This is an Accepted Manuscript of an article published by Taylor & Francis in Language, *Learning and Development* on 09/03/2018, available online at: https://doi.org/10.1080/15475441.2018.1427589

The Genetic and Environmental Etiology of the Association Between Vocabulary and Syntax in First Grade

Catherine Mimeau
Dalhousie University, Halifax, Nova Scotia, Canada
Ginette Dionne
Université Laval, Quebec City, Quebec, Canada
Bei Feng
Université Laval, Quebec City, Quebec, Canada
Mara Brendgen
Université du Québec à Montréal, Montreal, Quebec, Canada
Frank Vitaro
Université de Montréal, Montreal, Quebec, Canada
Richard E. Tremblay
University College Dublin, Dublin, Ireland
Université de Montréal, Montreal, Quebec, Canada
Michel Boivin
Université Laval, Quebec City, Quebec, Canada
Tomsk State University, Tomsk, Russian Federation

Author Note

Catherine Mimeau, Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia, Canada; Ginette Dionne, École de psychologie, Université Laval, Quebec City, Quebec, Canada; Bei Feng, École de psychologie, Université Laval, Quebec City, Quebec, Canada; Mara Brendgen, Département de psychologie, Université du Québec à Montréal, Montreal, Quebec, Canada; Frank Vitaro, École de psychoéducation, Université de Montréal, Montreal, Quebec, Canada; Richard E. Tremblay, School of Public Health, Physiotherapy and Population Science, University College Dublin, Dublin, Ireland; Département de pédiatrie and Département de psychologie, Université de Montréal, Montreal, Quebec, Canada; Michel Boivin, École de psychologie, Université Laval, Quebec City, Quebec, Canada; Institute of Genetic, Neurobiological and Social Foundations of Child Development, Tomsk State University, Tomsk, Russian Federation.

We are grateful to the children who participated in this study. We also thank Marie-Josée Côte, Carol-Ann Thouin, Laura Thivierge, Sophie Blais-Michaud, and Myriam Gagnon-Couture for their contribution to transcription and codification. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Canadian Institutes of

Health Research, the Canadian Language and Literacy Research Network, the Canada Research Chair Program, the Canadian Institute for Advanced Research, the National Health Research and Development Program, and the Fonds de recherche du Québec.

Correspondence concerning this article should be addressed to Ginette Dionne, École de psychologie, Université Laval, Pavillon Félix-Antoine-Savard, 2325 rue des Bibliothèques, Quebec City, Quebec, G1V 0A6, Canada. E-mail: ginette.dionne@psy.ulaval.ca

Abstract

This study examined the genetic and environmental etiology of vocabulary, syntax, and their association in first graders. French-speaking same-sex twins (N = 555) completed two vocabulary tests, and two scores of syntax were calculated from their spontaneous speech at 7 years of age. Multivariate latent factor genetic analyses showed that lexical skills were influenced mainly by the environment shared between the twins, whereas syntactic skills were influenced exclusively by genes and unique environment. Moreover, the moderate association between vocabulary and syntax was mostly due to common genetic factors. These novel findings may be attributable to the use of latent factors and the population studied. More research is needed to determine the specific factors involved in lexical and syntactic skills at this developmental period.

Learning language is a multi-component task for children. Among other things, they need to acquire vocabulary – make the correspondence between sounds and their meaning – in order to understand and produce intelligible words. They also need to acquire syntax – make the correspondence between the position of words and their function – in order to understand and produce intelligible sentences. As the mechanisms through which children acquire these lexical and syntactical skills are not yet completely understood, researchers still debate whether they stem mainly from genetic processes, environmental sources, or a combination of both. It also remains unclear whether the factors that underlie vocabulary and syntax are common or distinct, in other words, what is the nature of the association between these two components of language. Answering these questions is fundamental to build accurate theories of language development and to successfully help children who struggle to develop one or more language skills adequately. The challenge is heightened by the wide variety of languages spoken around the world and by the ongoing development of language skills throughout the lifespan. The objective of this study is thus to examine the relative contribution of genes and environment to lexical and syntactic skills, as well as to the association between the two components of language, in a population not previously studied on this topic: French-speaking first graders.

The Study of Vocabulary and Syntax

The developmental sequence of lexical and syntactic skills has been vastly studied during the last decades. A first major finding from this research is that of the great variability in those linguistic skills among children at a given age (Fenson et al., 1994; Siegler, 1996). This variability can be explained by different factors. For instance, some researchers have discovered specific genes (e.g., *KIAA0319*) that are related to general language ability (Newbury et al., 2011). Others have also shown that language is associated with cognitive factors known to have genetic origins (Bearden et al., 2012; Friedman et al., 2008; Wright et al., 2001). As an example, Lum, Conti-Ramsden, Page, and Ullman (2012) found that school-aged children's linguistic skills were correlated with their declarative and procedural memory. Similarly, vocabulary was shown to be associated with executive functions in school-aged children (Joseph, McGrath, & Tager-Flusberg, 2005) and with speed of recognition in toddlers (Fernald, Perfors, & Marchman, 2006).

In parallel with these genetic explanations to variability in language, some environmental factors have been examined. For example, in the first years of life, children's linguistic skills were found to be related to socio-economic status and parenting (Pungello, Iruka, Dotterer, Mills-Koonce, & Reznick, 2009), quality of formal instruction (Burchinal et al., 2008), and peers' linguistic skills (Mashburn, Justice, Downer, & Pianta, 2009).

A second major finding from the research on the development of vocabulary and syntax has been that these skills are associated throughout childhood. Indeed, researchers have found phenotypic correlations between these two components of language ranging from .40 to .82 in toddlers (Dale, Dionne, Eley, & Plomin, 2000; Dionne, Dale, Boivin, & Plomin, 2003), preschoolers (Hayiou-Thomas et al., 2006), and younger (DeThorne, Harlaar, Petrill, & Deater-Deckard, 2012) and older (Dale, Harlaar, Hayiou-Thomas, & Plomin, 2010) school-aged children. In line with these findings, some researchers have suggested that the same factors underlie vocabulary and syntax processing. For example, MacWhinney (1987) proposed a competition model whereby all components of language are governed by a single mechanism: competition between cues. Indeed, children use cues such as perceptual attributes to name objects, just as they use cues such as word order to identify the agent of verbs. Other theories such as bootstrapping (Gleitman, 1990; Pinker, 1984) also stipulate a close relation between the different components of language. Indeed, according to this hypothesis, children rely on their skills in one component to develop another one. As an example, knowing the syntactic properties of a word could help understand its meaning (syntactic bootstrapping), and vice versa (semantic bootstrapping).

In contrast, other researchers believe that vocabulary and syntax are processed by different mechanisms. For example, Ullman's (2004) declarative/procedural model claims that children access the pronunciation and the meaning of words through declarative memory, which is responsible for the explicit memorization of facts and events. By contrast, they would compute the regularities of language (e.g., syntactic rules) through procedural memory, which is responsible for the implicit memorization of sequences and procedures.

