

## ***CATEGORIES IN THE MAKING***

### ***Assessing the role of semantics in the acquisition of noun and verb categories<sup>1</sup>***

Caroline Rossi

Université Lyon 2. Laboratoire Dynamique du Langage.

Christophe Parisse

MoDyCo, INSERM, CNRS, Université Paris Ouest Nanterre La Défense

#### ABSTRACT

Little is known about what guides children in their acquisition of grammatical categories. This paper investigates how semantic knowledge could be involved in discovering these categories, thus confronting two competing hypotheses: are semantic categories innate, or are they developed in a piecemeal fashion? We tested for regular associations between basic semantic dimensions and the development of the founding categories of noun and verb. Six perceptually based semantic dimensions (Parisse and Poulain, 2010), shared by nouns and verbs but potentially distinctive, are coded in the productions of three children aged 1;06 to 2;06. Our results suggest that semantic dimensions do not offer an entry into the early differentiation of noun and verb categories.

---

<sup>1</sup> This work was supported by two grants from the Agence Nationale pour la Recherche, France (ANR COLAJE and ANR POLYCAT).

## KEYWORDS

noun and verb categorization, semantic bootstrapping, semantic / perceptual categories, language acquisition

## 1. INTRODUCTION

How do children manage to grasp the content and boundaries of grammatical categories? To the linguist, the distinction between noun and verb is not binary, but involves a complex set of notions (Black and Chiat, 2003) including semantic classes, phonological properties, as well as grammatical notions such as the distributional and morphological properties of words (e.g. Bassano, 2000 for a discussion of these properties in French). But are all these notions involved in early child language, where verbs can be used rather early on, and are not mistaken for nouns?

(1) *Madeleine* 2;04.15: line 2,432.

\*CHI: On lit sur mon lit ! [Let's read on my bed!]

Here Madeleine uses and clearly distinguishes two homophones: the verb *lit* (read), and the noun *lit* (bed) probably without even noticing the phonological similarity, /li/. However, as one- to two-year-olds utter their first words, they start by producing forms that are devoid of syntactic marking, or incorporate ambiguous marks. The syntactic difference between nouns and verbs is made obvious only gradually (Veneziano and Sinclair, 2000; Veneziano and Parisse, 2010). If children do not –or only very partially- rely on syntactic information to acquire grammatical categories (Pinker, 1994: 384), the first cues that they use could then be semantic ones. Macnamara (1984:126) suggests that ‘grammatical categories have an

intuitive basis, namely a semantic one' which could eventually 'help the child get syntax acquisition started' (Pinker, 1994:385).

Pinker thus posits that children have innate knowledge of the semantic properties of nouns and verbs, allowing them to relate elements of the world to what are syntactic nouns and verbs in their input. In other words, children would rely on 'certain contingencies between perceptual categories and syntactic categories, mediated by built-in semantic categories' (Pinker, 1994:385) to discover the abstract principles of syntax. Focusing on the categories of nouns and verbs, Pinker suggests that children have innate knowledge of a relation between the noun word class and reference to a person or thing on the one hand, and between the verb word class and such semantic categories as 'action' or 'change of state' on the other hand (Pinker 1984:41).

Another approach to language development has appeared over the past twenty years, rooted in Braine's initial work (1963): the so-called usage-based and construction-based approach (see Tomasello, 2003, and Goldberg, 2006). This approach holds that linguistic knowledge is based on constructions, i.e. form-function associations with varying degrees of openness. The most specific constructions correspond to words of the lexicon and lexical collocations, whereas open ones correspond to grammatical rules such as subject/verb agreement (in languages in which this applies). A key feature in this approach is that categorical knowledge

and construction openness (or generalisation) develop gradually in a piecemeal fashion (see Tomasello, 2000). The approach also derives from functionalism, where all linguistic knowledge is construed as form-function associations. In line with such a framework, we hypothesize that semantic categories must develop gradually and in harmony with form categories –a proposal which would go against both Macnamara's and Pinker's.

The goal of the present paper is to test the hypothesis of a gradual development of semantics against the traditional innate-based approach of Macnamara and Pinker. In order to do that, we investigate the acquisition of the basic verb/noun opposition, which is central to the argument of innate semantic knowledge and has been the object of much discussion and emphasis in works about early language acquisition (e.g. Bassano, 2000; Black & Chiat, 2003; Gentner 1982, 2006). More specifically, we examine the possibility that early but gradual emergence of semantic organization could either parallel or interact with the emergence of noun and verb categories.

In line with usage-based accounts of the acquisition of nouns and verbs, we do not assume that children have existing categories that they are specifying on a semantic basis. Rather, we would like to see whether semantics could be used as a wedge in the discovery and shaping of those categories. Those claims are tested thanks to fine-grained semantic analyses. We start from six basic semantic dimensions (Parisse and Poulin, 2010) that

may be used to define and contrast nouns and verbs, and check their distribution in the first productions of nouns and verbs in three French-speaking children: *Antoine*, *Madeleine*, and *Théophile*, from age 1;06 to 2;06, to see whether any dimension –or set of dimensions- is (or becomes) distinctive in profiling categories of nouns and verbs.

## 2. SETTING THE STAGE: INSIGHTS INTO THE EMERGENCE OF CATEGORIES

This paper focuses on the semantic interpretation of young children's production. 'Semantics' here refers to the regularities and generalisations of linguistic functions, whereas 'syntax' captures the regularities and generalisations of linguistic forms. The word 'lexicon' refers to elements produced by children that are considered close enough to an adult lexical target to be identified (and coded). Importantly, the term 'grammatical categories' is here used to refer to categories that are defined by both syntactic and semantic information. Our 'semantic dimensions' provide labels for the semantic properties of words or word groups produced by a child: not all of them are binary, as they correspond to sets of values in a semantic domain. For example, a dimension such as 'distance' can be reduced, medium, or long; a dimension such as 'number' can be singular, plural, or uncountable. Finally, we will use the term 'universal' to refer to semantic dimensions that can be found in all languages, either as

grammaticalized forms, or as lexical or compound lexical constructions: e.g. ‘distance’.

