Research Article

Raquel Fernández Fuertes*, Eduardo Gómez Garzarán, Sonja Mujcinovic, and Iban Mañas Navarrete

When teaching works and time helps: Noun modification in L2 English school children

https://doi.org/10.1515/opli-2022-0219
received September 10, 2021; accepted October 17, 2022

Abstract: The study focuses on the interaction between length of exposure and instruction in the L2 English acquisition process of L1 Spanish school children. Two target structures involving noun premodification are targeted: noun–noun (NN) compounds and adjective–noun (AN) strings. Four groups of participants have been studied for 3 years: a group that has been exposed to a specifically designed teaching program targeting NN compounds and a group that has received the regular English instruction program which does not address this structure as part of the curriculum. Two age subgroups appear in each case. The longitudinal judgment data elicited show that performance improves in the cooperation between length of exposure and the exposure to the NN instruction program. Furthermore, it is this last issue that actually takes the lead in that the NN instruction program directly impacts on not only NN compounds but also AN strings. This points to instruction being determinant in the L2 learning process; that is, a consciously and carefully directed instruction is proven to be more effective than length of exposure itself. This study on longitudinal experimental data contributes to shed light on the factors involved in instructed L2 acquisition.

Keywords: length of exposure, instruction, language learning, noun–noun compounds, adjective–noun strings, L1 Spanish–L2 English, acceptability judgments

1 Introduction

When learning a second language (L2) in an institutional setting, there are a number of issues that have been subject to much debate. In particular, the age of onset of the L2 learning process, the nature of L2 input (in terms of quantity and quality), the type of instruction (explicit and implicit), and the length of exposure to the L2 (MacSwan and Pray 2005) are said to shape the final attainment in the learning process. In the present study, the focus is placed on length of exposure – *i.e.*, exposure to the L2 in a school context measured in years – and instruction and, particularly, on whether the interaction between length of exposure and instruction has a direct impact on learning.

The role of exposure in instructed L2 acquisition has long been studied (Akhtar et al. 2012, Al-Zoubi 2018, García Mayo 2003) with exposure being defined as the quantifiable amount of time a learner is exposed to the L2 in an institutional setting; in other words, the amount of time a learner has received

*** Corresponding author: Raquel Fernández Fuertes,** English Department, University of Valladolid, University of Valladolid Language Acquisition Lab, Plaza Campus Universitario, s/n, Valladolid 47011, Spain, e-mail: raquelff@uva.es
Eduardo Gómez Garzarán: English Department, University of Valladolid, University of Valladolid Language Acquisition Lab, Plaza Campus Universitario, s/n, Valladolid 47011, Spain, e-mail: eduardo.gomez.garzaran@uva.es
Sonja Mujcinovic: English Department, University of Valladolid, University of Valladolid Language Acquisition Lab, Plaza Campus Universitario, s/n, Valladolid 47011, Spain, e-mail: sonja.mujcinovic@uva.es
Iban Mañas Navarrete: Linguistic and Literature Education Department, University of Barcelona, Plaza Campus Universitario, s/n, Barcelona 08035, Spain, e-mail: iban.manas@ub.edu

input in the L2. Much less attention has been paid to a more focalized type of exposure, that is, when input is manipulated by means of instruction in such a way that instruction targets a specific L2 structure (*i.e.*, noun–noun compounds as in Fernández Fuertes et al. 2020, Gómez Garzarán and Fernández Fuertes 2020). In this case, by conflating general exposure to the L2 into the specific exposure to the L2 structure being targeted via instruction, we create an arena where length of exposure and instruction interact. This constitutes the perfect ground to study how exposure and instruction connect and shape the learners' L2 knowledge – something that is well known but not often targeted in instructed L2 studies (Lambelet and Berthele 2015). Furthermore, and most importantly, the combination of general and explicit instructed exposure can provide valuable information with respect to effective language teaching understood as teaching leading to an improvement in L2 proficiency.

We aim at exploring these interactions between length of exposure and instruction by using noun modification, as in (1) and (2), as the specific L2 target structures. These show a noun modified by either another noun (1) or an adjective (2):

(1) a. *police dog* [*perro policía* 'dog police' in Spanish; noun + noun]
b. *shoe box* [*caja de zapatos* 'box of shoes' in Spanish; noun + preposition + noun]
(2) a. *big house* [*casa grande* 'house big' in Spanish; noun + adjective]

We investigate the effectiveness of length of exposure and its interaction with instruction by using (1) and (2) as landmark structures in a longitudinal study on the L2 English performance of native Spanish children. The selection of these structures lies in the cross-linguistic difference they exhibit in the participants' first language (L1) and L2 and in the different approach followed in the English classroom to address these two structures. Both target structures in (1) and (2) involve nominal premodification in English, by means of another noun, rendering a noun–noun (NN) compound (1), or by means of an adjective, rendering an adjective–noun (AN) string (2). The corresponding Spanish translation for both structures involves postmodification (*i.e.*, the reverse directionality) (*e.g.*, Cinque 2010, Kayne 1995). While both English and Spanish have productive AN–NA strings, respectively (with AN order in Spanish also possible but non-canonical or with alternative uses), NN compounds are more productive in English than in Spanish with alternative constructions like that in (1b) being more frequent in Spanish. In the case of instruction, while explicit instruction on AN strings is typically part of the English curriculum in the English as a foreign language classroom in Spain (cross-linguistic word order difference with L1 normally targeted in textbooks), that on NN compounds is not. And this is so, in spite of the frequency and productivity of NN compounds, not only in the language itself (*e.g.*, Piera 1995, Snyder 2001) but also in the materials used in the Spanish schools corresponding to the English classroom as well as to other subjects taught in English (Gómez Garzarán and Fernández Fuertes 2020).

Previous acquisition studies dealing with word order and productivity in NN compounds and AN strings show that virtually no problems arise in the acquisition of L1 English. These studies have also pointed out that at the age of 1 year and 8 months (1;8) monolinguals have already acquired the basic grammatical properties of these structures (*e.g.*, Nicoladis 2002, 2006, Snyder 2001). However, L2 studies have systematically shown that, when typologically dissimilar languages are in contact (*i.e.*, Germanic and Romance), cross-linguistic influence from the L1 to the L2 appears resulting in negative transfer. That is, if the participants' L1 is a Romance language where the NN compound order and AN string order are noun-modifiers, these participants show a high error rate in terms of directionality when producing the corresponding English structures (modifier-noun). This has been shown to be so for children and adults as well as for speakers of a Romance L2 and of a Germanic L2 (*e.g.*, Fernández Fuertes et al. 2008, 2020, Gómez Garzarán and Fernández Fuertes 2020, Liceras and Díaz 2001, Trías and Villanueva 2013).

Most previous works focus on word order and on only one of these structures (either NN compounds or AN strings). They also leave aside productivity, that is, whether the structures under discussion frequently appear in the language (*i.e.*, they are productive) or not (*i.e.*, they are not productive). We believe that addressing these two issues in both related structures can provide further insight into how L2 acquisition proceeds in the school context as one of the most common language learning scenarios.

The present study is part of a longitudinal project where data on the target structures have been elicited in three consecutive years, by means of an acceptability judgment task. The data presented in this study correspond to year 2 and year 3 of testing, while the analysis of year 1 has been addressed in a previous work (Gómez Garzarán and Fernández Fuertes 2020). Thus, the present study is a follow-up to Gómez Garzarán and Fernández Fuertes (2020), but it differs from the previous one in two respects which directly point to the present work's innovative contribution: the data under analysis correspond to a different time-frame in the L2 development (years 2 and 3); and this study adopts a longitudinal approach which is absent from the previous work as it compares data from year 2 and year 3.

Four participant groups differing in age and instruction have been analyzed for a period of 2 years by means of an acceptability judgment task. Two participant groups (a younger and an older group) constitute the groups in which both general exposure to the L2 and NN instruction converge because, in addition to the regular instruction in English AN strings, they have received an especially designed pedagogical intervention program on NN compounds. Two participant groups of the same age ranges as the previous ones have received the regular instruction in AN strings but not in NN compounds and are, therefore, the groups in which only general exposure to the L2 can be considered. The four groups have been tested in two consecutive years (year 2 and year 3). This experimental design allows us to address the interaction between exposure and instruction. More specifically, it helps to determine whether it is the overall length of exposure to the L2, or the more specific exposure via explicit instruction (or a combination of the two) that is behind these L1 Spanish children's performance in L2 English. Furthermore, it allows us to study whether an improvement in L2 proficiency is conditioned by general exposure or whether explicit direct instruction is a necessary factor.

Section 2 deals with length of exposure and types of instruction as key elements in the L2 learning process and both are targeted with a view to determining the role they actually play in reaching a better command of the L2. In Section 3, the research methodology is presented. This includes the description of the participants and their linguistic and learning contexts, as well as the materials used to elicit the data, followed by the research question and hypotheses that guide the present investigation. The results obtained are presented and discussed in Section 4. In Section 5 conclusions are given, which also points to venues for further investigation.

2 Effects of length of exposure and types of instruction

One of the topics that has focused the attention in L2 acquisition studies with regard to the increase in L2 proficiency is whether the starting age of exposure to the L2 and the time invested in the process, that is, length of exposure, play a crucial role (*e.g.*, Bialystok 1997, Hatch 1983, Scovel 2000, Singleton 1995).

For some authors, age seems to be determinant, with assertions such as "the younger the better" (Johnson and Newport 1989) or "the longer the exposure, the more proficient" (Burstall 1975) being frequently resorted to. However, other more recent instructed L2 acquisition works put a caveat to these assertions: age of onset is not necessarily determinant, and so a learning process that is early initiated may not lead to better results (Baumert et al. 2020, García Mayo 2000, García Mayo et al. 2006, Muñoz 2011). In fact, if the rate of learning is at play, it is a general assumption that older learners perform better. This is argued to be so since, in matched exposure situations, they attain a higher level of proficiency faster than their younger peers, when morphosyntactic issues are tested (*e.g.*, Blom and Baayan 2012, Krashen et al. 1979). However, in naturalistic settings, younger learners have been shown to be more advantaged and efficient – even reaching native-like levels (Krashen et al. 1979, Muñoz 2006). Thus, an asymmetry appears with younger learners showing an advantage in some respects and older learners in some other respects depending on the context. Therefore, the assertion "the younger the better" cannot be overgeneralized, especially in the L2 context, as there is no clear evidence (Muñoz 2006, 2010, 2011).