The Twin Method

To quantify the relative contribution of genetic and environmental factors to a given skill, researchers often compare samples of monozygotic (MZ) and dizygotic (DZ) twins. MZ (or identical) twins share 100% of their genes, whereas DZ (or non-identical) twins share on average 50% of their genes, as do non-twin siblings. Furthermore, both MZ and DZ twins share a portion of their environment with their co-twin that fosters similarities between them (shared environment). For example, socio-economic status, parents' language skills and parenting style, and reading habits at home are all likely to influence both twins of a pair similarly. However, twins also have unique experiences that make them different (unique environment). In that sense, friends' language skills and teacher's teaching style, for instance, are likely to influence both twins of a pair differently, given that the twins have different friends and teachers.

The logic behind twin studies is that the only difference between MZ and DZ twins is the proportion of genetic similarity between the twins of a pair. Therefore, the extent to which MZ twins are more similar than DZ twins on a given trait can be logically attributed to genetic factors

(A)[1]. By contrast, the extent to which both twins of a pair are similar on a given trait, regardless of whether they are MZ or DZ twins, can be attributed to shared environmental factors (C), considering that these factors have a similar influence on both types of twins. Finally, the remaining variance can be attributed to unique environmental factors and error (E), which make both MZ and DZ twins different. It should be noted that since these three types of factors represent proportions of variance, they add up to a total of 1 when standardised.

In addition to informing as to the sources of individual differences in a given trait, twin studies also allow uncovering the sources of the association between two traits. Following the reasoning just stated, the extent to which the association between one trait in one twin and the other trait in the other twin is greater in MZ than DZ twins indicates the relative contribution of genetic factors to the association (i.e., bivariate heritability). By contrast, the extent to which one trait in one twin is associated with the other trait in the other twin, regardless of whether the twins are MZ or DZ, indicates the relative contribution of shared environmental factors to the association. Finally, the remaining covariance between the two traits can be explained by unique environmental factors and error. As the three types of factors represent proportions of covariance, they add up to a total of 1 when standardised, as for a single trait.

Furthermore, twin studies also allow calculating genetic (*r*A), shared environmental (*r*C), and unique environmental (*r*E) correlations between two traits. Unlike the relative contribution of genes, shared environment, and unique environment to a given trait or to the association between two traits, these correlations are not cumulative. A strong correlation indicates that the specific factors influencing one trait are largely the same as those influencing the other trait. For instance,

[1] There are two types of genetic factors, additive (A) and dominant (D), but only additive factors were considered in the present study (see Note 2).

a strong genetic correlation between two traits would mean that most of the genes influencing the two traits are the same.

Previous Twin Studies With Toddlers

In the early 2000s, researchers used parental questionnaires to study the etiology of the productive linguistic skills of British toddlers (2-3-year-olds). They observed that both vocabulary and syntax were influenced mainly by shared environmental factors ($c\mathrm{s}^2$ = .46-.84). However, when they compared the two components of language, distinct patterns emerged: Vocabulary was more influenced by shared environment than syntax ($c\mathrm{s}^2$ = .69-.84, for vocabulary; $c\mathrm{s}^2$ = .46-.56, for syntax), and syntax was more influenced by genes ($a\mathrm{s}^2$ = .10-.25, for vocabulary; $a\mathrm{s}^2$ = .29-.42, for syntax) and unique environment ($e\mathrm{s}^2$ = .03-.07, for vocabulary; $e\mathrm{s}^2$ = .12-.19, for syntax) than vocabulary (Dale et al., 2000; Dionne et al., 2003).

Moreover, Dale et al. (2000) showed that the phenotypic covariance (r = .66) between vocabulary and syntax at age 2 years could be explained mainly by shared environmental factors (c^2 = .69). In other words, if toddlers' lexical and syntactic skills are strongly associated, it is likely due to exposure to environmental influences that make children of a same family more similar, such as socio-economic status.

Furthermore, although genetic factors were found to influence vocabulary, syntax, and their association to a lesser extent than shared environmental factors, researchers discovered that the two components of language mostly implicated the same genes ($r\mathrm{s}_\mathrm{A}$ = .61-.89). They also observed that the shared environmental factors underlying toddlers' lexical and syntactic skills greatly overlapped ($r\mathrm{s}_\mathrm{C}$ = .54-.78). However, unique environmental sources of influence were shown to be distinct for vocabulary and syntax ($r\mathrm{s}_\mathrm{E}$ = .07-.25; Dale et al., 2000; Dionne et al., 2003).

Previous Twin Studies With Older Children

Studies conducted with preschoolers (4-5-year-olds) led to somewhat different conclusions than studies conducted with toddlers. Taken together, they suggest that children's lexical and syntactic skills, as assessed directly with receptive and productive tasks, are influenced by genetic, shared environmental, and unique environmental factors equally (as^2 = .29-.53, cs^2 = .09-.60, es^2 = .08-.50). However, results vary greatly across studies and across measures: Whereas Kovas et al. (2005), who studied British children using numerous individual measures, found low contributions of shared environment for both vocabulary (cs^2 = .09-.13) and syntax (cs^2 = .21-.26), Samuelsson et al. (2005), who studied American, Australian, and Scandinavian children using latent factors, found high contributions of shared environment for both vocabulary (c^2 = .60) and syntax (c^2 = .59).

Turning to the association between the two components of language, no study has yet examined, to our knowledge, the relative contribution of genes, shared environment, and unique environment to this association in preschoolers. Furthermore, the only set of genetic correlations available in the literature for this age group indicated a substantial overlap between the genetic factors involved in vocabulary and syntax in British children (rs_A = .39-.86; Hayiou-Thomas et al., 2006), but no shared nor unique environmental correlations were reported.

To our knowledge, only two twin studies have examined the etiology of lexical and syntactic skills among school-aged children. First, DeThorne et al. (2012; see also DeThorne et al., 2008) assessed the productive vocabulary and the productive syntax of American 7- and 8-year-olds using three measures computed from spontaneous speech (one for vocabulary and two for syntax). The authors showed that both lexical and syntactic skill were mostly influenced by genes and unique environment (as^2 = .24-.55, cs^2 = .00-.06, es^2 = .45-.71, for vocabulary; as^2 = .08-.53, cs^2 = .00-.30, es^2 = .39-.64, for syntax), although the estimates varied across measures and ages. However, they did not investigate the underlying association between the two components of language.

Second, Dale et al. (2010) assessed the receptive vocabulary and the receptive syntax of British 12-year-olds using two web-based measures (one for vocabulary and one for syntax). Like DeThorne et al. (2012), they found that individual differences in both components of language were accounted for mainly by genetic and unique environmental factors (a^2 = .30, c^2 = .13, e^2 = .58, for vocabulary; a^2 = .30, c^2 = .15, e^2 = .54, for syntax). Moreover, while they did not report the relative contribution of genes, shared environment, and unique environment to the association between lexical and syntactic skills, they observed strong genetic (r_A = .71) and shared environmental (r_C = .86) correlations between the two skills, suggesting common influential factors, but no significant unique environmental correlation (r_E = .12).

