisition experiment for Spanish, taking into account that, unlike in English, GS are typically expressed by means of definite plurals, (1-b), and UQS include the definite determiner, (2-b). We show that, while GS may be considered to be easier than UQS at first sight, a more thorough look at the results of our experiment raises some doubts in that respect. While young children (4-5 years old) seem to be more tolerant to exceptions for generics than older children (8-9 years old), they do not have an adult-like behavior in the interpretation of GS. As to UQS, we make a case for a difference in the acquisition of universal quantifiers depending on whether their quantificational domain is explicitly established or it is implicitly treated as the entire domain.

2. Theoretical and experimental background. If we abstract away from the fact that generalization is conveyed through different linguistic structrues, and start with the logical universal quantifier $\forall$, we know from first order predicate logic that it binds a variable and sets every assignment to true without expressing existential commitment.

Concerning GS, they have been widely analyzed as a sort of quantified statement including a covert modal operator, called GEN since Dahl (1995), with an interpretation similar to a quantificational adverbial like *usually* in the sense of Lewis (1975), and hence as an unselective binder of any free variable (typically individuals and situations), (3).

(3) From Zamparelli (2002)
a. Dogs bark at the moon.
b. $\text{GEN}_{x,s}$[**dog**(x) in s][**bark-at-the-moon**(x) in s]
c. For each appropriate situation s, if x is a dog in s, then x barks at the moon in s.

Under this account, GEN is a dyadic modal operator that does not have a phonological exponent, thus making it possible for different realizations of this operator both within the same language and across languages.

As stated above, one of the main properties of GS is that they tolerate exceptions. In the literature on generics, several types of generics have been proposed. For instance, Leslie et al.

(2011) consider the following:

(4) a. Quasi-definitional: Triangles have three sides.
  b. Majority characteristic: Tigers have stripes.
  c. Minority characteristic: Lions have manes.
  d. Majority: Cars have radios.
  e. Striking: Sharks attack people.

The modal/quantificational approach to the semantics of GS has been recently under attack, based both on data from acquisition and on the fact that it fails to account for striking generics (Leslie 2007, 2008). Leslie and collaborators (Gelman 2010) propose a very different view on generics, especially in view of the so-called 'Generic Overgeneralization' (GOG) effect, an error consisting in interpreting or recalling UQS as GS (*All ducks lay eggs* would be recalled as *Ducks lay eggs*), and that several studies on adults have observed a tendency to treat UQS as true even in the face of exceptions. In view of these results, Leslie (2007, 2008) and Gelman (2010) have endorsed the Generics as Default (GaD) hypothesis, according to which GS are simpler and hence more easily acquired and processed than UQS, which would explain an asymmetry that is not otherwise observed or predicted by semantic theories. Such generic bias is told to be founded on the dual view of cognition as proposed by Kahneman & Frederick (2002), among others, such that GS would be part of System 1, while UQS part of System 2. According to Leslie, issuing and verifying generic generalizations (i.e., GS) is a matter of checking whether a certain category exhibits a certain feature by accessing a conceptual representation of such a category, whereas issuing and verifying universal generalizations (i.e., UQS) involves working memory and checking statements against possible exceptions. In Leslie's view, the overgeneralization of UQS as GS is an example of the "lazy" overuse of System 1. Relatedly, she interprets the lack of a dedicated overt operator to express GS as following from the fact that System 1 is simpler. Hence, lack of "markedness" is associated with less complexity.

3. Our study.

3.1. Research questions and hypotheses. Given this learned view in cognitive psychology, according to which GS are defaults, GS should be easier to process and easier to learn, while UQS are considered more difficult. Hence, we would expect higher accuracy for GS compared to UQS. Additionally, given our age variable (two age groups in children, and the adult controls), we would think that UQS would be more accurate in the older group, while we do not expect any age differences between the accuracy of GS of the younger and the older group. Moreover, given the results in Gelman et al. (2016), differences in morpho-syntactic markedness should not mirror processing complexity. Hence, while GS in Spanish are more complex than GS in English, cross-linguistic differences should not arise (on this, see the recall experiment by Gelman et al. (2016)).[1] Concerning the background on the acquisition of UQS (Katsos et al. 2016, Barberán-Recalde 2019), we should expect 4-year-olds to master universal quantification. However,

