EARLY DETERMINATION[*]

Ana T. Pérez-Leroux and Tanya Battersby
University of Toronto

1. Introduction

The focus of developmental linguistics is to explain the ontogeny of language with the full complexity of the structures and computational processes of the human language in mind. The field's basic assumption is that language acquisition is the result of the interaction between an organism, the human language learner, and experience. The child's native language abilities are composed of a set of initial representations, which are generally assumed to be language specific, as well as a set of perceptual and representational biases that guide category formation, which are most commonly assumed to be general to many modules of cognition (Hauser et al. 2002; Ullman 2001).

The present study examines what characteristics of the input guide the children in the acquisition of complex distributional patterns, and what the course of development can tells us about general issues of the interaction between cognitive modules, the general and local effect of input frequency on the developing grammar. Our study presents data from the early stages of productive syntax in 2 children from the Grinstead (1998) corpus, and the focus of our study is of the distribution of bare nouns (BNs) and determiners in Spanish, a language where argumental bare nouns have a mixed distribution, that is, in Spanish, like in other languages such as Italian, Catalan, bare nouns are allowed, but they appear in restricted positions.

2. Distribution of determiners and bare nouns

Spanish allows the use of bare nouns in argument positions. These BNs show a distribution constrained by grammatical function and position, lexical class, and number. Spanish disallows bare subjects in either position, as shown in (1), but postverbal subject of infinitives (2) pattern with objects (3) in allowing count plural and mass nouns to appear without a determiner. Other contexts that allow BNs include objects of preposition (4)-(5), and experience-denoting objects of light verbs. It seems clear that nominal interpretation plays a role, for bare count singular with predicative use are possible in these contexts, as shown in (7a), whereas referential uses are subject to the familiar lexical class/number restrictions, as the ungrammaticality of (7b) demonstrates. Last, as shown in
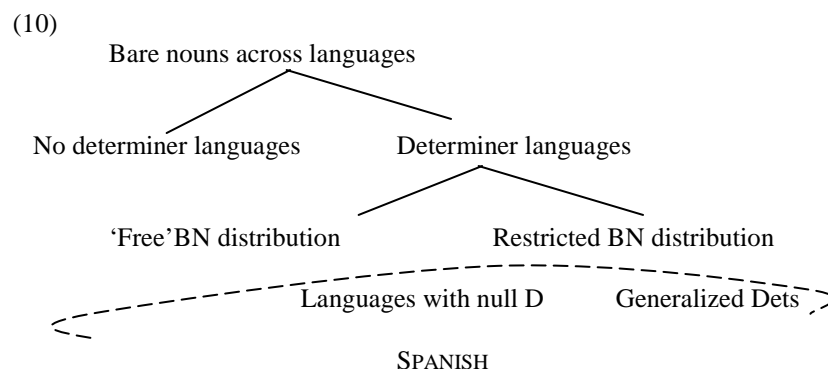
Chierchia (1998), and elsewhere, modification and conjunction provide an escape hatch for these restrictions. The standard assumption is that these complex nominal projections include a D-layer although they do not include a lexically realized D.

Bare nouns in Spanish

(1)	a. Juegan los niños en el parque./ Los niños juegan en el parque. '(The) children play in the park.' b. *Juegan niños en el parque./ *Niños juegan en el parque. 'Children play in the park.'	*Bare subjects
(2)	a. Vinieron clientes. 'Customers came.' b. *Clientes vinieron. 'Customers came.'	Unaccusatives
(3)	a. Compraron fruta y verduras. 'They bought fruit and vegetables. b. *Compraron libro. 'They bought book'	Plural and mass objects
(4)	Café sin azucar/ con leche 'Coffee without sugar/ with milk'	Obj of Prep
(5)	Una bailarina sin zapatos 'A dancer without shoes'	Obj of Prep
(6)	Tener miedo 'Be afraid', Hacer daño 'To hurt', Tener hambre 'To be hungry'	Light verb w/ psych objects
(7)	a. Un perro con sombrero 'A dog with (a) hat.' b. *Lo puse dentro de sombrero. 'I put it inside (the) hat.'	Bare singular count with predicative sense
(8)	A la reunión asistieron **trabajadores** de todas partes del país. 'Labourers from all over the country attended the meeting.'	Modification
(9)	A la reunión asistieron **trabajadores y gerentes**. 'Workers and managers attended the meeting.'	Coordination

