1997 Susana López-Ornat
Contemporary Perspectives on
the Acquisition of Spanish.
Volume I: Developing
Grammars. Ed Ana Teresa
Pérez-Leroux and William
R. Glass, 3-20. Somerville,
MA: Cascadilla Press.

What Lies in between a Pre-Grammatical and a Grammatical Representation? Evidence on Nominal and Verbal Form-Function Mappings in Spanish from 1;7 to 2;1

Susana López-Ornat
Universidad Complutense de Madrid

1. The problem

The simplest Spanish Noun Phrase, as produced by children around 26 months (Hernández-Pina 1984, Aguado 1988, 1995), combines a determiner marked for gender harmony and a noun, which has gender. For our research into the earliest grammaticalization processes, we have analyzed the acquisition process of those simplest NP productions which are rarely plural, do not include a gender-marked adjective, have unambiguous gender marking, do not have gender harmony as the neutral case, and have an article for determiner.

In these earliest NPs, gender is marked in both the indefinite (*un-una*) and the definite (*el-la*) articles: *[l]a mano* 'the hand, irregular feminine,' *u[n]a quéca* 'a doll, feminine,' *e[l] pié* 'the foot, masculine,' *u[n] camión* 'a lorry, masculine.' The indefinite determiner *otro/otra* 'another+gender' is very frequently produced by the youngest Spanish speakers, always in the singular form: *óto cóte* 'another+masculine car,' *óta fó* 'another+feminine flower.' Nevertheless we have excluded this determiner from our analyses as it can be produced alone, pronominally.

Neutral case phrases aren't produced at that age; they're pronominal and thus of later acquisition, and did not appear in the data. Adjectives at that age are very limited (*mamá guapa* 'mummy pretty+fem.,' *nene bueno* 'child+masc. good+masc.,' *papá malo* 'daddy bad+masc.') or they are morphologically invariant (*gande* 'grand/big') with both feminine and masculine nouns. Ambiguous nouns such as *el mar/la mar* 'the sea' are not used at that age and did not appear in the data.

From a morphological point of view, the gender harmony rule for the Spanish NP is a group of regulated exceptions to the intuitive basic "rule" that classifies nouns ending in *-a* as feminine and ending in *-o* as masculine. This agreement rule would also inaccurately mark final *-o* for determiners and adjectives of masculine nouns, and final *-a* for determiners and adjectives of feminine nouns (Alarcos LLorach 1984; R.A.E. 1994). The problem for the

© 1997 Susana López-Ornat. *Contemporary Perspectives on the Acquisition of Spanish, Volume I: Developing Grammars*, ed. Ana Teresa Pérez-Leroux and William R. Glass, 3-20. Somerville, MA: Cascadilla Press.

learner is knowing from the noun what is the appropriate gender-marked determiner form.

Experimental studies have established the intralinguistic nature of this acquisition process – its relative independence from pragmatic, semantic, or conceptual knowledge. These studies include evidence from Spanish children after 3;0 years of age (López-Ornat 1977) and from French children (Karmiloff-Smith 1979), replicated with Spanish children (Pérez Pereira 1990). Recent experimental-longitudinal studies with young Spanish children (1;6-2;6) can be found in López-Ornat 1996 and Mariscal 1996. Of course, semantically consistent cues to gender harmony exist for animate objects with natural gender, e.g., *el tío, la tía* 'aunt, uncle.' Thus the problem of knowing the gender of the noun is pertinent only for referents of inanimate objects: "the correlation of gender to semantic content is problematic and neither constant nor regular" (R.A.E. 1994).

By age 26-30 months, Spanish children are producing simple grammatical NPs, with gender agreement that is correctly marked in inanimate nouns though there are some, notoriously scarce, errors. In Rafael (Hernández-Pina 1984) at age 2;0 we find six reported errors:

"una pájaro"	(*un pájaro* 'a bird')	"un llave"	(*una llave* 'a key')
"el silla"	(*la silla* 'the chair')	"una camión"	(*un camión* 'a lorry')
"una pez"	(*un pez* 'a fish')	"una gato"	(*un gato* 'a cat')

From a phonoprosodic point of view, the article+noun productions constitute a single phonic group: the melodic unit in Spanish. Articles are non-stressed and there is articulatory continuity linking them to the typically stressed first syllable of the Spanish two-syllable noun (García Calvo 1989, Quilis and Hernández Alonso 1990). (For further details, see section 4.4.) Cues to gender subregularities are phonological, but complex; i.e., the best known of those subregularities is that nouns ending in *-a* are usually not masculine, except for Greek etymology: *el fantasma, el poeta, el taxista*, etc. (see section 6.2).

2. Goals and designs

This study examines the phonoprosodic cues which might be used by the child to bootstrap into morphology. It is based on longitudinal production data and explores first the possibility of there being a constant set of phonoprosodic features that characterize preNPs against a set of features that characterize preVerbal phrases. The constant preNP phonoprosodic form would include a gender specification feature. This is evaluated by analyzing, one by one, the phonoprosodic form of all the preNPs produced by María from ages 1;7 to 2;1 and of all her earlier (1;7-1;9) preVerbs.

Our first hypothesis is that the change from pregrammatical to grammatical NP representations is aided by previous phonoprosodic knowledge and grows step by step. To evaluate this we will compare earliest (1;7-1;8), intermediate (1;10-1;11), and later (2;0-2;1) production data. A second complementary hypothesis tests the linguistic categorization inference that is implied here: if

there is a phonoprosodic definition of the function type "NP," a different phonoprosodic definition should be in use, at the same developmental moment, for a different type: the preVerbal phrase.