For the present study, the focus is placed on the amount of time devoted to the learning of an L2, that is, on length of exposure rather than on age alone. Despite the fact that there does not seem to be a linear

correspondence between length of exposure and final attainment (Murphy 2001), several studies have targeted this correspondence and have put it to the test when it comes to instructional settings. A certain advantage of those foreign language learners who receive more exposure over those receiving less has not always been observed (see below).

García Mayo (2003) focused on age, length of exposure, and metalinguistic awareness on the L3 English of two groups of Basque-Spanish children learning English in two different moments in a school context: 60 children had received 4 years of instruction and 44 had received 6 years of instruction. She tested different issues related to the pro-drop parameter such as subject–verb inversion and the *that*-trace effect by using an acceptability judgment task. Her results regarding length of exposure showed that those participants with the longest time of instructed input performed better.

On a different scale study given the amount of data, Muñoz (2011) dealt with age and with what she referred to as input, that is, the participants' amount of exposure to the foreign language. Data were elicited from 162 L1-Spanish, L1-Catalan, and bilingual speakers (around 30 years of age or younger), after at least 10 years of instructed English exposure (a mean of 13.9 years or over 2,400 hours). A general proficiency test, a lexical test, and a phonetic test were used. With regards to input (*i.e.*, length of exposure), she concluded that in the long term it "[…] does not cease to have a significant effect on foreign language outcomes." Therefore, input emerges as one of the influencing factors in L2 development. Moreover, she asserted that "the results concerning the effects of exposure indicate that this factor has a significant influence on proficiency outcomes" (Muñoz 2011, 129).

Shojamanesh et al. (2018) tested 11 different morphosyntactic structures in a timed acceptability judgment task. Four different groups with 120 18-year-old female Iranian students participated in their investigation. They were classified according to the type of classroom exposure and the amount of exposure (as measured in years – 7 to 11 years – and in hours of L2 exposure – 792 to 1,272 hours or more depending on the group). Their findings pointed to a positive effect of length of exposure concluding that the more time invested in learning, whether at school or at language centers, the better the results in the morphosyntactic structures tested.

As opposed to studies like those of García Mayo (2003), Muñoz (2011), and Shojamanesh et al. (2018), other works have not found that length of exposure clearly correlates with better performance in the case of instructed L2 acquisition when other issues are also taken into consideration.

Mujcinovic (2020) analyzed the effects of length of exposure in the L2 English grammatical subjects produced by 13 L1 Spanish, 13 L1 Bosnian, and 13 L1 Danish children. Two subgroups with different length of exposure to L2 English appeared: 2 years (aged 9–10) and 4 years (aged 11–12). Data showed that the longer the participants have been instructed, the better results they obtained, although this was significantly so for the L1 Spanish participants. Nonetheless, when not only grammaticality but also adequacy was considered, length of exposure did not play a role, except for the L1 Bosnian participants. Thus, length of exposure did not necessarily correlate with better performance. The rationale is that grammaticality and adequacy combined reflect more precisely the knowledge learners have of the way subjects work in English, more than grammaticality alone. And it is in this interface where vulnerability shows and "the longer, the better" principle does not apply.

Fernández Fuertes et al. (2020) analyzed semi-spontaneous oral production data from 84 participants divided into four groups: two groups differing in the instructed input received (as in the present investigation; see Section 3.1) subdivided into two age groups each. Results showed that it is direct explicit NN instruction rather than length of exposure, which results in better performance with respect to NN compounds as the target structure. The older participants with no explicit NN instruction showed a higher rate of accuracy than their younger counterparts. Such a difference was not found between the younger and older instructed groups, pointing, in this case, to direct explicit NN instruction as the differencing factor.

These last two studies showed that some length of exposure effects reported in previous works might not really be in fact length of exposure effects as such. That is, other issues involved in the learning process might, in fact, be masked in length of exposure effects.

Fernández Fuertes et al. (2020) also take into consideration the effectiveness of explicit instruction as opposed to implicit instruction. In fact, a vast number of studies have put to the test this instructional dichotomy (*i.e.*, explicit *versus* implicit), although with a wide range of methodologies. The meta-analyses

conducted by Norris and Ortega (2000), Goo et al. (2015), and Spada and Tomita (2010) are a comprehensive attempt to address the different role that the two instruction types can play in the learning process.

In order to test the effectiveness of instruction, Norris and Ortega (2000) analyzed 49 empirical studies carried out between the years 1980–1998. In their review, explicit instruction had more permanent effects than implicit instruction and, thus, explicit instruction emerged as a more efficient instruction type. The authors point out that these conclusions, however, need to be toned down because the variability in the type of instruction used in the different L2 environments may also play a role in the results. Variability is, of course, intrinsic to this type of metanalyses.

Taking as a point of departure, Norris and Ortega's (2000) meta-analysis, Goo et al. (2015) reanalyzed 11 of their 49 studies and added 23 new ones, which were carried out between the years 1999–2011. Their findings are in line with those of Norris and Ortega in that explicit instruction is favored over implicit instruction.

Similar results were found by Spada and Tomita (2010) in their meta-analysis comparing 41 studies. In their case, the focus is placed on instruction type as well as on the degree of complexity of a number of English grammatical rules present in the L2 English classroom. Larger effect sizes of explicit over implicit intervention were found, regardless of the complexity of the rule at stake.

To sum up, the intention of our investigation is to operationalize and test the implicit/explicit instruction distinction in our experiment insomuch that we intend to make explicit concrete language aspects of NN structures in English to part of the child participants of our study, and test if for the others it is actually learned from input, therefore in an implicit way.

The variability of certain aspects such as the way in which instruction was operationalized in the L2 contexts, or the different grammatical rules tested, complicates the overgeneralization of all the findings. However, the results attested in the studies considered in these three meta-analyses point, in general, to the effectiveness of explicit instruction over implicit, with durable effects in time.

Taking into account the effect of possible issues such as length of exposure or type of instruction, the question is whether these effects disappear when other issues factor in, as in the studies referred to above, or rather whether they are kept constant in which case length of exposure will indeed be explaining differences across participant groups. This more refined approach to length of exposure that includes the role of instruction is the one followed in the present study. Our aim is to shed more light on the issues that intervene and shape the instructed L2 acquisition process.

3 Research methodology

This study is part of a large-scale work on the instructed L2 acquisition process followed by Spanish children learning English as an L2 in a school context in Spain. Longitudinal data on the target structures (NN compounds and AN strings) have been elicited in three consecutive years by means of different experimental tasks. In the case of the acceptability judgment task, as the one at stake in the present analysis, out of the three years for which data have been gathered, two of them are subject to investigation now: year 2 and year 3.

3.1 The participants

The data in the present study have been elicited longitudinally for a period of 2 years from 95 L1 Spanish children learning English as primary school students in a bilingual school in the region of *Castile and León* (Spain). The school follows the CLIL (Content and Language Integrated Learning) methodology and so, other than the English as an L2 subject, two or three more subjects are also taught in English, depending on the school year level. As in Gris-Roca's (2017) policy of inclusion, a combination of teaching methods is used: the communicative language teaching, the CLIL methodology in the content subjects, and task-based

language teaching. As per the language background questionnaire filled out by the children's parents, children have never been exposed to English in a naturalistic context, that is, they have not spent long periods abroad in an English-speaking country nor have been in contact with native speakers of the language out of the school context. Informed consent from parents was obtained and the school collaborated for the whole process of data collection. Ethical approval from the University of Valladolid was obtained before data collection as part of the UVALAL (University of Valladolid Language Acquisition Lab) research activities (protocol approval ref. PI 19-1461).

Table 1 provides information on the four experimental groups, comprising different participants each: the non-NN instructed groups (younger and older) and the NN instructed groups (younger and older).

The two instruction groups (*i.e.*, non-NN instructed and NN instructed) differ in terms of whether they have followed or not the pedagogical intervention program (see Section 3.2).

In each instructed group, two subgroups were considered so that length of exposure could be addressed. In this case, length of exposure is reflected in the years that participants have been exposed to L2 English in a school context. That is, participants that have been exposed to English for 2 years constitute the younger groups and participants that have been exposed to English for 4 years are referred to as the older groups. As for the children's age, the range goes from 7 to 10 years old.

As for the amount of exposure to L2 English in the school context, children received on average 6.5 hours per week (35 weeks in a normal academic year). This amounts to the following ranges, as in Table 1: while the younger groups have received between 341.25 and 568.75 hours of English input, the older groups have received between 796.25 and 1023.75. This counting takes into account that testing was implemented half-way through the academic year. That is, for instance, in the case of the younger groups, whose L2 exposure amounts to 341.25, 227.5 hours corresponding to the first year of exposure have been counted; and 113.75 hours corresponding to half of the second year of exposure have been added to those of the first year, because this is when they were tested. Out of school exposure was not controlled for, but none of the children were involved in English immersion contexts of any kind and, in this respect, the participants' exposure to English is considered to be quite homogenous.

The four participant groups have been tested in two consecutive years (*i.e.*, year 2 and year 3). The longitudinal nature of this study is captured in Table 2.

This experimental design, as reflected in Table 2, allows us to actually target the role of instruction (*i.e.*, non-NN instructed and NN instructed groups) and its interaction with length of exposure (*i.e.*, younger and older groups; year 2 and year 3).

