Mariscal, Sonia; Gallego, Carlos
The Relationship between Early Lexical and Grammatical Development in Spanish: Evidence in Children with Different Linguistic Levels
Universidad Complutense de Madrid
Madrid, España

Available in: http://www.redalyc.org/articulo.oa?id=17223141011
This study analyzes the relationship between lexical and grammatical development in Spanish children. The (European) Spanish version of the MacArthur-Bates CDI was used and administered to 593 Spanish-speaking children between the ages of 16 and 30-months-old. Regression analysis was applied to evaluate the relationship between age, vocabulary (total vocabulary, nouns, and verbs) and grammatical scores on two subsections of the Grammar Part. Total vocabulary explained a significantly greater proportion of variance in grammatical outcomes than age did. However, noun and verb vocabularies did not explain a greater proportion of variance in their respective morphologies than total vocabulary did. Additionally, the predictive relationship between vocabulary and grammar was found to be weaker for children whose scores were below the 10th percentile, although this could be due to the minor variability in this group and to extreme cases. We discuss the implications of these results in relation to the question of continuity between early vocabulary and grammar development in children.

**Keywords:** lexical-grammatical relationship, Spanish language acquisition, MacArthur-Bates Inventories, late talkers.

---

En este estudio se analiza la relación entre el desarrollo léxico y gramatical en niños españoles. La versión española de los Inventarios MacArthur-Bates fue administrada a 593 niños de 16 a 30 meses de edad. Se aplicaron análisis de regresión para evaluar la relación entre la edad, el vocabulario (total, nominal y verbal) y las puntuaciones en dos sub-secciones del apartado gramatical. El vocabulario total explica una proporción significativamente mayor de la varianza de las puntuaciones gramaticales que la edad. Sin embargo, los vocabularios parciales, nominal y verbal, no explican proporciones significativamente mayores de la varianza de sus puntuaciones en morfología, que la edad. Por otro lado, la capacidad de predicción del vocabulario sobre la gramática resulta menor en los niños cuyas puntuaciones los sitúan por debajo del percentil 10; sin embargo, este resultado puede atribuirse a la menor variabilidad y a casos extremos hallados en este grupo. Se discuten las implicaciones de todo este conjunto de resultados en relación a la hipótesis de continuidad entre el vocabulario y la gramática en el desarrollo temprano.

**Palabras clave:** relación léxico-gramática, adquisición de la lengua española, Inventarios MacArthur-Bates, hablantes tardíos.

---

Correspondence concerning this article should be addressed to Sonia Mariscal. Departamento de Psicología Evolutiva, Facultad de Psicología, Universidad Nacional de Educación a Distancia (UNED). C/ Juan del Rosal, 10. 28040 Madrid (Spain). E-mail: smariscal@psi.uned.es
During the last decade, a growing body of research has provided evidence that there is continuity between linguistic milestones: from babbling to first words to grammar. Focusing on the transition from early vocabulary to grammar, different studies have assumed that early lexical development not only occurs prior to, but is actually a prerequisite for the emergence of morphosyntactic constructions (Bates, Bretherton, & Snyder, 1988; Bates & Goodman, 1997; Bates et al., 1994; Marchman & Bates, 1994). Such an assumption not only challenges the linguistic hypothesis of the independence and superiority of syntax (i.e., the Chomskian view), but it is also posits a continuous view of the process of language acquisition.

The vast majority of studies to date have dealt with the issue of lexical and grammatical development and how they relate in the case of normally developing children (ND). Very few have explored this matter in children with late language emergence or language disorders (e.g., Lyytinen & Lyytinen, 2004; Moyle, Weismer, Evans, & Lindstrom, 2007; Windfuhr, Faragher, & Conti-Ramsden, 2002). Furthermore, research aimed at exploring the interdependency of lexical and morphosyntactic development, both in ND and late-talking children, has been conducted mainly in English. Only a few studies have been published on languages that have rich morphological systems (see for example, Bassano, 2000; Casselli, Casadio, & Bates, 1999; Devescovi et al., 2005; Serrat, Sanz-Torrent, & Bel, 2004; Stolt, Haataja, Lapineimu, & Lehtonen, 2009). With that in mind, the aim of the present study is to approach these theoretical and empirical issues using the large norming database of the European Spanish adaptation of the MacArthur-Bates Communicative Development Inventories (ESMBCDI) (López-Ornat et al., 2005).