[1] Although we cannot dwell on this issue due to space limitations, let us point out that there is previous work by Serratrice et al. (2009) that studies the interpretation of the definite determiner in Spanish and English in two groups of children (3;5-5;3 and 6;0-6;7). The results show that (a) there is a preference for the generic and, interestingly, (b) children did not go through a stage where the generic interpretation was not available.

we can only be sure about UQS whose domain restriction is made explicit by the utterance context, i.e., when the total number of objects being quantified over are visibly present. Lastly, from Lazaridou-Chatzigoga et al. (2019), we should not be surprised to find around 30% of adult participants who accept UQS even in the face of exceptions that are explicitly presented to them. From Lazaridou-Chatzigoga et al. (2019) we expect that adults will accept GS in the face of exceptions, but performance might not be at ceiling as the presence of exceptions lowers acceptance rates for GS (a difference from 99% to 87% in English and 92% to 77% in Greek).

Our research questions are the following, paired up with our research hypotheses:

RQ1) Are children sensitive to the reported differences between GS and UQS?

- H_0: Accuracy of GS = Accuracy of UQS in the two groups.
- H_1: Accuracy of GS > Accuracy of UQS in the two groups.

RQ2) Is there an interaction between NP type and age? Is the joint effect of NP type and age on accuracy predictable from the effect of the two factors individually?

- H_0: There is no interaction.
- H_1: There is an interaction.

3.2. Method.

3.2.1. Participants. A total of 55 Spanish-speaking children (30 male; 25 female) divided into two age groups, a 4/5-year-old group (N = 31, M = 68.16 months, SD =6.8, henceforth "young") and an 8/9-year old group (N = 24, M = 108.75 months, SD = 6.3, henceforth "old"), were recruited from a local (primary) school in Vitoria-Gasteiz (Spain). The participants in the young group belonged to different school years: 2nd and 3rd years of "Infantil" (child-care).

Alongside, we included a third group of adult controls (N = 26, 12 male; 14 female, henceforth "adults"). We did not test college undergraduates, but rather adults of various ages (M = 38, SD = 13.92) and in various schooling degrees (ranging from elementary education to a BA university degree). All participants were Spanish speakers and residents in Vitoria-Gasteiz. All volunteered to take part in the experiment. This study was carried out in accordance with the recommendations of the Human Beings Research Ethics Committee ("CEISH: Comité de Ética de Investigación con Seres Humanos de la UPV/EHU").

3.2.2. Design. Building on Lazaridou-Chatzigoga & Stockall (2013) and Lazaridou-Chatzigoga et al. (2019), the design we proposed placed the role of exceptions at the center of the discussion. To keep the experiment simple, we also focused only on majority characteristic statements. We manipulated NP into a generic (DET.PL) and a universal (*todos/as* DET.PL 'all') condition. By contrast, instead of manipulating the context as in Lazaridou-Chatzigoga et al. (2019), we did not manipulate the context variable between a supportive and a contradictory condition; all our critical items were uttered on the basis of a contradictory context, i.e., in the face of exceptions. Below is an example.

(5) Critical condition
   a. Context: Participant sees the picture of a rabbit with one ear and hears the utterance "a rabbit with one ear".

b. Target question:
¿Dirías que {todos los, los} conejos tienen dos orejas?
say.COND.2SG that all DET.PL DET.PL rabbits have.3PL two ears
'Would you say that {all, ∅} rabbits have two ears?'

However, to prevent participants from establishing a pattern, we included supportive contexts in the fillers.

Each participant in the study saw 16 critical items, 32 distractors. In the critical items, NP type (GS vs. UQS) was a within-individual independent variable. We also had a between-individual variable, namely the three age conditions: young, old and adult. The dependent variable was a yes/no response to the question prompted as part of the experimental procedure.

Given the manifest existence of exceptions and following the semantic-logical properties of the different expressions, GS and UQS were true in different scenarios. For instance, given the context described in (5) and the learned truth-conditions, the accurate answers are the ones in (6).