Regarding the proficiency of the participants, Cambridge Young Learners Exams of the corresponding age were used in order to set homogeneous groups. The results were analyzed by means of two independent sample *t*-tests with Welch correction for unequal variances. For the two younger learners' groups, the test showed that there was no statistically significant difference in the proficiency scores between groups ($t(42.832) = -1.71$, $p = 0.09$, $d = 0.4$, 95% CI [−1.80, 0.12]). For the two older learners' groups, no statistically significant difference was found either ($t(42.414) = 0.07$, $p = 0.93$, $d = 0.02$, 95% CI [−0.87, 0.94]). These results ensure that the groups are comparable in terms of age and length of exposure and that, within each age group, differences only appear in the type of instruction received. This addresses one of the drawbacks pointed out by Muñoz (2011), when referring to how typical studies that aim at measuring L2 ultimate

Table 1: Participant groups

	Non-NN instructed group		**NN instructed group**	
Length of exposure (in years) (grade in year 2–year 3)	Younger (2) (2nd–3rd)	Older (4) (4th–5th)	Younger (2) (2nd–3rd)	Older (4) (4th–5th)
# Of children (female + male)	25 (16 + 9)	23 (10 + 13)	23 (9 + 14)	24 (14 + 10)
Age (year 2–year 3)	7–8	9–10	7–8	9–10
Hours of L2 exposure (year 2–year 3)	341.25–568.75	796.25–1023.75	341.25–568.75	796.25–1023.75

Table 2: Longitudinal study: length of exposure and NN instruction

	Year 2 of testing	Year 3 of testing
Non-NN instructed younger	■■ ○	■■■ ○
Non-NN instructed older	■■■■ ○	■■■■■ ○
NN instructed younger	■■ ●●	■■■ ●●●
NN instructed older	■■■■ ●●	■■■■■ ●●●

Length of exposure
■ each square represents a year of exposure to L2 English in school context
■ marks the year they were tested

NN-instruction
● each circle represents a year
● marks the year they were tested
○ marks the absence of NN instruction

attainment are carried out: "learners have not been matched in terms of experience with the L2, and so input across participants has not been comparable" (Muñoz 2011, 114). This drawback has been controlled in the present study: all the experimental participants have been carefully selected in order to have the most homogenous and similar groups possible, except for the length of exposure and the NN instruction variables. At the time of the last testing (year 3), and during the two previous years (years 1 and 2), all participants have been enrolled in the same school, taught by the same teacher and with the same textbooks for all the subjects taught in English. Therefore, there are reasonable grounds to assert that, as far as the school context is concerned, all the participants received the same amount of instructional input, so they are comparable: the age-matched groups (younger non-NN instructed and younger NN instructed, on the one hand; and older non-NN instructed and older NN instructed, on the other), and the instruction type groups (younger non-NN instructed and older non-NN instructed; and younger NN instructed and older NN instructed).

Therefore, combining the general exposure to L2 English (younger *versus* older groups, year 2 *versus* year 3) with the specific NN intervention program implemented for two years, or lack thereof (NN instructed *versus* non-NN instructed groups), the role played by length of exposure and by the specific instruction received could be explored and teased apart.

3.2 The pedagogical intervention program

The taxonomy of the two different instructed groups (*i.e.*, the non-NN instructed and NN instructed groups) derives from the specific pedagogical intervention program used with half the participants during three academic years. As indicated before, in the present study we focus on the last two years of this longitudinal study (year 2 and year 3). This explicit intervention program was implemented targeting one of the structures under analysis, NN compounds, since the other target construction in this study (*i.e.*, AN strings) is already part of the curriculum in Spanish schools. Thus, AN strings are addressed via a set of varied activities and exercises included in the textbooks as well as explicit interventions on the part of the teacher. This AN instruction was exactly the same for all the participants in the different instructed groups and more importantly, no connection was ever established between AN strings and NN compounds. The NN instruction program, undergone by the NN instructed groups only, comprises a set of varied in and out-classroom activities aimed at overtly teaching the main properties and uses of English NN compounds (paying special attention to word order and productivity). It included different interpretation and production activities, planned beforehand (*e.g.*, creating, naming, drawing, and explaining novel NN compounds, systematic interpretation of every compound appearing in the sessions), that were carried out and integrated in the regular school lessons in all the subjects taught in English. In these activities, the differences between

English and Spanish compounds were highlighted. Besides the word order difference with Spanish as the most noticeable cross-linguistic difference, NN compounds, as previously mentioned, are highly productive in English, but not so much in the participants' L1 (Snyder 2001, Gómez Garzarán and Fernández Fuertes 2020). Due to these two differences, NN compounds thus present an additional interest from a cross-linguistic point of view as well as a challenge for these learners.

The NN instruction program can be considered highly systematic and intrusive enough as it permeated all the L2 English curriculum. It is not possible to precisely quantify the number of hours this intervention was used given the different nature of the activities involved. As mentioned above, some activities were systematic and repetitive such as commenting each and every NN compound appearing in the textbooks. To illustrate this, the younger participants have been exposed to English for 341.25 hours in year 2 (Table 1) and, during this time, whenever a compound appeared, an activity that focused on word order was used. Other activities, however, were integrated in the didactic units and, for instance, involved making up novel realities by using NN compounds, or using songs to interpret NN compounds, *etc*. In this case, the younger participants in year 3 (568.75 hours of general exposure to English) performed seven different activities of this type and each was used at least three times. In either case, this pedagogical intervention took place systematically during the three years of the longitudinal study in the school context. Additionally, the teacher in charge of implementing the NN instruction program is a co-author of the present study, and he is a university-trained linguist and teacher. Furthermore, during the longitudinal study, he was the coordinator of the CLIL program at the school. As he was also the teacher for the non-NN instruction groups, all this ensures that teaching practices were highly controlled for in the four participant groups.

Such explicit instruction in NN compounds is not normally part of the usual classroom practices in the English as an L2 course in Spanish schools (Fernández Fuertes et al. 2008, Gómez Garzarán and Fernández Fuertes 2020). This situation reflects a contrast when NN compounds are compared to AN strings. In spite of sharing a similar underlying syntactic structure as far as word order is concerned (Gómez Garzarán and Fernández Fuertes 2020), NN compounds are not typically part of the curriculum while AN strings are typically included in the usual teaching practices of the Spanish school context.

3.3 The experimental task

An untimed acceptability judgment task has been designed, which includes 12 NN compounds and 12 AN strings randomized with 10 fillers. Fillers were disregarded for the analyses (*e.g.*, copulative sentences with noun or adjective subject complements). In year 3 of testing, four extra experimental items of each type were added to the task so that, given that participants were already familiar with the task, more data were collected. Thus, the task comprises a total of 34 items and 42 items in year 2 and year 3, respectively. The task was preceded by a set of practice items, half of them grammatical and half of them ungrammatical and none involving the experimental structures (*i.e.*, spelling errors, determiner–noun word order errors, copulative sentences with adjectives and determiner–noun phrases). Thus, two data sets were gathered, one in year 2 and one in year 3, for the four experimental groups (*i.e.*, NN instructed and non-NN instructed groups and the corresponding two age groups in each case). This procedure follows a common practice in previous instructed L2 acquisition studies where participant groups are tested at two different moments in time using the same task (*e.g.*, García Mayo 2003, Ortega and Iberri-Shea 2005).

The main criterion used to select the components of the two-word target structures was that all the lexical items were part of the vocabulary the participants are frequently exposed to. NN compounds were mostly taken either from the textbooks used in the different subjects or built up from nouns well known by the participants, including a few novel NN compounds such as "*pirate bike*" (grammatical NN compound) or "**umbrella frog*" (ungrammatical NN compound given the picture referent provided). In the case of the adjectives in the AN strings, half of the adjectives referred to size or emotions, and the rest to colors. Nouns were balanced in terms of animacy, too (*i.e.*, [±animate]).

Figure 1: Grammatical experimental items: AN string and NN compound.

In the task, the target structures were part of a question, as in Figure 1 (*i.e.*, Would it be ok to say in English "*a brown dog*"/"*a pirate bik*e" to describe this picture?). "*a pirate bike*" would be an example of a novel NN compound.

Participants were asked to provide an answer using a four-option Likert scale, for which they were asked to act as teachers and mark the option they thought was more appropriate. Emoticons with evaluative words were used and later translated into a numerical scale for data processing ("*perfecto*/perfect = 4, *casi bien*/almost ok = 3, *regular*/so so = 2, *mal*/bad = 1"). With the practice items, participants were guided to use the options 4–3 for correct items, and the 2–1 for incorrect ones, considering that "perfect" is not the same as "almost ok," and that "so so" does not correspond exactly to "bad." When data were codified, the four-option Likert scale was reduced to a 2-value scale to obtain the correction rates; that is, options 1 and 2 were collapsed for judgments assigned to incorrect experimental structures and options 3 and 4 were so for correct experimental structures. Correction rates were then translated into proportions for the analyses and comparisons carried out.

Out of the total number of experimental items (24 in year 2 and 32 in year 3), half were grammatical and half ungrammatical. Ungrammaticality implied a word order violation, as in Figure 2.

The task was administered, as part of the regular classroom activities, via Power Point presentations, providing each participant with their respective answer sheet.

Data from a group of English monolingual children from an elementary school in Edmonton (Canada) have been used in order to determine the validity of the task, especially in the case of some novel NN compounds. Native participants performed nearly at ceiling in all conditions with an NN accuracy rate of

Figure 2: Ungrammatical experimental items: AN string and NN compound.

86.11% (*SD* = 0.074) and an AN accuracy rate of 95.15% (*SD* = 0.075). These high accuracy rates ensure the robustness of the task. Furthermore, and in order to check the reliability of the task, the response patterns from the native control group were examined by a Chi-squared goodness of fit test. The results for participants' tendency toward grammatical and ungrammatical responses showed that the differences between observed and expected frequencies were not significant for these English monolinguals ($X^2 = 0.193$, df = 1, $p = 0.659$, $V = 0.024$). Thus, it can be concluded that the participants were not biased toward accepting the sentences as grammatical or ungrammatical in this study.

3.4 Research question and hypotheses

The present work is a 2-year longitudinal study (year 2 and year 3, as in Table 2) involving participants classified into two main comparable groups that only differ in the instructional input received on NN structures. Each group comprises two subgroups that differ in the length of exposure to L2 English that they have received at school (older and younger, as in Table 2).