### 2.1. *Categories in the making*

In keeping with developmental approaches of categorization, one major problem addressed in this study is whether noun and verb categories are different in early child vs. adult speech. Clark (1982:395) examines the possibility that young children might overextend nouns for lack of the appropriate, specific verbs, e.g. using *car* and *foot* instead of *drive* and *kick*, respectively. But as she notices it, this would make children rather difficult to understand. Strikingly enough, nouns and verbs are not mistaken for one another in child speech, thus lending support to Pinker’s (1984) claim that children differentiate nouns and verbs very early, (i.e. at least by age two), a claim that was recently confirmed experimentally by Veneziano *et al.* (2010). The latter study showed that even in the case of polycategorical units, children were not misled by the similarity in form and meaning, and that by age two, 50% of the children displayed full knowledge of the syntactic information associated with the noun/verb distinction.

Another explanation for children’s early differentiation of nouns and verbs, however, is that noun and verb semantics may form distinctive sets that children rely on, even in the absence of morpho-syntactic marking (e.g. in example (1) above). This assumption is the theoretical starting point of the present study. It does not amount to denying the use of syntactic

information. Rather, it implies that semantic properties could be used as earlier and more fundamental cues in the acquisition of noun and verb categories.

### *2.2. Rethinking semantic bootstrapping with a usage-based model*

According to Pinker's (1984; 1994) semantic bootstrapping hypothesis, the 'discovery' of noun and verb categories depends on word meaning, which is established early on, possibly through the child's interactions with her material and human environment, and their regulation by innate rules. Pinker's proposal might be correct, but to date, 'naturalistic analyses of early child language have failed to find any evidence for innate linking rules' (Tomasello, 2000:233). Besides, what naturalistic analyses may help us understand is precisely what Pinker's hypothesis remains 'agnostic' about (Pinker, 1994:186), namely: how children's semantic knowledge is achieved. We suggest that basic and perceptually-based regularities in children's productions might help us bridge this gap. It should be noted however that such a proposal does not merely complement Pinker's view: its usage-based starting point goes against it. Indeed, either the child's grammar is usage-based, or she has innate principles to rely on.

According to a non-nativist point of view, children do not discover syntax on the basis of the innate principles of a 'universal grammar' or 'language acquisition device' (Chomsky, 1965; 1986), but through repeated use in distinctive constructions (Tomasello, 2003; Goldberg 1998; 2003).

Construction grammar's perspective on syntax and semantics differs from the Chomskyan model in that it adopts a 'what you see is what you get' approach to syntactic form, and then looks for pairings of forms (i.e. constructions, including morphemes or words) with semantic or discourse function (Goldberg, 2003: 219). There is no such thing as a bootstrapping problem for usage-based models, since they do not see adult linguistic competence as an abstract set of rules. Instead, they provide us with both a 'more child-accessible' account of the adult endpoint, and 'new ways of thinking about how children learn and construct abstract cognitive entities' (Tomasello, 2000: 235).

On these grounds, a number of arguments have been produced to justify the gradual emergence of forms, but little work has been conducted into first categories of meaning (or functions) and how they are related to forms. Tomasello (2000, 2003) argues that children learn the category of nouns earlier than the category of verbs and that this takes place in a piecemeal fashion. What is unclear is the status of semantics at each developmental step. Indeed, the semantics of early words is difficult to delineate (Bloom, 1991), and verb semantics is even harder to grasp (Gleitman, 1990; Golinkoff et al., 1995) –thus accounting for their later acquisition (Gentner, 2006).

### *2.3. Contrasting semantic dimensions for nouns vs. verbs*

A major difficulty in analyzing semantic differentiation between noun and verb is avoiding the use of semantic dimensions that are so specific to the properties of nouns and verbs that they only reject explanation to another level. Thus the proposal that nouns are often linked to physical objects and verbs to actions gives no clue as to how children learn to distinguish (or discover the difference between) objects and actions.

To avoid this pitfall, we chose to use semantic dimensions that are shared by nouns and verbs, and that can be identified by the child by means of simple perception (i.e. children do not need to create cognitive categories to identify these dimensions), while potentially applying them differently to nouns and verbs. An example of such a dimension is movement. Any word used by a child, and which refers to something in the child's surrounding, can be judged as referring to something that is moving or to something that is immobile. This measure is absolutely independent from the form of the word produced by the child (whether it is a noun, a verb or other), but if indeed nouns prototypically refer to objects and verbs to actions (as suggested by Pinker, 1984; Croft, 2001:89), then we should find more words that are nouns referring to static referents and more words that are verbs referring to mobile referents. If such a difference were to be found, it would suggest that children are following prototypical use of nouns and verbs and could use prototypical values to distinguish the two categories from a semantic point

of view. Such a difference should appear early if Pinker's hypothesis is valid, but should appear only gradually if our hypothesis is valid.

Beside motion, the other semantic dimensions that we use below are: number, concreteness, animacy, distance, and determination (see section 3.1 for details on the coding principles for each dimension). In many languages (including French), number applies to nouns and verbs alike. However, as shown by Macnamara (1984), number is closely related to sortals (i.e. concepts underlying the logical work of identifying and individuating count nouns), which Macnamara puts forward to explain how children learn the very concept of object. Besides, reference to count vs. mass elements is understood on the basis of objects rather than actions, so that reference to multiple or massive elements is more likely to develop with nouns before applying to verbs. Thus we should expect a tendency in young children to refer to number with nouns more than with verbs.