This will be evaluated by analyzing the phonoprosodic form of all the preVerbal phrases produced by María from ages 1;7 to 1;9. This analysis will also show any gradual changes in the phonoprosodic structure of the preVerbal productions. Both preNP and preVerbal tokens will be analyzed in terms of the following phonoprosodic features: number of syllables, stressed versus unstressed syllables and their relative order in the productive strings, presence/absence of phonemes which do not belong to the noun or the verb in the adult model and relative position of those.

All data are extracted from the database of the child María. This database contains the longitudinal corpora of a Spanish child from Madrid.[1] She is an only child who was videotaped from ages 1;07 to 4;00 every fortnight in sessions of about 30 minutes. Those sessions took place at home during bath, play, or feeding interactions with her parents, who belong to a professional middle class.

3. Background

3.1. The nature of premorphological representations

In this study, both the nature of pregrammatical representation and the issue of phonoprosodic bootstrapping into morphology are framed in a step-by-step view of grammatical acquisition in which the acquisition process is seen as continuous and its steps consist of acquiring the partial linguistic subregularities which define the rule. A main theoretical point is that the pregrammatical system is mapping a phonoprosodic form onto a pragmatic-semantic function. This mapping would allow storage of candidate NP exemplars, which would then be analyzed by the system in search of linguistic subregularities. In this sense phonoprosodic knowledge would be a cue to morphological representation.

Our research aim is not describing the nature of the underlying representations but the dimensions along which they change. The longitudinal microanalysis allows us to infer the nature of this representational shift, irrespective of whether or not the object called representation corresponds exactly to the form we infer from production. In turn, this background raises several theoretical and methodological questions. In the first place, what is the nature of pregrammatical representations? Answers to that question are based on the nature of amalgams (Bloom 1973, MacWhinney 1978, López-Ornat 1994) and Non-Analyzed-Units (Bowerman 1982, López-Ornat 1990). Those are labels for pregrammatical productions with an underlying pregrammatical representation. In addition, these pregrammatical productions have been described by Peters (1985, 1995) and Peters and Menn (1993) as formulaic expressions and as productions with "schwa" fillers or protomorphemes. Plunkett (1993b) discusses them and proposes the presence/absence of phonological assimilation for the differentiation of words vs. amalgams.

In these latter studies, phonoprosodic features are used to describe pregrammatical productions. Contrary to other interpretations (Gerken and McIntosh 1993), I claim that form-function definition takes place simultaneously at some particular level, since functions, pragmatic-semantic ones in this case, cannot be categorized in the absence of a form, i.e., some stable type of linguistic sound. As a consequence, we should find degrees of growth of grammaticalization along longitudinal evidence instead of a sudden or a generalized grammaticality shift.

As an initial assumption in this work a preNominal amalgam is defined as a non-morphological production functioning in context as a primitive NP (preNP), with a form that is acknowledged by the adults to be the child's version of a particular (future) NP. This form allows the child to discriminate between different Nominal meanings during the pregrammatical phase of this development (from about 1;6 to about 2;0 in Spanish children). I'll term the form an "amalgam" because it fuses phonology which corresponds to the article and to the Noun.

If our first hypothesis is supported by the data, this will raise a second issue – a question on preNP versus preVerbal categorization: If a particular child, María from 1;7 to 2;1, is using particular phonoprosodic knowledge to produce, and probably also to segment, primitives of NPs (preNPs) from her environmental language and mapping them onto discriminated meanings, then she should be using a different type of phonoprosodic form to segment her primitives of Verbal Phrases (as given in her preVerbal productions) from the same environmental model, and she would be mapping them onto a different set of discriminated meanings at the same point in development.

The evaluation of these two hypotheses demands, in our view, longitudinal microanalysis, because microacquisition processes are involved: i.e., the child's definition of a linguistic rule (in our case the gender harmony rule of the Spanish NP) would result from the gradual discovery of a number of smaller (micro) subregularities along time. The developmental effect we call a "rule" would result from the gradual acquisition of a number of partial form-function pairings: feminine, masculine, neutral, ambiguous, minus their many exceptions. And, last but not least, in this process other types of linguistic knowledge (semantic, syntactic, certainly pragmatic, etc.) are bound to be used by a child like María for the ultimate goal of grammaticalization. María's functional knowledge of those phonoprosodic forms would stem from those other types of analysis her system is doing simultaneously. There are several studies that can be consulted for other bootstrapping mechanisms, i.e., structural, morphosyntactic, or distributional cues (Naigles 1990, Gerken and McIntosh 1993), and semantic cues (Bowerman 1982, Pinker 1984, 1989).

3.2. Representational shift: methodological issues

One limitation of data in production is their dependence on attention. In general, our production-based inferences on underlying linguistic representations can be misled by the performance variables affecting speech editing (Levelt 1989). We need to take attention for granted in production, assuming that the

inferred representation has been successfully used by the system to generate that particular production. The system will primarily have used other types of representation, including non-linguistic ones, in the generation of any particular production.

