

Article



# Parents' reports of lexical and grammatical aspects of toddlers' language in European Portuguese: Developmental trends, age and gender differences

First Language  
2017, Vol. 37(3) 267–284  
© The Author(s) 2017  
Reprints and permissions:  
[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)  
DOI: 10.1177/0142723716689274  
[journals.sagepub.com/home/fla](http://journals.sagepub.com/home/fla)



**Carla Silva, Irene Cadime,  
Iolanda Ribeiro and Sandra Santos**

Universidade do Minho, Portugal

**Ana Lúcia Santos**

Universidade de Lisboa, Portugal

**Fernanda Leopoldina Viana**

Universidade do Minho, Portugal

## Abstract

The results from a large-scale study on toddlers' language acquisition in European Portuguese are presented. Toddlers' lexical and grammatical competencies were assessed using the MacArthur–Bates Communicative Development Inventory: Words and Sentences. The results, based on 3012 reports completed by parents, indicate an increase in the lexical size and on five measures of grammatical development (production of regular morphology, irregular morphology, over-regularizations, mean length of utterances and sentence complexity) across age groups. A main effect of gender was found for lexical size, production of regular and irregular morphology, production of over-regularizations and sentence complexity, with girls obtaining overall higher scores than boys. All lexical and grammatical development measures are positively correlated,

---

## Corresponding author:

Irene Cadime, Centro de Investigação em Estudos da Criança, Instituto de Educação, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal.

Email: [ireneCadime@ie.uminho.pt](mailto:ireneCadime@ie.uminho.pt)

even after controlling for age and gender effects. These findings are discussed in terms of their consistency with those obtained for other languages.

### **Keywords**

European Portuguese, gender differences, grammatical development, lexical development, MacArthur–Bates Communicative Development Inventory: Words and Sentences

Vocabulary growth is the main indicator of language acquisition in the early years, followed by grammatical development. In recent decades, the MacArthur–Bates Communicative Development Inventories (CDI), which are based on information given by parents (Fenson et al., 2007), have been used as assessment tools of communication and language development for children from ages 8 months through 30 months. These inventories, developed in the United States in the early 1990s, are the most widely used parent report instrument (Law & Roy, 2008; Simonsen, Kristoffersen, Bleses, Wehberg, & Jorgensen, 2014) and are primarily used as a research tool and as a clinical measure. They have already been translated and adapted for more than 50 different languages, spoken and signed (Dale & Penfold, 2011). The CDI has multiple applications to research such as studies about language developmental differences attributable to gender (Bauer, Goldfield, & Reznick, 2002; Reese & Read, 2000; Stennes, Burch, Sen, & Bauer, 2005), socioeconomic status and education (Arriaga, Fenson, Cronan, & Pethick, 1998; Hamilton, Plunkett, & Schafer, 2000; Hart & Risley, 1995; Locke, Ginsborg, & Peers, 2002), as well as linguistic skills in children with developmental disabilities and other syndromes (Caselli, Casadio, & Bates, 1999; Charman, Drew, Baird, & Baird, 2003; Luyster, Lopez, & Lord, 2007; Thal, Bates, Goodman, & Jahn-Samilo, 2009; Westerlund, Berglund, & Eriksson, 2006).

The original version of the CDI: Words and Sentences (CDI:WS) (Fenson et al., 1994, 2007) was developed to assess American English-speaking children between 16 months and 30 months. It can be used to obtain information on productive vocabulary and early grammatical development. The latter type of information is obtained by measuring the frequency of certain word combinations, early morphological development, mean length of utterances (assessed by applying a MLUw measure which takes into account the three longest sentences produced by children) and syntactic complexity. This inventory has two parts: the first focuses on lexical development and the use of language to refer to past and future events, and absent objects or people; the second analyses the production of selected regular and irregular morphology (such as the use of regular plural, possessives, progressive and past tense, the use of irregular plural nouns and irregular past tense verbs and the use of over-regularized nouns and verbs), as well as the production of multiword utterances and grammatical complexity.

In the reliability studies of the American CDI version two methodologies have been used: the calculation of internal consistency (Cronbach's alpha) and test-retest. Nonetheless, in most adaptation studies, only the first was used. Cronbach's alpha was computed for two scales – vocabulary production and sentence complexity – in the

American version of CDI, and very high coefficients were found for both (.96 in vocabulary production and .95 on sentences complexity). Similar results were found for other languages, with Cronbach's alpha ranging approximately between .97 and .99 for vocabulary production (Bleses et al., 2008; López-Ornat et al., 2005; Maital, Dromi, Sagi, & Bornstein, 2000; O'Toole & Fletcher, 2010; Pereira & Soto, 2003; Simonsen et al., 2014) and for sentence complexity (López-Ornat et al., 2005; O'Toole & Fletcher, 2010; Pereira & Soto, 2003).

Regarding reliability indicators based on stability, the test-retest correlations have ranged approximately between .74 and .99 for vocabulary production (Fenson et al., 1994, 2007; López-Ornat et al., 2005; Reese & Read, 2000; Simonsen et al., 2014) and approximately between .59 and .98 for sentence complexity (López-Ornat et al., 2005; Reese & Read, 2000).