Direction of the Relationship between Lexical and Grammatical Development

Many studies in English that maintain the continuous view of lexical and grammatical development have found a strong correlation between the two components. The pioneering study by Bates et al. (1988) provided evidence of a correlation of .54 and .83 in a sample of children between 20 and 28-months-old. A later cross-sectional study, where vocabulary and grammatical measures from the MacArthur-Bates Inventories (MBCDI) were included (Fenson et al., 1994), reported a correlation of .84. Subsequent studies that were conducted using adaptations of these inventories into numerous Romance languages (Caselli et al., 1999; Devescovi et al., 2005; Jackson-Maldonado et al., 2003; López-Ornat et al., 2005; Pérez Pereira & García Soto, 2003) and non-Romance languages (Barreña et al., 2008) contributed evidence that there is a high correlation between vocabulary and grammar, and that it surpasses the correlation between age and grammar. A more recent study of bilingual children learning English and Spanish indicated strong correlations between vocabulary and grammar within each language (Marchman, Martinez-Sussmann, & Dale, 2004).

In general, most of these results have been interpreted in favor of the continuity hypothesis regarding lexical and grammatical continuity, and particularly the critical mass hypothesis (Marchman & Bates, 1994). Those authors analyzed the relationship between verbal vocabulary, the emergence of regular and irregular morphology and the first overgeneralization errors in English. They concluded that the acquisition of a certain number of lexemes, in other words a critical mass, was needed to start the process of creating abstract morphological patterns. Moreover, there would be no need to postulate a specific mechanism underlying grammar acquisition, contrary to the criteria proposed by other researchers (Marcus et al., 1992), and as other connectionist models have highlighted (Elman, 1999; Plunkett & Marchman, 1993; Smith, Nix, Davey, López-Ornat, & Messer, 2003).

Obviously, correlational data can not reveal the directionality of the relationship between vocabulary and grammar. From the point of view of process, the sequence observed, which leads from the first words spoken to the first combinations of those words and the emergence of morphology, has been associated with a mechanism called lexical bootstrapping wherein children’s grammatical ability develops by being pulled up by lexical knowledge, as if “by their bootstraps”. Nevertheless, the opposite mechanism, syntactic bootstrapping, is another possibility, i.e., the use of syntactic information to derive and access the meaning of words (Gleitman, 1990; Naigles, 1990).

In fact, recent research (Dionne, Dale, Boivin, & Plomin, 2003) has proposed the term bi-directional bootstrapping to clarify that the process of creating abstract grammatical patterns from the lexicon (i.e., lexical bootstrapping) arises during language acquisition, while at the same time, acquired grammatical knowledge facilitates lexical acquisition (i.e., syntactic bootstrapping). On that note, Dionne et al. (2003) found genetic and phenotypic evidence of bi-directional bootstrapping between lexical and grammatical development in a study of more than 2,500 pairs of twins, who were studied at 2:00 and 3:00 years of age. Another study of a very different nature longitudinally examined the linguistic development of 30 children with normal development (Moyle et al., 2007) and provided evidence of a bi-directional relationship between vocabulary and grammar development between 2:00 and 3:06 years of age. This last study also included a comparative sample of late-talking children, introducing the question of what differences exist between the patterns of lexical and grammatical development of children with normal development, and those with language delay or other linguistic impairments. Several previous studies have shown that children with Specific Language Impairment (SLI) require more exposure to learn new lexical
items, and need to have acquired a greater number of lexical types in order to generate abstract morphosyntactic regularities from an input (Conti-Ramsden & Jones, 1997; Windfuhr et al., 2002). That is, in addition to the delay in vocabulary acquisition, children with SLI exhibit a different pattern of lexical acquisition from that of normally developing children, and that difference influences their syntactic development. Along those lines, longitudinal work by Moyle et al. (2007) found correlational evidence in favor of the lexical-grammatical relationship in children with both delay and normal development. Meanwhile, in children with normal development, the nature of this relationship was clearly bi-directional, while late-talkers “exhibited a predominance of lexical bootstrapping and less evidence of syntactic bootstrapping” (Moyle et al., 2007, p. 525). Those authors also proposed that the relationship between vocabulary and grammar development seems to be relevant only in the earliest stages of the process of language acquisition (in their study, it was until 3;06 years of age), not necessarily in later stages.

To summarize the information above, the body of research examined clearly reveals that to study the relationship between the lexicon and grammar, along with each one’s development, is of obvious theoretical interest, but also offers interesting prospects for applied research as to the early detection and prevention of problems during linguistic development. The size of a child’s vocabulary appears to more accurately predict grammatical development than their age. Moreover, the hypothesis of continuity between the two developmental processes allows one to predict that a delay in lexical development will lead to slowed grammatical development, and that a low lexical level will constitute a reliable index or indication of delayed or impaired linguistic development (see Thal, O’Hanlon, Clemmons, & Fralin, 1999; Hulman, Weismer, Evans, & Hollar, 2005).