Minimizing the first of these problems is usually done by analyzing a very high number of different productions of the same structure during one age period, so that the effects of attention-driven performance variability are counteracted. But the main problem, our building inferences on the nature of representations underlying production, remains. As Bates et al. (1995) observe, this problem has no solution within longitudinal methodology but needs a combination of experimental approaches to solve it. Nevertheless, if our research focus is not on the nature of the underlying representations, but, instead, on their change process, then, longitudinal analysis can reliably infer representational shift, regardless of whether the object called representation corresponds exactly to the inference we made from the produced form, or not.[2]

4. Study I: representation of preNominal phrases

4.1. Hypothesis

We hypothesize that the change from pregrammatical to grammatical representations happens step by step, implying that representations for language become "more and more" grammatical through time. The process of acquiring a rule should be analyzable in sequenced microlearning processes for each of the grammatical subregularities which make it up as a rule. If this is the case, this grammaticalization process would consist in slowly adding new analyzed subregularities until a grammatical rule could be defined for a set of exemplars ([pre]NPs) even if in a simplified version (Art+N).

The evaluation of this hypothesis requires an operational definition of the initial and end states of this acquisition process. We're taking the morphologically correct Art+N productions, e.g., *e[l]/u[n] coche*; *[i]a/u[n]a muñeca*, as our end state. This form of the NP is only its simplest form but it is fully grammatical and contrasts significantly with the pregrammatical (initial state) productions: *[e,o,u]cóte*; *[a,ú]quéca*, where vowels preceding the noun are not productive. The phonomorphological structure of the Art+N is clearly defined. But what, if any, is the structure of the pregrammatical form?

A first analytical task is, then, establishing what are the relevant structural (phonoprosodic) features of the pregrammatical form as produced by the 1;7 to 2;1 Spanish child in the process of acquisition of gender harmony. This was based on the observation of the protocols of our experimental subjects, aged 1;6 to 2;2 (López-Ornat 1996). In this procedure the child is at home playing with experimenter 1, while experimenter 2 is videotaping the scenes.

Whenever the child produces a preNP, coded as such from context analysis, experimenter 1 immediately asks *¿Qué?* 'What?' as if he hadn't heard the child well. The child's task, in this control group, is only to repeat (=R2) its previous (=R1) production. This task is supposed to obtain from the child an identical response to his previous one. But identity was predicted at the morphological

level; and the protocols show there are some phonoprosodic variations from R1 to R2, as well as a set of phonoprosodic features which do not change from R1 to R2. As an example, let us look at the protocol of the child Fuencisla, age 1;6.

Table 1.
Fuencisla 1;6: analysis of R1/R2 variability

R1	What ?	R2	
abó		abó	(huévo)
achú		achú	(súgus)
tá		tá	(táza)
acó		acó	(muñéco)
cá		acá	(váca)
cá		acá	(cása)
pí		pié	(pié)
epá		apá	(el pán)
má		má	(máno)
acó		có	(conéjo)
cúca		cá*	N.W. (cúca)
acóco		cóco	N.W. (cóco)

Constant R1 Form=Pregrammatical=[v]stressed sb.[sb]
* stress change: sb final → v final
(N.W. = New Word = cúca/o): the experimental procedure included for generalizability purposes the child's learning of two new words (here: *cúca/cúco*) with masc./fem. formal values which, on first spontaneous production, underwent the same elicitation by E. (here, the control condition: *¿Qué?*).

As can be seen in Table 1, if we analyze the prosodic structure (Echols 1993) of her preNPs we find that one stable (constant) formula, coded in phonoprosodic terms (see section 6.1), accounts for all her preNPs: "start with an optional determiner's vowel sound, articulate a compulsory stressed syllable and end either there or adding a last nonstressed one": [v]sb'[sb]. As the basic Spanish syllable is CV (see section 4.4), her preNP amalgams tend to start and end in a vowel sound. Fuencisla also manipulated the optional prefixed determiner's vowel combinatorially, either dropping or adding it in her response (R2).

We can then take the above formula as a form definition in the initial state:

1. Optional phonological knowledge about the nature of the phonemes which can accompany a preNoun: a/e/o/u and of those which can't: /i/ (see section 4.4).
2. Order knowledge: those phonemes, if produced, precede the "noun."
3. A compulsory stressed syllable on its own or followed, but not preceded, by a non-stressed one.

Having defined our initial and end states in this grammaticalization process, let us now proceed to the evaluation of our first hypothesis. This hypothesis stated that the growth of grammatical representations should be sequenced in complexity, by adding partial bits of grammatical knowledge to the preexisting preNominal form, and that those "bits" would consist of different NP+gender harmony subregularities.

4.2. Methods

We tested the first hypothesis across our longitudinal database by looking at the structure of all of María's pregrammatical NPs during the 1;7 to 2;1 period. We extracted from our longitudinal database all the preNPs that María had produced from 1;7 to 2;1. We thus intended to scan all her amalgams and also all her intermediate forms up to the first Art+N stable productions, which are taken here as an end state.[3]

All instances of NPs were extracted using the COMBO program in the CLAN pack (MacWhinney 1991, 1995), adding "+y" to the commands as it is a non-CHAT format. Thus we obtained:

Ages	Total N of preNPs	masculine	feminine
1;7 & 1;8	N=194	123	71
1;9 & 1;10	N=260	162	98
1;11 & 2;0	N=137	66	71
2;1	N=106	56	50
Total	N=697		

All strings containing a preNoun or a Noun were included. Determiners were reduced to the basic type: the article in its feminine and masculine forms and in its definite/indefinite form: *el-la/un-una* (and, rarely, the plurals: *los-las/unos-unas*). Repetitions of adult utterances were excluded from the analysis. The program was used first to obtain all the masculine strings, then the feminine ones. By manual analysis we separated the strings which contained Determiner omission errors (0N), from those which prefixed a vowel (v) to the preNoun (vN). Data analysis proceeded in chunks of two months. We extracted from the total [pre-]NPs (=697) a subtotal of those Nouns which were produced more than once. This allowed us to analyze form variability within the same preNP (where same indicates identity of meaning).