The inter-correlations among the subscales in the American CDI are large (between .71 and .83), with the highest correlations (.82 and .83) obtained for the relationship between vocabulary production and the two grammar scales: mean length of the three longest utterances (M3L) and sentence complexity (Fenson et al., 1994, 2007). Slightly lower correlations (between .52 and .77) were obtained with the same subscales in the Hebrew version, with the highest correlations observed between sentence complexity and vocabulary production (.77), and between sentence complexity and M3L (.60) (Maital et al., 2000). The work developed for the Spanish version (López-Ornat et al., 2005) analysed the inter-correlations among all the subscales of CDI and obtained correlations between .43 and .92, with the highest correlations obtained for the relationship between word production and sentence complexity (.92) and between word production and word endings (.91). The lowest correlations were found for the relationship between vocalizations and over-regularized words (.43).

Research on the CDI:WS has also focused on the developmental trends, on the variation and the composition of early vocabulary and on the relationship between vocabulary and grammar. The results have shown similar developmental trends and range of variation across languages, with vocabulary and grammatical complexity increasing with age (Andonova, 2015; Bleses et al., 2008; Feldman et al., 2000; Fenson et al., 1994, 2007; López-Ornat et al., 2005; Reese & Read, 2000; Simonsen et al., 2014; Trudeau & Sutton, 2011); they equally show strong correlations between vocabulary and grammatical complexity, ranging between .82 in the American version (Fenson et al., 1994, 2007) and .92 in the Spanish version (López-Ornat et al., 2005).

The results concerning gender differences in the CDI components have been mixed. On the one hand, some studies suggest clear differences favouring girls. Eriksson et al. (2011) studied these differences in 10 non-English language communities and found that girls obtain better results than boys in productive vocabulary (use of more types of words) and that a higher percentage of girls in each age group (especially in the group between 19 and 27 months) were starting to combine words. Similar results were found in studies using the American CDI (Feldman et al., 2005), the New Zealand CDI (Reese & Read, 2000) and the Quebec-French CDI (Trudeau & Sutton, 2011). Other studies found that girls obtain better results also in the sentence complexity subscale (Bleses et al., 2008; Feldman et al., 2000; Fenson et al., 2007; Pereira & Soto, 2003; Simonsen et al., 2014; Trudeau & Sutton, 2011). On the other hand, some other studies did not show gender

differences (Maital et al., 2000), others showed only small differences favouring girls (Szagun, Steinbrink, Franik, & Stumper, 2006), and finally, others showed small differences in vocabulary production but not in grammar acquisition (Andonova, 2015).

In Portugal, the period from birth until 24 months is not covered by any assessment tool at the national level. The existing instruments only allow the assessment of children older than 24 months. In addition, they only assess subdomains of linguistic knowledge, thus not allowing a broader evaluation of the emerging language (Afonso, 2011).

The CDI is a versatile, efficient and valid measure of language development in young children (Law & Roy, 2008), and its adaptation and validation for European Portuguese (EP) proves to be of great importance as it fills a gap in the set of available tools for language assessment. Furthermore, and although there is a large number of CDIs in different languages, the field of child language research will benefit from the widest possible amount of evidence: the European Portuguese adaptation contributes to this field.

In this article, we present the results from a large-scale study on toddlers' (16–30 months) acquisition of European Portuguese. Their lexical and syntactic competencies were assessed using the European Portuguese adaptation of the CDI:WS. Therefore, this study was conducted to accomplish the following:

1. Document the adaptation of the CDI:WS to European Portuguese, attesting the reliability of the instrument;
2. Test the existence of age and gender effects on lexical development;
3. Assess the existence of age and gender effects on five measures of grammatical development: production of regular morphology, irregular morphology, over-regularizations, mean length of utterances and sentence complexity;
4. Identify the relationships between the five previous grammatical developmental measures, also after controlling for age and gender effects;
5. Identify the relationships between lexical development and the five previous grammatical developmental measures, also after controlling for age and gender effects.

## Method

### *Participants*

The sample was composed of 3012 participants between 16 and 30 months. We used a stratified sampling method. Strata were defined according to the distribution of the population in the different geographic areas of Portugal, using the statistical data from the 2011 Census, the largest national data source about population, family and housing. Table 1 describes the sample by area, gender, sibling status, day-care attendance and parents' education and presents the same type of information for the general population, if available.

For the sake of comparison, the distribution of the sample per area is close to the general population in Portugal. The final sample was balanced with respect to gender (52.4% boys and 47.3% girls), and the sibling status of the children matches what is observed in the population. The number of children per month of age ranges between 144 and 238,

**Table 1.** Demographic characteristics of the population in the Portuguese CDI study.

Characteristics of the population	Portuguese CDI study (age 16–30 months)		General population in Portugal
	N	%	%
<b>Area</b>			
North	1083	36.0	34.9
Centre	583	19.4	21.8
Lisbon	825	27.4	27.1
Alentejo	159	5.3	7.0
Algarve	195	6.5	4.3
Azores	76	2.5	2.4
Madeira	91	3.0	2.5
<b>Gender</b>			
Boys	1578	52.4	47.6
Girls	1424	47.3	52.4
No information	10	0.3	–
<b>Siblings status</b>			
Only child	1557	51.7	55.0
At least 1 sibling	1437	47.7	45.0
No information	18	0.6	–
<b>Day-care attendance</b>			
Yes	2875	95.4	–
No	57	1.9	–
No information	80	2.7	–
<b>Mother education</b>			
Grades 1 to 4	56	1.8	6.7
Grades 5 to 6	231	7.7	13.2
Grades 7 to 9	573	19.0	20.7
Grades 10 to 12	912	30.3	29.1
Higher education	1181	39.2	30.3
No information	59	2	–
<b>Father education</b>			
Grades 1 to 4	97	3.2	7.8
Grades 5 to 6	326	10.8	18.7
Grades 7 to 9	782	26.0	26.3
Grades 10 to 12	867	28.8	28.1
Higher education	768	25.5	19.1
No information	172	5.7	–

and the number of female and male participants is similar in each monthly stage,  $\chi^2(14) = 22.86, p = .06$ . The majority of children attended day-care centres (95.4%). About one-third of the parents completed high school (30.3% of the mothers and 28.8% of the fathers) and about another third has a college degree (39.2% of the mothers and 25.5% of the fathers).