The literature review of the leading studies in this area revealed that the majority of research has been conducted in English and with typically-developing children. Very few studies have carefully analyzed the relationship between lexical and grammatical acquisition in languages with different typologies, such as the Romance languages group. In light of the morphological richness of these languages, research in this area would be especially interesting where these issues are concerned, as this trait could contribute to an earlier start of grammar acquisition in comparison to poorly inflected languages, such as English (see Devescovi et al., 2005). Besides, the comparison between typically developing children and late-talkers deserves a further study.

Empirical Evidence in Romance Languages

The empirical evidence available so far seems to support a relationship between lexical and grammatical development in richly inflected languages, which the Romance languages are. However, most of this data was of a correlational nature or was obtained from longitudinal studies with very few participants. In this way, by studying early development of nouns and verbs in French in a longitudinal case-study, Bassano (2000) discovered a relationship between the grammaticalization process in both categories, as well as the quantitative development of the child’s noun and verb lexicons. She remarks that the ‘explosion’ of grammaticalization indices observed at age 2;04, which happened slightly after the increase in lexical frequencies of nouns and verbs that was observed at age 2;03. This study’s author interpreted this temporal relationship to be in line with the critical mass hypothesis and the interdependence of lexical and grammatical development.

In Catalan and in Spanish, Serrat et al. (2004) specifically studied the interrelatedness of verbal vocabulary and the morphosyntactic aspects closely related to verb grammar (i.e., morphological productivity and syntactic complexity). Six children took part in this study, two of whom were Catalan monolingual, two Spanish monolingual and two bilingual. Among the results obtained, it was particularly noteworthy that contrary to Bassano’s (2000) results, the children learned verbs at a constant pace, and no period of lexical acceleration was observed. However, a significant increase in verbal morphology was reported to emerge at a certain point in development. They contributed evidence for a period of morphological acceleration in verb acquisition within the context of constant lexical increase. The authors proposed that their results were not incompatible with the critical mass hypothesis because acquiring a certain number of verbs (i.e., critical mass of verbs) may be a prerequisite to the observed morphological explosion. This aforementioned quantity of verbs, along with the period of time needed before being able to generate abstract regularities, would both be necessary for morphological development to advance.

Another important source of data within the study of the lexical-grammatical relationship in Romance languages is the normative data taken from adaptations of the MBCDI into some of these languages. Specifically, Pérez-Pereira and García-Soto (2003) in Galician, Jackson-Maldonado et al. (2003) in Mexican Spanish and López-Omat et al. (2005) in European Spanish observed high correlations between vocabulary scores and different measures of grammatical development that these inventories provide (MLU, measures of morphological development and grammatical complexity). The following table summarizes the correlational data obtained in those studies.

A study by Devescovi et al. (2005) focused on the relationship between vocabulary size and four different measures of MLU in Italian and English and yielded interesting results beyond the correlational data from the studies in French, Spanish and Galician mentioned above. Aside from the considerable importance of the crosslinguistic
comparison between two, typologically speaking, very different languages, the results of that study indicated that vocabulary accounted for more than five times the amount of unique variance explained by age (Devescovi et al., 2005, p. 773).

Additionally, the authors suggested that from a methodological standpoint, their results indicate that vocabulary size may constitute the best basis for comparing samples of children in crosslinguistic studies of grammatical development. That study contributed another very interesting result that was obtained by analyzing the shape of change in the vocabulary-MLU relationship in both languages. For three of the MLU measures taken (of total words and morphemes), their relationship with vocabulary was non-linear in the case of English (i.e., slow increases in MLU up to 400 words followed by a sharp burst) and, conversely, it was linear in the case of Italian (constant increase in MLU at every level of vocabulary). The study’s authors interpreted this result as indicating that grammar seems to get off the ground earlier in a richly inflected language such as Italian. That is to say, a rich morphological and relatively consistent system would make the task of extracting regularities and learning grammar easier; learners would need to acquire fewer exemplars and a smaller vocabulary before starting to analyze grammar and learn the system.

One potential problem with this interpretation is the kind of data on which it is based: measures of children’s three longest sentences as reported by parents, and counting MLU in words and morphemes. No access to more relevant and critical data was available, such as data related to the morphological productivity of the children’s language.

Objectives of the Present Study

Stemming from the findings of prior research, the present study focused on the interrelationship between the lexicon and grammar in Spanish, a language with rich morphology. The following are its primary aims:

1. To study the relationship between the development of vocabulary and grammar in a broad sample of monolingual Spanish children (n = 593) under 30 months of age, by analyzing both the relationship between total vocabulary and grammar, and the relationships between nominal and verbal vocabularies, and their respective morphologies.