Out of the Lexical total (total different Ns) we analyzed the following categories, for every age period:

- [pre-]NPs produced with a determiner omission error = 0N
- [pre-]NPs produced with a prefixed vowel = vN
- NPs produced in their simplest version = Art+N

This analysis was also done for the subtotal of NPs which had been produced more than once, obtaining the following elements:

- [pre-]NPs which were produced, in one same age period, with 2 or 3 alternative forms: 0N, vN, Det+N.
- preNPs which were only produced as 0N
- preNPs which were only produced as vN
- NPs which were only produced as Art+N

4.3. Results: from preNPs to simple NPs

For this analysis we considered as masculine or feminine the [pre-]NP forms which are masculine or feminine in the adult model. The total number of exemplars was 697. Out of those we separated the Subtotals of masculine/feminine exemplars. The evolution of the masculine and feminine forms of [pre-]NPs in María is in Table 2.

Table 2.
María: evolution of the masculine and the feminine forms of [Pre]NPs

		Ages			
		1;7 & 1;8	1;9 & 1;10	1;11 & 2;0	2;1
Form	0N	47.1% masc. 33.8% fem.	32.1% masc. 32.6% fem.	16.7% masc. 16.9% fem.	10.7% masc. 6% fem.
	vN	51.2% masc. 63.4% fem.	62.3% masc. 64.3% fem.	57.6% masc. 33.8% fem.	30.3% masc. 10% fem.
	Det+N	1.6% masc. 2.8% fem.	5.5% masc. 3.1% fem.	25.7% masc. 49.3% fem.	59% masc. 84% fem.
	Totals	masc.=123 fem.=71	masc.=162 fem.=98	masc.=66 fem.=71	masc.=56 fem.=50

0N = the omission of a determiner in an obligatory context plus the production of a preNoun: (1;7):"pí" for "el lápiz",when "el" was compulsory.
vN = the production of a Det.'s vowel (v) prefixed to a preNoun:(1;7): "acóte" for "el coche."
Det+N = the production of a simple NP where: Det = Article.

The data in Table 2 show the following:

1. The three forms: 0N, vN, Det+N are developmentally sequenced. 0N and vN are two different types of pregrammatical forms, and Det+N is a linguistic structure.
2. The feminine value of the morpheme in the gender harmony rule seems to be regulated, in María, slightly ahead the masculine one, at all age periods except 1;9 and 1;10.
3. Linguistic representations are being acquired with greater ease for one of the conditions of the rule (the feminine subregularity, in María's case). This could depend on higher form-function transparency for feminine in Spanish (see section 6.2), and/or on articulatory facilitation in the case of the feminine Det+N groups (Mariscal 1996).
4. The longitudinal nature of these data limits our inferences to the nature of the representational shift. Change occurs by using phonoprosodic analysis to segment and produce preNP forms. Mapping these forms onto their preNominal pragmatic function allows their storage. This is the prerequisite for their inner analysis in search of higher order, i.e., morphological subregularities.

4.4. Preliminary conclusions: the structure of vN forms and the question of uniqueness of phonoprosodic units

Syllables are the rhythmic units of the Spanish language (Alarcos LLorach 1984, García Calvo 1989, Quilis and Hernández Alonso 1990, and R.A.E. 1994). A vowel is defined as the sort of phoneme which can constitute a syllable on its own, as opposed to consonants, which can't. The melodic unit in Spanish is the "grupo fónico" or "grupo melódico." I'll term it "phonic group." This group is marked by pauses in articulation. Every pregrammatical form (be it preNP or preVerbal) in the longitudinal samples we used was a phonic group.

A detailed way of coding a phonic group, if we are not to include musical notation, could be, as we've done for preNPs and for preVerbs, representing their order, number, and type of syllables, their stress pattern, and the order and existence of a [vocalic] phoneme [v] which does not belong to the Noun nor to the Verb in the model language.

This prefixed vowel forms part of the phonic group but it does not only represent an articulatory support but a linguistic constriction. Going through all of María's [pre-]NP productions we find v is most frequently /a/ with a feminine [pre-]Noun; it is either /e/, /o/, or /u/ with a masculine [pre-] Noun; but v is never /i/. One linguistic subregularity captured by the child is that /i/ is the only vowel which does not exist in the Spanish determiner system: neither in articles in any position, nor in any other determiners in any position.

However, the vN group also implements, as seen above, a gender-harmony lawful distribution: /a/ is the suffix in all of the model language's feminine determiners (*la*, *una*, *esta*, *esa*, *aquella*, and *otra*), while /e/, /o/, and /u/ belong

to the masculine determiners (*el*, *un*, *este*, *ese*, *aquel*, and *otro*), in different positions. So the v in vN is a phoneme of morphological relevance, a protomorpheme (Peters and Menn 1993), because it captures some linguistic subregularities.