(6) a. All rabbits have two ears. → No
b. Rabbits have two ears. → Yes

Coming back to the distractors, we created two conditions: (a) similar sentences with proper name subjects (henceforth "Name"), (7), which was a mix of true and false statements, to control for the participants' *yes*-bias and for their understanding of the procedure, and (b) generic-like statements that adults tend to reject in generic form even though they predicate prevalent properties of a kind. These statements have been called "false generalisations" by Leslie et al. (2011). We used them as control generalisations with a supportive context (henceforth "Controlgen"), (8).

(7) Name condition
a. Context: Participant sees a picture of Plaza de la Virgen Blanca and hears a voice uttering "Plaza de la Virgen Blanca".
b. Target question:
¿Dirías que la Plaza de la Virgen Blanca está en Vitoria?
say.COND.2SG that DET.SG Plaza de la Virgen Blanca is in Vitoria
'Would you say Plaza de la Virgen Blanca is in Vitoria?'

(8) Controlgen condition
a. Context: Participant sees a picture of a black roof and hears a voice uttering "A black roof".
b. Target question:
¿Dirías que los tejados son negros?
say.COND.2SG that DET.PL roofs are black.PL
'Would you say roofs are black?'

3.2.3. MATERIALS. As previously mentioned, participants saw 16 critical items (8 GS and 8 UQS) and 32 distractors (16 Names and 16 Controlgen). They were distributed in two lists that were randomly assigned to the participants. So, each critical item came in two different conditions (GS or UQS), and participants in the different lists did not see the same item in the same condition. All participants saw the same set of distractors. Apart from the set of 16 critical items and 32 distractors, there were 4 training items, also common to all participants. The materials were counterbalanced across participants. Also, experimental items were randomized every time a participant started a new experimental session.

The 16 critical items consisted of majority characteristic statements like *Cats have whiskers* and *Horses have four legs*. Special care was taken to select properties about which young children would be knowledgeable. Also, to avoid the possible ambiguity in the GS condition (DET.PL can be ambiguous between a generic and an exemplar reading in Spanish), the picture that was shown always depicted a single individual holding an exceptional property.

Also, bear in mind that, for each critical generalization, a picture describing an exception was presented. To create these materials we used two strategies: we either selected a subkind of the species (a sphinx cat, which does not have whiskers, or a black pig, which is not pink) or an animal that, for an accidental reason, may have suffered a mutation (a three-legged horse or a hen with four wings).

Regarding Controlgen, unlike GS, they described characteristics held of only some of the instances of a kind. For instance, square-shaped pizzas, black roofs or blue butterflies. In these cases, accompanying pictures were supportive, so the participant would see a picture of a square-shaped pizza while she was asked whether pizzas are square-shaped.

3.2.4. PROCEDURE. Children were tested individually in a quiet room in their school. They had been previously told that they would play a game on the computer. In the case of adults, they were administered the study at a quiet place of their convenience.

Participants sat in front of a computer screen with the investigator beside them. In the case of children, the experimenter was in charge of clicking the left or right key on the computer's mouse, corresponding to a "yes" or "no" answer, while adults handled the mouse themselves.

The software used was E-prime 3 ("Psychology Software Tools") on a PC running Windows. No feedback was given during the main task. The testing process took approximately 15 minutes to complete.

In the study, each target item was composed of two parts. In the first part, an image occupying the center of the screen presented one individual contradicting or supporting the generalization to be judged. For instance, a cat without whiskers. At the same time, a pre-recorded audio of a female voice said: "un gato sin bigotes" ('a cat without whiskers'). By pressing a key, we moved to the second part, where an image of a girl, a cartoon character, appeared on the right-hand-side of the screen and asked "¿Dirías que (todos) los gatos tienen bigotes?" ('Would you say that (all) cats have whiskers?'). The pre-recorded audio with the question was played twice. It was a two-alternative forced choice task, whereby participants were instructed to choose "yes" or "no" in light of the picture that had been presented to them.

3.2.5. DATA ANALYSIS. We used the statistic tools from IBM SPSS 26 to analyze the variance of the means of accurate responses per item by running repeated measure ANOVAs with NP type as a within-individuals factor and age as a between-individuals factor.

3.3. RESULTS. Table 1 summarizes the mean accuracy and standard deviation of responses by NP type (GS vs. UQS) and age group (young, old, adult) in a by-item analysis. Figure 1 graphically illustrates the critical differences.