Children born prematurely with low weight (less than 32 weeks of pregnancy and 1500 grams), children with neither parent who spoke European Portuguese and children with severe medical conditions that could result in language impairment (e.g. Down syndrome) were not included in the sample. Children with documented ear infections remained in the sample, regardless of the number of infections.

## **The Portuguese adaptation of the CDI (PT-CDI:WS)**

The present adaptation for European Portuguese was made from the second edition of the American version of the CDI (Fenson et al., 2007) but also taking into account the Spanish adaptation (López-Ornat et al., 2005), due to the cultural and linguistic proximity between Portugal and Spain.

The process started with the adaptation to European Portuguese of the content of the American and Spanish versions and the discussion about the inclusion or substitution of items. In this discussion, some linguistic and cultural factors and the opinion of experts in language development were considered. To define the set of words to be included in the word production checklist, we used a very recent lexicon of child speech – CEPLEXicon (Santos, Freitas, & Cardoso, 2014). CEPLEXicon (Santos et al., 2014) is a child lexicon based on two previous corpora (approximately 86 hours of child–adult interaction in total), which integrate samples of spontaneous speech of seven children (aged 1;02.00 to 3;11.12). The lexicon was obtained after automatic tagging, which comprised the lemmatization and morphosyntactic classification of the speech produced by the seven children included in the corpora. The lexicon contains information pertaining to lemmas and syntactic categories as well as the absolute number of occurrences and frequencies in three age intervals (< 2 years; ≥ 2 years and < 3 years; ≥ 3 years).

A pilot study was performed with a sample of 636 children between 16 and 30 months of age. On the basis of this pilot study, word frequencies were calculated for each item and the thinking aloud method was used to decide the relevant adjustments, both for the instructions and for the inclusion or substitution of items. The vocabulary list was the subscale which was subjected to more adjustments, with the exclusion of words selected in fewer than 10% of the questionnaires and the addition of words suggested by parents in at least 10% of the completed inventories. Words corresponding to the names of TV programmes, cartoon characters, brands and persons' first names were not added.

The PT-CDI:WS includes two main sections: (1) first words and (2) morphology and syntax. The first section is divided into two parts: (a) word production and (b) how the child uses and understands language. The second section is divided into five parts: (a) regular morphology, auxiliaries used to refer to future and ongoing events and copula verbs; (b) difficult verbs (irregular morphology); (c) over-regularizations and non-standard forms; (d) sentence length: MLU<sub>w</sub> (mean length of utterances measured in words) of the three longest sentences produced by children and reported by parents (from now on MLU<sub>w3</sub>); and (e) sentence complexity.

A note should be added concerning the adaptation of parts (a) and (e) in the second main section of the questionnaire. Part (a) in the original CDI:WS (Fenson et al., 2007) contains only four items assessing the child's performance in the use of regular morphology. The items include the regular plural of nouns, past verbal morphology or the

morphology associated with the present continuous (*-ing*). In EP, we refer to an ongoing event also using a periphrastic form (see (1)), but in this case this form combines an auxiliary (*estar a*) with the infinitive of the main verb. This is indeed a very frequent structure in EP and one of the first structures in which children use auxiliaries in EP, according to available descriptions based on case studies (see Santos, 2009, p. 167). The same descriptions show early production of periphrastic forms with the future auxiliary *ir* (see (2)). Although EP has a morphological future form, it is rarely used in informal speech, the periphrastic form being preferred.

- (1) O cão está a ladrar.  
the dog is PREP bark.INF  
'The dog is barking.'
- (2) O cão vai ladrar.  
the dog goes bark. INF  
'The dog will bark.'

Given the relevancy of assessing children's production of these type of forms, which allow reference to ongoing and future events, we added them to part (a) in the second section of the questionnaire. We also added an item concerning the use of *estar* 'be' as a copula verb, which can be used to form stative passives, as well as the use of the copula verb *ficar* 'get', associated with resultative passives. Therefore, the part (a) in the PT-CDI:WS contains 12 items and combines the assessment of regular morphology with the assessment of some syntactic structures referring to basic temporal or aspectual contrasts, which, in some cases, may correspond to morphological contrasts in other languages.

Part (e) in the PT-CDI:WS is more similar to the Spanish version of CDI:WS (López-Ornat et al., 2005) in the form of the section, but the scoring has more similarities with the American CDI:WS (Fenson et al., 2007). It contains 26 hypothetical situations with three different options of structures that could be produced in the context (the options increase in length and grammatical complexity), but only the third option (the more complex) is counted in scoring. Additionally, the option of stating that the child simply does not produce the structure is also provided.

Table 2 shows a comparison between the sections and the number of items of the three versions: American, Spanish and European Portuguese.

## Procedures

Data collection was preceded by the request of an authorization to the Portuguese National Committee of Data Protection; in addition, all parents who filled the PT-CDI:WS also signed an informed consent form, stating that they knew the purpose of the study and agreed to participate. Data collection was performed using two types of supports: (a) online data collection; and (b) data collection using a paper version of the questionnaire. In the first case, a webpage was created and parents were recruited through social networks and through email lists to fill in the online version. Only a very small percentage of the data were collected online (approximately 2%). In the second case, 250 institutions

**Table 2.** Comparison between the sections and number of items of the American, Spanish and European Portuguese versions of the CDI.