The prefixed v is then growing into a linguistically constrained unit. In addition, we saw how Fuencisla could manipulate that v, adding or suppressing it in a single form, a manipulation which is in agreement with the linguistic (combinatorial) character of this unit. The form vN is an intermediate value in representational change. This value is closer to grammatical than 0N, because it already implements some grammatical subregularities:

- vN preserves grammatical word order in the Spanish NP.
- vN roughly discriminates masculine/feminine values in gender harmony subregularities carried in the protomorpheme v.
- vN is not just chronologically ahead of 0N, but is also formally ahead.

Returning to the analysis of Table 2, in order to consider an early phonoprosodic categorization in María's acquisition of gender harmony in the NP, we need to show that she was mapping a different (and also stable) phonoprosodic unit onto preVerbs at that particular age period (1;7 to 1;9). This led to the longitudinal analysis of all those preVerbal forms.

5. Study II: representation of preVerbal forms and functions

5.1. Hypothesis

Since our first hypothesis is supported by the data, we should find other different types of phonoprosodic forms being mapped onto preVerbal functions at the same age. Otherwise the linguistic categorization role we are inferring for phonoprosodic knowledge would be untenable.

It would be harder to argue that María's vN productions were part of a grammaticalization process if we found that she uses the same phonoprosodic form [v]sb'[sb] for her preVerbal forms in the same period. Our hypothesis would be strengthened if we found her preV utterances differ in phonoprosodic structure not only from preN utterances but also from each other depending on the pragmatic function played by the preV form. Within each pragmatic role we would like to find some constant phonoprosodic features.

5.2. Methods

As before, the COMBO (+y) program was used to extract all of María's preVerbal forms from the database. Again, the subtotal of forms which were different lexical items (not repeated) were separated from the total. If a form had been imitated (repeated immediately after an adult's use of it) it was excluded from the data. The pragmatic context-dependent criteria we had used in the original coding of the database enabled us to allocate the preVerbal utterances to four functions:

1. Infinitive:
   From 1;7 María produces a primitive form of the periphrastic future preceding an action, e.g., *a jubá* for *vamos a jugar*. This is a common form along young Spanish children (see Hernández-Pina 1984, Fernández 1994, and Aguado 1995 for this and the following empirical references). Although these preVerbals are not true infinitives, we will use that term for convenience.
2. Imperative:
   From 1;7 María, like Spanish learners in general, produces a primitive imperative ([*m*/*v*]*en*). These forms are used only in their second person singular inflection whenever children are demanding action from their environment. These are not true imperative forms.
3. Present:
   From 1;7 María, like Spanish children in general, uses a third person singular present indicative form in, again, an infinite version: its use is impersonal and it appears in enunciation contexts: *cóme*, *cáe*. It is not a verbal form yet, although we've coded it Present.
4. Present Perfect:
   From 1;9 María, like other Spanish learners, uses a simplified form of a third person singular present perfect (*ha roto*, *ha caído*). As with the other forms it is used in an impersonal fashion to refer to a recent past. The form used is a vowel plus the past participle (*a lóto* for *ha roto* or *a caío* for *ha caído*). The vowel, as before, belongs to the grammatical model. I've termed this Present Perfect (P.Perfect).

All of María's preVerbal forms produced during this three month period were analyzed by hand. The same phonoprosodic criteria used for her preNPs were applied now. The total number of preVerbal exemplars was:

Age	Total N of preVerbs	Ratio of lexically different preVerbs
1;7	N=82	14/82
1;8	N=33	11/33
1;9	N=70	20/70
Total	N=185	

For every preVerbal production (N=185), its phonoprosodic form and its preVerbal function were obtained. Next, each function group was analyzed for its constant form, if one existed. Variations in form within one repeated preVerbal item are coded [ ] as optional units. Phonoprosodic units without brackets represent the forms which are always produced, the formal constant across lexical repetitions and across different lexical items for a single function, such as present.

5.3. Results: preVerbal versus preNominal mappings in María from 1;7 to 1;9

The results are shown in Table 3. The first column gives the phonoprosodic form with brackets denoting optional elements – elements that were sometimes present and sometimes absent in an utterance with that function. The second column gives the function the utterance played in the child's discourse, the third column shows what proportion of different lexical items fell into that functional category, and the final column gives the proportion of total utterances, including repetitions of lexical items, in that category.

Table 3.
Form-function mappings for preVerbal productions

María 1;7-1;9

Age	Phonoprosodic Form	PreVerbal Function	Lexical Frequency	Total Frequency
1;7	v [sb] sb'	Infinitive	71.4%	53.6%
	[sb] sb'	Present	14.2%	37.8%
	sb' [sb]	Imperative	14.2%	8.5%
Subtotal			(N=14)	(N=82)
1;8	v [sb] sb'	Infinitive	45.4%	39.3%
	[sb] sb'	Present	27.2%	33.3%
	sb' [sb]	Imperative	27.2%	27.2%
Subtotal			(N=11)	(N=33)
1;9	v [sb] sb'	Infinitive	20%	12.8%
	[sb] sb' sb	Present	60%	62.8%
	sb' sb	Imperative	15%	22.8%
	v sb' sb	P.Perfect	5%	1.4%
Subtotal			(N=20)	(N=70)

[ ] = optional unit in María's functional production.
sb = non-stressed constant syllable.
sb'= stressed constant syllable.
v = a "linguistically relevant" vowel, a protomorpheme. In Spanish this vowel is also a syllable, but coding "sb" would have missed its linguistic relevance (see section 4.4).