	GS	UQS
young	0.92 [SD 0.08]	0.30 [SD 0.18]
old	0.72 [SD 0.08]	0.64 [SD 0.25]
adults	0.79 [SD 0.19]	0.72 [SD 0.20]

Table 1: Descriptive statistics (Critical)

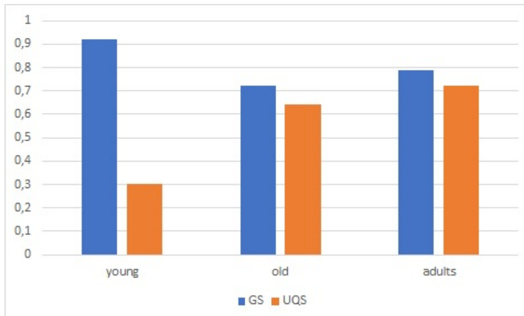

Figure 1: Mean accuracy per NP type and age group

We conducted a repeated measure ANOVA on the control categories, crossing NP type (GS vs. UQS), and age (young, old, adult). The analysis revealed a significant effect of age (F(2.45) = 8.057, p = 0.001, η^2 = 0.264) and a significant effect of NP type (F(1.45) = 35.709, p $<$ 0.001, η^2 = 0.442). Further, the analysis yielded a significant interaction between age and NP Type (F(2.45) = 17.357, p $<$ 0.001, η^2 = 0.435). The magnitude of the age effect is small, but the effect size of NP type and the interaction is close to medium.

To be able to properly interpret this interaction, we performed pairwise comparisons. The pairwise comparisons between the two levels of the factor NP type (GS vs. UQS) corrected for Bonferroni, yielded the result that the accuracy of GS is significantly higher than the accuracy of UQS. If we compare the age factor within the two levels of the NP type factor, we observe that only in the young group there is a significant difference between GS and UQS, such that the accuracy of GS is significantly higher than UQS. Additionally, accuracy in GS differs significantly between the young group on the one hand, and the old group (p = 0.000) and adults (p = 0.025) on the other hand, in favor of the young group. Finally, with respect to UQS, there is a significant difference (p = 0.000) between the young group on the one hand, and old group and adults on the other hand, this time in favor of the latter groups. To obtain results from the pairwise comparisons within the between-participants factor (namely, age), we ran the Games-Howell test (since we could not assume equality of variance), and the results yielded that accuracy differs significantly only between the young and adult conditions (p = 0.000), but not between young and old or between old and adult.

Moving to the distractor items, Table 2 presents the mean accuracy and standard deviations of the three age groups depending on filler type. This is graphically represented in the bar graph in Figure 2. From the results in the Name condition, it is obvious that all participants understood the task, were paying attention and had some amount of world knowledge (it increases with age). More revealing is the Controlgen condition, which shows very low measures in the children groups, whereas a comparatively high value in the case of adults.

	Controlgen	Name
young	0.49 [SD 0.20]	0.81 [SD 0.10]
old	0.44 [SD 0.14]	0.92 [SD 0.09]
adults	0.84 [SD 0.10]	0.97 [SD 0.04]

Table 2: Descriptive statistics (Distractors)

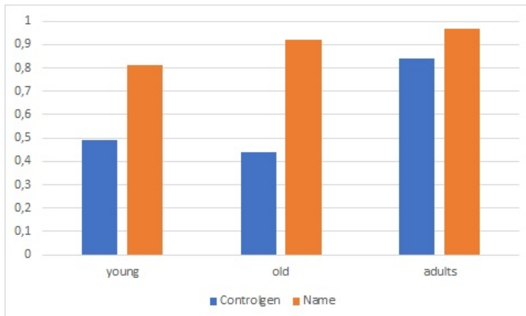

Figure 2: Mean accuracy per NP type and age group (Distractors)

Even though GS were counterbalanced in two different lists while the list of Controlgen was common for all participants, we decided to conduct a repeated measure ANOVA to compare GS and Controlgen across the age factor and thus explore the significance of differences, even if only as a tentative measure. The analysis revealed a significant interaction between type of generic ("GenType") and age. The analysis revealed significant differences of GenType (F(1.45) = 47.736, $\eta^2 = 0.515$), Age (F(2.45) = 22.802, $\eta^2 = 0.503$) and their interaction (F(2.45) = 19.822, $\eta^2 = 0.468$). Regarding pairwise comparisons within GenType, the young and old group differ significantly in their accuracy of GS vs. Controlgen (p = 0.000), but not the adult group. With respect to the age factor, there were significant differences in all of them. The difference between young and old group yielded a p = 0.004, between young children and adults, the comparison yielded a p = 0.013, and the difference between older children and adults had a p = 0.000.