Sections	American CDI	Spanish CDI	European Portuguese CDI
1. Vocalizations	–	12	–
2. First words			
Starting to talk	–	2	–
Vocabulary development	–	1	–
Word production	680	588	639
Interjections, animal and object sounds	12	19	22
Animals	43	43	47
Vehicles	14	16	13
Toys	18	29	15
Food and drinks	68	59	63
Clothing	28	28	40
Body parts	27	24	34
Small household items	50	–	49
Furniture and rooms	33	59	26
Outside things	31	32	20
Places to go	22	–	18
People	29	22	27
Games, routines and greetings	25	31	36
Verbs	103	86	84
Descriptive words	63	35	47
Word about time	12	8	11
Demonstratives, possessives and personal pronouns	25	29	21
Question words	7	8	15
Prepositions and adverbs	26	24	21
Articles and quantifiers	17	24	20
Modal and auxiliary verbs	21	5	5
Connecting words	6	7	5
How children use words	5	4	4
3. Morphology and syntax			
Regular morphology, auxiliaries and copula verbs	4	12	15
Difficult verbs/words (irregular morphology)	25	19	32
Over-regularizations and non-standard forms	31	2	20
Sentences: MLU of the 3 longest utterances (MLUw3)	3	3	3
Sentence complexity	37	34	26

were invited to collaborate in the project, and a subset of these (200 institutions) agreed to participate. In this case, data collection was performed using one of two procedures: (a) inventories were delivered, filled and collected in predetermined meetings with the parents; or (b) paper copies of the inventories were sent to the institutions, were distributed to the children's parents by teachers and later collected by the same teachers.



## Results

### Reliability

Reliability was measured in terms of internal consistency, calculating Cronbach's alpha. Internal consistency was high for the subscales of word production (639 items,  $\alpha = .99$ ), regular morphology, auxiliaries and copula verbs (15 items,  $\alpha = .93$ ), irregular morphology (32 items,  $\alpha = .96$ ), over-regularizations (20 items,  $\alpha = .85$ ) and sentence complexity (26 items,  $\alpha = .96$ ).

### Lexical development

A multiple linear regression analysis was performed, taking age and gender as predictors of the lexical development of the toddlers. The interaction term between both predictors was also tested. The model was statistically significant,  $F(3, 2998) = 841.902$ ,  $p < .001$ , and explained approximately 46% of the variance observed in lexical development ( $R^2 = .457$ ; Adjusted  $R^2 = .457$ ). Age and gender were both significant predictors ( $p < .001$ ). The regression results also indicated that age made the largest contribution in the model,  $\beta = .82$ ,  $sr^2 = .065$ . Gender made only a small contribution to the observed variance in lexical development,  $\beta = -.121$ ,  $sr^2 = .014$ , with girls performing better than boys. The results also indicated a significant interaction of age  $\times$  gender,  $\beta = -.163$ ,  $sr^2 = .002$ ,  $p < .001$ . As seen in Figure 1, the gender gap occurs mainly at more advanced ages.

### Grammatical development

The acquisition of European Portuguese grammar by toddlers was assessed taking into account the following scores: (1) regular morphology, auxiliaries and copula verbs, (2) irregular morphology, (3) over-regularizations, (4) MLUw3 and (5) sentence complexity. The following analysis will focus on the differences in gender and age for the different grammatical development measures.

**Regular morphology, auxiliaries and copula verbs.** A multiple linear regression analysis was performed, taking age and gender as predictors of the use of regular morphology, auxiliaries and copula verbs. The interaction term between both predictors was also tested. The model was statistically significant,  $F(3, 2240) = 627.951$ ,  $p < .001$ , and explained approximately 46% of the variance observed in the use of regular morphology, auxiliaries and copula verbs ( $R^2 = .457$ ; Adjusted  $R^2 = .456$ ). Age and gender were significant predictors (age and gender:  $p < .001$ ). Age was the main predictor,  $\beta = .81$ ,  $sr^2 = .062$ , followed by gender,  $\beta = -.126$ ,  $sr^2 = .015$ . Once again girls had higher scores than boys. The interaction term age  $\times$  gender was significant,  $\beta = -.151$ ,  $sr^2 = .002$ ,  $p < .01$ . Once again, the plot of the interaction effect indicates that the differences between gender increase with age (see Figure 2).

**Irregular morphology.** Age and gender as predictors of irregular morphology were analysed using a multiple linear regression analysis. We also tested the interaction term

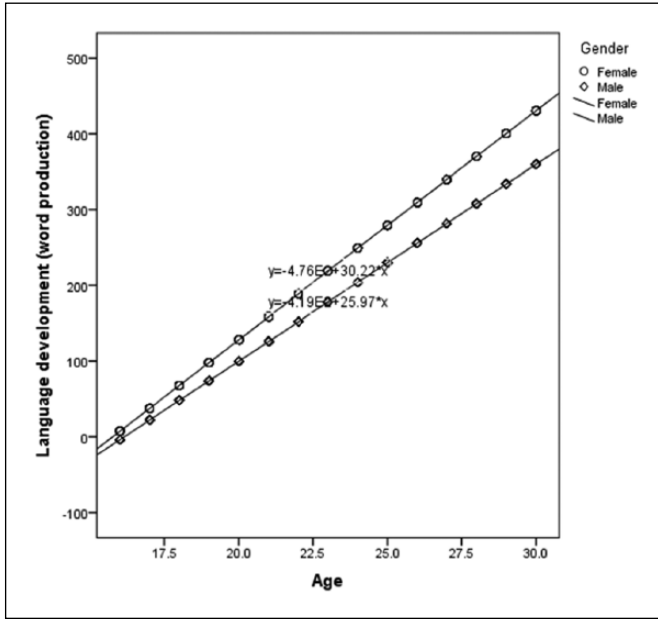


Figure 1. Interaction between gender and age on lexical development (word production).