To take an example to illustrate how these data arise: at 1;7 María produces 14 examples of *a tapar* 'let's cover [it] up' and she uses two forms: *a tapá*, coded as v sb sb' and *apá* coded as v sb'. Putting these together as v [sb] sb' accounts not only for all these 14 examples but also for all the other 26 examples of Infinitives at this age.

If we compare these preVerbal phonoprosodic forms (see Table 4) with the phonoprosodic form produced for preNouns at 1;7 and 1;8, [v] sb' [sb], we see that this pattern is not being used at that age for any preVerbal function. This is encouraging because it supports our categorization hypothesis.

Table 4.
María: Compared form-function mappings from 1;7 to 1;9

Function	Form at 1;7 & 1;8	Form at 1;9	Form Change
preVerbal Infinitive	v [sb] sb'	v [sb] sb'	0
preVerbal Present	[sb] sb'	[sb] sb' sb	add end sb
preVerbal Imperative	sb' [sb]	sb'sb	end[sb] →end sb
preVerbal P.Perfect	0	v sb' sb	(new)
preNP	[v] sb' [sb]	v sb' [sb]	pre[v]→pre v

If we look at the forms used for each preVerbal function at each age, the functions take distinct forms. Moreover the forms change over age in an orderly manner: the Infinitive form is stable across this three month period; the Present form remains stable over 1;7 and 1;8 and is then expanded by the addition of an unstressed syllable; the Imperative is stable over the first two months and at 1;9 the second unstressed syllable changes from optional to always present; and the P.Perfect, which appears for the first time at 1;9, takes a novel form that has not been used for any of the previous functions.

The phonoprosodic analysis of María's productions of both preNPs and preVPs strongly suggests that these forms are well on the way to playing a grammatical role before 2;0. If this is the case, then the obvious questions are what is the origin of these functionally distinct phonoprosodic forms and how are the changes from pregrammatical to grammatical structures best characterized.

6. Conclusions

6.1. Phonoprosodic representation and bootstrapping into morphology

Across all analyzed functions we find that, for example, stressed syllables are never optional – they are common to every form. This feature is then being selected by the grammar learner as a productive structure. So what are the features which determine form differences? Some phonoprosodic features have

been isolated in this study which are being used by the child as cues into morphology. Those features are:

- Syllable order: pre-position and post-position of stressed-unstressed syllables (see the difference between Present and Imperative).
- Inclusion of a premorphological filler (Peters and Menn 1993, Echols 1993) in a phonic group: presence of a prefixed vowel which in 5 months will be substituted by a grammatical determiner, to which it belongs. For example, preposed /u/ in vN (*úpáto* 'a duck') will be substituted by a preposed masculine article in Det+N (*un pato*). Also preposed /a/ in preinfinitive forms will be substituted by preposition *a* in a periphrastic future verbal clause.
- Optionality: optional syllables are made compulsory, as in preVerbal Imperative.
- Adding new "filler" units to the phonic group: a 'to-be' finite form, preVerbal Present at 1;9, develops an extra final non-stressed syllable. Later in acquisition process, the verbal inflections will be non-stressed suffixes.

These are form differences which allow us to infer that a phonoprosodic analysis of linguistic input may be coding, storing and accessing relevant linguistic exemplars before they can be represented in morphological terms. We now raise the question of how these distinctions can aid morphological learning.

One of the mechanisms which serve the grammaticalization process is certainly manipulating the order of stressed/unstressed syllables in the pregrammatical phonic groups and mapping the resulting variations onto different pregrammatical functions. We've seen how the system can also substitute its optional prosodic units by compulsory ones, thus lengthening its phonic groups.

Lengthening (i.e., adding a "filler") provides the space needed for future inflections. One way of lengthening is, as we've seen, defining a prefixed premorphological unit within the phonic group (vN). This unit shares some of the linguistic properties of the corresponding morpheme in the adult model (the determiner's gender-marker and its position before the Noun). Another method of lengthening is adding a suffix, like the final unstressed syllable in the pre-Present; this suffix shares some of the linguistic properties of the model's morphology (Spanish verbal inflections are suffixed and unstressed).

This process would certainly not exclude other types of bootstrapping operations which could be acting simultaneously on the same grammaticalization process. As Peters (1995: 463) puts it: "at times and for some children phonology acts as a kind of "handle" for getting hold of grammatical morphemes. At other times and/or for other children it is the function or 'meaning' of the morpheme that serves as the initial nucleus for collecting information about a particular functor."

In our case, the intermediate unit vN is not produced by all children. An overview of the protocols of our experimental children (López-Ornat 1996) has raised a question as to how gender harmony would be acquired by children who seem to jump from 0N forms to Det+N forms. Before establishing the existence of such a jump and labeling the jumpers as different we will need to microanalyze

their 0N forms along time, to establish whether there is any subtle change in the form of 0Ns used before the use of Det+N in those children.

This demands experimental-longitudinal work, in a combined fashion, as discussed in section 3.3. In the meantime, readers interested in fine-grained approaches to subject-bound differences in language acquisition can look at Plunkett 1993b, Peters and Menn 1993, and Bates et al. 1995.