Thus, the two variables for this investigation are the amount of exposure to L2 English at school, as seen in the different age groups (younger and older) and in the two years of testing (year 2 and year 3), and the NN exposure received by half of the participants as part of a specific instruction program (NN instructed and non-NN instructed groups).

By analyzing the participants' answers in the experimental task regarding the two structures discussed – NN compounds and AN strings – and taking into account the two variables mentioned above, the research question that guides this study is the following: what shapes the L2 English production of L1 Spanish children: length of exposure or explicit instruction? That is, in order to minimize the negative effect of cross-linguistic influence (*i.e.*, transfer) from Spanish into English, is length of exposure enough? Cross-linguistic influence would be reflected in the transferring of Spanish word order into English: noun-modifier instead of modifier-noun (as in Gómez Garzarán and Fernández Fuertes 2020, please see the comparative account in Section 1). More specifically, transfer is expected to occur in both AN strings and NN compounds, as they share the same underlying structure. Differences between the two, with a higher degree of transfer in NN compounds, could be attributed to NN compounds being unproductive structures in Spanish and not typically part of the English curriculum as such.

If amount of L2 exposure could be measured in two ways (*i.e.*, in terms of overall exposure to L2 English in the school context and in terms of NN exposure as part of a specific instruction program), then cross-linguistic influence could be affected by either or both. In particular, three hypotheses can be put forward.

Hypothesis 1 states that L2 exposure will account for the improvement of the participants when judging NN compounds and AN strings. This way, older students will outperform younger students; and data from year 3 will show higher accuracy rates when compared to those from year 2. In both cases, whether length of exposure is seen in terms of age (*i.e.*, younger/older, between group comparison) or in terms of year of testing (*i.e.*, year 2/year 3, within group comparison), this difference will equally be shown regardless of whether data come from the non-NN instructed group or the NN instructed group. Improvement will affect the two structure types (*i.e.*, NN compounds and AN strings) in the same proportion.

Hypothesis 2 states that intervention through the NN instruction program will explain the L2 improvement and so the NN instructed groups will outperform the non-NN instructed groups. Difference across instruction groups will be seen in NN compounds as it was the target structure of the NN instruction program; and it will equally be shown regardless of age (*i.e.*, younger/older, between group comparison).

Hypothesis 3 states that a combination of overall exposure and explicit NN instruction will create a hierarchy among the four participant groups, with NN instructed groups taking the lead, so that the older NN instructed group will perform the best and the younger NN instructed group will outperform both non-NN instructed groups. In this last scenario, the cooperation of length of exposure and NN instruction will bootstrap the participants' L2 final attainment with a more native-like performance.

If hypothesis 1 is confirmed, a difference between the answers of year 2 and year 3 groups will appear, as in García Mayo (2003), Muñoz (2011), and Shojamanesh et al. (2018). This difference will be so regardless of whether participants belong to the NN instructed or the non-NN instructed groups. A hierarchy or progression from lower to higher accuracy rates will be seen as in (3). In this hierarchy, groups are ordered in terms of the amount of exposure (from the lowest to the highest):

(3) younger in year 2 (341.25 hours) < younger in year 3 (568.75 hours) < older in year 2 (796.25 hours) < older in year 3 (1023.75 hours)

If hypothesis 2 is confirmed and NN instruction is the key factor, the difference will be spotted between the NN instructed and the non-NN instructed groups, as in Fernández Fuertes et al. (2020) and Gómez Garzarán and Fernández Fuertes (2020). The hierarchy in (4) shows the accuracy rates expected for the NN compounds (from the lowest to the highest):

(4) non-NN instructed groups in NNs < NN instructed groups in NNs

If hypothesis 3 is confirmed and a cooperation between length of exposure and NN instruction is behind the results, a hierarchy or progression from lower to higher accuracy rates will be seen as in (5). In this hierarchy, groups are ordered in terms of the relative amount of cross-linguistic influence (from the highest to the lowest):

(5) younger non-NN instructed > older non-NN instructed > younger NN instructed > older NN instructed.

By addressing what is behind cross-linguistic influence effects and the relation between length of exposure and explicit direct NN instruction, we want to contribute to the ongoing debate on the factors impacting L2 acquisition in the school context through a longitudinal experimental study.

4 Results and discussion

Data have been analyzed considering the two variables targeted in our study. These appear in two different subsections. In the first one (*Length of exposure: year 2 versus year 3*), testing times are targeted, that is, results from years 2 and 3 are compared. In the second one (*Explicit instruction: NN instruction versus non-NN instruction*), instruction is focalized and so results from the NN instructed and the non-NN instructed groups are compared. In each case, three analyses have been conducted: overall correct judgments; correct judgments corresponding to grammatical and ungrammatical experimental items; and correct judgments corresponding to NN compounds and AN strings. The analyses of the two variables will allow us to address the three hypotheses put forward with comparisons between the experimental groups in terms of length of exposure (*i.e.*, testing times and age) and in terms of specific exposure through the NN instruction program followed by two of the four participant groups. Statistical analyses were performed in R environment, Version 4.0.3 (R Core Team 2018). Robust analyses based on trimmed means were performed using the *WRS2* package (Mair and Wilcox 2020). The significance level for multiple comparisons was adjusted using the *post-hoc* Tukey test.

4.1 Length of exposure: year 2 versus year 3

The results corresponding to testing times (year 2 *versus* year 3) appear in Figures 3–5. Overall, for year 2 the mean correction rate is lower than the mean correction rate for year 3 (year 2: $M = 62.66$, $SD = 16.33$; year 3:

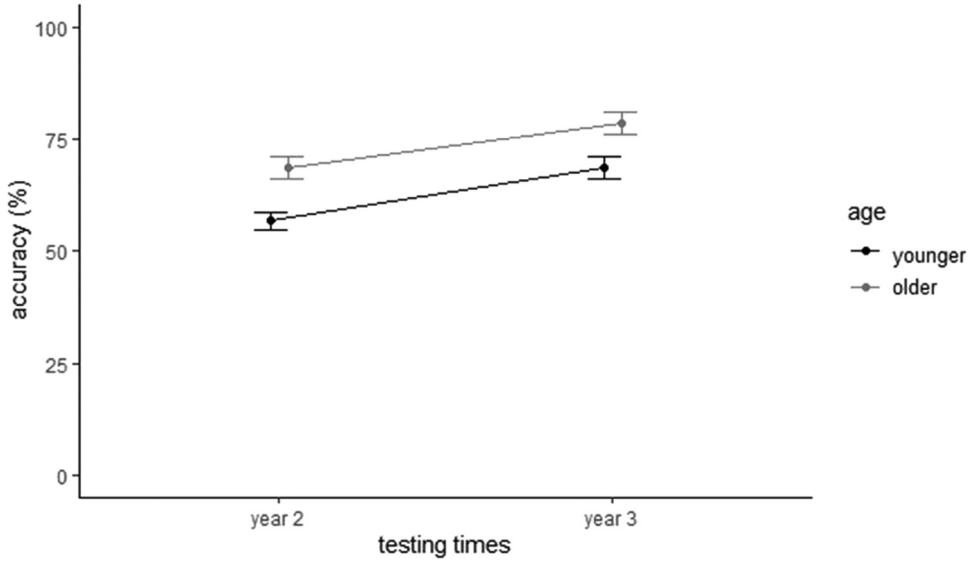

Figure 3: Correct judgments at the two testing times.

M = 73.65, *SD* = 17.79). As data are not normally distributed, robust tests have been used to compare these paired samples. Results from the Wilcoxon Signed-Rank Test show a significant difference between years 2 and 3 with a large effect size ($V = 571.5$, $p < 0.001$, $r = 0.63$, $n = 95$). That is, the longer the exposure, the better the performance, as in Burstall (1975).

Where overall accuracy rates are shown (Figure 3), the higher the length of exposure, the better the performance in both age groups. That is, participants tested in year 3 have higher accuracy rates (younger: *M* = 68.72, *SD* = 17.19; older: *M* = 78.68, *SD* = 17.14) than participants in year 2 (younger: *M* = 56.77, *SD* = 13.14; older: *M* = 68.88, *SD* = 17.19). This increase in accuracy suggests that overall exposure does have an impact on these speakers' performance (as suggested in hypothesis 1).

A robust mixed analysis of variance (ANOVA) with trimmed means at 20% was conducted to explore the interaction and main effects of age, as a between-subjects factor, and testing times, as a within-subjects factor. The ANOVA results show a significant main effect for testing times with a large effect size ($F_{(156.81)}$ =

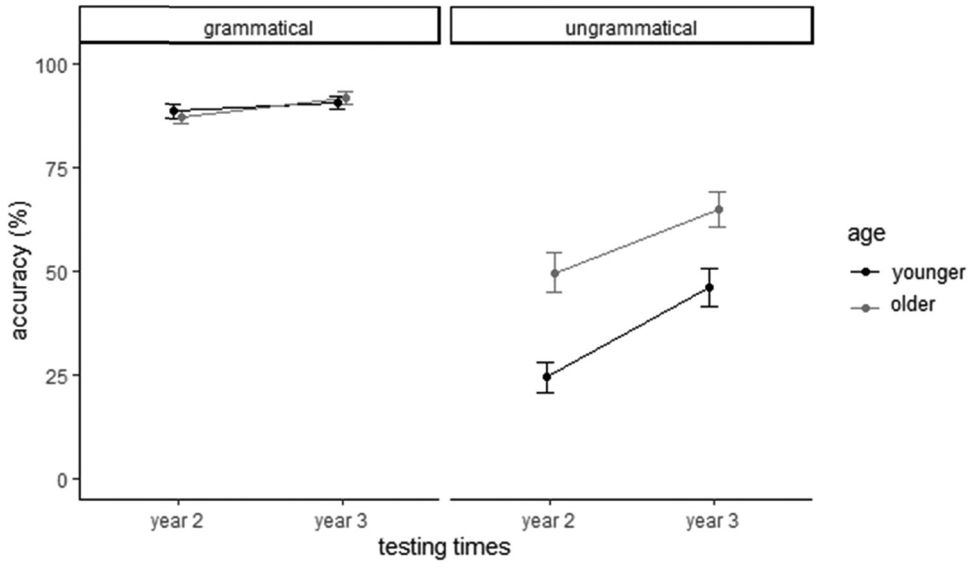

Figure 4: Correct judgments at the two testing times: grammatical versus ungrammatical items.