Chierchia (1998) proposes a tripartite typology of bare nouns across languages. On one hand there are determinerless languages such as the Chinese languages, where nominals directly map as semantic arguments (type *e*). On the other, languages where nominals are mapped as semantic predicates (type *<e,t>*) and require for the nominal projection to merge with D in order to or both. In his proposed Nominal Mapping Parameter, these are labeled argumental and predicative languages. The third type is considered +argumental and +predicative, because it contains determiners, but allow bare nouns in all positions. English and other Germanic languages, fall into this category. Within the NMP typology, languages like Spanish, where BNs can serve as semantic arguments but in restricted distributions, are considered predicative

languages, and assumed to require generalized use of determiners. The appearance of argumental BNs in an asymmetric pattern for objects and subjects is explained under the additional assumption that the language contains a special empty determiner δ, with a distribution comparable to those of traces (i.e., in governed positions such as objects). This classification is represented in (10).

(10)

Bare nouns across languages

No determiner languages   Determiner languages

'Free' BN distribution   Restricted BN distribution

Languages with null D   Generalized Dets

SPANISH

3. Determiners and bare nouns in children's language

The nominal mapping parameter inspired a series of studies comparing rates of acquisition of determiners across languages. These various studies have shown that children acquire determiners at different rates across different languages. These results clearly aligned along typology: predicative languages such as French and Greek have early acquisition of determiners, in comparison with the Germanic languages, with the mixed-distribution languages such as Italian and Catalan falling in the middle range (Guasti and Rizzi 1999; Kupisch 2006; Marinis 2003). To date, there are several cross-sectional studies of Spanish that suggest early use of determiners and protodeterminers (Aguado 2000; López-Ornat 1999, 2003), but there is no extensive longitudinal analysis of development in the language.

We conducted a longitudinal study of early production to determine, when determiners first appear, whether they have distributional patterns comparable to the adult grammar. Our goal was not to compare rates of development across languages, but to compare rates of development across the various relevant contexts and categories within the same language, using longitudinal data. In other words, we ask the following question: Do children exhibit sensitivity to the overall system of determiner use in a mixed language? To do so, they must demonstrate awareness of the various syntactic (position/function) and lexical-semantic (mass/count/number) factors. Learning determiner use thus has multiple components, summarized in (11):

(11)   Components of the learning task
         a. Lexical: attend to mass/count syntax and determiner which nouns belong to the mass or count category.

      b. Functional: learn the interaction between number and BN distribution.
      c. Syntactic: detect the presence of subject/object asymmetry.

By examining these components in turn, we can see if the whole system appears globally, at once, as the parametric hypothesis would predict, or whether the emergence of determiner use in each context is determined by its relative local variability. What is at stake here is the interrelatedness of the different elements in the system; if grammar develops on a purely lexical basis, as Tomasello (2000) would expect, each context and category is expected to emerge in its own timetable. If, as the parametric hypothesis suggests, the system is the result of abstract typological differences, and the distribution is determined by these in conjunction with abstract computational constraints, one should see global rather than isolated patterns. One issue under consideration is the question of variability in the input. The traditional assumption is that more robust, more stable forms are acquired earlier that forms that are in variable or less frequent distribution. Recent approaches to variability under the generative model suggest this can be approached in various ways. One is articulated in Yang's (2004) variational model, where relative frequency of the relevant types of sentences determines speed of development. In Yang's model, learners have multiple grammars, and learning happens by means of an algorithm that rewards grammars that successfully parse a given input sentence, and penalizes grammars that fail to parse. As a result the overall frequency of the relevant triggering contexts can predict not only which parameters are set early or late (Yang 2004), but also acquisition of functional categories (Legate and Yang 2007). A different variant of this approach is developed in Miller (2007), who links variability within a grammatical category to rates of development. Her data shows that rates of phonological realization of plural –s determines the ages of mastery of plural in comprehension and production across dialects of Spanish.

