ON NATIVE SEMANTIC ROLES – COMPARATIVE STUDY BASED ON DATA FROM CHILD LANGUAGE ACQUISITION OF ENGLISH AND FRENCH

Dr. Velina Slavova, New Bulgarian University, Department of Computer Science, Bulgaria

E-mail: vslavova@nbu.bg

ARTICLE INFO

ABSTRACT

This study explores statistically child language-acquisition using data extracted from large collections for acquisition in two languages – English and French. Comparison of the two collections reveals that the advancement in acquiring vocabulary displays very big differences when the children's speech is classified by the parts of speech deployed, as these are formally defined in the two languages, despite there being no reasons to suppose that the two language groups of children should show significant differences in cognitive development. The hypothesis put forward is that there exist general classes of meaning-representation and the challenge is to obtain evidence corroborating this. A specific set of classes is proposed, derived according to their different contributing roles in the mental representation of the world, considered from the perspective of an "Actor in the environment" cognitive model. The identified parts of speech from the two languages are sorted into the proposed classes. It is shown statistically that when children's speech is discriminated to these classes, the acquisition processes in the two languages are very alike. Examining the data, the use of these classes is evident from the onset of language production. Some particularities related to factors influencing the use of communicators, interjections and onomatopoeias in children's speech are discussed in addition to the study's overall findings.

1. INTRODUCTION

How the brain forms and organizes meaningful concepts and how the faculty of language permits structured expressing of their meaning are questions of fundamental theoretical importance to linguists and cognitive scientists alike. Thanks to the emergence of powerful brain-imaging technologies, neuroscientific research has made significant strides in revealing anatomically the brain's activity pertaining to the cytoarchitectural organization of concepts and their labelling with words. The neuronal networks found to be involved in this activity appear to implicate the entire brain. All of these studied phenomena can be assumed to be universal biological properties of human brains.

Considering this wide-ranging body of research permitted Slavova and Soschen (2015 a, b) to amalgamate its findings with a hierarchical information-treatment model and propose a general theory explaining how perceptual experiencing of environmental phenomena and interacting with them provides a working basis for forming concepts and subsequently associating them with particular words. The model presented in Slavova and Soschen describes the process by which humans in general progressively acquire the faculty of language during infancy. This model supposes that syntax is founded on concept semantics. It suggests that the internal creation of semantic description of the world and the mental treatment of language syntax are products of one and the same principles of information processing. The underlying mechanisms were identified as based on multimodal perception, interoception, proprioception, the mirror neuron network and default mode network — all of them ready to run in a synchronized manner

Corresponding Author
Dr. Velina Slavova, New Bulgarian University, Department of Computer Science, Bulgaria
E-mail: vslavova@nbu.bg

at birth. Following this model, the process of establishing semantic description of the world initially ensues automatically as the result of interacting with it, and in accord with some underlying principle of structuring meaning.

When children learn their first language, the meaning of the words used in spontaneous communication can suggest the structure of their mental world. At the same time, the language input that children are exposed to is of crucial importance. That is why investigating the structure of the primary semantic description of the world and its constituents necessitates analysing a wide diversity of languages. As a first effort toward such an investigation, the present paper's author undertook a large-scale study of corpora of recorded utterances collected from English and French infants over the formative period of language-acquisition. This paper presents the procedure of the study and the results of its analysis.

2. PROCEDURE AND DATA

Data from 42 corpora containing 1,515 free dialogues with child speech in English and in French, annotated with part of speech and grammar, were extracted from CHILDES (Child Language Data Exchange System) and used for the statistical analyses presented here (Appendix A).

Child speech dialogues (written, audio and video recordings) are stored with their transcripts and available on-line in the CHILDES data repository. They are collected (in separate corpora) and transcribed by researchers in language acquisition using the standard developed over the course of several decades especially for the Exchange system (see Mac-Whinney and Snow, 1985). The transcription is performed using CLAN (Computerized Language ANalysis), a computerized system designed specifically for the Exchange system's standardized format (Appendix B). Important for the study presented here is that the stored transcripts include for each speech utterance a separate line marked with "mor%", created by the transcribers using the computerized tools, developed for supporting the annotation in a large number of target languages. This line contains the system's standardized symbols for the parts of speech (POS), based on Hausser's MORPH system (Hausser, 1989, see MacWhinney, 2012).