2. To analyze whether the predictive value of vocabulary to grammar is the same for all children in the sample, particularly for those whose vocabulary scores fall in the lowest percentiles.

To obtain empirical evidence as to the function (linear or non-linear) that best defines the relationship between vocabulary and grammar development.

Method

Participants

The parents of 593 children (296 females and 297 males) between 1;04 and 2;06 years of age participated in this study. Table 2 displays the distribution of boys and girls as a function of their sex and age in months.

The children were Spanish monolingual, but due to the fact that they came from allover Spain, 24.5% of the sample had some sort of contact with another language (via TV, speakers outside the home, etc.), with other languages spoken in Northern Spain (Galician, Basque and Catalan), or foreign languages. These last data reflect the everyday linguistic experience of children growing up as monolingual in Spanish. All the children exhibited normal development. Children with major birth complications, diagnosed developmental disabilities and/or hearing loss were excluded from participation and their weight at birth had to be above 2.100 gr., the cut-off point established by Fenson et al. (1993). The population represented spans all of Spain, although the highest percentage (43.5%) live in Madrid. Contact with participants was made possible by the

---

1 Devescovi et al. (2005) used two measures of MLU in morphemes, a ‘conservative’ measure treating pronouns as simple uninflexed lexical forms, and an ‘expanded’ measure.
cooperation of different educational and pediatric institutions throughout the country. Both mothers (49.1%) and fathers (45.4%) reported having a college education; that percentage is higher than the average educational level of Spain’s adult population (21.22%), according to 2003-2004 official records. Despite this bias toward a population with university degrees, the analysis of variance performed to test the effect of mothers’ educational levels on the scores of vocabulary and syntactic complexity revealed no significant effect of this variable, as described in the technical manual of the ESMBCDI (López-Ornat et al., 2005).

In the vast majority of cases (87.5%), the mother was the one who filled out the inventory.

Language Measures

This study’s data was collected by administering the vocabulary and grammar sections of the Inventario de Desarrollo Comunicativo MacArthur (López-Ornat et al., 2005), which is the (European) Spanish adaptation of the MacArthur-Bates CDI (ES MBCDI). This adaptation was created by starting with the American English original (Fenson et al., 1993) and modifying it culturally and linguistically, while taking into account the available data about the process of Spanish language acquisition by children living in Spain. In doing the adaptation work, preliminary and final versions of the Mexican-Spanish MBCDI were taken into consideration (Jackson-Maldonado, Thal, Marchman, Bates, & Gutiérrez-Clellen, 1993; Jackson-Maldonado et al., 2003), along with the version developed in Galician (Pérez Pereira et al., 2003).

The structure of the inventories on the ES MBCDI adaptation was basically the same as the original, but it did incorporate certain significant novelties and modifications. The ES MBCDI is no mere translation; it is a linguistic instrument, so new items were created based on data obtained in previous longitudinal (Karousou, 2003; López-Ornat, Fernández, Gallo, & Mariscal, 1994; Mariscal, 2001) and cross-sectional (Gallo, 1990; López-Ornat, 1990, 1992) studies. After two pilot versions of the ES MBCDI, the final version was set (see technical details in López Ornat et al., 2005).

We will here present, in brief, the main characteristics of the Grammar Part, as this section underwent substantial modification from the original inventory (see Lopez-Ornat et al., 2005).

Measures of Grammatical Development on the (European) Spanish MBCDI (Inventory II: Vocalizations, Words and Sentences)

The Grammar Part of the ES MBCDI was built on the same structure as the original American English version, but due to the typological characteristics of Spanish—a richly inflected Romance language—, the following modifications were made: 1) the section devoted to regular inflectional morphology was modified considerably (word endings, or terminaciones de las palabras, section) to include more items (17 questions about noun gender and number morphology, diminutives and various items about verbal morphology—as opposed to the 4 included in the original American English version); 2) the section devoted to overgeneralization errors was adapted by selecting examples of this kind of error that affect verbs (for example: ‘*se ha rompido’ vs ‘it has broken’) and nouns (for example: ‘*un flor’, that is, a masculine indefinite article + flower, a feminine noun); and 3) all the grammatical complexity items were reconstructed (the morphosyntactic complexity, or complejidad morfosintáctica, section). This last section incorporates important novelties concerning the items’ structure, instructions and presentation. First of all, all items were obtained from the longitudinal database at the team’s disposal (López-Ornat et al., 1994). New instructions were included and the items were inserted in real conversational