The Present Study

In sum, the etiology of lexical and syntactic skills appears to shift from mainly shared environmental influences in the first years of life (Dale et al., 2000; Dionne et al., 2003) to genetic and unique environmental influences later in development (Dale et al., 2010; DeThorne et al., 2012). In addition, whereas vocabulary and syntax show somewhat different etiological patterns in toddlerhood, with vocabulary tending to be driven more by shared environment, and syntax, more by genes and unique environment (Dale et al., 2000; Dionne et al., 2003), the influences on the two components of language seem to be similar during the preschool and school years (Dale et al., 2010; DeThorne et al., 2012; Kovas et al., 2005; Samuelsson et al., 2005). Furthermore, the phenotypic covariance between lexical and syntactic skills was found to be explained mainly by shared environmental factors in toddlers (Dale et al., 2000), and the specific genetic and shared environmental factors responsible for individual differences were shown to be similar across components of language in 2-3-year-olds (Dale et al., 2000; Dionne et al., 2003), 4-5-year-olds (Hayiou-Thomas et al., 2006), and 12-year-olds (Dale et al., 2010).

Yet, some unaddressed issues remain. First, as illustrated by the divergent findings of Kovas et al. (2005) and Samuelsson et al. (2005) in preschoolers, the use of latent factors may have an impact on the etiological patterns reported. Indeed, latent factors take into account only what is common to different measures, and so exclude specific measurement error, freeing more variance to be explained by genetic and environmental factors. As both cited studies conducted with school-aged children only used individual measures of vocabulary and syntax (Dale et al., 2010; DeThorne et al., 2012), the high influence of unique environment – which includes error – found by the authors could mask greater contributions of genes or shared environment.

Second, to our knowledge, no study has yet investigated the association between vocabulary and syntax in children who have recently begun formal instruction: DeThorne et al. (2008, 2012) only reported estimates for the two components of language separately. However, school entry is an important transition during which several genetic and cognitive changes occur. Indeed, formal instruction has been shown to have a direct positive impact on some general cognitive skills such as short-term memory (Morrison, Smith, & Dow-Ehrensberger, 1995), but also specifically on lexical and syntactic skills (Huttenlocher, Levine, & Vevea, 1998). For instance, children are likely to be exposed to infrequent vocabulary words and complex syntactic constructions as they learn about new abstract concepts (Snow, 2010). Moreover, genetic factors associated with linguistic skills are thought to become increasingly influent from 7 years of age, with new genes coming into play during this period (Hayiou-Thomas, Dale, & Plomin, 2012).

Third, all cited studies but one (Samuelsson et al., 2005) examined English-speaking children only. However, language learning may vary across languages. For instance, Thordardottir (2005) has shown that French-speaking toddlers had higher syntactic skills but lower lexical skills than English-speaking toddlers of the same age. Thus, it is possible that the etiology of vocabulary and syntax varies across these languages.

Therefore, the objective of this study is to use latent factors to examine the relative contribution of genes and environment to lexical and syntactic skills, as well as to the association between the two components of language in French-speaking first graders.

Method

Participants

Participants came from the Quebec Newborn Twin Study (QNTS). Parents of all twins born without any major complication in the greater Montreal area (Quebec, Canada) between April 1995 and December 1998 were contacted to take part to the QNTS. The twins whose parents gave their consent have been assessed annually on cognitive, behavioural, social, and environmental components of their development starting at age 6 months (initial N = 622 pairs; see Boivin et al., 2013, for more details). Ethical approval was obtained before each data collection. In the present study, we analyzed data collected when the twins were 7 years old (M = 7.08, SD = 0.27, N = 476 pairs). Only French-speaking same-sex twins were included (142 male pairs and 146 female pairs). Most of these twins had a mother with a bachelor's degree, and mother's education level was comparable for MZ and DZ twins, $\chi^2(8) = 4.32$, $p = .83$. For 74% of the pairs, the twins were in different classrooms. The exact number of participants for each language measure is presented in Table 1.

(Table 1 here)

Procedure

When participants were in Grade 1, they completed the Vocabulary subtest of the French version of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991) and the French version of the Peabody Picture Vocabulary Test (PPVT; Dunn, Thériault-Whalen, & Dunn, 1993). Furthermore, participants' answers to the Vocabulary subtest of the WISC were recorded and transcribed by hand by four trained assistants and the first author, who also revised all of the transcripts. Then, mean length of utterance (MLU; Brown, 1973) and clause density (Scott & Stokes, 1995) were calculated by hand from the transcripts.

Transcription. To facilitate the calculation of MLU and clause density, the transcripts were divided into utterances. Given the nature of the Vocabulary subtest of the WISC, two types of answers were considered as utterances. First, an utterance could be a sentence containing minimally a subject and a verb, and optionally a maximum of one additional coordinate clause (Lee, 1974) and/or an unlimited number of subordinate clauses. For example, each of the following sentences would be considered as a single utterance: *Elle nage* (It swims), *Elle nage et elle sort de l'eau* (It swims and it comes out of the water), *Elle nage et elle sort de l'eau parce qu'il faut qu'elle respire* (It swims and it comes out of the water because it is necessary that it breathes).

Second, a sequence of words that was not a sentence but that was separated from the rest of speech by pauses of at least one second was also considered as an utterance. However, hesitations or reflection pauses within a clause did not divide the clause into more than one utterance (Rondal, 1997). For example, each of the following sequences of words would be considered as a single utterance: *Un animal* (An animal) as an answer to the question *Qu'est-ce qu'une baleine ?* (What is a whale?), *Et elle sort de l'eau* (And it comes out of the water) said at least one second after *Elle nage* (It swims). However, *Sort de l'eau* (Comes out of the water) said at least one second after *Elle nage et elle* (It swims and it) would not be considered as a single utterance because the pause represents a hesitation or a reflection period within a clause.

Finally, utterances of only one uninflected word were not included in the calculation of MLU or clause density (Rondal, 1997), as well as utterances that were unintelligible or that contained an unintelligible segment, repeated utterances (Lee, 1974), and utterances not related to the task. All of these segmentation and exclusion criteria are in accordance with Mimeau, Plourde, Ouellet, and Dionne's (2015) transcription procedure for different types of tasks, including the Vocabulary subtest of the WISC.

In the present study, the transcripts included 32.77 utterances on average (*SD* = 19.41). For 41 participants, the transcriptions were completed by the first author and at least one trained assistant. As the main challenge of transcription resides in the segmentation into utterances, the number of utterances was recorded for each participant for each transcriber. The intra-class correlation coefficients between the first author and the trained assistants' number of utterances ranged from .987 to .996.

Materials

Vocabulary. The Vocabulary subtest of the WISC assesses 6- to 17-year-old children's vocabulary knowledge by asking them to define a list of 30 words. Score is determined by the accuracy of the definitions provided: 0 point is assigned to an incorrect answer, 1 point is assigned to a partially correct answer, and 2 points are assigned to a completely correct answer (maximum total score = 60). The task ends after four consecutive scores of 0. Inter-rater reliability is good for this measure, with an intra-class correlation coefficient of .98 (Wechsler, 1991). Raw scores were used in the present study.

The PPVT assesses 2 ½- to 18-year-old children's vocabulary knowledge by asking them to choose, out of a set of four black and white pictures, the one that best describes a word. It includes 170 words, each worth 1 point (maximum total score = 170). The task ends after the occurrence of six errors out of eight items. Stability is good for this measure, with a test-retest correlation coefficient of .84 (Bracken & Murray, 1984). Raw scores were used in the present study.

Syntax. MLU is a measure of productive syntax that indicates utterance length. In this study, it was calculated by dividing the total number of words by the total number of utterances. Groups of words and expressions considered as a single unit (e.g., *parce que* [because]) were counted as only one word (Thordardottir, 2005).