    One last dimension to consider is the possibility of modular interaction. Grinstead (2004) identifies a stage of subject delay in children acquiring a pro-drop language. While these children freely generate topic/comment structures, they seem to avoid producing subjects at a stage in which these are frequent in non-prodrop languages. He attributes this subject delay to the special challenge of developing the interface between the syntactic module and the pragmatic module. Young children give indication of possessing the relevant knowledge base in both modules, but the subjects in prodrop languages require the development of cross-modular interactions between these two representational systems.

4. Hypothesis

With these considerations in mind, we develop the following three hypotheses:

H1: Parametric hypothesis

The distribution of bare nouns in a mixed is the result of UG-driven principles of computation, as configured in the specific settings of a mixed language. Therefore, the main syntactic characteristics (subject-object asymmetries and mass/plural –count noun distinction) need not be learned, but should emerge in the grammar as a whole. Two possibilities:

> H1A: From the onset of productive determiner use (Very Early Parameter Setting)
> H1B: Following a (+argument) determinerless stage, as proposed by Chierchia

H2: Interface delay hypothesis (IDH)

There is delay in the syntax-discourse pragmatics interface, as proposed in Grinstead (2004) where children avoid using syntactic subjects in some languages if subject position is determined by discourse. Since determiner choice and determiner absence/presence is linked to common ground in the same way as subjects, the IDH predicts radical changes at the same time as children start producing sentential subjects.

> H2: Clear developmental shifts in determiner use at the target age (around 2;3 for Carlos and 1;10, for Eduardo)

H3: Variable input hypothesis

Degree of variability in input determines speed of acquisition (Legate and Yang 2007; Miller 2007; Yang 2004). Does it affect the global parameter, or the actual variable items?

> H3A: Effect of variability is local (Singular>Plural>Mass)
> H3B: Effect of variability is global (affecting the speed of the language, not the item)

5. Study

To test these hypotheses, we analyzed a longitudinal corpus of two monolingual Spanish-speaking children from the Grinstead 1998 corpus. We looked at monthly speech samples from the age of 1;4 to 3;6 for Carlos and 1;5 to 3;6 for Eduardo, for a total of between 12,000 and 18,000 utterances per child. This yielded between 3,000 and 4,000 NPs per child. We also analyzed their parents' speech samples at six-month intervals, which yielded totals of around 1500 to 1700 utterances per parent and around a thousand noun phrases each. The extracted NPs were classified by lexical type as mass, count, pronominal or proper nouns. Then, potentially-determined NPs were classified by number and function, that is, subject or object. This yielded approximately 400-500 potentially determined NPs for each child and parent.

For the children's production of determiners, we identified three stages of development. In the initial stage, production of determiners was less than ten per session, and this stage lasted until 2;3 for Carlos and 1;11 for Eduardo. The intermediate stage was characterized by production of ten to 30 determined NPs per session, which lasted until 2;10 for Carlos and 2;8 for Eduardo. In the third

stage the children used determiners in 30 or more NPs per session, and this phase continued until 3;6 for both children.