---

2 Departing from the original version where parents are provided a list of frequent examples of overgeneralization errors, the (European) Spanish adaptation includes examples in two items, one devoted to verb errors and the other to nominal errors. It is well-known that the main interest in the process of morphological acquisition lies in knowing when a child makes errors of this nature—an indicator of his/her control of the underlying morphological rule—and not so much in how many errors there are, or which exactly they are.
Table 3
Sample Item and Scoring of the Grammar Complexity (Complejidad Morfosintáctica) Section of the (European) Spanish Adaptation of the CDI

Item 1: María está en el salón y quiere coger un objeto frágil. Usted no le deja. María responde [María is in the living room and would like to pick up a fragile object. You don’t let her do it. María answers]:

- Todavía no dice nada parecido (she doesn’t say anything like this yet) 0 point
- Mamá má [=mala] (mommy ba [=bad]) 1 point
- Mamá e má (mommy its ba(d)) 2 points
- Mamá es mala (mommy is bad) 3 points

Note: Total score of this section is computed by summing up the points for all the items. Minimum score 0; maximum score 102 points (34 items x 3 points)

Table 4
Comparison of Part II (Grammar) of the Original MacArthur-Bates Inventory and the European Spanish Version

<table>
<thead>
<tr>
<th>Sections and Items in the Original Version</th>
<th>Sections and Items on the Spanish Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Word endings I: regular morphology; 4 items.</td>
<td>I. Word Endings (Terminaciones de las palabras): regular morphology; 17 items.</td>
</tr>
<tr>
<td>C. Word Endings II: overgeneralization errors; 14 nouns and 31 verbs.</td>
<td>III. Surprising Words (Palabras sorprendentes): verbal and nominal overgeneralization errors; clustered into 2 items.</td>
</tr>
<tr>
<td>D. Three longest sentences.</td>
<td>IV. Combining Words (Combilación de Palabras).</td>
</tr>
<tr>
<td>E. Grammar Complexity: 37 items; 2 response options.</td>
<td>V. Morphosyntactic Complexity (Complejidad Morfosintáctica): 34 items; new structure and instructions and 4 response options.</td>
</tr>
</tbody>
</table>

contexts in order to make it easier for parents to report accurately on this language component. Finally, four response options were included, as opposed to the two offered in the original version, aiming to capture how gradual the natural process of morphosyntactic acquisition is (see example in Table 3 along with the scoring system). Table 4 presents information that allows for a comparison between the grammar section in the original MBCDI (Inventory II) and the (European) Spanish version.

Several studies of the reliability and the concurrent and predictive validity of this grammar section have demonstrated that it is both a reliable and valid evaluation tool (López-Ornat et al., 2005; Mariscal, López-Ornat, & Nieva, 2010).

Procedure

Parents in the sample were asked to fill out the inventories, which were personally distributed to them by people cooperating in the standardization of the ES MBCDI. The parents received all clarifications necessary to complete the inventories. In general, parents reported feeling comfortable with giving data about their children’s communicative-linguistic development.

For the purposes of this study, the primary measures used to analyze the relationship between the lexical and grammatical components were total scores on the Vocabulary, Word Endings and Grammar Complexity sections. In order to study the specific relationships between verbal and nominal vocabularies and their respective morphologies, partial scores were calculated drawing from the items that measure each of the two aspects in the Vocabulary and Word Endings sections. Hence, to calculate the nominal morphology score, items 1, 2 and 3 on the Word Endings section were considered; those items ask about noun number, gender and diminutiv morphemes. To calculate a verbal morphology score, on the other hand, items 4, 5, 6, 7, 9, 10, 11 and 12 were considered; those items refer to verb tense, person, modality and aspect. To derive partial scores of nominal and verbal vocabulary, nouns and only verbs from the Vocabulary section, respectively, were considered (312 nouns and 91 verbs out of 588 total vocabulary items). Various analyses were carried out with these data, all related to objective 1 (see Results).

In order to determine whether or not the predictive value of vocabulary to grammar is the same for all children in the sample, especially those who score in the lowest percentile (see objective 2), the percentile corresponding to each child’s vocabulary score was calculated. The percentile tables published in López-Ornat et al. (2005) were used in doing so. The full sample of children was divided into five groups
according to their percentiles in vocabulary production, as shown in Table 5. So, for example, 16 and 30 month-old children whose scores were above 76 and 518 respectively, were assigned to group five (high percentiles), in accordance with the available data on standardization.