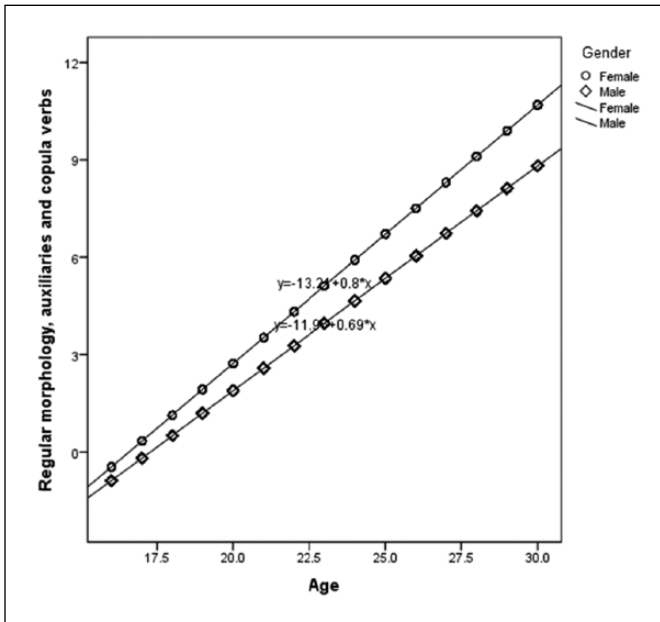


Figure 2. Interaction between gender and age on regular morphology, auxiliaries and copula verbs.

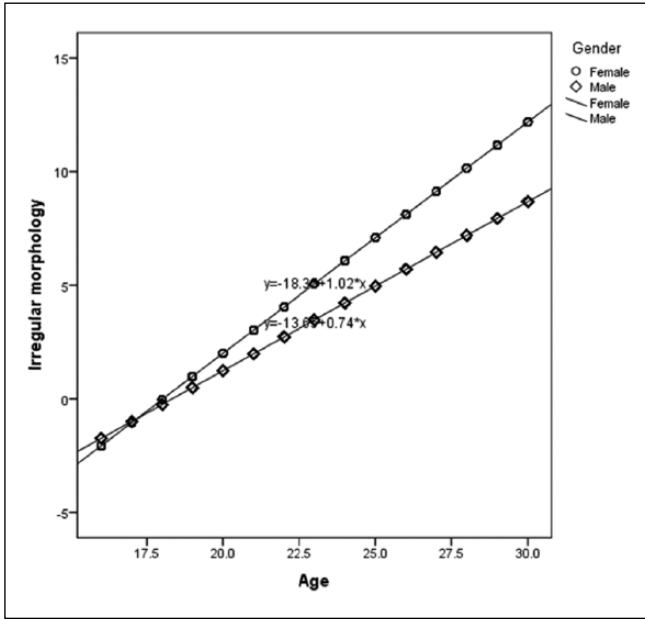
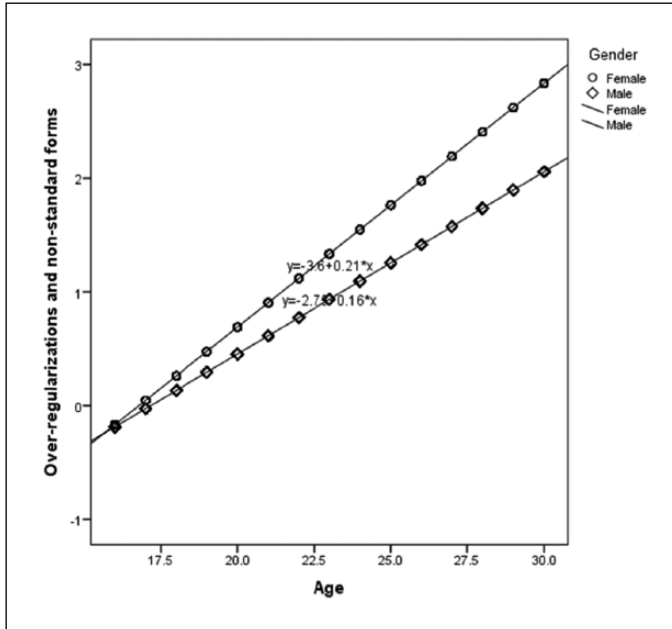


Figure 3. Interaction between gender and age on irregular morphology.

between both predictors. The model was statistically significant,  $F(3, 2998) = 370.380$ ,  $p < .001$ , and explained approximately 27% of the variance observed ( $R^2 = .270$ ; Adjusted  $R^2 = .270$ ). Age and gender were significant predictors ( $p < .001$ ). The largest contribution in the model was made by age,  $\beta = .74$ ,  $sr^2 = .053$ . Gender contribution was smaller,  $\beta = -.115$ ,  $sr^2 = .013$ , with girls having higher results than boys. An interaction effect of age  $\times$  gender was also found,  $\beta = -.253$ ,  $sr^2 = .006$ ,  $p < .001$ , with the differences between gender again increasing with age, as shown by Figure 3.