3.4. Discussion. Let us start by addressing the research questions that we spelled out in 3.1.

RQ1) Are children sensitive to the reported differences between GS and UQS?
RQ2) Is there an interaction between NP type and age? Is the joint effect of NP type and age on accuracy predictable from the effect of the two factors individually?

In both cases, the null hypothesis can be rejected. First, considering all age groups, the accuracy in GS is greater than the accuracy in UQS. Second, there is an interaction between Age and NP type such that the difference in accuracy between GS and UQS is much larger in the case of young children than old children and adults.

Now, do these data support the GAD view? One of its main claims is that generics are held to be true even if we are aware that there are exceptions. As we have mentioned above, this may seem supported by the data at first sight. However, we have observed that generics are not always verified as they should, so this is not true across development. Even in the case of majority characteristic generics, the ones we have tested, generics are not universally taken to be true. In fact, some of our reported data seem to go against it. Especially, the high success of the young group in the GS condition (in fact higher than the old group and adults) shows there is a significant decline in the accuracy of GS between the young and old groups and even extending to adults. If the greater accuracy of GS in young children, both with respect to the UQS condition and other ages, were to be interpreted as evidence in favor of GS being easy, we would be forced to entertain the idea that GS become more difficult across development, which is something that we would not want to argue for. Moreover, the low rates in the Controlgen condition in the children groups clearly show that they behave unlike adults, so there seems to be a development trajectory in the adult-like understanding of GS, which is not predicted under the hypothesis that verifying a GS involves System 1 and, as such, the correct interpretation of GS should be preserved across ages. In fact, given the poor performance of young children in the UQS condition, the old group is adult-like in their interpretation of UQS, while the output in the generic types taken together (GS and Controlgen) suggest the two children groups are non-adult-like in the interpretation of generics. Finally, the mere fact that the old group and adults do not have at ceiling results for GS can also be viewed as data that go against the GAD (i.e., if GS are easy, we would expect higher accuracy than the one we observe). On the other hand, it should be noted that the centrality of exceptions and the fact that the participants had to reason in the face of a counterexample might have influenced these non-at-ceiling effects.

Remember that GAD is an answer to the Generic Overgeneralization (GOG) effect that was observed in various experiments. Now, do we also find a GOG effect in our data? The low values in UQS in the young group could be evidence in favor of this, since it seems they are interpreting them as tolerating exceptions. The GOG effect would be attenuated in the older group and in adults, but the low performance in UQS in these two groups may also suggest that they tend to interpret some UQS as GS.

As a final remark, we may be tempted to make sense of this developmental cline with respect to GS as describing a U-shape (as has been claimed e.g. by Berko (1958) for the acquisition of English past tense morphology); that is, one could imagine that young children easily acquire GS and then, when they become proficient with UQS, they start having doubts about the relevance of exceptions in generalizations, so they start rejecting GS when they should not. We believe the collected data suggest otherwise, especially if we take into account the results of Controlgen and the lack of ceiling effects in adults. In fact, the at-chance rate of acceptance observed in the young group with Controlgen (remember that Controlgen includes items that the literature on generics would consider not to be acceptable as generic generalizations) seem incompatible with the idea that young children have an adult-like command of generics altogether. We also can't interpret our results as providing evidence that the old group behaves differently from young and adults taken together. The old group and adults pattern together in GS, UQS and Name, while the young and old groups pattern together in Controlgen. Since even in the odd case of a U-shaped curve the GAD theory would not be able to account for the data we present, we need to find other theories that are compatible with them.

4. Conclusions. In the present paper we have carried out an investigation about generalizations in Spanish-speaking children of two age groups (and a corresponding group of adult controls).

Building on previous work by Lazaridou-Chatzigoga & Stockall (2013), Lazaridou-Chatzigoga et al. (2019), which addressed Leslie and Gelman's "Generics as Default" (GAD) hypothesis, we have proposed a design that could test differences between generic statements (realized as definite plurals in Spanish) and unrestricted universal quantified statements, when the participants were faced with a photograph of an individual failing to support the generalization.