6. Results

### 6.1. Lexical and functional factors

The first analysis considered the production of determiners in NPs according to lexical class and number. We present the parental data in Figure 1, where it is clear that there are consistent differences in the production of determiners among the various types of NPs. Singular count nouns showed the highest rates of determiner insertion, with a 97% average for Carlos' mother and a 93% average for Eduardo's mother. Plural count nouns showed slightly lower rates of determiner insertion, with a 76% average for Carlos' mother and an 80% average for Eduardo's mother. The percentages of determiner use for both singular and plural count nouns showed a narrow range of variation, with 8% across files for count singular and 21% across files for count plural nouns. However, mass nouns had both the lowest rates of determiner insertion and a wider range of variation in the parents' speech. Carlos' mother used determiners with mass nouns an average of 46% of the time and Eduardo's mother an average of 52% of the time. These rates varied from 25 to 67% for Carlos' mother and 13 to 68% for Eduardo's mother.

Figure 1: Percentage of determiner use across NP types for Carlos' mother and Eduardo's mother at different points in the child's development
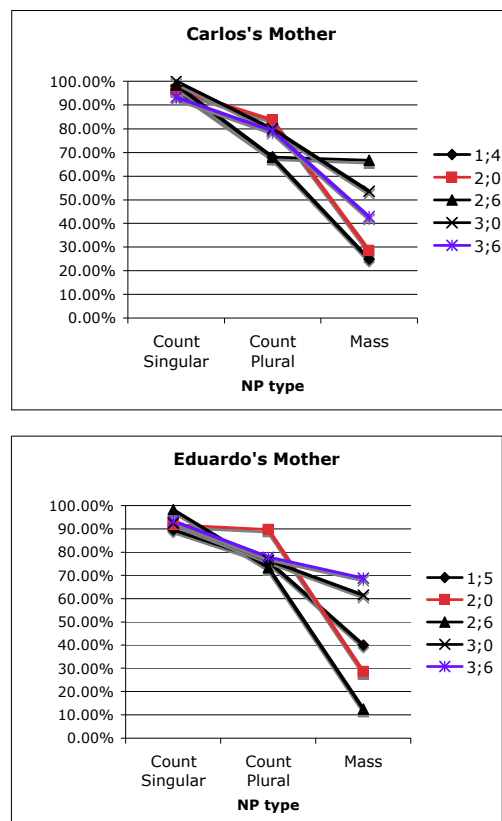
In Figure 2 we present the children's levels of determiner production by lexical category across the stages of development. For singular count nouns, both children showed high rates of determiner production, but still fall short of the parental target. For plural count nouns, there is a lot of fluctuation for both children. In Carlos' last two files, his determiner use is higher than the parental averages, whereas Eduardo's data shows that his determiner use in lower than the parental averages for plural count nouns. For mass nouns, Carlos showed the lowest levels of determiner use of all types of nouns, and his production was lower than the parental target. Eduardo, by contrast, showed 100% realization of determiners with mass nouns until 3;0, when production drops to levels comparable to the parental target, which is about 40%.

Figure 2: Percentage of determiner use across NP types for Carlos' and Eduardo's data
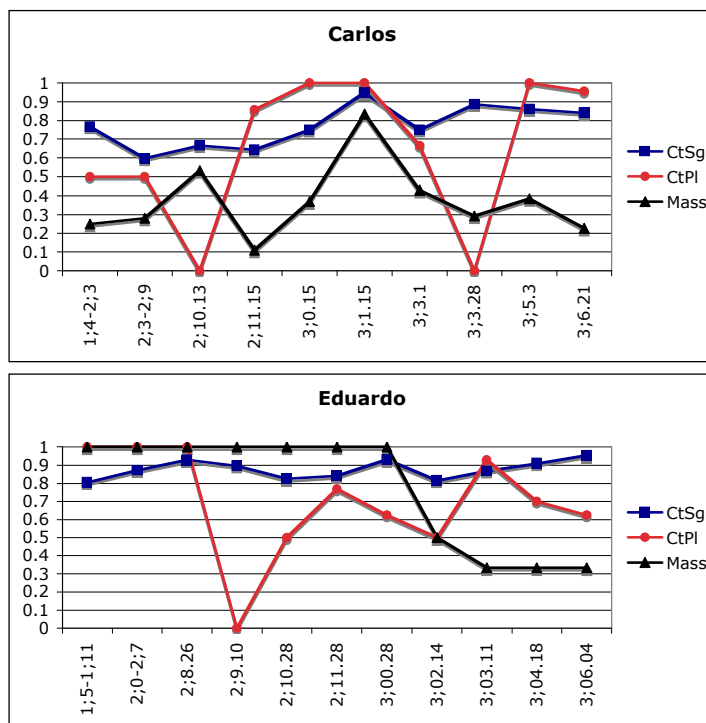
6.2. Syntactic factors