*Over-regularizations and non-standard forms.* Using a multiple linear regression analysis, age and gender, as well as an interaction between them, were analysed as predictors of the use of over-regularizations and non-standard forms. The model was statistically significant,  $F(3, 2998) = 135.200$ ,  $p < .001$ , and explained approximately 12% of the variance observed ( $R^2 = .119$ ; Adjusted  $R^2 = .118$ ). Age and gender were significant predictors ( $p < .001$ ). Age made the largest contribution in the model,  $\beta = .48$ ,  $sr^2 = .022$ . Gender had only a small contribution to the observed variance in the use of over-regularizations and non-standard forms,  $\beta = -.089$ ,  $sr^2 = .007$ , with girls once again having higher scores than boys. The results also indicated a significant interaction effect of age  $\times$  gender,  $\beta = -.156$ ,  $sr^2 = .002$ ,  $p < .01$ . The gender gap appears at more advanced ages, as seen in Figure 4.

*MLUw3.* A multiple linear regression analysis was performed to analyse age and gender as predictors of the use of MLUw3. The interaction term between both predictors was



**Figure 4.** Interaction between gender and age on over-regularizations and non-standard forms.

also tested. The model was statistically significant,  $F(3,2690) = 420.164$ ,  $p < .001$ , and explained approximately 32% of the variance observed in the use of MLUw3 ( $R^2 = .319$ ; Adjusted  $R^2 = .318$ ). Age was a significant ( $p < .001$ ), and the main predictor,  $\beta = .679$ ,  $sr^2 = .044$ . Gender was also significant ( $p < .001$ ) and gives a smaller contribution to the observed variance,  $\beta = -.081$ ,  $sr^2 = .006$ , again with girls having higher scores than boys. The interaction term age  $\times$  gender was also significant ( $p = .013$ ,  $\beta = -.126$ ,  $sr^2 = .001$ ). Similarly to the other linguistic dimensions, the gender differences are not evident in the first months of age (see Figure 5).

**Sentence complexity.** To analyse age and gender as predictors of sentence complexity, we once more performed a multiple linear regression analysis; the interaction term between both predictors was also tested. The model was statistically significant,  $F(3, 2119) = 378.320$ ,  $p < .001$ , and explained approximately 35% of the variance observed in the use of sentence complexity ( $R^2 = .349$ ; Adjusted  $R^2 = .348$ ). Age and gender were significant predictors ( $p < .001$ ). The regression results indicated that age is the main predictor,  $\beta = .919$ ,  $sr^2 = .081$ . Gender contribution to the observed variance was smaller,  $\beta = -.108$ ,  $sr^2 = .011$ , with girls having better results than boys. The interaction term age  $\times$  gender was significant,  $\beta = -.368$ ,  $sr^2 = .012$ , with the plot of the interaction effect indicating that the differences between gender increase with age, as seen in Figure 6.

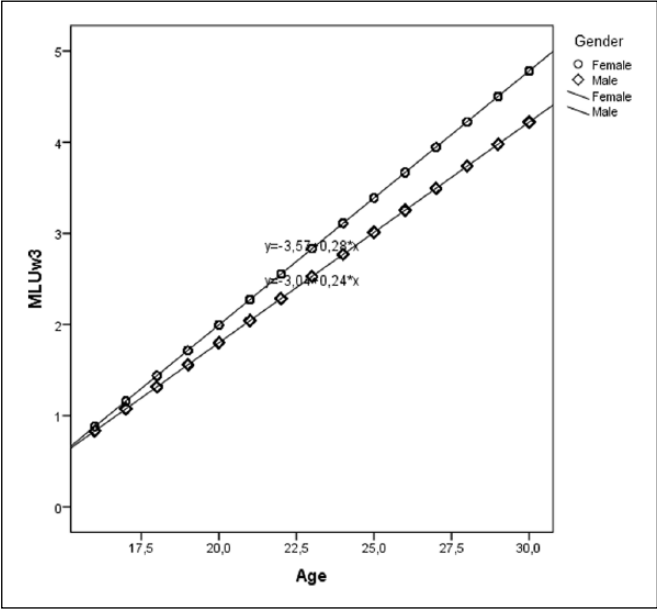


Figure 5. Interaction between gender and age on MLUw3.

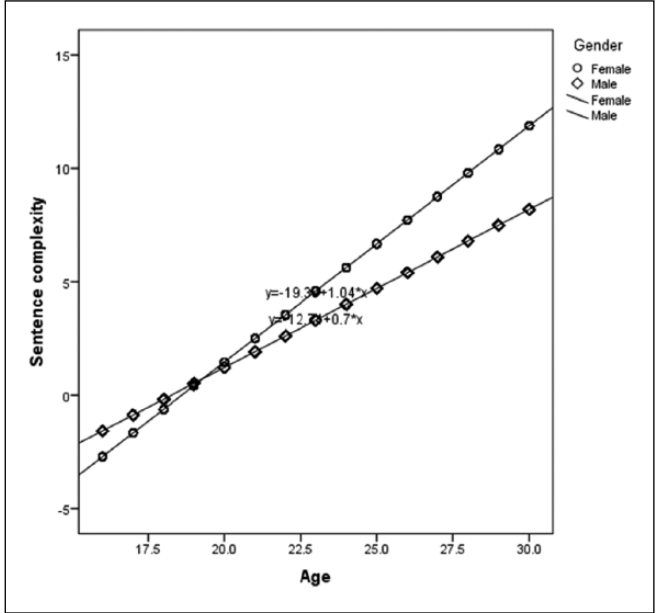


Figure 6. Interaction between gender and age on sentence complexity.