For the purposes of this study the transcripts, with the entire linguistic annotation, were stored locally. Additionally, a number of tools were developed for extracting the transcripts from CLAN format and organizing them in a relational database where each dialogue and each utterance is tagged with a unique identifier (Appendix B). A more detailed description of the procedure, the tools, the data treatment and the technical aspects of the data organization and representation are provided in Slavova (2016).

The English data collection used in this study contains 620 dialogues (with 62 girls, 66 boys, and 7 children with gender not specified in the source); the French collection contains 895 dialogues (with 157 girls and 141 boys). Some children are "recorded" during several successive months and some are not. The parameters of the dialogues are as follows: the English data contains in average 520 utterances of different participants in a dialogue, where 202 child utterances; the French data contains in average 388 participants' utterances in a dialogue, where 171 are utterances of the child.

Next, a large number of queries were elaborated to select, regroup and calculate parameters of the child speech utterances. In the present study, the results for two large corpora collections — 125,873 child speech-utterances for English language production and 153,824 for French language production — were treated statistically. On average 2,400 utterances per child-month were treated, taken from different corpora, belonging to different children aged between 6 months and 62 months and originating from dialogues taking place in different circumstances and selected by different researchers over the course of the last four decades.

Examples of the children's utterances extracted from the dialogues, with their POS annotation, are given in Appendix B. The language-related parameters presented further in the statistical analysis are obtained by parsing the mor% annotation extracted from the annotated dialogues in CHILDES.

Observation of children's utterances in the two languages confirmed several known facts regarding language acquisition: The first pronounced distinguishable word- forms appear around 10-13 months in single-word expressions, and with the development of the child's overall capacities, utterances become longer, expressing increasingly complex ideas. At approximately 26 months all the analysed utterances have a phonological content comprising at least one word-form identifiable as belonging to the given language (Fig. 1). Following the collected data, after the age of 62

months, child speech starts to contain complex and subordinate sentences in a single communication utterance.
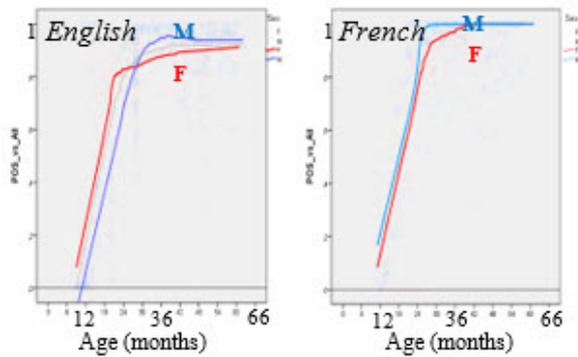
Figure 1. Utterances with phonological content recognizable as word-forms in the given language. Ratio over all utterances. (girls – red, boys – blue)

The task in this study is to judge how the global content of the speech develops in terms of mental images that underlie the meaning-expression.

3. STATISTICAL ANALYSIS OF THE POS ACQUISTION

For studying the use of POS, the statistical analysis relies on the annotation performed by the authors of the respective data-corpora in the data collections (Appendix A). In the data-annotation scheme deployed in the CHILDES source, the word-forms produced by children are classified by POS as they are distinct in the corresponding languages. This annotation uses 33 POS for the English collection and 30 for the French.

In order to obtain a measure for the contribution-weight of a given POS within a speech-utterance, the following formula for calculating the Ratio per Utterance (RU) of the given POS within a dialogue was applied:

$$RU\left(POS_{ij}\right) = \frac{NPos_{ij}}{N_j} \qquad (1)$$

where :

POS_i is one of the POS annotated in the corpora,

j is the dialogue,

N_j is the number of utterances with recognizable POS in the dialogue j.

$NPos_{ij}$ is the number of the POS_i in the dialogue j.

By applying the formula (1) a RU was obtained for all the POS for each of the 620 dialogues in English and 865 dialogues in

French. The RU show the extent of use of the given POS for expressing the child's notions within an exemplary utterance "averaged" for the dialogue. They can be seen as weights of the use of given POS for expressing the meaning communicated by the child within the dialogue. The RUs were used in the further analyses of the similarities.