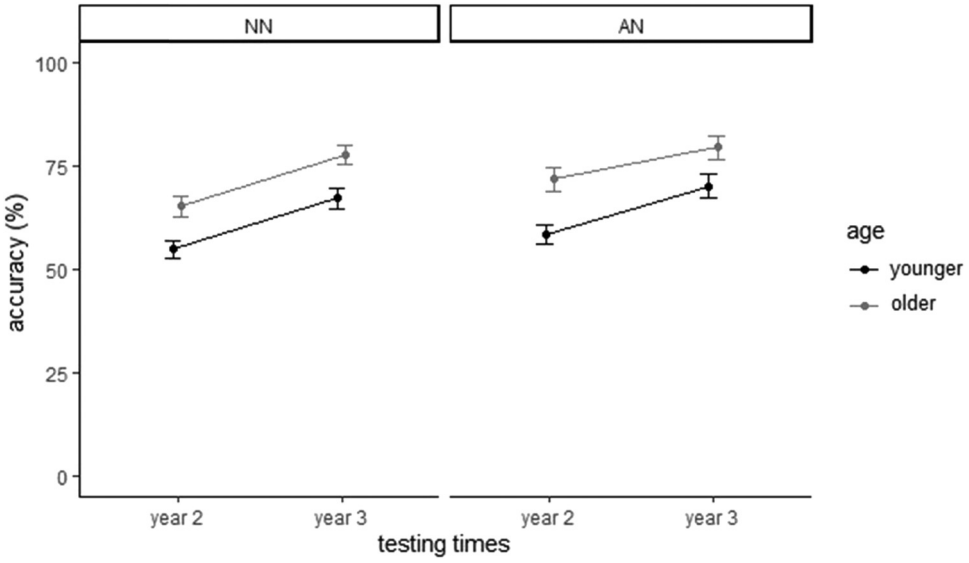

Figure 5: Correct judgments at the two testing times: NN compounds versus AN strings.

43.04, $p < 0.001$, $\eta_p^2 = 0.36$) and for age with a moderate effect size ($F_{(149.74)} = 13.18$, $p < 0.001$, $\eta_p^2 = 0.13$). However, there is no interaction effect between age and testing times ($F_{(156.81)} = 0.03$, $p = 0.86$, $\eta_p^2 = 0.004$). This points again to a clear effect of length of exposure, that is, year 3 participants outperform year 2 participants in both age groups. This is in line with previous works on the role of length of exposure (García Mayo 2003, Muñoz 2011, Shojamanesh et al. 2018).

When considering the correct judgments given to grammatical and ungrammatical experimental items, participants seem to have no problem with grammatical items with their performance being almost at ceiling. The mean correction rate for year 2 ($M = 88.22$, $SD = 11.43$) was compared with that for year 3 ($M = 91.57$, $SD = 10.14$), using the Wilcoxon Signed-Rank Test. The results showed a significant difference between years 2 and 3 but with a small effect size ($V = 3522$, $p = 0.007$, $r = 0.26$). Ungrammaticality, however, is more difficult for them to detect in both testing times (year 2: $M = 36.98$, $SD = 32.38$; year 3: $M = 55.68$, $SD = 31.68$). Results from the Wilcoxon Signed-Rank Test ($V = 3025$, $p < 0.001$, $r = 0.57$, $n = 95$) show a significant difference between years 2 and 3 with a large effect size. This reconfirms hypothesis 1 in the sense that year 3 participants outperform year 2 participants in the judgment of both grammatical experimental items and, especially so, ungrammatical experimental items.

Accuracy rates considering age show the same pattern (Table 3 and Figure 4). Grammatical items are correctly judged, with the lowest mean proportion corresponding to the older group in year 2 (87.55%). Ungrammatical items show lower accuracy rates, the highest mean proportion corresponding to the older group in year 3 (65.17%).

Table 3: Accuracy rates per testing times and grammaticality condition

Group	Testing time	Grammaticality condition	*M*	*SD*
Younger	Year 2	Grammatical items	88.88	12.07
Older			87.55	10.82
Younger	Year 3		91.01	10.62
Older			92.15	9.7
Younger	Year 2	Ungrammatical items	24.47	26.26
Older			49.75	33.29
Younger	Year 3		46.39	31.91
Older			65.17	28.78

A robust mixed ANOVA with trimmed means at 20% was conducted in order to identify interaction and main effects of age (between-subjects factor), grammaticality condition (within-subjects factor) and testing times (within-subjects factor). The results show a significant main effect for grammaticality with a large effect size ($F_{(1,372)} = 262.42$, $p < 0.001$, $\eta_p^2 = 0.49$), for testing times with a moderate effect size ($F_{(1,372)} = 21.62$, $p = 0.001$, $\eta_p^2 = 0.09$), and for age with a moderate effect size ($F_{(1,372)} = 21.91$, $p = 0.001$, $\eta_p^2 = 0.09$). An interaction effect between grammaticality condition and testing times was found ($F_{(1,372)} = 10.93$, $p < 0.05$, $\eta_p^2 = 0.06$). Pairwise comparisons using Tukey HSD tests attested a significant difference between years 2 and 3 when it comes to judging ungrammatical experimental items ($t = -3.034$, $p < 0.05$). No such difference appears with grammatical items ($t = 0.049$, $p > 0.05$). That is, when age is taken into consideration, significant differences are only found with ungrammatical items. Therefore, length of exposure (as seen in the two testing times) seems to have an impact on how participants detect ungrammaticality, while grammaticality is more stable across the two years. Furthermore, the large effect size found in the grammaticality condition is also reflected when comparing within each testing time (year 2: $t = 4.250$, $p < 0.001$; year 3: $t = 3.097$, $p = 0.011$).

In this case, the pattern in Figure 4 matches the one in Figure 3: the higher the length of exposure, the higher the accuracy rate, both in the younger group and in the older group, and both considering grammatical and ungrammatical items. This suggests that the pattern observed in Figure 3 for all the experimental items together is to be mainly attributed to ungrammatical items. That is, it is the judgment of ungrammatical items what significantly improves from the first to the second testing time (*i.e.*, from year 2 to year 3). Therefore, ungrammaticality seems to be a landmark for how the processing of noun premodification proceeds in the case of these L2 bilingual children and for these two target structures.

If correct judgment rates are separated for each of the two experimental structures, a similar rate of accuracy seems to be given to NN compounds and to AN strings. In the case of NN compounds, the mean correction rate for year 3 is higher than that for year 2 (year 2: $M = 60.12$. $SD = 16.48$; year 3: $M = 72.47$, $SD = 17.49$). Results from the Wilcoxon Signed-Rank Test show a significant difference between years 2 and 3 with a large effect size ($V = 526.5$, $p < 0.001$, $r = 0.63$, $n = 95$). In the case of AN strings, the correction rate increases from year 2 to year 3 (year 2: $M = 65.19$, $SD = 19.42$; year 3: $M = 74.83$, $SD = 20.22$). Results from the Wilcoxon Signed-Rank Test show a significant difference between years 2 and 3 with a moderate effect size ($V = 697.5$, $p < 0.001$, $r = 0.42$, $n = 95$).

Figure 5 shows these results in the two age groups.

A robust mixed ANOVA with trimmed means at 20% was applied on accuracy rates considering age (between-subjects factor), testing times (within-subjects factor), and structure types (within-subjects factor). Results show a significant main effect for age with a moderate effect size ($F_{(1,372)} = 30.82$, $p < 0.001$, $\eta_p^2 = 0.09$), for testing times with a moderate effect size ($F_{(1,1387.86)} = 30.31$, $p = 0.001$, $\eta_p^2 = 0.08$), and no significant main effect of structure type ($F_{(1,372)} = 2.05$, $p = 0.16$, $\eta_p^2 = 0.008$). No significant interactions were found.

The lack of a difference between the two target structure types for the two age groups at the two testing times seems to suggest that both structures do share a common underlying representation (Cinque 2010, Gómez Garzarán and Fernández Fuertes 2020, Harris 1991a,b, Kayne 1995, Piera 1995, Snyder 2001).

4.2 Explicit instruction: NN instruction versus non-NN instruction

When data are classified in terms of the two NN instruction type groups, the results are as shown in Figures 6–8. Overall, for the NN instructed group the mean correction rate is higher than that of the non-NN instructed group (NN instructed: $M = 74.58$, $SD = 18.35$; non-NN instructed: $M = 61.86$, $SD = 15.07$). As data are not normally distributed, a non-parametric test has been used to compare these paired samples. Results from the Wilcoxon Signed-Rank Test show a significant difference between the NN instructed and the non-NN instructed groups with a large effect size ($V = 11.95$, $p < 0.001$, $r = 0.63$, $n = 95$). That is, the group that has received a specific

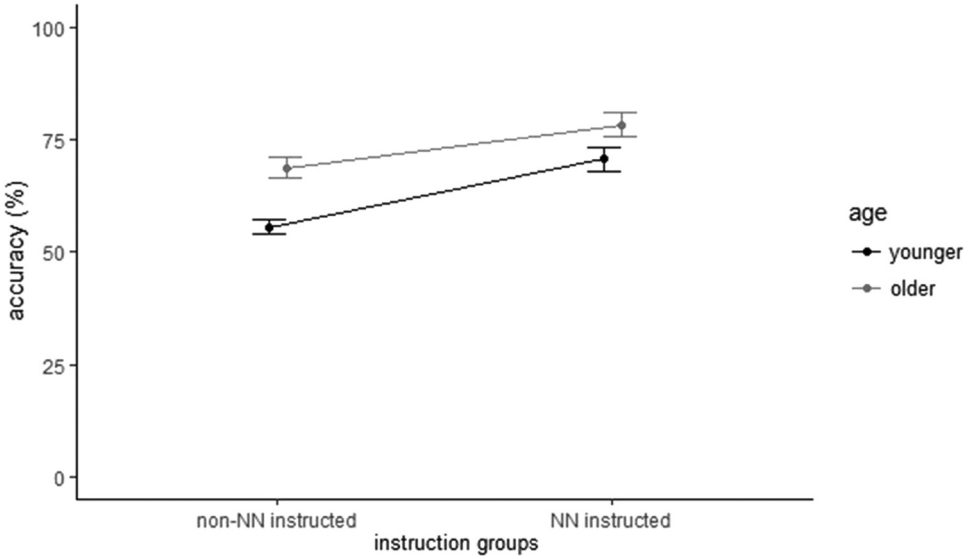

Figure 6: Correct judgments in the two NN instruction type groups: younger versus older.

pedagogical intervention program in NN compounds shows better results overall than the group that has received the traditional L2 instruction program. This lends support to hypothesis 2.