Because of their inherently relational content (Gentner, 1982; Black and Chiat, 2003), verbs have been described as more abstract than object labels (Bird, Lambon Ralph, Patterson, & Hodges, 2000; Breedin, Saffran, & Coslett, 1994. In Vigliocco et al., 2004), thus concreteness should be biased towards nouns.

Animacy applies to both nouns and verbs, but just like motion, the dimension should apply more often to verbs in young children's universe, where the concept of animacy most probably develops in a piecemeal

fashion (Rakison, 2005:187), and is primarily based on a notion of self-initiated vs. caused motion (Gelman & Gottfried, 1996). Also, reference to inanimate beings could be a characteristic of first nouns, if they prototypically refer to physical objects, but not of first (presumably egocentric) verbs.

Three levels were used to code distance: touching (proximal), visible (distal) or absent (distal + invisible). Although all three values can be grammaticalized on both nouns and verbs (at least in some languages, see Payne, 1997), distance (and especially visible vs. absent) is more closely associated with objects. For example, distance anchors definite vs. indefinite reference on nouns (see Maratsos, 1976). It is also at the root of the concept of object permanence. Finally, objects are more likely to be in the child's hands or close to him when he talks about them. However, the first actions named have also been analysed as egocentric, i.e. as referring to the child's own action, so that differences here might be toned down.

Similar arguments might be used for determination (specificity/genericity), which, although it applies to nouns and verbs alike, should be more fundamental for nouns (Macnamara, 1984: 144-156). What is more, if first nouns are known to be specific and context-bound, this is not necessarily true for first verbs. Indeed, the importance of "light verbs" such as *go*, *do*, *make* or *give* (Ninio, 2006), which are used to refer to broad

categories of actions in early productions, suggests that generic uses could be characteristic of early verbs, as opposed to nouns.

#### *2.4. Goal of the paper*

The goal of this paper is to see whether there is a gradual emergence of semantic differences between words in children's speech that are considered as nouns or verbs by adults, before the children acquire syntactic categories. Alternatives would be that such differences exist from the start, or do not exist at all.

In order to test our hypothesis, we have defined six basic sets of semantic dimensions, and coded early nouns and verbs for those six dimensions in three longitudinal studies of French-speaking children. As explained above, all six dimensions apply to both nouns and verbs. However, they should be used differentially in child language. Thus, if children do develop a semantic difference in profiling nouns and verbs, this difference should be seen in the preferential association of one dimension to nouns while another dimension would be preferred with verbs. The details of our coding system is explained and discussed in the next section.

### 3. METHODS

We coded all first linguistic productions in the longitudinal corpora of three French-speaking children (*Antoine*, *Madeleine* and *Théophile*), providing we could attribute them an adult target that was either a noun or a verb.

Coding ranges from 1;06 to 2;06, with approximately one session coded every two months. Antoine (12 sessions) produced 680 nouns and 328 verbs, Madeleine (8 sessions) 2252 nouns and 1474 verbs, Théophile (13 sessions) 481 nouns and 50 verbs. Detailed figures are presented in Table 1<sup>2</sup>.

Insert Table 1 here.

### *3.1. Coding principles*

Coding was done according to the adult target, i.e. to the word, grammatical category and semantic dimension that an adult (the person doing the coding) would attribute to the child, if he were interacting with her as a conversational partner. We relied on mothers' (or observers') interpretations insofar as they matched the coder's decision, which was also based on the general situation and context of the interaction. Thus, our coding categories were devised to take into account all the elements that children may rely on as they discover language: perceptual elements as well as adult interpretations and feedback. According to our hypotheses, whether the child does indeed master the categories of nouns and verbs does not matter – in fact, we are suggesting that he does not master them at first, and that these categories only gradually become productive. This, however, does not

---

<sup>2</sup> Table 1 includes mean values and detailed values at the first and last point studied. These values show significant individual differences, based on mean length of utterance in words (MLU) and number of words produced –but see description of the corpus (this issue) for a more in-depth presentation of the children and definitions of these measures.

preclude the use of adult categories as independent variables (Bates *et al.*, 1994: 90). Although we rely of the adults' interpretation, we have adopted the child's point of view, *as understood by the interlocutor*, on the general situation and discursive context: semantic dimensions were determined according to a combination of environmental clues that could be taken as representative of what the child is operating with –namely, social pragmatic, environmental and linguistic cues (Tomasello and Bates, 2001: 283).

All the relevant items were coded for syntactic category (noun vs. verb), and according to six semantic dimensions defined below:

- 1) **Distance from the child:** we coded whether or not the referent of the noun or verb was in the child's hands, and if the action was performed by the child (touch); if it was visible (audible) to the child; or absent.
- 2) **Number:** without taking into account grammatical number, we looked at whether referent(s) was (or were) single – multiple – massive.
- 3) **Concreteness:** we coded for concrete versus abstract reference.
- 4) **Animacy:** for both nouns and verbs, referents could be either animate or inanimate.
- 5) **Motion:** nouns could refer to static or mobile elements, as could verbs with the dynamic vs. static events they referred to.

- 6) **Determination:** if one item, action or process was singled out by the child, it was coded as specific; if the noun or verb referred to a whole category, i.e. a series of instances considered together, it was coded as generic.

Insert Table 2 here.