Clause density is a measure of productive syntax that indicates clause embedment. In this study, it was calculated by dividing the total number of independent and dependent clauses by the number of independent clauses. A dependent clause was defined as a clause that is embedded in an independent clause. Thus, relative clauses (e.g., *Qui vit dans l'eau* [That lives in water]), noun clauses (e.g., *Que ça veut dire drôle* [That it means funny]), and adverbial clauses (e.g., *Quand il pleut* [When it is raining]) were counted as dependent clauses. However, non-embedded clauses (e.g., *C'est comme un poisson* [It is like a fish]), coordinate clauses (e.g., *Et elle te dit le temps* [And it tells you time]), and utterances with no inflected verb (e.g., *Quitter quelqu'un* [To leave someone]) were counted as independent clauses. As an example, the utterance *Ça veut dire que c'est vieux* (It means that it is old) would receive a score of 2 because it has one independent clause (*Ça veut dire* [It means]) and one dependent clause (*Que c'est vieux* [That it is old]).

Both MLU and clause density calculated from a definition task such as the Vocabulary subtest of the WISC have been shown to be valid and reliable measures to assess French-speaking school-aged children's syntactic skills. Indeed, these measures were found to increase as a function of age, to be correlated with other components of language such as vocabulary knowledge and narrative skills, and to be correlated with MLU and clause density calculated from a narration task (Mimeau et al., 2015).

In the present study, MLU was calculated by five different raters, and scores for 30 participants were calculated by all five of them. The intra-class correlation coefficient was .998.

Clause density was calculated by two different raters, and scores for the same 30 participants were calculated by both of them. The intra-class correlation coefficient was .990.

Statistical Analyses

Data preparation. WISC score, PPVT score, MLU, and clause density were all distributed normally. To make sure that the sample of 7-year-olds was representative of all the QNTS participants, earlier language scores of the twins who participated in the study at 7 years of age were compared with earlier language scores of the twins who did not participate at that age. Language scores were computed at 2 ½ years of age from an abbreviated version of the French adaptation of the MacArthur-Bates Communicative Development Inventory: Words and Sentences (Frank, Poulin-Dubois, & Trudeau, 1997). Differences between the two groups of participants were examined in SPSS 22 with two generalised estimating equations (for receptive and productive vocabulary) and one chi-square test (for productive syntax), using the Huber–White robust sandwich estimator for standard errors to control for the non-independence of the observations. The differences were not significant for any of the scores tested (*p*s > .15), which indicates that the participants included in the present study are representative of the QNTS.

Since both twins of a pair share the same age and the same sex (only MZ twins and same-sex DZ twins were included in the study), similarities among twins of a same family may be overestimated (McGue & Bouchard, 1984). Therefore, all scores were corrected for age and sex and the resulting standardized residuals were used in the correlational, genetic, and factor analyses. The syntactic scores of one participant were excluded from the analyses, as they were more than 7 standard deviations from the mean. Furthermore, Little's test indicated that missing data (less than 4% for each measure) were missing completely at random, $\chi^2(11) = 11.78$, $p = .38$, so the full information maximum likelihood method was used in the genetic and factor analyses to handle missing data.

Univariate genetic analyses. To investigate the sources of individual differences in vocabulary and syntax, univariate genetic analyses were performed on the four measures with Mplus 7.11.

Confirmatory factor analysis. To determine whether WISC score and PPVT score could be grouped within a Vocabulary factor, and MLU and clause density within a Syntax factor, a confirmatory factor analysis with an Oblimin rotation was performed with Mplus 7.11.

Multivariate latent factor genetic analyses. To investigate the sources of the association between vocabulary and syntax, a Cholesky decomposition model was performed with Mplus 5.2. Then, a correlated factors model was derived from the Cholesky decomposition model, as illustrated in Figure 1. In this model, genes, shared environment, and unique environment explaining the total variance in vocabulary and syntax are represented by A, C, and E, respectively, above each corresponding latent factor. Moreover, genes, shared environment, and unique environment explaining the residual variance not accounted for by the latent factors are represented by A, C, and E, respectively, below each measured variable. A power analysis (Preacher & Coffman, 2006) revealed sufficient power for this model (.88).

Results

Phenotypic Analyses

Means, standard deviations, range of values, and number of participants for the two measures of vocabulary and the two measures of syntax are presented as a function of zygosity (MZ or DZ) in Table 1. No mean (ps > .15) or variance (ps > .21) differences were found between the two groups for any measure.

Table 2 shows the phenotypic correlations between the four measures for one twin per pair (in order to preserve independence of data). Both within-component (within-vocabulary and within-syntax) correlations were strong, whereas vocabulary-syntax correlations were modest to moderate.

(Table 2 here)

Univariate Genetic Analyses

Table 3 shows the intra-class correlations, the standardized *a*, *c*, and *e* parameter estimates, and the model fit indices for all measures. Regarding vocabulary, for both WISC and PPVT scores, the intra-class correlations were stronger within MZ than within DZ twin pairs, but moderate to strong in both types of twins. Therefore, contributions of genes, shared environment, and unique environment were expected. Indeed, for both measures, *a*, *c*, and *e* parameter estimates were significant (*p*s < .001). Regarding syntax, the intra-class correlations for MLU and clause density were also stronger within MZ than within DZ twin pairs, but they were overall much lower than for vocabulary[2]. Therefore, contributions of genes and unique environment, but not of shared environment, were expected. Indeed, for both measures, *a* and *e* parameter estimates were significant (*p*s < .001), and *c* was equal to 0 (*p*s > .999).

(Table 3)

Confirmatory Factor Analysis

The confirmatory factor analysis revealed that WISC score (factor loading = .90) and PPVT score (factor loading = .61) could be grouped within a Vocabulary factor, and that MLU (factor loading = .82) and clause density (factor loading = .85) could be grouped within a Syntax factor. Indeed, the fit of the model was excellent: $\chi^2(1) = 0.99$, $p = .32$, CFI = 1.00, RMSEA = .00.

[2] Even though this pattern may suggest an ADE model, an ACE model was used to enable the comparison with the model used for vocabulary.

Multivariate Latent Factor Genetic Analyses

In accordance with the univariate analyses, the multivariate latent factor analyses (see Figure 1) indicated that *a*, *c*, and *e* parameter estimates were significant for the Vocabulary latent factor (*p*s < .001), whereas only *a* and *e* parameter estimates were significant for the Syntax latent factor (*p*s < .001). Shared environmental factors explained about half of the variance in vocabulary, and genetic factors explained most of the remaining variance. Genetic and unique environmental factors each explained approximately half of the variance in syntax.

(Figure 1 here)

Concerning measure-specific residual parameter estimates, *a* was significant for PPVT score and MLU (*p*s < .001), which indicates that the latent factors accounted for all the genetic variance in WISC score and clause density. Furthermore, *c* was not significant for any measure (*p*s > .54), which indicates that the latent factors accounted for all the shared environmental variance in all measures. Finally, *e* was significant for all measures (*p*s < .001), which indicates that the latent factors did not account for all the unique environmental variance and error in any measure.