In the case of overall judgments in the two age groups (Figure 6), the NN instructed groups perform better than the non-NN instructed groups (for the younger groups, NN instructed: $M = 70.63$, $SD = 17.47$ and non-NN instructed: $M = 55.49$, $SD = 11.26$; and for the older groups, NN instructed: $M = 78.36$, $SD = 18.55$ and non-NN instructed: $M = 68.80$, $SD = 15.73$).

A robust mixed ANOVA with trimmed means at 20% was applied on accuracy rates considering age (between-subjects factor) and NN instruction type groups (within-subjects factor). Results show a significant main effect of NN instruction type groups with a moderate effect size ($F_{(1,186)} = 20.84$, $p = 0.001$, $\eta_p^2 = 0.12$) and a significant main effect of age with a moderate effect size ($F_{(1,186)} = 17.06$, $p = 0.001$, $\eta_p^2 = 0.10$). No significant interaction was found between age and instruction group ($F_{(1,186)} = 0.92$, $p = 0.34$, $\eta_p^2 = 0.005$). These results confirm that the NN instructed group achieved higher accuracy rates than the non-NN

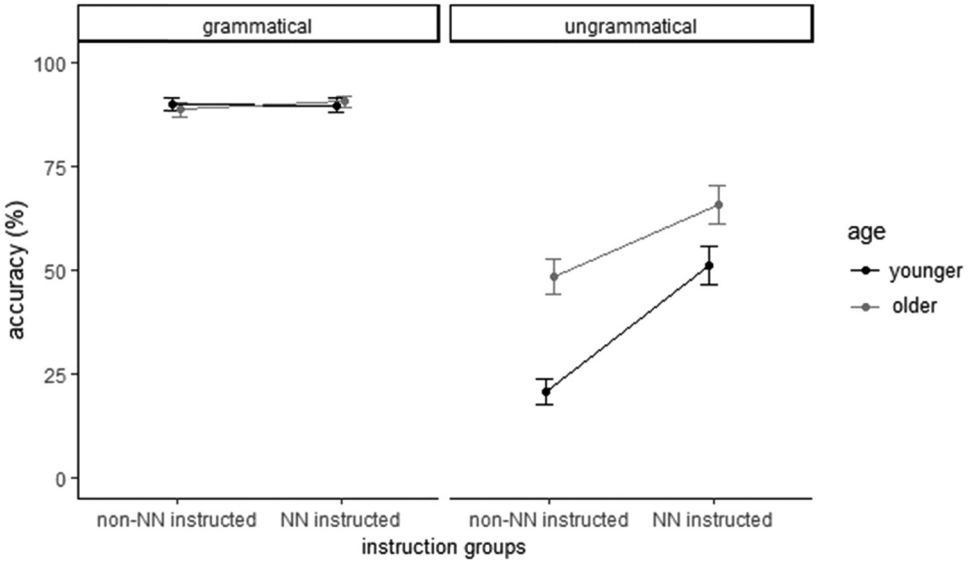

Figure 7: Correct judgments in the two NN instructed groups: grammatical versus ungrammatical items.

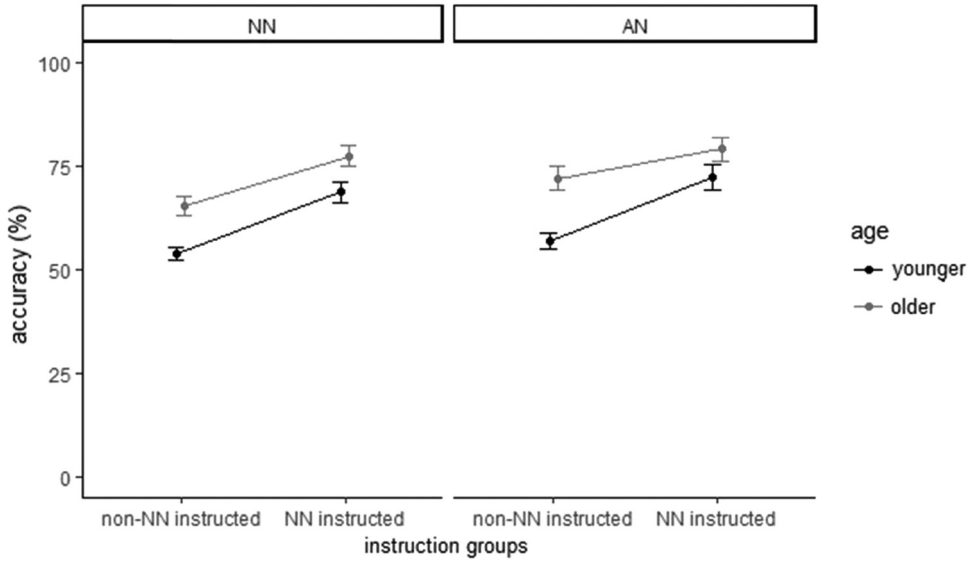

Figure 8: Correct judgments in the two NN instructed groups: NN compounds versus AN strings.

instructed group ($t = -5.17$, $p < 0.001$). The main effect of age shows that the older group always outperforms the younger group without taking instruction into account ($t = -4.70$, $p < 0.001$).

A similar pattern is seen in the case of the overall correct judgments when grammatical and ungrammatical items are separated. In the case of grammatical experimental items, the comparison between the correction rate for the NN instructed group ($M = 90.34$, $SD = 10.86$) and for the non-NN instructed group ($M = 89.47$, $SD = 10.99$), calculated with the Wilcoxon Signed-Rank Test, is found to be non-significant ($V = 4188.5$, $p = 0.383$, $r = 0.06$). For ungrammatical stimuli, however, the Wilcoxon-Rank Test revealed significantly lower accuracy rates ($V = 2575$, $p < 0.001$, $r = 0.37$) in the non-NN instructed group ($M = 34.18$, $SD = 28.84$) than in the NN instructed group ($M = 58.74$, $SD = 33.10$). These results confirm that instruction plays a salient role when judging ungrammatical items.

Accuracy rates considering age show the same pattern (Table 4 and Figure 7). Grammatical items are correctly judged, with the lowest mean proportion corresponding to the older non-NN instructed group (88.90%). Regarding ungrammatical items, accuracy rates are lower, with the highest accuracy rate corresponding to the older NN instructed group (65.90%).

The results from the robust mixed ANOVA show significant main effects for the three factors at alpha-level <0.001 (age and NN instruction groups as between-subjects factors and grammaticality condition as a within-subjects factor). In addition, significant interaction effects were found between NN instruction groups and the grammaticality condition ($F_{(1,186)} = 18.47$, $p < 0.001$, $\eta_p^2 = 0.06$), and between age and the grammaticality condition ($F_{(1,186)} = 20.12$, $p < 0.001$, $\eta_p^2 = 0.05$).

Table 4: Accuracy rates per NN instruction groups and grammaticality condition

Group	NN instruction	Grammaticality condition	*M*	*SD*
Younger	No	Grammatical items	89.99	10.90
Older			88.90	11.18
Younger	Yes		89.90	11.96
Older			90.76	9.79
Younger	No	Ungrammatical items	20.86	21.33
Older			48.65	29.13
Younger	Yes		51.26	32.43
Older			65.90	32.45

Grammatical items are judged correctly with high accuracy rates, with no differences between age groups (t = 0.02, p = 0.95) or instruction groups (t = −0.25, p = 0.58). Ungrammatical items are more correctly judged in the NN instructed groups than in the non-NN instructed groups (t = −7.28, p < 0.001), which suggests that it is NN instruction, rather than age, what plays a role. These results suggest that the specific pedagogical intervention program targeting NN compounds has a positive effect making the NN instructed group perform better than the non-NN instructed group.

When considering only ungrammatical items, an interesting observation can be made. Even if no significant differences appear between the older non-NN instructed group and the younger NN instructed group (non-significant interaction between age and instruction: $F_{(1,186)} = 1.31$, $p = 0.25$, $\eta_p^2 = 0.004$), this last group obtains higher accuracy rates for ungrammatical items (older: M = 48.65, SD = 29.13; younger: M = 51.26, SD = 32.43). This is so in spite of their having had 1 year less of English NN instruction when compared to the participants in the older non-NN instructed group. Therefore, NN instruction seems to accelerate sensitivity to ungrammaticality.

The distribution of correct judgments across the two experimental structures is rather balanced. In the case of NN compounds, the mean correction rate for the NN instructed group is higher than that of the non-NN instructed group (NN instructed: M = 73.29, SD = 18.05; non-NN instructed: M = 59.45, SD = 15.26). Results from the Wilcoxon Signed-Rank Test show a significant difference between the two instruction groups with a moderate effect size (V = 2577.5, p < 0.001, r = 0.37). In the case of AN strings, the correction rate increases from the non-NN instructed to the NN instructed groups (NN instructed: M = 75.86, SD = 21.18; non-NN instructed: M = 64.28, SD = 17.84). Results from the Wilcoxon Signed-Rank Test show a significant difference between the two instruction groups with a small effect size (V = 3078, p < 0.001, r = 0.27).