6.2. Step-by-step grammaticalization and connectionist modeling

One general consequence of this study is that in order to explain the grammaticalization process our research designs are required to capture "more or less grammatical" steps in acquisition. This is in contrast to "0/1" designs where the possibility of continuity is not contemplated (López-Ornat 1991). It was that sort of design which rendered results that prompted this research, and it was the detailed coding of longitudinal data (López-Ornat et al. 1994) which has allowed us to recover these transitional forms.

For María, the grammaticalization of gender harmony in the NP starts after 1;7, when she starts modifying her 0N preNP phonic groups into vN groups which capture some linguistic subregularities. This process lasts until 2;0, when she starts producing Det+N structures. In addition, at 2;0 what we have is not the acquisition of the gender harmony rule but the acquisition of a defective (reduced) version of it (López-Ornat 1990, 1994). This version is defective because by age 2;1, María only produces simple Det+N NPs with gender harmony.

Her grammaticalized representation is really a simplified version of the rule. It excludes adjectives and pronouns, as well as neutral case and ambiguous nouns. It rarely contemplates plural-marked productions in any gender. The gender harmony acquiring process goes on after grammaticalization is achieved. It continues both by adding new units to the clause, such as an adjective, which challenges generalization of the gender harmony rule, as well as by adding new morphological subregularities, such as neutral case or plurals, to the NP.

Connectionism provides a powerful tool for the modeling of continuous learning processes like this (Abrahamsen 1993, Clark and Karmiloff-Smith 1993, Plunkett 1993a). A strong test of the phonoprosodic bootstrapping hypothesis presented in this paper would be to account for the lawful changes seen, for example, in María's preNoun productions over 1;7 to 2;1 purely on the analysis of regularities in the phonoprosodic input to the language learner. In collaboration with colleagues (Nix 1997), I am currently exploring a neural network model in which the input is a phonoprosodic representation of noun phrases in the adult's speech to María over the period presented in this paper.

The architecture of the net forces it to produce representations of all the inputs within the same n-dimensional space (in this project, 10 dimensions). What is of interest is to see how the regularities, subregularities, and exceptions related to gender harmony are represented as the net is increasingly exposed to the input as María was. Of course, the database on which the model is based is only that speech which was directed to María during the observations and in our model it lacks all semantic and pragmatic context.

In this case, the poverty of the input is its strength as a test of the hypothesis that phonoprosodic features may enable a learning system to detect and represent distinctions which will provide a scaffold for morphological structure. If this modeled growth process proceeds by gradually adding new partial subregularities to the NP representations, we must inquire as to which subregularity goes first; e.g., why would María start off from feminine?

Although we already have evidence from Mariscal (1996) that not all children start off from feminine, but from masculine exemplars which rate very high in frequency in their particular experience, feminine is the morphological value which rates higher in form-function transparency in Spanish. This will predictably determine a significant number of early acquisition processes, as it seems to determine María's.

We argue that the feminine rates higher in form-function transparency for the following reasons:

- nouns ending in *-a* are rarely masculine
- feminine noun determiners carry an /-a/
- /a/ is a suffix (*la, una, otra, esa, esta, aquella*)

In contrast, the form-function correspondences for masculine would be opaque because in the model language:

- the masculine noun's final vowel can be almost any vowel
- the masculine noun's final vowel mostly isn't /o/
- in the determiner system *-o* endings mark neutral (*esto, eso, aquello*), but not masculine
- in the determiner system, masculine is marked /e/: *este, ese aquel, el*
- even that generalization is not constant: *un, otro*
- the /e/ determiner marker for masculine is not always a suffix (el)

The degree of form-function transparency within the various conditions of a rule could be one of the direct determiners of María's grammaticalization sequence. In this case, the features which make up transparency and opacity can be made explicit and, consequently, represented in a formal model such as a neural net.

The empirical limitation (one child and production data only) of the results in the present study has only allowed building hypotheses on the nature and the sequence of the grammaticalization shift. We've seen how both longitudinal and experimental data require microanalyses if the goal is explaining, rather than labeling, transitional steps into grammar.

Notes

* This research was supported by a Spanish DGICYT's grant to López-Ornat's research project: PB-91-0352, as well as by the Acción Integrada 314 B (Nix, Davey, López-Ornat, Messer, and Smith), for its neural net simulation part.

I thank Pamela Smith, Annette Karmiloff-Smith, Ana-Teresa Pérez-Leroux and Ann Peters for their very helpful comments, suggestions and friendly revisions of this work's earlier form and content.

1. The corpus is analyzed for linguistic and for psycholinguistic information. Those analyzed files are available from our electronic database. The full natural language corpus (TRAMO.TXT) has been included in a book (López-Ornat et al. 1994) which also draws conclusions on the early grammaticalization process. The files are not in CHAT format, but most CLAN programs (MacWhinney 1991, 1995) run correctly with them, adding the "+y" symbol to commands. All of them are available with anonymous ftp from our database in Madrid at sol.sis.ucm.es (or 147.96.2.3), in the subdirectory: pub/Spanish.

2. The same hypotheses of this study have been explored in experiments (López-Ornat 1996) and neural net simulation models (Nix 1997) dealing with the same hypotheses of this study.

3. Sonia Mariscal has actively contributed to the database electronic analysis and to the discussion of this work in general. Her research collaboration has been very productive.

References

Abrahamsen, Adele. 1993. Cognizer's innards and connectionist nets: a holy alliance? *Mind and Language* 8: 520-530.