A 'not coded' choice was available for all semantic dimensions, and was considered an important coding option. It was used whenever no other choice was possible, for lack of information, because it did not apply to the situation, or when we could not decide, thus avoiding the pitfalls of over-interpretation as much as possible. Inter-coder reliability, measured using Cohen's kappa, was 0.91 for distance, 0.89 for number, 0.95 for concreteness, 0.83 for animacy, 0.79 for motion, and 0.82 for specificity. The values were high enough to confirm that the coding categories were valid and operational.

### *3.2. Coding perceptual categories in early child speech*

The distinction of verb- versus satellite-framed languages in Talmy (1985)'s typology of motion events clearly show that verb semantics may vary considerably from one language to the other: verbs may conflate different aspects of a scene, and as a result, lexically-based semantic universals will be hard to find. What is more, early referential processes are not as strictly tied to the conventions of language as they are in adult speech: children are indeed likely to use, for example, abstract terms with concrete reference.

Because of the variability of referential processes in child speech, interpretation is bound to rely mostly on perceptual factors, which are at the heart of early lexical acquisition (Golinkoff and Hirsh Pasek, 2008; Pruden et al., 2006). That is why our coding categories are essentially perceptual.

### *3.3. Coding problems and decisions*

As shown by the general description (section 3.1), coding criteria were rather straightforward. Some items, however, proved more difficult to code than others. We had to make additional decisions in order to rule out the items that included very little information, and to adjust to the characteristics of child language.

#### *3.3.1. Nominal and verbal constructions*

First of all, it should be mentioned that we coded nominal and verbal constructions rather than bare nouns and verbs, so as to take into account the (at least partially) unanalysed chunks or amalgams (Peters, 1983; Aksu-Koç and Slobin, 1985) often found in early productions. This implies, for example, that we considered determiners together with nouns and auxiliaries together with verbs –both elements often being produced as filler syllables (i.e. syllables that do not directly correspond to conventional, adult markers, but are first signs of morpho-syntactic organisation –see e.g. Veneziano and Sinclair, 2000).

In first combinations and multi-word utterances, however, we focused on each specific item rather than on the overall meaning of the

utterance, so that within the same utterance, two different items could be coded differently. For example, a generic action could be performed with a specific object or vice versa, as shown in our coding of *Madeleine's* utterance presented in Table 2. In the same way, the use of two different verbs or nouns in the same utterance was more likely to be distinctive than repetitive.

### 3.3.2. *Animates*

Young children's understanding of animacy has been shown to be only partial to start with: we therefore based our decisions primarily on our knowledge of each child's development. For instance, one child, *Madeleine*, showed early understanding (at 2;02) that the characters in a story were animates, i.e. that they had feelings and could be understood in just the same way as the people she was regularly interacting with. In this session, she thus offers to lend her bunny to one of the characters in the story who looks sad (Morgenstern and Sekali, 2009). From 2;02 onwards, we therefore coded the characters she named in a story as animates. When it was less obvious, however, i.e. when children were only naming the characters upon request, those nouns were not coded for animacy. Moreover, Striano *et al.* (2001) have shown that before two years of age, children have symbolic skills with gestures but not with objects. Here again, we used our knowledge of the children's development, together with a close consideration of context, to decide on the coding. Thus for example, when children talked

about toys figuring people or animals, but which were manipulated and put into containers, the referents were not coded as animates.

On the other hand, in our coding of animacy for verbs, we took into account the (most often implicit) agent, patient or instrument that the verb applied to: their semantics was considered as an integral part of verbal constructions.

### 3.3.3. *Motion*

Our coding of motion based on perceptual clues led us to code the dimension of motion for many nouns. Surprising as it may seem, this finding is perfectly in keeping with infants' predominant interest for things in motion (Slater 1989: 59; Casasola et al., 2006; Pulverman et al., 2006). Here again, however, some of our coding decisions are more obvious than others. For nominal constructions, the difference between static and mobile referents very much corresponded to mobile figures vs. static grounds. Designated objects, however, could be static when shown from a distance, and were more likely to be mobile when in the child's hands (i.e. manipulated objects). For instance when Madeleine (2;04) says: *C'est pas avec l'aspirateur* (you don't do this with the vacuum-cleaner), while putting it down, *l'aspirateur* (vacuum-cleaner) was straightforwardly coded as mobile. This also applies to cases where nouns were not used in single-word utterances -e.g. in (2) below:

(2) *Antoine* 2;01: line 348 & sq., as he watches the water flow from the hose

\*OBS: c'est l'eau ? [This is water, right?]

\*CHI: oui. [Yes.]

%pho: wi

\*OBS: et pourquoi qu'est-ce qu'elle fait ? [Why? What is it doing?]

\*OBS: hein qu'est-ce qu'il se passe là ? [Look, what's happening here?]

%com: children playing in the yard can still be heard

\*CHI: a@fs di@fs l'eau. [There's water.]

%pho: a di lo

%xpol: NOUN/eau/

/visible(audible)/uncountable/concrete/inanimate/mobile/specific/

We chose to code *l'eau* (water) as mobile because to the child, it does designate a mobile referent, and our coding of a single nominal element reflects the absence of clearly verbal elements in the child's utterance (in spite of the observer's deliberate prompts).

## 4. RESULTS

Variations in the absolute number of lexical elements produced in the various sessions would have masked any interesting trend in the children's production. Therefore, all our analyses had to be done using percentages.