In Figure 8, data corresponding to NN compounds and AN strings are presented separately by age groups.

The robust mixed ANOVA results confirm a significant main effect for the two between-subjects factors, instruction with a moderate effect size ($F_{(1,372)} = 37.11$, $p < 0.001$, $\eta_p^2 = .10$) and age with a moderate-large effect size ($F_{(1,372)} = 28.67$, $p < 0.001$, $\eta_p^2 = 0.08$). Non-significant main effects were found for the within-subject factor structure types ($F_{(1,372)} = 28.67$, $p < 0.001$, $\eta_p^2 = 0.08$). No significant interaction effects were detected at alpha-level 0.05.

In fact, no significant differences are found between the accuracy rates corresponding to the two structures in any of the two NN instructed groups regardless of age. What makes groups differ is the total accuracy (Figure 8). This keeps on pointing to the connection between NN compounds and AN strings (Gómez Garzarán and Fernández Fuertes 2020) in that explicit instruction in NN compounds (*i.e.*, the complex predicate structure) boosts structures having the same underlying representation (*i.e.*, noun pre-modification either by another noun or by an adjective). That is, it is the NN instruction program what explains the difference between the NN instructed and non-NN instructed groups, with the difference between NN compounds and AN strings being constant throughout.

The analyses presented above show that both length of exposure (*i.e.*, comparing year 2 *versus* year 3; and older *versus* younger groups) and the specific NN instruction program (*i.e.*, comparing the NN instructed group *versus* the non-NN instructed group) have an impact on these child English L2 learners. That is, the longer they have been exposed to L2 English, the better their performance in the acceptability judgment test; and, if they have been exposed to the pedagogical intervention program targeting NN compounds, their performance is better when compared to those who have not been exposed to such program. This is so for both age groups (younger and older) and for both structure types (NN compounds and AN strings).

We further explore the relation between instruction type groups and testing times. A robust mixed ANOVA on accuracy rates revealed a statistically significant main effect for instruction with a large effect size ($F_{(151.53)} = 17.43$, $p < 0.001$, $\eta_p^2 = 0.14$) and for testing time with a moderate-large effect size ($F_{(153.71)} = 54.80$, $p < 0.001$, $\eta_p^2 = 0.12$). However, no significant interaction effect was found ($F_{(153.71)} = 1.82$, $p = 0.18$, $\eta_p^2 = 0.006$). This means that the differences found between the NN instructed and the non-NN instructed groups are kept constant throughout the two testing times, even if both groups significantly improve their performance from testing time 2 (*i.e.*, year 2) to testing time 3 (*i.e.*, year 3).

When grammaticality condition is examined in the relation between instruction type groups and testing times, the robust mixed ANOVA results showed a statistically significant main effect for the grammaticality condition with a large effect size ($F_{(1,372)} = 251.60$, $p < 0.001$, $\eta_p^2 = 0.50$) and for instruction ($F_{(1,372)} = 27.41$, $p < 0.001$, $\eta_p^2 = 0.11$) and for testing times ($F_{(1,372)} = 25.15$, $p < 0.001$, $\eta_p^2 = 0.09$), both with a moderate effect size. The grammaticality condition, in turn, generates a significant interaction effect with instruction ($F_{(1,372)} = 24.16$, $p = 0.001$, $\eta_p^2 = 0.08$) and with testing times ($F_{(1,372)} = 13.51$, $p < 0.001$, $\eta_p^2 = 0.05$) with a moderate effect size. There was no three-way significant interaction effect, which entails that ungrammatical items are challenging for the NN instruction group and specially so for the non-NN instruction group, but no differences appear between year 2 and year 3 in this respect. This suggests that the NN instruction program is more determinant than the length of exposure.

When distinguishing between NN compounds and AN strings, the robust mixed ANOVA results confirmed a significant main effect for instruction ($F_{(1,372)} = 50.40$, $p < 0.001$, $\eta_p^2 = 0.11$) and testing times ($F_{(1,372)} = 41.84$, $p < 0.001$, $\eta_p^2 = 0.09$) both with a moderate effect size. No significant main effect was found for structure type ($F_{(1,372)} = 1.64$, $p = 0.20$, $\eta_p^2 = 0.009$) and no interaction effects at alpha-level < 0.05 appeared. Participants seem to be treating both structure types in the same way regardless of their respective productivity in their L1 (*i.e.*, NN compounds are not productive in Spanish while AN strings are). This further points to both structures having a common underlying representation, a fact that seems to be behind these speakers' L2 judgments.

5 Conclusions

By focusing on noun premodification, this study addresses the interplay between length of exposure and instruction, two of the main defining factors impacting on the instructed L2 acquisition process. The results from the present investigation show that length of exposure (as seen by considering two testing times: year 2 and year 3; and two age groups: younger and older) has an impact on these L2 English children's performance, as seen in previous works. However, this impact is superseded by that of instruction, which suggests not only an interplay between length of exposure and instruction, but most importantly, a predominant role of instruction derived from the NN intervention program implemented.

Hypothesis 1 relied on length of exposure. Results show that older groups outperformed younger groups and that participants' data from year 3 are more accurate than those from year 2. This is reflected in the hierarchy in (3) and in the data in Table 3 as well as in Figures 3–5 (*i.e.*, overall, per grammaticality condition and per structure type). This suggests a confirmation on the role of length of exposure; that is, the higher the amount of exposure to the L2, the better their performance. This is in line with the results obtained by previous works on the role of length of exposure (*e.g.*, García Mayo 2003, Muñoz 2011). In this particular case, our study makes a further contribution to explore the role of length of exposure by measuring length not only in terms of age but also in terms of testing times. That is, as opposed to previous works which have only focused on cross-sectional methods, we have combined both cross-sectional data (*i.e.*, older versus younger groups) and longitudinal data (*i.e.*, year 2 versus year 3). This makes our conclusion more robust.

Hypothesis 2 focused on NN instruction. Results point to the NN instructed group outperforming the non-NN instructed group in the case of NN compounds. This is reflected in the hierarchy in (4) and in the data in Table 4 as well as in Figures 6–8 (*i.e.*, overall, per grammaticality condition and per structure type). This points to a positive effect of the role of the NN intervention program. This is evident from the better performance shown by the NN instructed groups when compared to the non-NN instructed groups and not only in the case of the structure targeted in the specific instruction program that makes the two groups differ, but, most importantly and equally so, in the two structures under investigation (*i.e.*, NN compounds and AN strings). Our results are in tune with those obtained by Fernández Fuertes et al. (2020), as well as in the three meta-analyses performed by Norris and Ortega (2000), Goo et al. (2015), and Spada and Tomita

(2010). That is, explicit instruction has a positive effect in the participants' performance and makes them perform better than those participants who have not received this specific explicit instruction.

Given that both hypotheses 1 and 2 have received confirmation, the next step is to explore which of the two factors, if any at all, outranks in accounting for the group differences that appear in the data. On this behalf, hypothesis 3 explored the interaction between length of exposure and instruction, confirming this way the hierarchy in (5). Given the results between the two NN instructed groups in the two years of testing, in the case of the ungrammatical stimuli (Table 4), evidence is provided to ascertain that the instruction factor prevails upon length of exposure. Our data show that a combination of the two factors (*i.e.*, length of exposure and explicit instruction) accounts for the participants' performance and that, in fact, it is explicit instruction that takes the lead.

These conclusions point to two relevant contributions: on the side of formal description, our findings lend support to analyses arguing for a common underlying representation for both nominal premodification structures (*i.e.*, NN compounds and AN strings); on the side of learning and teaching, our findings suggest that manipulating classroom input by means of explicit instruction and intervening in the curriculum with a cardinal structure such as NN compounds has indeed a positive outcome.

However, the conclusions we have obtained need to be taken cautiously in the sense that we have analyzed a specific grammatical construction, a specific group of participants, in a specific instruction context. Our results could be, therefore, limited to a context like the one presented here and their applicability to other grammatical rules and participants with other linguistic profiles would require further investigation. However, our results do point to the importance of empirical research and to how it can inform on effective teaching practices (Ferrero 2020).

Further investigation on the interaction between length of exposure and instruction could be addressed by analyzing data elicited via a different methodology in order to combine processing data and production, so that the English-Spanish cross-linguistic differences could be further explored. This could help complete the picture of the L2 acquisition proceeds in the school context as one of the most common language learning scenarios.

Abbreviations

L2	second language
L1	first language
NN	noun–noun
AN	adjective–noun
CLIL	content and language integrated learning

Acknowledgments: We would like to thank the school and specially the children that have participated in the longitudinal study. We would also like to thank Lucía González Calvo for the design and drawing of the pictures for the experimental task and Estíbaliz Vaquera Tabera for her help in the data collection from native speakers.

Funding information: This research was funded by the Castile and León Regional Government and ERDF (European Regional Development Fund) (Ref. VA009P17), and by the Spanish Ministry of Science, Innovation and Universities and ERDF (PGC2018-097693-B-I00) as part of the UVALAL (University of Valladolid Language Acquisition Lab) research group activities.

Author contributions: The authors applied the sequence-determines-credit approach (SDC) for the sequence of authors. RFF and EGG are responsible for the conceptualization of the study and the methodology. EGG is responsible for data collection and transcription. RFF, EGG, and SM have analyzed the data and performed the

formal analysis. They have written, edited, and revised the manuscript. IMN has performed the statistical analysis. RFF is the supervisor, the project administrator, and responsible for funding acquisition.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: The datasets generated during and/or analyzed during the current study are available in the University of Valladolid Office 365 repository in the following links: Raw data: https://uvaes-my.sharepoint.com/:x:/g/personal/raquelff_uva_es/Ebtlf2mOeAxNiNYNT2BYHCIBv4d4DkySOlbWz5QuF1UH0Q?e=ptveLL. List of stimuli: https://uvaes-my.sharepoint.com/:x:/g/personal/raquelff_uva_es/EZtU70m7g-hIvWzS69ei4eMB25B3dRz68rNNuuBQM4G5LA?e=A82Jj7&wdLOR=c6C02233C-F10C-E841-B234-D4F6973AA3C1.