Results

Age and Total Vocabulary as Predictors of Grammar Complexity

In the present study, simple and stepwise regression analyses were carried out, using age and vocabulary size as predictors of grammar, specifically, predictors of morphosyntactic complexity. Stepwise regression adds to the correlational data already available (see Table 1) the possibility of determining which percentage of variance each of the variables accounts for by introducing them into the model step by step (in this study, age and vocabulary). Table 6 indicates that total vocabulary scores were better predictors of grammatical complexity than the children’s age was. This last variable accounted for 58% of variance in grammar complexity, whereas vocabulary size explained 84.4% of variance. The stepwise regression revealed that while vocabulary scores explained 84.4% of variance in grammar scores, adding the effect of age to the model only slightly increased this percentage to 86.6%.

Age and Noun Vocabulary as Predictors of Noun Morphology

In order to examine the relationship between the size of noun vocabulary and nominal morphology, a regression analysis was performed with age and noun vocabulary size as predictors. The results indicate there is practically no difference between the predictive value of total and noun vocabulary to nominal morphology. Total vocabulary accounted for a significant amount of unique variance (62.4%, as \( R^2 = .624; p < .001 \)) in noun morphology, while noun vocabulary accounted for 61.8% (\( R^2 = .618; p < .001 \)).

Age was found to account for only 38.8% of that variance (\( R^2 = .388; p < .001 \)).

Age and Verb Vocabulary as Predictors of Verb Morphology

In order to study the relationship between vocabulary size and its respective morphology, a regression analysis was performed using age and verb vocabulary size as predictors. Similarly, in the case of noun morphology, no difference was observed between the predictive value of total and verb vocabulary to verbal morphology. Total vocabulary accounted for a significant amount unique variance in verb morphology (81.8 %, as \( R^2 = .818; p < .001 \)), whereas verb vocabulary explained 80% (\( R^2 = .800; p < .001 \)). Nevertheless, total and verb vocabularies accounted for a higher percentage of variance than age did (49.4%, as \( R^2 = .494; p < .001 \)).

The Developmental Pattern of the Vocabulary-grammar Relationship as a Function of Children’s Linguistic Levels

With the purpose of analyzing whether the predictive value of vocabulary to grammar is the same for all children in the sample, the sample was organized into five groups according to linguistic level from the lowest percentile (P0-10) group to the highest (P76-99). Table 7 presents descriptive statistics including age (in months), vocabulary and grammar scores for the full sample organized into five groups according to percentile (\( N = 593 \)).
Table 7
Mean, Standard Deviation (SD) and Age Range, Vocabulary and Grammar Scores for the Full Sample Organized into Five Groups According to Percentile (N= 593)

<table>
<thead>
<tr>
<th>Group (N)</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>P0-10 (74)</td>
<td>19.78 (4.45)</td>
<td>18.77 (19.96)</td>
</tr>
<tr>
<td>P11-25 (83)</td>
<td>21.75 (4.66)</td>
<td>64.22 (47.78)</td>
</tr>
<tr>
<td>P26-50 (130)</td>
<td>23.22 (4.82)</td>
<td>155.15 (104.10)</td>
</tr>
<tr>
<td>P51-75 (160)</td>
<td>24.28 (4.93)</td>
<td>277.87 (125.33)</td>
</tr>
<tr>
<td>P76-100 (146)</td>
<td>24.10 (4.91)</td>
<td>413.10 (133.72)</td>
</tr>
</tbody>
</table>

Table 8
Regression Analysis: Vocabulary as a Predictor of Grammar Complexity as a Function of Children’s Linguistic Levels (the Data in Parentheses Excludes Children Who are in Contact with Other Languages)

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>n</th>
<th>R²</th>
<th>% Variance of GC Explained by Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>74 (50)</td>
<td>.574 (.752)</td>
<td>57.4 <em>(75.2</em>)</td>
</tr>
<tr>
<td>11-25</td>
<td>83 (59)</td>
<td>.657 (.847)</td>
<td>65.7 <em>(84.7</em>)</td>
</tr>
<tr>
<td>26-50</td>
<td>130 (96)</td>
<td>.719 (.859)</td>
<td>71.9 <em>(85.9</em>)</td>
</tr>
<tr>
<td>51-75</td>
<td>160 (123)</td>
<td>.779 (.887)</td>
<td>77.9 <em>(88.7</em>)</td>
</tr>
<tr>
<td>76-99</td>
<td>146 (120)</td>
<td>.764 (.875)</td>
<td>76.4 <em>(87.5</em>)</td>
</tr>
</tbody>
</table>

* p < .0001

Table 8 summarizes the results of the regression analysis that was carried out to determine the predictive value of vocabulary size to the measure of grammar complexity, for each of the 5 groups of participants.