The statistical result displayed in Fig. 2 is consistent with specialized studies in language acquisition. For example, it has been shown (Bassano, 2000) that in French acquisition, between the ages of 14 and 30 months, nouns clearly predominate over verbs, but that verbs are however produced in the early stages. The statistical result shown on plot of POS-acquisition does not contradict either the time-scale of acquisition of different POS reported in the specialized studies for English.
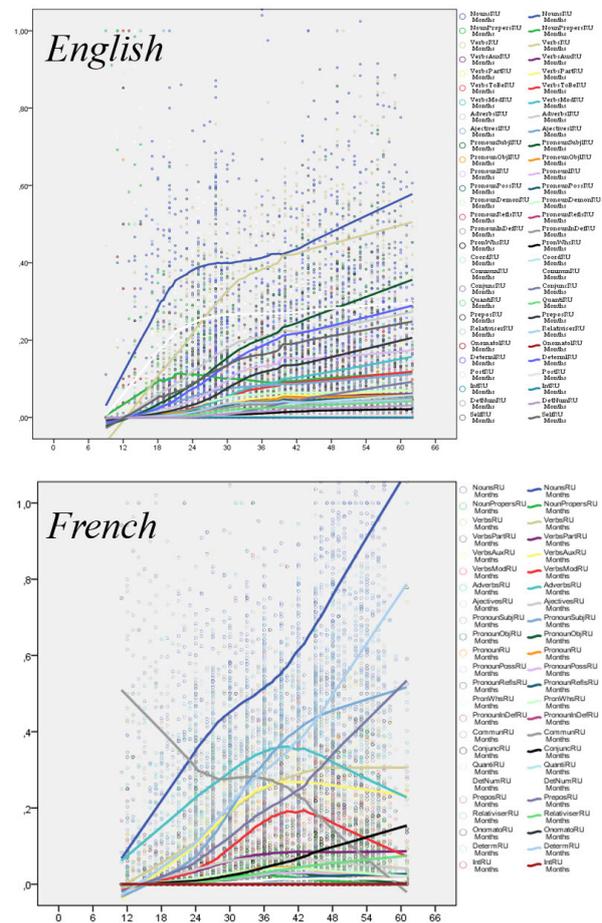
Figure 2. Development with children's age of the RU of the POS in the English and French data

As shown in Fig. 2, the developmental paths for the use of different POS in the utterances produced by English-acquiring and French-acquiring children are quite different. This difference can be attributed to language-

specific features, as obviously the grammatical particularities of English and French are different

Table 1 gives the correlations of the use of identical POS in the two languages (between-languages correlation). This correlation is small, taking into account that the use word-forms in the considered period is described by an increasing function.

The between-languages correlation presented in Table 1 is for the 25 identically labeled POS in the two language-corpora (in descending order following the correlation values), for the 50 months (from 11 to 61) which have data collected for both languages. The p – values which suggest statistically unreliable result are given in italic. The between-languages correlations are obtained based on month-to-month correspondences. That is, each POS RU of all the dialogues belonging to equally aged (in months) children are averaged within the same month, for each of the languages, and after that compared.

The average between-languages correlation for the use of identical POS is only 0.46. At the same time, for the period investigated, the Sums of the POS RU (roughly - the length of the utterances) develop in a very similar way, correlated at 0.88, i.e. higher than the maximal POS-to-POS correlation.

There are no reasons to suspect that the two languages are very different at the level of POS-structure as they are close representatives of one and the same family and are of the same morphological type. As seen in the correlation table, the POS which are used in the most comparable way are the prepositions and the determiners.

There are also no reasons to suspect that, in terms of conveyed meaning, the 1-. 2-, 3- years old English-acquiring and French-acquiring children have very different ideas to communicate.

Specialized studies discover interactions between semantic and grammatical development. For example, Bassano (2000) found that between 14 and 30 months, verb- and noun-grammati-calization in French is related to the production of concrete action verbs and to concrete object nouns. Bassano proposes that "These findings, discussed in a cross-linguistic perspective, suggest that both conceptual and grammatical packaging are important and interacting factors in noun and verb development". This idea is strongly supported by all the contemporary research presented in the book edited by Hirsh-Pasek and Golinkoff (2006), "Action meets word: How children learn verbs."