References

Akhtar, Nameera, Jennifer Menjivar, Elena Hoicka, and Mark. A. Sabbagh. 2012. "Learning foreign labels from a foreign speaker: The role of (limited) exposure to a second language." *Journal of Child Language* 39, 1135–49. doi: 10.1017/S0305000911000481.

Al-Zoubi. 2018. "The impact of exposure to English language on language acquisition." *Journal of Applied Linguistics and Language Research* 5(4), 151–62.

Baumert, Jürgen, Johanna Fleckenstein, Michael Leucht, Olaf Köller, and Jens Möller. 2020. The long-term proficiency of early, middle, and late starters learning English as a foreign language at school: A narrative review and empirical study. *Language Learning* 70(4), 1091–135. doi: 10.1111/lang.12414.

Bialystok, Ellen 1997. "The structure of age: In search of barriers to second language acquisition." *Second Language Research* 13(2), 116–37. doi: 10.1191/026765897677670241.

Blom, Elma and Harald R. Baayan. 2012. "The impact of verb form, sentence position, home language, and second language proficiency on subject–verb agreement in child second language Dutch." *Applied Psycholinguistic* 34, 777–811. doi: 10.1017/S0142716412000021.

Burstall, Clare. 1975. "Factors affecting foreign-language learning: A consideration of some recent research findings." *Language Teaching* 8(1), 5–25. doi: 10.1017/S0261444800002585.

Cinque, Guglielmo. 2010. *The syntax of adjectives*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262014168.001.0001.

Fernández Fuertes, Esther Álvarez de la Fuente Raquel, Isabel Parrado, and Susana Muñiz. 2008. "La bici pirata se convierte en pirate bike o en bike pirate: la composición nominal en datos de adquisición de niños (The bici pirata turns into pirate bike or in bike pirate: Noun compounding in children's data acquisition)." In *Estudios de metodología de la lengua inglesa IV*, edited by Leonor Pérez, Isabel Pizarro and Elena González-Cascos, p. 421–38. Valladolid, Spain: Centro Buendía.

Fernández Fuertes, Raquel, Eduardo Gómez Garzarán, and Iban Mañas Navarrete. 2020. "Noun-noun compounds in a game task: what child data can tell us about teaching practices." In: *Current perspectives in language teaching and learning in multicultural contexts*, edited by Margarita Planelles, Alice Foucart and Juana M. Liceras, p. 481–501. Cizur Menor, Spain: Thomson Reuters Aranzadi.

Ferrero, Marta. 2020. "Can research contribute to improve educational practice?" *The Spanish Journal of Psychology* 23, 1–6.

García Mayo, María Pilar. 2000. "Juicios de gramaticalidad y el factor edad en inglés como lengua extranjera (Grammaticality judgments and age factor in English as a foreign language)." *Plenary lecture presented at the XIX Summer Courses in the University of the Basque Country, Spain*.

García Mayo, María Pilar. 2003. "Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language." In *Age and the acquisition of English as a foreign Language*, edited by María Pilar García Mayo and María Luisa García Lecumberri, p. 94–114. Clevedon, UK: Multilingual Matters.

García Mayo, María Pilar, Amparo Lázaro Ibarrola, and Juana M. Liceras. 2006. "Agreement in the interlanguage of Basque/Spanish bilingual children: A minimalist farewell to pro." *ITL-International Journal of Applied Linguistics* 151(1), 83–98. doi: 10.2143/ITL.151.0.2015223.

Gómez Garzarán, Eduardo and Raquel Fernández Fuertes. 2020. "Is explicit instruction effective? The learning of English noun-noun and adjective-noun structures by L1 Spanish school children." *Culture and Education* 32(2), 299–339. doi: 10.1080/11356405.2020.1741872.

Goo, Jaemyung, Gisela Granena, Yucel Yilmaz, and Miguel Novella. 2015. "Implicit and explicit instruction in L2 learning." *Studies in Bilingualism* 48, 443–82. doi: 10.1075/sibil.48.18goo.

Gris-Roca, Juan. 2017. *Explicitness and implicitness as perceived by EFL teachers. A study based on the analysis of the explicit and implicit teaching potential of teaching materials. PhD thesis*. Murcia, Spain: University of Murcia.

Harris, James W. 1991a. “The exponence of gender in Spanish.” *Linguistic Inquiry* 22(1), 27–62.
Harris, James W. 1991b. “The form classes of Spanish substantives.” In *Yearbook of morphology 1991*, edited by Geert Booij, and Jaap van Marle. Yearbook of Morphology. Springer: Dordrecht. doi: 10.1007/978-94-011-2516-1_6.
Hatch, Evelyn M. 1983. *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House. Publishers, Inc.
Johnson, Jacqueline S., and Elissa L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21(1), 60–99. doi: 10.1016/0010-0285(89)90003-0.
Kayne, Richard S. 1995. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
Krashen, Stephen. D., Michael H. Long, and Robin C. Scarcella. 1979. Age, rate, and eventual attainment in second language acquisition. TESOL Quarterly 13(4), 573–82.
Lambelet, Amelia, and Raphael Berthele. 2015. *Age and foreign language learning in school*. Basingstoke, UK: Palgrave Macmillan. doi: 10.1057/9781137525901.
Liceras, Juana M. and Lourdes Díaz. 2001. “Triggers in L2 acquisition: the case of Spanish NN compounds.” *Studia Linguistica* 54(2), 197–211. doi: 10.1111/1467-9582.00060.
MacSwan, Jeff and Lisa Pray. 2005. “Learning English bilingually: Age of onset of exposure and rate of acquisition among English language learners in a bilingual education program.” *Bilingual Research Journal* 29(3), 653–78. doi: 10.1080/15235882.2005.10162857.
Mair, Patrick and Rand Wilcox. 2020. “Robust Statistical Methods in R Using the WRS2 Package.” *Behavior Research Methods* 52, 464–88. doi: 10.3758/s13428-019-01246-w.
Mujcinovic, Sonja. 2020. *English subjects in the linguistic production of L1 Spanish, L1 Bosnian and L1 Danish speakers: typological similarity and transfer. PhD thesis*. Valladolid, Spain: University of Valladolid.
Muñoz, Carmen. 2006. “The effects of age on foreign language learning: The BAF Project.” In *Age and the rate of foreign language learning*, edited by Carmen Muñoz, p. 1–40. Clevedon: Multilingual Matters.
Muñoz, Carmen. 2010. “On how age affects foreign language learning.” In *Advances in research on language acquisition and teaching: Selected papers*, edited by Angeliki Psaltou-Joycey and Marina Mattheoudakis, p. 39–49. Thessaloniki, Greece: Greek Applied Linguistics Association.
Muñoz, Carmen. 2011. “Input and long-term effects of starting age in foreign language learning.” *International Review of Applied Linguistics in Language Teaching* (*IRAL*) 49(2), 113–33. doi: 10.1515/iral.2011.006.
Murphy, Elizabeth. 2001. “The relationship of starting time and cumulative time to second language acquisition and proficiency.” *The morning watch: Educational and social analysis* 28(3–4). http://www.mun.ca/educ/faculty/mwatch/win21/murphy.htm (13.08.2020).
Nicoladis, Elena. 2002. “What’s the difference between ‘toilet paper’ and ‘paper toilet’? French-English bilingual children’s crosslinguistic transfer in compound nouns.” *Journal of Child Language* 29, 43–863. doi: 10.1017/S0305000902005366.
Nicoladis, Elena. 2006. “Cross-linguistic transfer in adjective-noun strings by preschool bilingual children.” *Bilingualism: Language and Cognition* 9, 15–32. doi: 10.1017/S136672890500235X.
Norris, John and Lourdes Ortega. 2000. “Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis.” *Language Learning* 50, 417–528. doi: 10.1111/0023-8333.00136.
Ortega, Lourdes, and Gina Iberri-Shea. 2005. “Longitudinal research in second language acquisition: Recent trends and future directions.” *Annual Review of Applied Linguistics* 25, 26–45. doi: 10.1017/S0267190505000024.
Piera, Carlos. 1995. “On compounding in English and Spanish.” In: *Evolution and revolution in linguistic theory*, edited by Hector Campos, and Paula Kempchinsky, p. 302–15. Washington: Washington University Press.
R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.
Scovel, Thomas. 2000. “A critical review of the Critical Period Hypothesis.” *Annual Review of Applied Linguistics* 20, 213–23. doi: 10.1017/S0267190500200135.
Shojamanesh, Vafa, Tan Kim Hua, and Khazriyati Salehuddin. 2018. “The temporal role in grammatical performance of foreign language learners of English.” *The Southeast Asian Journal of English Language Studies* 24(4), 57–70. doi: 10.17576/3L-2018-2404-05.
Singleton, David. 1995. “A critical look at the Critical Period Hypothesis in second language acquisition research.” In *The age factor in second language acquisition*, edited by David Singleton and Zsolt Lengyel, p. 1–29. Clevedon, UK: Multilingual Matters.
Snyder, William. 2001. “On the nature of syntactic variation: Evidence from complex predicates and complex word-formation.” *Language* 77, 324–42. doi: 10.1353/lan.2001.0108.
Spada, Nina and Yasuyo Tomita. 2010. “Interactions between type of instruction and type of language feature: A meta-analysis.” *Language Learning* 60, 263–308. doi: 10.1111/j.1467-9922.2010.00562.x.
Trías, Marta and Ana María Villanueva. 2013. “The compounding parameter in the acquisition of Spanish as a foreign language considering the influence of the mother tongue and the learning context.” *Colombian Applied Linguistics Journal* 15(1), 57–75. doi: 10.14483/udistrital.jour.calj.2013.1.a04.