**Table 3.** Correlations between grammar measures by gender.

	Regular morphology	Irregular morphology	Over-regularizations	MLUw3	Sentence complexity
Regular morphology	–	.764**	.485**	.766**	.771**
Irregular morphology	.783**	–	.559**	.649**	.779**
Over-regularizations	.512**	.533**	–	.362**	.404**
MLUw3	.692**	.602**	.380**	–	.736**
Sentence complexity	.777**	.745**	.383**	.724**	–

Note: Coefficients below the diagonal are for girls and above the diagonal are for boys.

\*\* $p < .001$ .

**Table 4.** Partial correlations between lexical and grammar development measures after controlling for gender and age.

	Word production	Regular morphology	Irregular morphology	Over-regularizations	MLUw3	Sentence complexity
Word production	–	.790**	.681**	.376**	.587**	.652**
Regular morphology		–	.678**	.402**	.576**	.639**
Irregular morphology			–	.463**	.473**	.665**
Over-regularizations				–	.237**	.260**
MLUw3					–	.601**
Sentence complexity						–

\*\* $p < .001$ .

### *Relation between the grammatical development measures*

Table 3 shows the correlations between grammar measures considered for boys and girls. These variables were significantly and positively correlated, with coefficients ranging from .362 to .783.

The analyses of partial correlations between grammatical development measures (after controlling for gender and age) reveal that variables were significantly positively inter-correlated, with correlation coefficients ranging from .237 to .678, as shown in Table 4. All correlations were significant at  $p < .001$ .

### *Relation between lexical and grammatical development*

This relationship was examined in the European Portuguese sample in terms of vocabulary size and the following different grammatical development measures: regular morphology (we are simplifying the reference to part (a) of section 2, i.e. ‘regular morphology,

auxiliaries and copula verbs'), irregular morphology, over-regularizations, MLUw3 and sentence complexity, each for girls and boys. For girls, moderate to high correlations were found between word production and regular morphology,  $r = .88$ , irregular morphology,  $r = .78$ , over-regularizations,  $r = .50$ , MLUw3,  $r = .72$ , and sentence complexity,  $r = .80$ . For boys, moderate to high correlations were also found between vocabulary size and regular morphology,  $r = .89$ , irregular morphology,  $r = .77$ , over-regularizations,  $r = .47$ , MLUw3,  $r = .76$ , and sentence complexity,  $r = .77$ . All correlations were significant at  $p < .001$ . The partial correlations between word production and all grammatical development measures (after controlling for gender and age) indicate that variables were significantly positively inter-correlated, with correlation coefficients ranging from .376 to .790, as shown in Table 4.

## Discussion

The data analysis of a large representative sample on the development of lexical and grammatical aspects of toddlers' production in European Portuguese is presented in this article. This study aimed at documenting the adaptation of the CDI:WS to European Portuguese, attesting to its reliability. This study also aimed to assess the existence of age and gender effects on lexical development and on five measures of grammatical development; and finally, the study aimed to determine correlations between lexical and grammatical development and between the five grammatical development measures.

All of the results obtained in the reliability study show high internal consistency of the PT-CDI:WS. The Cronbach's alpha obtained in word production and in sentence complexity are comparable to those obtained in other CDI versions (Bleses et al., 2008; Fenson et al., 1994, 2007; López-Ornat et al., 2005; Maital et al., 2000; O'Toole & Fletcher, 2010; Pereira & Soto, 2003; Simonsen et al., 2014).

The results of this study indicate an age effect in the lexical development and in all grammatical development measures. Vocabulary increases with age and the present version of the PT-CDI:WS is sensitive and captures this developmental effect. The trajectories of Portuguese children's lexical and grammatical development are similar to those observed in many other studies for other languages using the CDI (Andonova, 2015; Fenson et al., 1994, 2007; López-Ornat et al., 2005; Pereira & Soto, 2003; Trudeau & Sutton, 2011).

Contrary to the findings obtained with the Hebrew version of the CDI (Maital et al., 2000), which show no gender differences, the results obtained with PT-CDI:WS show small gender differences, with girls having better results than boys in all lexical and grammatical measures. These findings are consistent with the results obtained for other CDI versions, such as the Quebec-French version (Trudeau & Sutton, 2011), which show results slightly higher for girls on productive vocabulary; they are also consistent with the American (Fenson et al., 2007), the Spanish (López-Ornat et al., 2005) and the German (Szagun et al., 2006) versions, which show results slightly higher for girls also in productive vocabulary and sentence complexity.

The interaction effect between age and gender found in all lexical and grammatical measures, with gender differences increasing with age, are also consistent with the

results obtained for other CDI versions (Bleses et al., 2008; Fenson et al., 2007; Simonsen et al., 2014).

The correlations between lexical and grammatical measures are all moderate to high, showing a strong lexical-grammatical continuity in early acquisition (Marchman, Martínez-Sussmann, & Dale, 2004), with the highest correlations being identified between word production and regular morphology, irregular morphology and sentence complexity. The correlations across the five grammatical development measures are also moderate to high, being especially high among sentence complexity and the production of regular morphology and irregular morphology, either for boys or girls, and revealing a common set of related grammatical knowledge (Andonova, 2015). The lowest correlations were found between over-regularizations and MLUw3. These results are consistent with the ones found in the English (Fenson et al., 1994, 2007), Spanish (López-Ornat et al., 2005) and Hebrew (Maital et al., 2000) CDI versions.