The data that we have collected does not talk in favor of the GAD view. In fact, it describes an interesting picture yet to be fully understood. We would like to emphasize that, while it was established that 4-year-olds were able to comprehend (restricted) UQS in an adult-like manner, we have found out that they are not adult-like in the interpretation of unrestricted UQS. In fact, by comparing three age groups, we have been able to spot an age group, namely 8/9-year-olds, as having certain adult-like behaviors (for instance in the interpretation of UQS) and child-like behaviors (for instance in the interpretation of the "false" generics).

Finally, in this study we have analyzed behavioral results from a forced-choice task. However, in view of the apparent mismatches observed in the literature between the information collected from behavioral tasks and e.g. reaction times, we believe that processing data should be key in further informing us on whether the status of the GAD hypothesis.

References

Barberán-Recalde, Tania. 2019. *The acquisition of Basque and Spanish quantifiers: an empirical study*: University of the Basque Country dissertation.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177. 10.1080/00437956.1958.11659661.

Dahl, Östen. 1995. The marking of the episodic/generic distinction in tense-aspect systems. In Greg Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 412–425. Chicago: Chicago University Press.

Gelman, Susan A. 2010. Generics as a window onto young childrenâs concepts. In Francis Jeffry Pelletier (ed.), *Kinds, things, and stuff: Mass terms and generics*, 100–120. New York: Oxford University Press New York. 10.1093/acprof:oso/9780195382891.003.0006.

Gelman, Susan A., Ingrid Sanchez Tapia & Sarah-Jane Leslie. 2016. Memory for generic and quantified sentences in Spanish-speaking children and adults. *Journal of Child Language* 43(6). 1231–1244. https://doi.org/10.1017/S0305000915000483.

Kahneman, Daniel & Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment*, 49–81. Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004.

Katsos, Napoleon, Chris Cummins, Maria-José Ezeizabarrena, Anna Gavarró, Jelena Kuvač Kraljević, Gordana Hrzica, Kleanthes K Grohmann, Athina Skordi, Kristine Jensen De López, Lone Sundahl et al. 2016. Cross-linguistic patterns in the acquisition of quantifiers. In *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113 33, 9244–9249. https://www.jstor.org/stable/26471413.

Lazaridou-Chatzigoga, Dimitra, Napoleon Katsos & Linnaea Stockall. 2019. Contextualising Generic and Universal Generalisations: Quantifier Domain Restriction and the Generic Overgeneralisation Effect. *Journal of Semantics* 36. 617–664. https://doi.org/10.1093/jos/ffz009.

Lazaridou-Chatzigoga, Dimitra & Linnaea Stockall. 2013. Genericity, exceptions and domain restriction: experimental evidence from comparison with universals. In Emmanuel Chemla, Vincent Homer & Grégoire Winterstein (eds.), *Proceedings of sinn und bedeutung 17*, 325–343. École Normale Supérieure. https://semanticsarchive.net/sub2012/LazaridouChatzigogaStockall.pdf.

Leslie, Sarah-Jane. 2007. Generics and the structure of the mind. *Philosophical perspectives* 21. 375–403. https://doi.org/10.1111/j.1520-8583.2007.00138.x.

Leslie, Sarah-Jane. 2008. Generics: Cognition and acquisition. *Philosophical Review* 117(1). 1–47. https://doi.org/10.1215/00318108-2007-023.

Leslie, Sarah-Jane, Sangeet Khemlani & Sam Glucksberg. 2011. All ducks lay Eggs: The Generic Overgeneralization Effects. *Journal of Memory and Language* 65(1). 15–31. https://doi.org/10.1016/j.jml.2010.12.005.

Lewis, David. 1975. Adverbs of quantification. In Ed Keenan (ed.), *Formal semantics in natural languages*, 3–15. Cambridge University Press. https://doi.org/10.1017/CBO9780511897696.003.

Serratrice, Ludovica, Antonella Sorace, Francesca Filiaci & Michela Baldo. 2009. Bilingual children's sensitivity to specificity and genericity: Evidence from metalinguistic awareness. *Bilingualism: Language and Cognition* 12(2). 239–257. https://doi.org/10.1017/S1366728909004027.