Aguado, Gerardo. 1988. Valoración de la competencia morfosintáctica en el niño de dos años y medio. *Infancia y Aprendizaje* 43: 73-96.

Aguado, Gerardo. 1995. *El desarrollo del lenguaje de 0 a 3 años*. Madrid: CEPE.

Alarcos LLorach, Emilio. 1984. *Estudios de gramática funcional del español*. 3rd ed. Madrid: Gredos.

Bates, Elisabeth, Philip S. Dale, and Donna Thal. 1995. Individual differences and their implications for theories of language development. In *The handbook of child language*, ed. P. Fletcher and Brian MacWhinney. Oxford: Blackwell.

Bloom, Louise. 1973. *One word at a time: the use of single-word utterances before syntax*. The Hague: Mouton.

Bowerman, Melissa. 1982. Reorganizational processes in lexical and syntactic development. In *Language acquisition: the state of the art*, ed. E. Wanner and L. Gleitman. Cambridge: Cambridge University Press.

Clark, Andy and Annette Karmiloff-Smith. 1993. The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language* 8.4: 487-519.

Echols, Catharine. 1993. A perceptually-based model of children's earliest productions. *Cognition* 46: 245-296.

Fernández, Almudena. 1994. El aprendizaje de los morfemas verbales. Datos de un estudio longitudinal. In *La adquisición de la lengua española*, ed. S. López-Ornat et al., 29-46. Madrid: Siglo XXI.

García Calvo, Agustín. 1989. *Hablando de lo que habla*. Madrid: Lucina.

Gerken, LouAnn and Bonnie J. McIntosh. 1993. Interplay of function morphemes and prosody in early language. *Dev. Psychology* 29.3: 448-457.

Hernández-Pina, Fuensanta. 1984. *Teorías psicosociolingüísticas y su aplicación a la adquisición del español como lengua materna*. Madrid: Siglo XXI.

Karmiloff-Smith, Annette. 1979. *A functional approach to child language: a study of determiners and reference*. Cambridge: Cambridge University Press.

Levelt, Willem J.M. 1989. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

López-Ornat, Susana. 1977. Un ovejo tiene cinco patas:interpretaciones evolutivas de la señal. Revista de Psicol. *Gral y Apl* 32.145: 211-224.

López-Ornat, Susana. 1990. La formación de la oración simple: las omisiones de categorías sintácticas (S/V/O) en la adquisición del español. *Estudios de Psicología* 41: 41-72.

López-Ornat, Susana. 1991. Book review: On the definition of binding domains in Spanish: evidence from child language, by José A. Padilla. *First Language* 12. 229-231.

López-Ornat, Susana. 1994. La adquisición del lenguaje:talón de Aquiles y poción mágica de la teoría cognitiva. *Cognitiva* 6.2: 213-239.

López-Ornat, Susana. 1996. Mecanismos de adquisición morfosintáctica. In *Estudios sobre la adquisición del castellano, catalán, euskera y gallego*, ed. M. Pérez Pereira, 175-193. Universidade de Santiago de Compostela: Servicio de Publicacions.

López-Ornat, Susana, Almudena Fernández, Pilar Gallo, and Sonia Mariscal. 1994. *La adquisición de la lengua española*. Madrid: Siglo XXI.

MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Resources in Child Development* 43, 174.

MacWhinney, Brian. 1991. *The CHILDES project*. Hillsdale, NJ: Erlbaum.

MacWhinney, Brian. 1995. Computational analysis of interactions. In *The handbook of child language*, ed. P. Fletcher and Brian MacWhinney. Oxford: Blackwell.

Mariscal, Sonia. 1996. Adquisiciones morfofonosintácticas en torno al Sintagma Nominal: el género gramatical en español. In *Estudios sobre la adquisición del castellano, catalán, euskera y gallego*, ed. M. Pérez Pereira, 263-270. Santiago de Compostela: Servicio de Publicacions de la Universidade.

Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of Child Language* 17: 357-374.

Nix, Andrew. 1997. A connectionist account of the acquisition of gender harmony in Spanish Noun Phrases. Ph.D. dissertation, Univ. of Hertfordshire.

Pérez Pereira, Miguel. 1990 ¿Cómo determinan los niños la concordancia de género? Refutación de la teoría del género natural. *Infancia y Aprendizaje* 50: 73-91.

Peters, Ann M. 1985. Language segmentation: Operating principles for the analysis and perception of language. In *The crosslinguistic study of language acquisition, vol 2*, ed. D.I. Slobin. Hillsdale, NJ: Erlbaum.

Peters, Ann M. 1995. Strategies in the acquisition of syntax. In *The handbook of child language*, ed. P. Fletcher and B. MacWhinney. Oxford: Blackwell.

Peters, Ann M. and Lise Menn. 1993. False starts and filler syllables: ways to learn grammatical morphemes. *Language* 69.4: 742-777.

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

Pinker, Steven. 1989. *Learnability and cognition*. Cambridge, MA: Harvard University Press.

Plunkett, Kim. 1993a. Making nets work hard. *Mind & Language* 8.4: 549-558.

Plunkett, Kim. 1993b. Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language* 20.1: 43-60.

Quilis, Antonio and César Hernández Alonso. 1990. *Lingüística española aplicada a la terapia del lenguaje*. Madrid: Gredos.

R.A.E. (Real Academia de la Lengua Española). 1994. *Gramática de la lengua española*. Madrid: E. Alarcos LLorach and Espasa Calpe.