The second analysis considered the effect of syntactic function, and for subjects, the effect of position. In the adult data, preverbal subjects showed no determiner omissions, postverbal subjects and oblique objects showed 13 to 20% determiner omissions, and accusative objects showed 21 to 23% determiner omissions. Carlos' production of determiners with subjects shows a clear

advantage for preverbal subjects in contrast with postverbal subjects emerging in the third stage at 2;10. From that point on, he reaches the 100% target for preverbal subjects, and his production of determiners with postverbal subjects reaches target levels at 3;3. Eduardo produced no determined preverbal subjects during the first stage, but then met the 100% target determiner use at the second stage. He remained at that target except for two files at 3;0 and 3;2, where he produced again no determiners in preverbal position, and then returned to his previous level of 100% determiner production. Eduardo's data for postverbal subjects shows an initial omission stage, followed by a gradual pattern of increased determiner use that reaches the target range at around 3;0. Determiners emerge more gradually with both types of objects in Carlos' speech, and their production still falls short of the parental range by the age of 3;6. As for Eduardo, direct objects show a steady trend that reaches the parental baseline by 3;0, whereas obliques show much more fluctuation, and no clear convergence. This is summarized in Table 1.

Table 1: Determiner use by syntactic function for Carlos' and Eduardo's data

Position	Carlos	Eduardo
SV	Reaches target by 2;10.	Initially absent because of lack of common noun subjects; these appear at 100% D insertion at 2;11, then he produces determinerless SVs for two files, then continues to produce Ds for all.
VS	Reaches target by 3;3.	Reaches target by 3;0
Objects	Increases gradually for both accusative and oblique but does not meet adult range by 3;6.	Increases gradually; accusative objects reach target baseline by 3;0, obliques fluctuate and do not show clear convergence.

6.3. Determiner production

An additional source of evidence about determiner emergence at the earliest stages of development is provided by truncated NPs, which make up most of children's early production of NPs. Truncated NPs are nominal sentential fragments that appear in discourse, or in isolation. It represents about 20% of parental data, and in sizeable proportions in children's initial utterances, as these children develop from the holophrastic stage to increased production of multi-word utterances. An analysis of determiner use in these contexts allows us to look at determiner emergence separate from issues of position and function, which is useful for evaluating two of our hypotheses. Figure 3 presents the proportion of determined NPs in truncated NPs in the speech of Carlos and Eduardo, in comparison with the parental data extracted. The parents' use in these contexts ranges between 65% and 35%, as is represented by the black

squares. Both children's use of determiners is very low, but by 2;5 for Carlos and 2;8 for Eduardo it approximates the parental average.