For *Théophile*, the number of lexical verbs coded was quite small (M. = 39.8 for nouns, M. = 5.5 for verbs). Even in the final session, he produced only 47 verbs for 124 nouns. So for *Théophile*, trends about verbs may not be highly significant and ought to be regarded with caution. For *Antoine* (M. = 52 for nouns, M. = 29 for verbs), the total number of lexical elements produced in the first three sessions (from age 1;05 to age 1;07) is low (less than 10), so caution with statistical analysis should apply. After age 1;09, at least 50 and up to more than 200 lexical elements could be analyzed, making our data more reliable. For *Madeleine*, the number of words analyzed is consistently large (M. = 246 for nouns, M. = 154 for verbs) and she never produced less than 30 verbs in one session – see Table 1 for more information about the children's production.

### *4.1. Differences across categories*

The percentages of nouns and verbs produced for each of the semantic categories analyzed is presented in Table 3 for all six coded dimensions, including the cases when no categorization was possible. The difference between mean percentage of 'noun' and 'verb' items coded was significant 36% of the time (the pairs of gray cells in Table 3 correspond to significant

differences between noun and verb). In only two cases, all three children had a similar significant tendency to prefer nouns for a specific semantic dimension: namely for ‘distance=absent’ and ‘animacy=inanimate’.

Insert Table 3 here

#### 4.2. Developmental trends

For some specific dimensions, there was a significant trend throughout all the sessions. Significant trends are presented with a \*, \*\*, or \*\*\* in Table 3. Increasing trends are indicated by superscript, decreasing trends by subscript. Out of the 28 dimensions coded (3 dimensions for distance, 3 for number, 2 for abstractness, etc.), trends were significant 4 times (see cells with asterisks) for *Antoine*, 8 times for *Madeleine*, and only one time for *Théophile*. For reasons of space, we cannot make an extensive description of all the observed tendencies here, so we will focus on the most salient ones and indicate whether they are increasing or decreasing. For *Antoine*, an increasing trend to use verbs with ‘distance=visible’ resulted from the absence of reference to visible elements with verbs in the first two sessions (they were only coded as ‘touch’), whereas in all of the following sessions, verbs were used 61% of the time with visible reference. This could correspond to an initial egocentric conceptualization of actions (Piaget, 1962; Nelson, 1974), which only gradually enables the child to distinguish and name the actions performed by others:

(3) *Antoine* (1;07): Assis [Sit.] –asking to be put on the chair

(4) *Antoine* (1;09): Assis! [Sit!] –asking the observer to sit next to him.

However, the small number of verbs produced in the first two sessions calls for caution. A more interesting trend was found for ‘motion=static’: with nouns and verbs alike, there is a steady increase of reference to static elements, and a decrease in the percentage of reference to mobile elements.

For *Madeleine*, the most interesting trends were found with ‘abstractness’ for verbs, ‘specificity’ for nouns and verbs, and ‘absence’ for verbs. First, verbs tended to become more abstract and there were fewer concrete items as *Madeleine* got older. The most frequent abstract verb forms produced by *Madeleine* included: *attends* (wait), *arrive* (come), *faut* (must), *veux* (want), and *sais* (know). No such trend was found with nouns, which were, one the whole, neither more nor less concrete than verbs. Secondly, both nouns and verbs tended to become less specific and more generic, although generic reference remains the unusual case. As for the third trend, verbs were not used to refer to absent situations before age two. After two, nearly a third of all verbs were coded for absent reference. Although the change happened rather suddenly, the trend was significant. Here again, the evolution in verb use could be linked to the child’s social and cognitive development and, for *Madeleine* more specifically, to the early development of narrative skills:

(5) *Madeleine* (2;04) : Elle est où Marie que j'avais vue ?

[Where is Mary that I saw earlier?]

(6) *Madeleine* (2;06) : Euh ah hum Martine elle avait écrasé un peu mon doudou. [Erm oh huh Martine, she'd squashed my teddy a little.]

#### 4.3. Differences between children

The differences between children were measured using a t-test for each pair of children. *Antoine* and *Théophile* were the most similar with differences found only three times out of 40 (20 dimensions x 2 syntactic categories). These differences applied to nouns only, in the following three dimensions: 'motion=not coded', 'distance=touch', and 'abstractness=concrete'. *Antoine* and *Madeleine* were very similar for verbs (one difference only for 'specificity=generic'), but quite dissimilar for nouns, with eight differences out of 20 (the largest difference for animacy and motion). *Madeleine* and *Théophile* were dissimilar ten times out of 40, but with an equal number of differences in nouns and verbs. Once again there were differences in animacy, but also in distance. No differences were found at all in the use of number: all three children overwhelmingly used singular reference.

#### 5. DISCUSSION

Our results did not confirm any of the hypotheses presented in the introduction. On the whole, no clear semantic differences between nouns

and verbs could be found, although we expected either a difference from the start, or a difference that would appear with language development.

According to the semantic bootstrapping hypothesis, perceptual semantic categorization should be guiding the acquisition process before syntax appeared in the children's productions. The assumption implies that, based on semantic dimensions, a clear line could be drawn between major categories such as verbs and nouns. In that case, semantic differences across categories should not change much with age, as syntax is considered to tune itself to semantics and not the other way around. There should be no significant variation between children either.

According to construction grammar, on the other hand, semantic categorization should be item-based and variable from one child to another. There should be no semantic tendencies to begin with, and categorization should develop over time. Syntax and semantics should develop together.

Overall, our results did not support either of the theories. There were few clear differences in semantic dimensions which could help the children to differentiate nouns and verbs. Surprisingly, differences in abstractness were found for one child only. The main difference between nouns and verbs was found with the "animacy" dimension. Nearly all verbs were used with reference to animates (counted as such even when the agent, patient or instrument was only implied). On the contrary, nouns were used 30 to 50% of the time with an animate referent, so reference to animates turns out to be

a negative criterion: in our children data, verbs are characterized by the fact that they are never inanimate. The only other systematic difference between nouns and verbs is for ‘distance=absent’ which is used more for nouns than for verbs. But this is not distinctive for most cases, i.e. when they refer to visible or touched elements. However, especially for *Madeleine*, it appears that the distance used with nouns is often wider than for verbs. This tends to change with age, which makes it a potential developing feature.