Moreover, the multivariate latent factor analyses indicated that 73% of the association between the Vocabulary and the Syntax latent factors ($r = .37$) could be accounted for by genetic factors ($p < .001$), 20% by shared environmental factors ($p = .25$), and 7% by unique environmental factors ($p = .28$). As shown, the contributions of the environmental factors were not significant. The correlated factors model also indicated that the genetic and the shared environmental correlations between vocabulary and syntax were perfect (*p*s < .001), while the unique environmental correlation was not significant ($p = .27$), indicating that the genetic and shared environmental factors influencing vocabulary and syntax are identical, but that the unique environmental factors influencing the two components of language are mostly distinct.

Discussion

The Etiology of Vocabulary and Syntax

The first aim of this study was to examine the etiology of lexical and syntactic skills in French-speaking first graders using multiple measures grouped into latent factors. The genetic analyses performed revealed that at 7 years of age, the relative contribution of genes and environment to vocabulary and syntax was comparable ($a\mathrm{s}^2$ = .38-.51; $c\mathrm{s}^2$+$e\mathrm{s}^2$ = .49-.62, for vocabulary; $a\mathrm{s}^2$ = .27-.42; $c\mathrm{s}^2$+$e\mathrm{s}^2$ = .58-.73, for syntax). However, we found differences in the type of environment at play: Shared environmental factors influenced vocabulary, but not syntax ($c\mathrm{s}^2$ = .26-.52, for vocabulary; $c\mathrm{s}^2$ = .00-.02, for syntax), whereas unique environmental factors (and error) influenced syntax more than vocabulary ($e\mathrm{s}^2$ = .09-.36, for vocabulary; $c\mathrm{s}^2$ = .56-.73, for syntax). This finding converges with those of researchers studying toddlers (Dale et al., 2000; Dionne et al., 2003). Unexpectedly, however, it contrasts with previous findings on school-aged children, which indicated an equally modest contribution of shared environment to the two components of language (Dale et al., 2010; DeThorne et al., 2012).

The novel finding that shared environment contributes substantially to vocabulary in school-aged children may be explained by the sophisticated procedure used to represent vocabulary more accurately in the present study: the grouping of multiple measures into a latent factor. Indeed, DeThorne et al. (2012) assessed this component of language with a single measure of lexical diversity, and Dale et al. (2010) assessed it with a single measure adapted from the Vocabulary subtest of the WISC. The reason why latent factors may increase the contribution of shared environment is that they exclude specific measurement error and by consequence, they free more variance to be explained by relevant sources of influence – such as shared environmental factors. In fact, this assumption is supported by our own results, with shared environmental factors tending to play a greater role in vocabulary in the multivariate latent factor analyses (c^2 = .52) compared with the univariate analyses (cs^2 = .26).

Another explanation that should be considered is that this study examined French-speaking children, while all previous twin studies but one (Samuelsson et al., 2005) examined English-speaking children. Since vocabulary seems to be acquired at a slightly slower rate in French (Thordardottir, 2005; see also Bornstein et al., 2004, and http://www.cdi-clex.org), it could be that shared environment continues to contribute to French learners' lexical skills during the school years whereas it no longer does for English learners. This hypothesis is consistent with Hayiou-Thomas et al.'s (2012) finding that shared environmental influences decrease as children increase their mastery of language (see also Olson et al., 2011). Nonetheless, it should be noted that the differences between English and French are very small, making this explanation unlikely. Regarding Samuelsson et al.'s (2005) study, which included Norwegian- and Swedish-speaking preschoolers (combined with English-speaking preschoolers), the contribution of shared environment to vocabulary was also important (c^2 = .60). However, as in the present study, the authors used a latent factor, making it impossible to distinguish whether the source of their (and our) distinctive finding is due to the choice of language or analyses.

That being said, the different etiological patterns of vocabulary and syntax are perhaps unsurprising. It might be the case that at the beginning of the school years, the language heard at home (usually similar for both twins) has a stronger impact on vocabulary, and that the language heard at school (usually different for both twins, who often have different teachers in Quebec) has a stronger impact on syntax. In line with this hypothesis, Weizman and Snow (2001) observed that in kindergarten and Grade 2, children who performed better in a receptive vocabulary task had a mother who had used more sophisticated words embedded in an instructive and helpful speech when they were 5 years old. Some other studies also showed that certain family practices such as shared reading predicted lexical skills better than syntactic skills (Lever & Sénéchal, 2011; Sénéchal, Pagan, Lever, & Ouellette, 2008). Contrastingly, Huttenlocher, Vasilyeva, Cymerman, and Levine (2002) found that preschoolers' receptive syntactic skills increased in one school year as a function of the complexity of the sentences produced by their teacher. Yet, further research is needed to confirm which environmental factors operate in school-aged children.

The Etiology of the Association Between Vocabulary and Syntax

The second aim of this study was to examine the etiology of the association between lexical and syntactic skills in French-speaking first graders. First, the multivariate latent factor genetic analyses performed revealed that at 7 years of age, 73% of the association between the two components of language could be accounted for by genetic factors, with only a minimal contribution of the environment. This finding is, to our knowledge, the first empirical evidence that school-aged children's lexical and syntactic skills are associated mainly through common genetic influences, contrary to toddlers' skills, which are associated mainly through common shared environmental influences (Dale et al., 2000). Still, this result does not come as a surprise, given the increasing contribution of genes to language across development. This etiological change could be attributed to a homogenisation of the environment during the school years, "leav[ing] more room for genetic factors to drive differences in the phenotype" (Hayiou-Thomas et al., 2012, p. 245). Second, the multivariate latent factor genetic analyses performed also revealed that the specific genetic factors responsible for individual variations in vocabulary and syntax were identical ($r_A = 1.00$), which is in accordance with the high genetic correlations reported previously at different time points (Dale et al., 2000, 2010; Dionne et al., 2003; Hayiou-Thomas et al., 2006).

Taken together, these findings suggest that similar genetic mechanisms are at play when children process vocabulary and syntax shortly after entering school. This is consistent with the idea of a single mechanism of acquisition for all the components of language (MacWhinney, 1987) and with theories of interaction between vocabulary and syntax (Gleitman, 1990; Pinker, 1984), but less consistent with the proposal that lexical and syntactic skills are learned through different memory systems (Ullman, 2004), assuming that these systems stem from different genes.

Some genes that could be involved in both lexical and syntactic skills have been identified. For instance, in members of families at risk for specific language impairment (SLI), the genes *KIAA0319*, *CNTNAP2*, *ATP2C2*, and *CMIP* were found to be associated with general linguistic skills (Newbury et al., 2011). More broadly, Plomin and Kovas (2005) proposed that generalist genes, that is, “all-purpose” genes, operate on cognition at different levels of proficiency (e.g., ability vs. disability), in different domains (e.g., language vs. mathematics), and in different components of a same domain (e.g., vocabulary vs. syntax). However, variance in identified genes accounts only for a small portion of variance in linguistic skills, despite the substantial contribution of genetic factors estimated in twin studies, which indicates that many genes still need to be discovered (Bishop, 2009). Therefore, one alternative way to address the question of why lexical and syntactic skills are associated is to find which lower-level general cognitive mechanisms – which are partly specified by genes – are involved in both components of language.