The table indicates that the predictive value of vocabulary size to the scores of grammar complexity was smaller for children in the lowest percentiles (P0-10). By further analyzing the differences between correlations, we were able to confirm there were significant differences (p < .005 and p < .01 respectively) between the P0-P10 group and the highest percentile groups (i.e., P51-75 and P76-99) in terms of the predictive value of vocabulary to grammar. The difference between P0-10 and P26-50 merely revealed a tendency toward significance (p < .08). Therefore, these results appear to confirm that for children who score below the 10 percentile, the predictive value of vocabulary scores to grammar scores is significantly lower than for children in the medium and high percentile groups (P51-75 and P76-99). These results are not due to floor effects (see Table 7, column with range of scores), but they could well be due (at least partially) to the levels of variance in the two variables, especially vocabulary; these levels are low compared to the variance in the medium and higher percentile groups. In fact, when regression analyses were performed grouping participants into only two percentile groups (P0-50 and P51-99), the amount of variance in grammar complexity explained by vocabulary was very similar (R² = .770 for P0-50 and R² = .773 for P51-99; p < .000 in both cases).

Although, as stated in the Participants section above, only monolingual families were chosen, the parents addressing their children in Spanish alone, a potential problem with the results discussed so far is the possible influence of contact with other languages (TV, contact with other speakers, etc.). A considerable number of the children (24% of our sample) live in communities where other languages are spoken (Galician, Catalanian and Euskera). To eliminate the effect of this condition on the results obtained so far, a new set of analyses was done.

The correlation between vocabulary and grammar scores for the whole sample of children was .921; the partial correlation, controlling for the effect of contact with other languages, was .920.

Next, regression analysis was used to assess the relationship between vocabulary and grammar scores excluding all children who have contact with other languages. These results for each group of children are displayed in parentheses in Table 8. By analyzing the differences between correlations, we were able to confirm that there were significant differences (p < .01 and p < .03 respectively) between the P0-P10 group and the highest percentile groups (i.e., P50-75 and P76-99) in terms of the predictive value of vocabulary to grammar. The differences between P0-10 and P26-50 indicated only a tendency toward significance (p < .08). Hence, these results mirror those of the full sample. Contact with other languages, as defined for the purposes of this study (see Participants section), does not seem to exert an important influence on children’s language development, at least at this early age.
The Shape of the Vocabulary-Grammar Relationship

Aside from exploring the relationship between vocabulary and grammar development, this study aimed to gather empirical evidence as to the function that best describes this relationship. Figure 1 offers a graphical representation of the children’s grammar development as a function of vocabulary size. It depicts the relationship between vocabulary and grammar scores on the Morphological Complexity (Complejidad Morfosintáctica) section by distinguishing—for the sake of clarity—between two scores: one related to items that establish an initial level of grammatical complexity (items 1 to 24, related to simple sentences; maximum score 72) and another whose items refer to complex sentence use (items 25 to 34; maximum score 30).

![Figure 1](image)

**Figure 1.** Evolution of Grammar Complexity (Items Assessing Simple vs. Complex Sentences) as a Function of Vocabulary Size (Plotted on the Same Graph).

Table 9

| Regression Curve Estimation Analysis: Vocabulary-grammar Complexity. Summary of the Model |
|---------------------------------|-----------------|-----------------|-----|
| [Header]                        | \( R^2 \)       | \( F \)         | Sig.|
| Linear                          | .848            | 3303.268        | < .001|
| Quadratic                       | .848            | 1648.79         | < .001|

that when linear and quadratic models were tested, no differences were discovered between the two model types.

The same sort of curve estimations was employed to model the predictive value of noun and verb vocabularies to scores on noun and verb morphology, to the same result. Again, no difference was found between the linear and quadratic models. We will consider these results in the next section.

Discussion

The first objective of this study was to analyze the relationship between lexical and grammatical development in a broad sample of Spanish children. Our results were consistent with the evidence obtained in other languages by indicating a strong relationship between vocabulary and grammar.

From a methodological standpoint, these results suggest that vocabulary size may provide the best basis for cross-language matching in comparative studies of grammar development, which has also been demonstrated in Italian by Devescovi et al. (2005). It may be the best basis for between-subjects comparisons, too. This suggests that more homogeneous samples would be selected if gathered as a function of linguistic level, rather than children’s chronological age.