Our results do not support developmental theories on the acquisition of categories any more than nativist theories, since very few developmental tendencies were found, even though at age 2;06 all the children’s MLUs were higher than two (or close to two for Théophile- see description of the children at the beginning of this issue). At an age when children are already able to produce clear syntactic regularities, one could expect that they should have developed semantic trends. In fact, only *Madeleine*, with an MLU higher than three, seems to get closer to such a point.

### *5.1. Intra- and inter-individual variability*

The explanation for these inconclusive results might be found in the great variability of the semantic use of language. This can be exemplified by the use of ‘motion=mobile’ by *Madeleine*, as shown in Figure 1, with the large difference found between her massive use of verbs for mobile reference and the (more unusual) use of nouns for this reference. Even though substantial differences across sessions forbid us to draw reliable conclusions, it does

seem possible that the first references to motion by Madeleine are both verbal and nominal, a trend which may be confirmed by the results for *Antoine* (Figure 2). *Antoine* appears to be similar to *Madeleine* before age 1;10. As his overall language development is not as quick as Madeleine's, he may still be at an early stage of semantic development towards the end of our analyses, which would account for this persisting ambiguity towards the end of our coded data.

Insert Figures 1 and 2 here

### 5.2. *Towards usage-based semantic universals*

Our results, however, could also be the consequence of our understanding of usage-based semantics. The theory is still in the making (see for example Croft, 2007; Cysouw, 2010, for proposals about such a theory) and leaves room for various interpretations. We focused on context, which is crucial to both usage-based theories and early language acquisition, in accordance with Cysouw's usage-based approach, which proposes to '[define] the meaning of a language-specific expression as the collection of all contexts in which the expression can be used' (2010: 71). Haspelmath (2007) also denies the existence of predefined syntactic categories, but not that of semantic universals. He considers that although 'there is ample evidence that meaning, too, is conventional and varies across languages', we could still use 'low-level notions' across languages (2007: 127). In order to solve the riddle of cross-linguistic understanding, which may not be so different

from parent-child understanding, Haspelmath suggests we take as semantic universals something which is very close to semantic exemplars: a notion yet to be generalized in cognitive grammar, but which is certainly compatible with the principles of usage-based linguistics. Our coding of semantic dimensions is therefore not ‘usage-based’ in that the dimensions were not derived from a semantic mapping of the meanings found in our data, but rather from our intuitive semantic analyses. Whether those intuitions actually match child-directed speech remains to be determined.

### *5.3. Fuzzy semantic categories*

Finally, our results point to the necessity of reconsidering the relationship between syntax and semantics. Although both the principle of a form/function (or signifier/signified) relationship and the basic postulates of cognitive linguistics call for a close correspondence between syntax and semantics, this strict correspondence is not as intuitive as it may seem. Indeed, the difference between nouns and verbs as syntactic elements is clear-cut in French (nouns are preceded by determiner and verbs by subject personal pronouns) and is functional very early in young French children, but the semantic difference is not as clear-cut, as evidenced by the fact that categorization inevitably produces fuzzy sets.

Our results evidence the need for a reevaluation of the relationship between syntax and semantics throughout development. Instead of semantics being more fundamental than syntax, the relationship between

syntax and semantics itself could be the result of a developmental process. In this case, only when a child has moved from items and one-word utterances to syntactic organization, and simultaneously evolved from item-based meaning to semantic organization (categorization), would she be able to use form-function properties and relationships to create (or understand) new linguistic properties, derive new meanings and walk the path towards full mastery of language.

## 6. CONCLUSIONS

Ultimately, the absence of clearly delineated semantic sets, at a stage when nouns and verbs are probably only superficially distinguished, should come as no surprise. Our results suggest that in her first predications, the child masters neither the semantics, nor the syntax of noun and verb categories. The construction of these categories seems to be achieved only through the gradual mastery of more and more complex and detailed language. In the end, then, our results confirm the proposals of distributional grammar, as well as those of construction grammar: it is only thanks to the understanding of constructions and of the relations between different elements in the utterance that noun and verb categories can be grasped.

## REFERENCES

- Aksu-Koç, A. and Slobin, D. (1985). Acquisition of Turkish. In: D. I. Slobin (ed.) *The crosslinguistic study of language acquisition*. Vol. 1: The data. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 839-878.

- Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language*, 27: 521-559.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznik, J., Reilly, J. and Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21.1: 85-123.
- Black, M. and Chiat, S. (2003). Noun–verb dissociations: a multi-faceted phenomenon. *Journal of Neurolinguistics*, 16.2-3: 231-250.
- Bloom, L. (1991). *Language development from two to three*. Cambridge: Cambridge University Press.
- Braine, M. D. S. (1963). The ontogeny of English phrase structure: The first phrase. *Language*, 39: 1-14.
- Casasola, M., Bhagwat, J. and Ferguson, K. (2006). Precursors to verb learning: infant's understanding of motion events. In: K. Hirsh-Pasek and R. M. Golinkoff (eds), *Action meets word: how children learn verbs*. New York: Oxford University Press, pp. 160-190.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1986). *Knowledge of Language*. Praeger.