One potential candidate known to be influenced by genes (Friedman et al., 2008; Wright et al., 2001) is working memory. This memory system allows mental retention of verbal information for a short period of time by repeating it through its phonological loop (Baddeley, Gathercole, & Papagno, 1998). Adams and Gathercole (2000) observed that 4-year-olds with better working memory skills produced words and syntactic constructions that were more diverse than those produced by children with poor working memory skills, which points out the close association between working memory and both lexical and syntactic skills. A study of English-speaking adults learning Welsh yielded a comparable conclusion (Ellis & Sinclair, 1996). Indeed, it showed that participants who repeated the target words and utterances during learning were better at translating them from English to Welsh than participants who occupied their working memory with articulatory suppression (counting from 1 to 5 repeatedly) during learning. These findings also suggest that working memory plays a fundamental role in both vocabulary and syntax because learning of both words and utterances was reduced when working memory was made unavailable.

Another general cognitive mechanism with a possible genetic basis (Lobo, Karsten, Gray, Geschwind, & Yang, 2006; Ullman, 2004) that could explain the persistent genetic association between vocabulary and syntax is procedural memory. Procedural memory (also referred to as statistical learning) is the ability to capture the transitional probabilities of sequences. For instance, as the sound /beɪ/ is very often followed by the sound /bi/ in the speech babies hear, they can figure out, using procedural memory, that these sounds belong together and create the word *baby* (Saffran, Aslin, & Newport, 1996). In a study of artificial language learning, Saffran and Wilson (2003) observed that infants could segment a continuous speech stream into words but also extract syntactic rules from it. Those findings, although in contradiction with Ullman's declarative/procedural model, suggest that infants use transitional probabilities to learn both vocabulary and syntax, highlighting the central role of procedural memory in the two components of language. Furthermore, the same research group (Evans, Saffran, & Robe-Torres, 2009) showed that school-aged children with SLI, which affects both lexical and syntactic skills, presented deficiencies in procedural memory compared with typically developing children. The authors proposed that a poor procedural memory might underlie the language delays observed in SLI, but procedural memory might contribute to individual differences in lexical and syntactic skills in the general population as well.

Still, even if working and procedural memory could explain why vocabulary and syntax are associated and accounted for by the same genes, more research is needed to clarify whether this applies to different age groups and populations. Researchers should also investigate which other cognitive mechanisms may be at the core of different components of language and how these several sources of genetic and cognitive influence are organized. Moreover, molecular research should be continued in the hope of finding additional genes involved in language development.

Limitations of the Present Study

Some limitations of this study need to be acknowledged. First, compared with other twin studies (e.g., Dale et al., 2010, *N* = 8638; Samuelsson et al., 2005, *N* = 1254), the number of participants included in this study is rather small (*N* = 555). Although statistical power was sufficient, a larger sample size could have produced smaller confidence intervals, enabling a better comparison of the parameter estimates.

Second, the measures we used were not entirely independent. Indeed, both measures of syntax were computed from the answers the children gave in the Vocabulary subtest of the WISC, which was used as a measure of vocabulary. This has the consequence of reducing the advantage of using latent factors, especially to represent syntax. Indeed, latent factors exclude the error that is specific to each measure, but it cannot exclude the error that is common to the measures (e.g., variance due to task engagement, mood, or tiredness). Because our two measures of syntax were likely influenced by more common error than our two measures of vocabulary, the extent to which unique environment influences syntax more than vocabulary could be overestimated.

However, it should be recalled that the high contribution of unique environment (and error) to syntax in the present study was not surprising, as this result is what is typically found when examining the linguistic skills of school-aged children (e.g., Dale et al., 2010; DeThorne et al., 2012). Moreover, the Syntax latent factor did reduce some error, as the contribution of E was generally higher in the univariate analyses (es^2 = .58-73) than in the multivariate latent factor analyses (e^2 = .56). Additionally, the fact that our measures yielded different etiological patterns in the univariate analyses (see Table 3) and that there was some residual variance in the multivariate analyses (see Figure 1) suggests that our measures were not completely dependent.

Third, our two measures of vocabulary assessed language with standardised tests, while our two measures of syntax assessed language with analyses of spontaneous speech. Since the latter need to be calculated from a limited number of utterances, they offer a narrow representation of a child's abilities, much influenced by timing and context. As for the issue stated above, this could increase measurement error, which might explain the high contribution of E to syntax. However, DeThorne et al. (2008) showed that lexical and syntactic measures of spontaneous speech were influenced by the same factors of unique environment and error as those influencing standardised lexical measures. Yet, the authors also found that the contribution of E was slightly higher for measures of spontaneous speech (es^2 = .33-.63) than for standardised measures (es^2 = .13-.55). This last finding leaves open the possibility that the different environmental influences observed for vocabulary and syntax in the present study are caused by differences in measurement error. As such, in future studies, researchers should include a better-balanced combination of measures if possible. For instance, measures of lexical diversity

(Malvern, Richards, Chipere, & Durán, 2004) and syntactic comprehension (e.g., Bishop, 2003) could be added to measures similar to those of the present study.

Finally, it is worth noting that our findings apply solely to individual differences. That is, they can explain why some children have better or worse lexical and syntactic skills than others. However, despite their focus on genetics, twin studies are not informative as to why only humans and not other species possess the faculty of language. In other words, they do not allow examining species universals.

Conclusion

In summary, the present study showed that during the critical period of school entry, children’s processing of words is associated with, yet distinct from their processing of sentences. Indeed, lexical and syntactic skills were found to share the same genetic factors of influence, but the contribution of environmental factors varied from one component of language to another: Lexical skills were influenced mainly by shared environment and syntactical skills were influences mainly but unique environment. Although this study fills a gap in the literature regarding the association between vocabulary and syntax in school-aged children, further research is needed to determine more precisely which factors are involved in language at this developmental period.

References

Adams, A.-M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders, 35*, 95–116. doi:10.1080/136828200247278

Baddeley, A., Gathercole, S., & Papagno. C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173. doi:10.1037/0033-295X.105.1.158

Bearden, C. E., Karlsgodt, K. H., Bachman, P., van Erp, T. G. M., Winkler, A. M., & Glahn, D. C. (2012). Genetic architecture of declarative memory: Implications for complex illnesses. *The Neuroscientist, 18*, 516–532. doi:10.1177/1073858411415113

Bishop, D. V. M. (2003). *The Test for Reception of Grammar* (2nd ed.). London, United Kingdom: Psychological Corporation.