Figure 3: Proportion of determiners in truncated NPs in the speech of Carlos and Eduardo
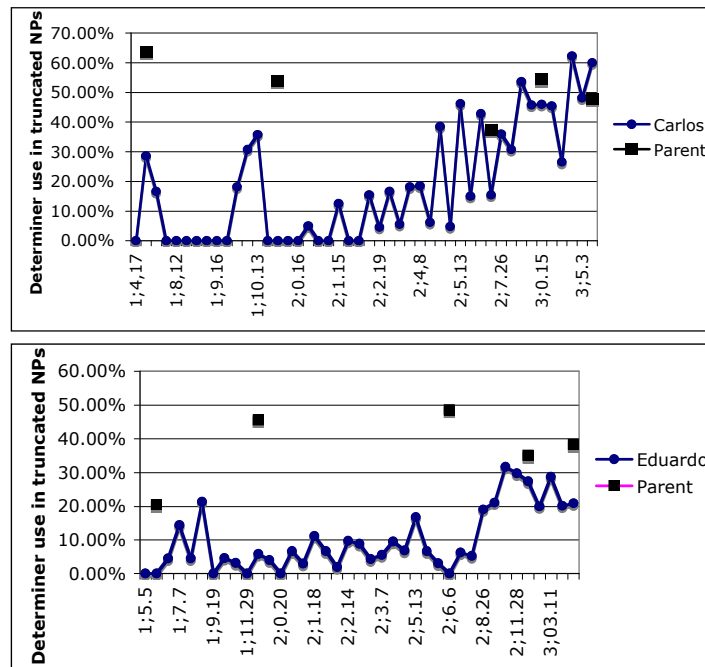
Carlos' production of determiners spikes at 1;10, but does not start to ascend until around 2;2, in a developmental curve that shifts upwards after 2;5. For Eduardo, there is a spike in use around 1;9, but production remains constrained until 2;8. These data suggest that the main developmental shifts occur around 2;8 for Eduardo, and 2;7 for Carlos, when they seem to enter the adult range of determiner production in truncated NPs.

7. Discussion and conclusion

The data presented above shows no evidence for a pure bare noun stage, since children used determiners from the outset. We observe a gradual pattern of convergence, with partial sensitivity to the syntactic and lexical factors that constrain BN distribution in the adult grammar. We evaluate each of the hypotheses discussed, in light of the data presented above.

H1: Parametric hypothesis

The first hypothesis predicted that children would exhibit awareness of the positional asymmetry, and of the count singular, mass and count plural distinction. The data suggest that both children in this study are sensitive to the function-position asymmetry. With regards to function and position, Carlos masters determiner use with preverbal subjects first, followed by objects. Eduardo masters determiner use for preverbal subject contexts first, and shortly after, for both postverbal subjects and objects. As for the mass/count distinction, both Carlos and Eduardo showed high rates of realization of determiners with count singular nouns. Carlos' plural data fluctuates widely, and is thus difficult to interpret, but Eduardo's plural count NPs had consistently lower rates of determiner use than his count singular nouns. Carlos' determiner production with mass nouns was consistently lower than for count nouns. However, for Eduardo there is no evidence of sensitivity to mass/count until 3;0.

H2: Interface delay hypothesis
As for the interface delay stage, the relevant ages are 2;3 for Carlos and 1;10 for Eduardo, dates which coincide with a rise in NP productivity. These is not surprising, since the higher the NP count, the more likely the probability that an overt subject will be produced. However, we see no clear association between the ages relevant to the Interface Delay Hypotheses, and increased production of determiners in truncated NPs. In Eduardo, the child with the clearest developmental pattern, productivity does not increase until 2;8, a full 10 months after the relevant age. In Carlos, the process of increase occurs in proximity to the relevant age, but in gradual process spanning between 2;2 and 2;6 that does not show the sharp patterns present in Grinstead (2004) data. Thus we argue that the determiner domain does not exhibit interface-dependent effects.

H3: Variable input hypothesis
The third hypothesis considered the relationship between domain variability and ages of development. The variational model predicts that variable and/or unreliable targets are slower to be acquired than systematic targets. This can be interpreted at the level of grammars (i.e., settings of a given parameter), or at the local level of input. If local variability rather than system variability matters, we predict an asymmetry in the ages of convergence for the lexical classes: CtSg>CtPl>Mass. Examining our results, we are limited by the fact that Carlos' plural data had too much variability to allow proper evaluation of actual age of convergence. For both his mass nouns and his singular count nouns, he approached the parental range from the beginning, but did not quite meet it. For Eduardo the results are much clearer: count singular nouns are in the target range from the outset, whereas mass nouns initially behave like count nouns, and then drop to the target range. Plural nouns stabilize around the target range later, at 3;4. We interpret the data to show that both children meet the parental target for singular count NPs first and produce high variability for plural count and mass NPs.