- Clark, E. (2001). Emergent categories in first language acquisition. In: M. Bowerman and S. Levinson, (eds.), *Language acquisition and conceptual development*, Cambridge University Press, pp. 379-405.
- Clark, E. V. (1982) The young word-maker: A case study of innovation in the child's lexicon. In E. Wanner and L. R. Gleitman (eds.), *Language acquisition: The state of the art*. Cambridge: Cambridge University Press, pp. 390-425.
- Croft, William. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, W. (2007). Exemplar semantics. Unpublished manuscript, available online at <http://www.unm.edu/~wcroft/Papers/CSDL8-paper.pdf>
- Cysouw, M. (2010). Semantic maps as metrics on meaning, *Linguistic Discovery*, 8, 1: 70-95.
- Gelman, S. A. and Gottfried, G. M. (1996). Children's Causal Explanations of Animate and Inanimate Motion, *Child Development*, 67:1970-1987.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In: S. A. Kuczak. (ed.), *Language development*, vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gentner, D. (2006). Why verbs are hard to learn. In: Hirsch-Pasek, K. and Golinkoff, R. (eds.) *Action meets word: How children learn verbs*. New York: Oxford University Press.

- Gleitman, L.R. (1990) The structural sources of verb meanings. *Language Acquisition* 1, 3–55
- Goldberg, A. (1998). Patterns of experience in patterns of language, In: M. Tomasello (ed.), *The new psychology of language*, Mahwah, NJ: Erlbaum, pp. 203-219.
- Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7.5: 219–224.
- Goldberg, A. (2006). *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
- Golinkoff, R. M. and Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Science*, 12: 397-403.
- Golinkoff, R. M., Hirsh-Pasek, K., Mervis, C. B., Frawley, W. and Parillo, M. (1995). Lexical principles can be extended to the acquisition of verbs. In: M. Tomasello and W. Merriman (eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Haspelmath, M. (2007). Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology*, 11, 1: 119-132.
- Macnamara, J. (1984). *Names for things: A study in human learning*. Cambridge, MA: Bradford.

- Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children: An experimental study of semantic acquisition*. Cambridge: Cambridge University Press.
- Morgenstern, A. and Sekali, M. (2009). What can child language tell us about prepositions? A contrastive corpus-based study of cognitive and social-pragmatic factors. In: J. Zlatev, M. Johansson Falck, C. Lundmark and M. Andrén (eds.) *Studies in Language and Cognition*, Cambridge Scholars Publishing, pp. 261-275.
- Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, 81:267-285.
- Ninio A. (2006) *Language and the learning curve: a new theory of syntactic development*. Oxford: Oxford University Press.
- Parisse, C., and Poulain, C. (2010). Shared semantic properties of nouns and verbs in French-speaking children before age two, *VI Congrès Internacional d'Acquisicio del Llenguatge*, Barcelona, Espagne.
- Payne, T. E. (1997). *Describing morphosyntax: a guide for field linguists*. Cambridge: Cambridge University Press.
- Peters, A. (1983). *The Units of Language Acquisition*. New York: Cambridge University Press.
- Piaget, J. (1962). *Le langage et la pensée chez l'enfant*, Neuchâtel, Paris : Delachaux et Niestlé.

- Pinker, S. (1984). *Language Learnability and Language Development*, Cambridge, MA, USA: Harvard University Press.
- Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, 92:377-410.
- Pruden, S.M., Hirsh-Pasek, K., Golinkoff, R.M. and Hennon, E.A. (2006) The birth of words: ten-month-olds learn words through perceptual salience. *Child Development*, 77: 266–280.
- Pulverman, R., Hirsh-Pasek, K., Golinkoff, R. M., Pruden, S. and Salkind, S. (2006). Conceptual foundations for verb learning: celebrating the event. In: K. Hirsh-Pasek and R. M. Golinkoff (eds.), *Action meets word: how children learn verbs*. New York: Oxford University Press.
- Rakison, D. H. (2005). Developing knowledge of motion properties in infancy. *Cognition*, 96: 183-214.
- Slater, A. (1989). Visual Memory and Perception in Early Infancy. In: A. Slater, and G. Bremner, (eds.) *Infant Development*. London: Lawrence Erlbaum Associates.
- Striano, T., Tomasello, M., and Rochat, P. (2001). Social and object support for early symbolic play. *Developmental Science*, 4(4): 442-455.
- Talmy, L. (1985) Lexicalization patterns: semantic structure in lexical forms. In: T. Shopen, (ed.) *Language Typology and Syntactic Description*, Cambridge: Cambridge University Press, pp. 57–149.

Tomasello, M. (2000) Do young children have adult syntactic competence?  
*Cognition* 74: 209-253.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*, Cambridge (Mass.); London: Harvard University Press.

Tomasello, M. and Bates, E. (eds) (2001). *Language Development: the essential readings*. Oxford: Blackwell publishers.

Veneziano E. and Parisse, C. (2010). The acquisition of early verbs in French: Assessing the role of conversation and of child-directed input. *First Language*, 30, 3-4: 287-311.

Veneziano, E., Parisse, C. and Delacour, A. (2010). The Early Comprehension of Noun-Verb Distinction in French. An Experimental Method. *International Conference on Infant Studies 2010*, Baltimore, USA, Mars 2010.

Veneziano, E. and Sinclair, H. (2000). The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language*, 27(3): 461-500.