Bishop, D. V. M. (2009). Genes, cognition, and communication: Insights from neurodevelopmental disorders. *Annals of the New York Academy of Sciences, 1156*, 1–18. doi: 10.1111/j.1749-6632.2009.04419.x

Boivin, M., Brendgen, M., Dionne, G., Dubois, L., Pérusse, D., Robaey, P., . . . Vitaro, F. (2013). The Quebec Newborn Twin Study into adolescence: 15 years later. *Twin Research and Human Genetics, 16*, 64–69. doi:10.1017/thg.2012.129

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., . . . Vyt, A. (2004). Crosslinguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75*, 1115–1139. doi:10.1111/j.1467-8624.2004.00729.x

Bracken, B. A., & Murray, A. M. (1984). Stability and predictive validity of the PPVT-R over an eleven month interval. *Educational & Psychological Research, 4*, 41–44.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Developmental Science, 12*, 140–153, doi:10.1080/10888690802199418

Dale, P. S., Dionne, G., Eley, T. C., & Plomin, R. (2000). Lexical and grammatical development: A behavioural genetic perspective. *Journal of Child Language, 27,* 619–642. doi:10.1017/S0305000900004281

Dale, P. S., Harlaar, N., Hayiou-Thomas, M. E., & Plomin, R. (2010). The etiology of diverse receptive language skills at 12 years. *Journal of Speech, Language, and Hearing Research, 53,* 982–992. doi:10.1044/1092-4388(2009/09-0108)

DeThorne, L. S., Harlaar, N., Petrill, S. A., & Deater-Deckard, K. (2012). Longitudinal stability in genetic effects on children's conversational language productivity. *Journal of Speech, Language, and Hearing Research, 55,* 739–753. doi:10.1044/1092-4388(2011/11-0014)

DeThorne, L. S., Petrill, S. A., Hart, S. A., Channell, R. W., Campbell, R. J., Deater-Deckard, K., . . . Vandenbergh, D. J. (2008). Genetic effects on children's conversational language use. *Journal of Speech, Language, and Hearing Research, 51,* 423–435. doi:10.1044/1092-4388(2008/031)

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*, 394–412. doi:10.1111/1467-8624.7402005

Dunn, L. M., Thériault-Whalen, C. M., & Dunn, L. M. (1993). *Échelle de vocabulaire en image Peabody : Adaptation française du Peabody Picture Vocabulary Test-Revised* [Échelle de vocabulaire en image Peabody: French adaptation of the Peabody Picture Vocabulary Test-Revised]. Toronto, Canada: Psycan.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 49,* 234–250. doi:10.1080/713755604

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52,* 321–335. doi:10.1044/1092-4388(2009/07-0189)

Fenson, L., Dale, P. S., Bates, E., Reznick, J. S., Thal, D. J., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59,* 1–173. doi:10.2307/1166093

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42,* 98–116. doi:10.1037/0012-1649.42.1.98

Frank, I., Poulin-Dubois, D., & Trudeau, N. (1997). *Inventaire MacArthur-Bates du développement de la communication : Mots et énoncés*. Retrieved from http://coder-eoa.eoa.umontreal.ca/Web/MacArthur_Bates/telechargements/IMBDC_mots_et_enonces.pdf.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201–225. doi:10.1037/0096-3445.137.2.201

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1,* 3–55. doi:10.1207/s15327817la0101_2

Hayiou-Thomas, M. E., Kovas, Y., Harlaar, N., Plomin, R., Bishop, D. V. M., & Dale, P. S. (2006). Common aetiology for diverse language skills in 4½-year-old twins. *Journal of Child Language, 33*, 339–368. doi:10.1017/S0305000906007331

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: a longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15,* 233–249. doi:10.1111/j.1467-7687.2011.01119.x

Huttenlocher, J., Levine, S., & Vevea, J. (1998). Environmental input and cognitive growth: A study using time-period comparisons. *Child Development, 69,* 1012–1029. doi:10.2307/1132360

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*, 337–374. doi:10.1016/S0010-0285(02)00500-5

Joseph, R. M., McGrath, L. M., & Tager-Flusberg, H. (2005). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology, 27*, 361–378. doi:10.1207/s15326942dn2703_4

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. M., & Plomin, R. (2005). Genetic influences in different aspects of language development: The etiology of language skills in 4.5-year-old twins. *Child Development, 76*, 632 – 651. doi:10.1111/j.1467-8624.2005.00868.x

Lee, L. L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians*. Evanston, IL: Northwestern University Press.

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology*, *108*, 1–24. doi:10.1016/j.jecp.2010.07.002

Lobo, M. K., Karsten, S. L., Gray, M., Geschwind, D. H., & Yang, X. W. (2006). FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nature Neuroscience, 9,* 443–452. doi:10.1038/nn1654

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*, 1138–1154. doi:10.1016/j.cortex.2011.06.001

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum Associates.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, United Kingdom: Palgrave Macmillan.

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development, 80,* 686–702. doi:10.1111/j.1467-8624.2009.01291.x

McGue, M., & Bouchard, T. J., Jr. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14*, 325–343. doi:10.1007/BF01080045

Mimeau, C., Plourde, V., Ouellet, A.-A., & Dionne, G. (2015). Comparison of measures of morphosyntactic complexity in French-speaking school-aged children. *First Language, 35*, 163–181. doi:10.1177/0142723715577320

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology, 31,* 789–799. doi:10.1037/0012-1649.31.5.789

Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., . . . Monaco, A. P. (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behavior Genetics, 41*, 90–104. doi:10.1007/s10519-010-9424-3

Olson, R. K., Keenan, J. M., Byrne, B., Samuelsson, S., Coventry, W. L., Corley, R., . . . Hulslander, J. (2011). Genetic and environmental influences on vocabulary and reading development. *Scientific Studies of Reading, 15*, 26–46. doi:10.1080/10888438.2011.536128

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*, 592–617. doi:10.1037/0033-2909.131.4.592

Preacher, K. J., & Coffman, D. L. (2006). *Computing power and minimum sample size for RMSEA*. Retrieved from http://quantpsy.org/rmsea/rmsea.htm.

Pungello, E. P., Iruka, I. U., Dotterer, A. M., Mills-Koonce, R., & Reznick, J. S. (2009). The effects of socioeconomic status, race, and parenting on language development in early childhood. *Developmental Psychology, 45*, 544–557. doi:10.1037/a0013917

Rondal, J. A. (1997). *L'évaluation du langage* [The evaluation of language]. Sprimont, Belgium: Mardaga.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928. doi:10.1126/science.274.5294.1926

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*, 273–284. doi:10.1207/S15327078IN0402_07

Samuelsson, S., Byrne, B., Quain, P., Wadsworth, S., Corley, R., DeFries, J. C., . . . Olson, R. (2005). Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology, 97*, 705–722. doi:10.1037/0022-0663.97.4.705

Scott, C. M., & Stokes, S. L. (1995). Measures of syntax in school-age children and adolescents. *Language, Speech, and Hearing Services in Schools, 26*, 309–319. Retrieved from http://lshss.asha.org

Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (2008). Relations among the frequency of shared reading and 4-year-old children's vocabulary, morphological and syntax comprehension, and narrative skills. *Early Education and Development, 19*, 27–44. doi:10.1080/10409280701838710

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328*, 450-452. doi:10.1126/science.1182597

Thordardottir, E. (2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. *International Journal of Language & Communication Disorders, 40*, 243–278. doi:10.1080/13682820410001729655

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270. doi:10.1016/j.cognition.2003.10.008

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*, 265–279. doi:10.1037/0012-1649.37.2.265

Wright, M., De Geus, E., Ando, J., Luciano, M., Posthuma, D., Ono, Y., . . . Boomsma, D. (2001). Genetics of cognition: Outline of a collaborative twin study. *Twin Research*, *4*, 48–56. doi:10.1375/1369052012146

Table 1

Means, Standard Deviations, Range of Values, and Number of Participants for Vocabulary (WISC and PPVT Scores) and Syntax (MLU and Clause Density) as a Function of Zygosity

	MZ				DZ			
Measure	*M*	*SD*	Range	*n*	*M*	*SD*	Range	*n*
WISC score	12.50	4.79	0-27	336	12.47	5.25	0-27	232
PPVT score	87.73	16.89	36-131	333	85.45	16.72	15-125	229
MLU	7.24	2.04	2.09-13.21	328	7.26	2.01	2.86-15.13	227
Clause density	1.39	0.23	1.00-2.02	328	1.36	0.21	1.00-2.13	227

Note. WISC = Vocabulary subtest of the Wechsler Intelligence Scale for Children; PPVT = Peabody Picture Vocabulary Test; MLU = mean length of utterance; MZ = monozygotic twins; DZ = dizygotic twins.