The findings described here are also the first resulting from the adaptation of the MacArthur–Bates Inventory for European Portuguese. The PT-CDI:WS shows sensitivity to language development in Portuguese children, and it shows a good linguistic and cultural adjustment, which makes it useful in the assessment of language development of European Portuguese speaking toddlers. The richness of the data and the quantity of information that can be obtained with this inventory will certainly contribute to a more profound knowledge of language acquisition and development and will be extremely useful in the identification of delays in language development.

In this study, apart from age and gender, no other demographic factor was analysed. Maternal education and birth order are some of the demographic factors which other studies (Fenson et al., 1994, 2007; Pereira & Soto, 2003) identified as having an effect in lexical and grammatical development. These factors should be analysed in future studies with the PT-CDI:WS.

The adopted design does not allow the study of individual trajectories in language development. Longitudinal studies with the PT-CDI:WS, to be developed in the future, will certainly enhance our understanding of some linguistic developmental patterns.

One of the advantages of the MacArthur–Bates Inventories is the fact that they have been adapted to a very large number of languages and are used as an assessment tool in many different countries, which makes it a good basis for cross-linguistic comparisons. The assessment of linguistic development is a very complex process, which needs more than one tool. The PT-CDI:WS can be a useful tool to assess acquisition of European Portuguese language, allowing us to fill the gap of the lack of robust measures validated for very young children acquiring European Portuguese and, at the same time, producing data that can be compared with the data obtained with similar instruments for other languages.

### **Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by European Regional Development



Fund (FEDER) through the European programme COMPETE (Operational Programme for Competitiveness Factors) under the National Strategic Reference Framework (QREN) - FCOMP-01-0124-FEDER-029556 and from FCT (Fundação para a Ciência e a Tecnologia PTDC/MHC-PED/4725/2012). The first and second authors are also supported by grants SFRH/BD/86795/2012 and SFRH/BPD/102549/2014 from FCT.

## References

- Afonso, M. A. (2011). *Elementos para a construção de um teste de rastreio de perturbações do desenvolvimento sintático* [Elements for the construction of a screening test of the syntactic development disorders] (Master's thesis). Universidade Nova de Lisboa, Portugal.
- Andonova, E. (2015). Parental report evidence for toddlers' grammar and vocabulary in Bulgarian. *First Language, 35*, 126–136.
- Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. J. (1998). Scores on the MacArthur Communicative Development Inventory of children from low and middle-income families. *Applied Psycholinguistics, 19*, 209–223.
- Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics, 23*, 313–335.
- Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). The Danish Communicative Developmental Inventories: Validity and main developmental trends. *Journal of Child Language, 35*, 651–669.
- Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*, 69–111.
- Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*, 213–236.
- Dale, P., & Penfold, M. (2011). *Adaptations of the MacArthur–Bates CDI into Non-U.S. English Languages*. Retrieved from <http://mb-cdi.stanford.edu/documents/AdaptationsSurvey7-5-11Web.pdf>
- Eriksson, M., Marschik, P., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., & Gallego, C. (2011). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology, 30*, 326–343.
- Feldman, H., Dale, P., Campbell, T., Colborn, D., Kurs-Lasky, M., Rockette, H., & Paradise, J. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*, 856–868.
- Feldman, H., Dollaghan, C., Campbell, T., Kurs-Lasky, M., Janosky, J., & Paradise, J. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development, 71*, 310–322. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10834466>
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242), 1–176.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories user's guide and technical manual*. Baltimore, MD: Paul H. Brookes.
- Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language, 27*, 689–705.
- Hart, B., & Risley, R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

- Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health, 13*, 198–206.
- Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language & Communication Disorders, 37*(1), 3–15.
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). *Inventario de Desarrollo Comunicativo MacArthur*. Madrid, Spain: TEA Ediciones, S.A.
- Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur–Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*, 623–654.
- Maital, S. L., Dromi, E., Sagi, A., & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language, 27*, 43–67.
- Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science, 7*, 212–224.
- O’Toole, C., & Fletcher, P. (2010). Validity of a parent report instrument for Irish-speaking toddlers. *First Language, 30*, 199–217.
- Pereira, M. P., & Soto, X. R. G. (2003). El diagnóstico del desarrollo comunicativo en la primera infancia: Adaptación de las escalas MacArthur al gallego. *Psicothema, 15*, 352–361.
- Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and Sentences. *Journal of Child Language, 27*, 255–266.
- Santos, A. L. (2009). *Minimal answers: Ellipsis, syntax and discourse in the acquisition of European Portuguese*. Amsterdam, The Netherlands: John Benjamins.
- Santos, A. L., Freitas, M. J., & Cardoso, A. (2014). *CEPLEXicon – A Lexicon of Child European Portuguese*. Lisboa, Portugal: Anagrama (Faculdade de Letras da Universidade de Lisboa).
- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jorgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language, 34*, 3–23.
- Stennes, L. M., Burch, M. M., Sen, M. G., & Bauer, P. J. (2005). A longitudinal study of gendered vocabulary and communicative action in young children. *Developmental Psychology, 41*, 75–88.
- Szagan, G., Steinbrink, C., Franik, M., & Stumper, B. (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language, 26*, 259–280.
- Thal, D. J., Bates, E., Goodman, J., & Jahn-Samilo, J. (2009). Continuity of language abilities: An exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology, 13*, 239–273.
- Trudeau, N., & Sutton, A. (2011). Expressive vocabulary and early grammar of 16- to 30-month-old children acquiring Quebec French. *First Language, 31*, 480–507.
- Westerlund, M., Berglund, E., & Eriksson, M. M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur–Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research, 49*, 237–247.