To summarize, these Spanish-speaking children have determiners from a very young age, but their initial use is not as generalized as in the adult's. Their overall rates of production in syntactically isolated NPs suggest they reach adult productivity around 2;7-2;8, beyond the ages predicted for the interface delay hypothesis. We also note that during the transitional stage, between first emergence and reaching parental baselines for context and category, children show syntactic asymmetries that go beyond the lexical groups. Their transitional data indicates sensitivity to the function-position asymmetry, and sensitivity to the categorical status of singular count nouns. There are indications that lexical learning is not yet in place, as one of the two children treated mass nouns as count nouns. The data also suggest sensitivity to input variability, as children were much slower in reaching adult baselines for plurals than for count nouns. We suggest that children attend to the categorical distinctions, and are able to use variability in the data to identify correct parametric stage. At the same time, local variability, such as introduced by the lexical mass/count distinction did not determine age of convergence. Both boys reached the mass noun baselines before plural noun baselines, although the rates of bare nouns were higher, and the parental data more variable for mass nouns. For one of them, the frequency differences between postverbal subjects and objects made no difference in age of convergence.

We conclude that this data shows that while children are sensitive to local variability in the input, developmental patterns are not fully determined by differences in the variability of domains and categories. Instead, the patterns observed in this study suggest that children interpret the asymmetries in the input, and select from the outset a system compatible with these asymmetries. Thus the results are compatible with the version of the parametric hypothesis where determiners are predicted to show asymmetries from the onset of productive use.

References

Aguado, Gerardo. 2000. *El desarrollo del lenguaje de 0 a 3 años*. Madrid: CEPE.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339-405.

Grinstead, John. 1998. *Subjects, Sentential Negation and Imperatives in Child Spanish and Catalan*. Doctoral Dissertation, UCLA, Los Angeles.

Grinstead, John. 2004. Subjects and interface delay in child Spanish and Catalan. *Language* 80(1):40-72.

Guasti, Maria Teresa, and Luigi Rizzi. 1999. *Agreement and Tense as distinct syntactic positions: Evidence from acquisition*. Ms. University of Siena.

Hauser, Marc, Noam Chomsky and W. Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve. *Science* 298(5598):1569 – 1579.

Kupisch, Tanja. 2006. *The acquisition of determiners in bilingual German-Italian and German-French children*. Doctoral Dissertation, Universitat Hamburg.

Legate, Julie Anne and Charles D. Yang. 2007. Morphosyntactic learning and the development of Tense. *Language acquisition* 14: 315-344.

López-Ornat, Susana. 1999. La adquisición del lenguaje: nuevas perspectivas. In *Psicolingüística del español*, coord. Manuel de Vega and Fernando Cuetos, 469-533. Madrid: Trotta.

López-Ornat, Susana. 2003. Learning earliest grammar: evidence of grammar variations in speech before 22 months. In *Linguistic theory and language development in Hispanic languages,* eds. Silvina Montrul and Francisco Ordoñez, 254-274. Somerville, MA: Cascadilla Press.

Marinis, Theodoros. 2003. *The Acquisition of the DP in Modern Greek.* Amsterdam: John Benjamins.

Miller, Karen L. 2007. *Variable input and the acquisition of plurality in two varieties of Spanish.* Doctoral Dissertation, Michigan State University.

Tomasello, Michael. 2000. Do Young Children Have Adult Syntactic Competence? *Cognition* 74(3):209-253.

Ullman, Michael T. 2001. *The declarative/procedural model of lexicon/grammar.* Journal of Psycholinguistic Research 30, 37-69.

Yang, Charles D. 2004. Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10):451-456.