Table 2

Phenotypic Correlations Between Vocabulary (WISC and PPVT Scores) and Syntax (MLU and Clause Density)

	WISC score	PPVT score	MLU	Clause density
WISC score	–			
PPVT score	.61***	–		
MLU	.35***	.22***	–	
Clause density	.43***	.26***	.68***	–

Note. The correlations were calculated with one twin per pair in order to preserve independence of data. WISC = Vocabulary subtest of the Wechsler Intelligence Scale for Children; PPVT = Peabody Picture Vocabulary Test; MLU = mean length of utterance.

***$p < .001$.

Table 3

MZ and DZ Intra-Class Correlations and Number of Pairs, Standardized a, c, *and* e *Parameter Estimates, and Model Fit Indices for Vocabulary (WISC and PPVT Scores) and Syntax (MLU and Clause Density)*

	MZ		DZ									
	ICC [95% CI]	*n*	ICC [95% CI]	*n*	a^2 [95% CI]	c^2 [95% CI]	e^2 [95% CI]	χ^2	df	*p*	CFI	RMSEA
WISC score	.63 [.54, .72]	170	.47 [.31, .64]	117	.38 [.08, .69]	.26 [-.01, .53]	.36 [.27, .44]	3.50	6	.74	1.00	.00
PPVT score	.77 [.70, .84]	171	.50 [.24, .76]	116	.51 [.24, .78]	.26 [.00, .52]	.23 [.17, .29]	11.19	6	.08	.97	.08
MLU	.44 [.31, .57]	170	.16 [-.04, .36]	116	.42 [.30, .54]	.00 [.00, .00]	.58 [.46, .70]	3.67	6	.72	1.00	.00
Clause density	.31 [.18, .44]	170	.01 [-.20, .23]	116	.27 [.14, .40]	.00 [.00, .00]	.73 [.60, .86]	7.59	6	.27	.89	.04

Note. MZ = monozygotic twins; DZ = dizygotic twins; WISC = Vocabulary subtest of the Wechsler Intelligence Scale for Children; PPVT = Peabody Picture Vocabulary Test; MLU = mean length of utterance; ICC = intra-class correlation; CI = confidence interval; a^2 = proportion of variance explained by additive genetic factors; c^2 = proportion of variance explained by shared environmental factors; e^2 = proportion of variance explained by unique environmental factors and error.

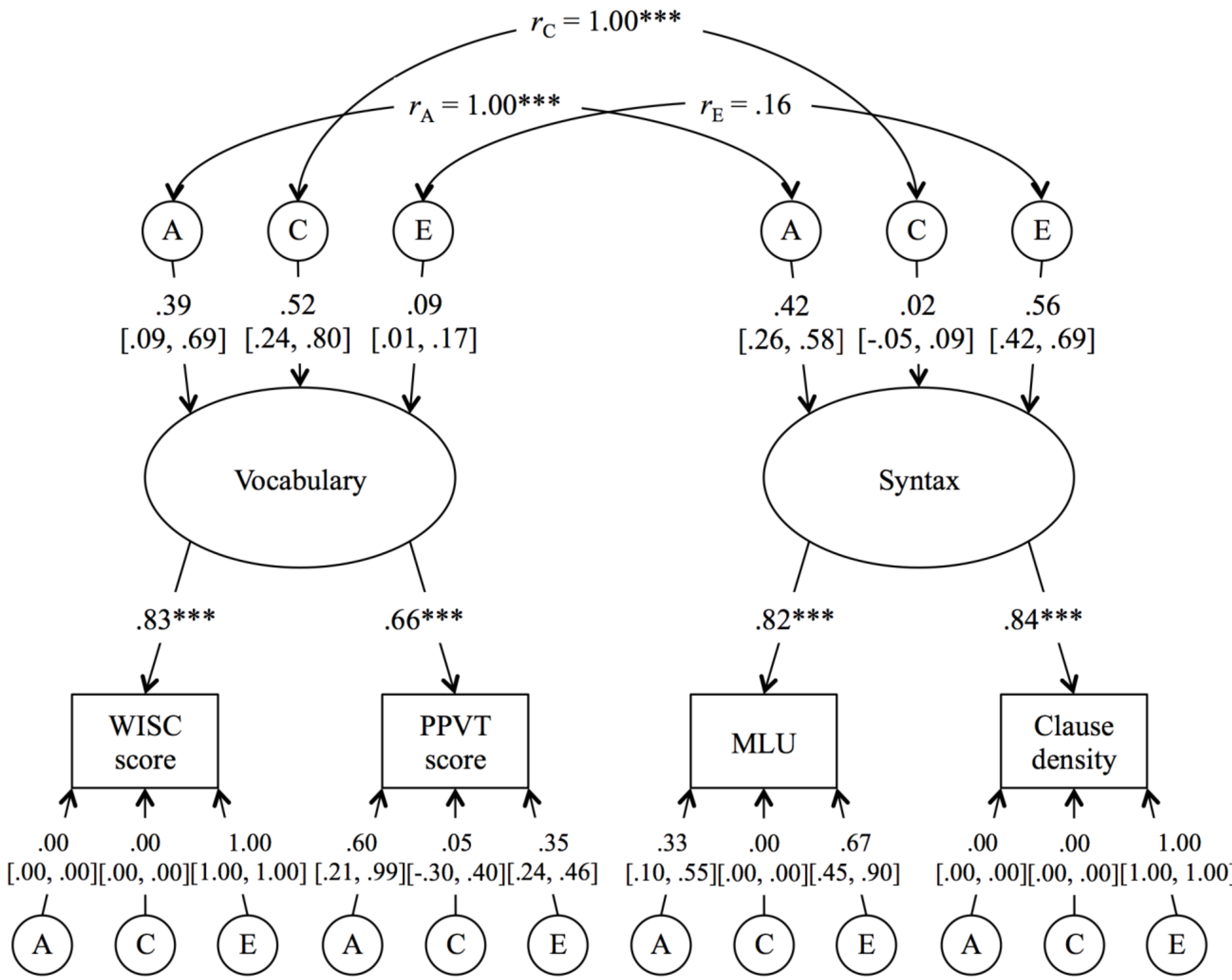

Figure 1. Correlated factors model of vocabulary and syntax with standardized parameter estimates [and 95% confidence intervals]. Model's fit: $\chi^2(72) = 94.69$, $p = .04$, AIC = 5405.03, BIC = 5463.64, CFI = .98, RMSEA = .05. A = additive genetic factors; C = shared environmental factors; E = unique environmental factors and error; WISC = Vocabulary subtest of the Wechsler Intelligence Scale for Children; PPVT = Peabody Picture Vocabulary Test; MLU = mean length of utterance.

***$p < .001$.