From a theoretical perspective, these results are compatible with an interactive view of language development. According to that view, complex relationships operate at many levels between different language components during the process of acquisition. Those components (lexicon and grammar, for example) are relatively independent according to the ‘adult’ definition of the process, but that is not necessarily so from the point of view of the language learner (López-Ornat, 1999; Marchman & Bates, 1994; Tomasello, 2000), at least initially (Bates & Goodman, 2001; Karmiloff-Smith, 1992). Nevertheless, empirical evidence found in this study does not allow us to draw specific conclusions about the direction of the relationship between lexical and grammatical development. For developmental reasons, the focus of this paper was to establish the relationship between vocabulary, age and grammar, considering the first two as independent variables (IV). Obviously, though, correlational data by no means allow to us establish directionality. Besides, the
cross-sectional nature of the data would limit our ability to address this question anyway. Other studies using different methodologies are clearly needed. For example, Dionne et al. (2003) concluded based on longitudinal data that early in children’s transition into grammar use, the major direction of influence is from lexical to grammatical development, although influence is evident in both directions between 3 and 4 years of age. Dixon and Marchman (2007) also provided an insightful analysis of why the CDI does not allow for conclusions about directionality.

In spite of the effort made to go beyond mere global measures of lexical and grammatical progress, our analyses of the relationships between noun and verb vocabularies and their respective morphologies (gender and number for nouns, and person, tense, modality and aspect for verbs) revealed that they are no more robustly associated than with total vocabulary. In the case of nouns, this result could be attributed to the composition of the ES MBCTI vocabulary list, which contains a very high percentage of nouns (53%), compared to other syntactic categories. The same cannot be said of verbs, however, which are less represented in the vocabulary list (15.48%). A more plausible explanation for these results has to do with the very nature of the morphological measures used in these kinds of inventories. If we analyze the items in the Word Endings section in detail, the difficulty of accessing data related to morphological productivity becomes very apparent. The format of the questions that comprise the MBCTIs (original and adaptations), along with the fact that parents are reporting, are not conducive to gathering critical evidence of the presence or absence of productivity in children’s utterances. If children produce any token at all, even if it is only one of the morphemes they are asked about in the Word Endings section, parents will inevitably report it. However, this does not provide enough information to evaluate whether children are producing morphemes in a productive way. For example, a child might produce the third person plural of certain verbs, but their use could still be linked to specific contexts of use, and not yet be not generalized to other verb roots; his/her parents would nonetheless report it on the inventory. However, the occurrence of these subtle variations of the forms of the words it is informative. It could be pointing to an intermediate step in the development of grammar before a strict combinatorial level is achieved in development.

A tool such as the MBCTI has numerous advantages to evaluate early linguistic development, as many studies have shown (Bleses et al., 2008; Fenson et al., 1993; Fenson et al., 2007; Westerlund, Berglund, & Eriksson, 2006), but it is also important to bear in mind its limitations. Instead of tapping on morphological productivity, this instrument could be credited with detecting grammatical-relevant vocabulary (i.e., those morphological variations on the word forms).

Another question addressed in this study in which we were particularly interested, was whether or not children with different linguistic levels, especially those in the lowest percentiles, exhibit the same patterns of lexical and grammatical growth. Data presented in Table 8 depicted a weaker relationship between vocabulary and grammar for children who fall into the lowest percentile group (P0-10). However, additional results indicated these differences could be attributed to the P0-10 group exhibiting less variability. The fact that these differences disappeared when the sample was reorganized could indicate that the relationship between vocabulary and grammar in early language development is similar for all children, regardless of their linguistic differences or levels. Other researchers have offered contrasting data such as, for example, the findings of a study by Conti-Ramsden and Jones (1997) about SLI children (see Introduction). These contrasting results could be attributed to the differing methodologies used. In this respect, more detailed analysis focused on particular grammar structures or categories would be needed to identify possible differences in the development of the lexical-grammatical relationship.

The kind of design chosen for our study does not provide data about the individual developmental trajectories of children in the lowest percentiles (P0-10). It would be very interesting to analyze, for example, the vocabulary composition of children at or below the 10 percentile and counting on longitudinal data to study its evolution. Some researchers have found that ‘at risk’ children produce fewer verbs and closed-class words compared to control groups, although no differences have been found in their overall production of nouns and predicates (Koster et al., 2005).

By way of synthesis, this study provides further evidence for a robust relationship between vocabulary and grammar complexity (broadly measured) in Spanish, and suggests some precautions regarding use of the MBCTI to obtain critical evidence on morphological development. Given that vocabulary explained an important amount of variance in grammar complexity across the whole sample of children, our results favor a perspective on language acquisition that emphasizes the interdependency of lexical and morphosyntactic development, at least early on (from 16 to 24 months). This finding adds evidence to the theories of language acquisition that “view lexical and grammatical development within a unified system that shares important representational and computational resources” (Dixon and Marchman, 2007, p. 209). Besides, it has implications crucial to evaluation methods.

References


Received April 25, 2011
Revision received June 13, 2011
Accepted July 14, 2011