Total number of words: 7677 words with abstract, keywords, references and words in figures

Revised manuscript: August 21, 2011

	Child	Age	MLU	Number of Nouns	Number of Verbs	Number of Words
Madeleine (7 sessions)	Mean values	2;00	2.44	246	154	1129
	First session	1;06	1.38	194	30	390
	Last session	2;06	3.02	199	169	1397
Antoine (12 sessions)	Mean values	1;11	1.48	71	51	330
	First session	1;06	1	3	0	23
	Last session	2;06	2.72	183	187	354
Théophile (13 sessions)	Mean values	1;11	1.60	41	10	323
	First session	1;05	1.05	4	0	59
	Last session	2;08	2.61	124	47	1297

Table 1: Data coded

Data	Utterance	Noun	Verb	1	2	3	4	5	6
<i>Antoine</i> 1;5	Bain	bain		visible	sing.	abst.	inanim	static	specif.
<i>Théophile</i> 2;2	Il est cassé		est cassé	touch	sing.	concr .	anim	static	specif
<i>Madeleine</i> 2;4	Je vais te raconter un livre		vais raconter	touch	sing.	concr .	anim	mobile	specif.
<i>Madeleine</i> 2;4	Je vais te raconter un livre	livre		absent	sing.	concr .	inanim .	-	generic

Table 2: Coding examples

	<i>Children</i>	<i>NOUN VERB</i>		<i>NOUN VERB</i>		<i>NOUN VERB</i>		<i>NOUN VERB</i>	
<i>Distance</i>		<i>not coded</i>		<i>Absent</i>		<i>Touch</i>		<i>Visible</i>	
	<i>Antoine</i>	8	4	22	6	16	38	54	42*
	<i>Madeleine</i>	1	23	34	12**	24	45*	41	21
<i>Number</i>	<i>Théophile</i>	0	1	15	2	28	50	57**	47
		<i>not coded</i>		<i>Plural</i>		<i>Singular</i>		<i>Uncountable</i>	
	<i>Antoine</i>	3	1	5	0*	91	88	1	0
<i>Abstractness</i>	<i>Madeleine</i>	0	6	10**	1	79	92	11	0
	<i>Théophile</i>	0	0	4	0	83	100	13	0
		<i>not coded</i>		<i>Abstract</i>		<i>Concrete</i>			
<i>Animacy</i>	<i>Antoine</i>	6	4	35	8	59	78		
	<i>Madeleine</i>	0	1	17	20***	83	79***		
	<i>Théophile</i>	0	0	16	10	84	90		
<i>Motion</i>		<i>not coded</i>		<i>Animate</i>		<i>Inanimate</i>			
	<i>Antoine</i>	3	0	41	87	56	2		
	<i>Madeleine</i>	0	0	27	97*	73	2		
<i>Specificity</i>	<i>Théophile</i>	0	0	53	100	47	0		
		<i>not coded</i>		<i>Mobile</i>		<i>Static</i>			
	<i>Antoine</i>	45	2	42	78	13*	10*		
	<i>Madeleine</i>	33	19	26	65	42	16		
	<i>Théophile</i>	15	0	60	79	26	21		
		<i>not coded</i>		<i>Generic</i>		<i>Specific</i>			
	<i>Antoine</i>	8	8	2	1	90	81		
	<i>Madeleine</i>	5	15	16**	11	78*	74		
	<i>Théophile</i>	1	0	5	0	94	100		

Table 3: Mean percentages of dimensions coded for all categories.

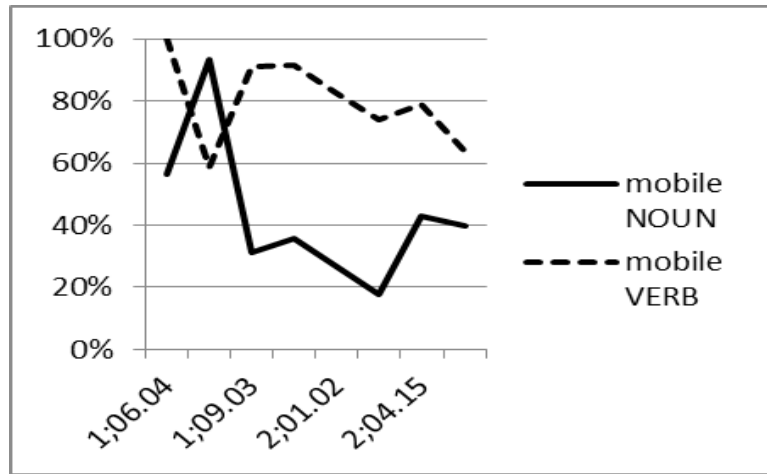


Figure 1: Percentages of nouns and verbs used for mobile references by *Madeleine*

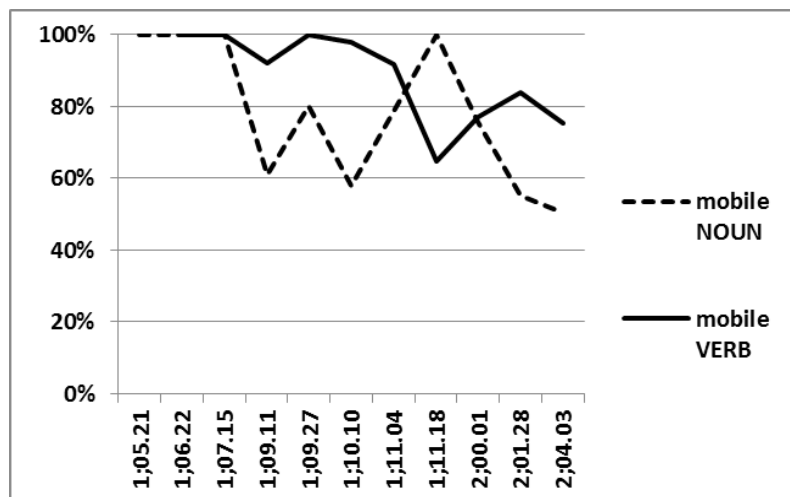


Figure 2: Percentages of nouns and verbs used for mobile references by *Antoine*