Table 1. Between-languages correlations for the development over the time of language acquisition of POS RU

POS:	Correlation	p – value
Preposition RU	0,840	0
Determiner RU	0,830	0
Pronoun Subjective RU	0,825	0
Verb RU	0,818	0
Conjunctions RU	0,812	0
Verb Auxiliary RU	0,795	0
Adverb RU	0,721	0
Noun – common RU	0,697	0
Verb Modal RU	0,660	0
Relativizer RU	0,553	0
Pronoun Demonstr. RU	0,541	0
Determiner Numeric RU	0,517	0
Pronoun RU	0,483	0
Pronoun Objective RU	0,471	0
Vbs. Participles RU	0,430	0
Quantifiers RU	0,376	0
Onomatopoeia RU	0,352	0,01
Pronoun Interrogat. RU	0,349	0,01
Pronoun Indefinite RU	0,331	0,01
Pronoun Reflective RU	0,327	0,02
Adjective RU	0,304	0,02
Pronoun Possessive RU	0,299	0,03
Interjection RU	-0,084	0,04
Noun - proper RU	*-0,197*	*0,56*
Communicators RU	*-0,415*	*0,17*
Average POS-to-POS	**0,465**	

Table 2. Example - expression of meaning and desires in a dialogue of a 15 months old child. (CHI – child, MOT – mother, FAT – father, SIS – sister)

Data extracted from the dialogue taz15.bw
*CHI: Mommy Mommy Mommy.
*CHI: tee [: tree] tee [: tree].
*CHI: Mommy tee [: tree].
*CHI: Dada.
*CHI: Mommy.
*MOT: what?
*CHI: out.
*CHI: baby.
*CHI: hi.
*FAT: hi, Laura.
*CHI: hi.
*SIS: hi, Laura.
*CHI: Mommy!
*MOT: what Laura.
*CHI: ah Dada.
*MOT: what's the matter?
*CHI: dee@b dee@b.
*MOT: oh my.
*CHI: out Dee baba [= bottle].
*MOT: what?
*CHI: hi, Dada hi, Dada Mommy.
*CHI: car car car car.
*CHI: car.
*CHI: key key key.

The core meaning-related question to be clarified concerns the very different paths of acquisition of the POS-constituents in the two languages (Fig. 2). The children's utterances in the corpora are most often incomplete and grammatically incorrect sentences, but they express the meant quite well (an example is provided in Table 2). In fact, the used measure – the POS RU reflects statistically the profile of the words produced within a dialogue and does not reflect the sentence level.

The next section presents the approach proposed here in order to find a common semantic organization which can explain the language acquisition processes as similar. Such a structure is supposed to equilibrate the difference displayed in the acquisition process measured at the POS level.

4. COGNITIVE MODEL

It is largely agreed in cognitive science that learning to assign meaning to sensory stimuli lies at the foundation of human cognition (e.g., Glezer et al. 2015). Unaided by language, infants from birth are able to begin forming meaningful conceptual knowledge about entities they perceive on a basis of interacting purposively with them, and to apply this to their interactions in increasingly structured ways and diverse contexts. Several contemporary findings suggest, as proposed herein, that the brain can execute internal meaning-related processing independently of language - that is to say, on a basis of processing of intact perceptual representations (here termed information-units) prior to, or in default of, their corresponding conceptual representations' lexicalization. A study by Moran and Tommerdahl (2011) of an 8-year-old child raised in a social environment, but without contact with spoken or signed language, found that the child exhibited no clear evidence of cognitive deficits. The celebrated case of Helen Keller provides similar testimony. In addition, several studies (e.g., Frishberg, 1987, Torigoe and Takei, 2002) have investigated cases of home-signing, that is, the inventing of sign-languages by groups of two or more hearing-impaired individuals who have not been taught a conventional sign-language. The phenomenon of home-signing suggests that humans are predisposed to elaborate systems exhibiting a language-structure in order to share their internally created concepts, and use them for communicating. The main question addressed here concerns the existence of some primary structure behind these internally created concepts.

The present study's approach focuses upon the progressively elaborated content of children's spoken utterances. The creation of meaningful representation is portrayed in the proposed model as a process in which inborn information-treatment mechanisms organize information-flows obtained in interaction with the world so as to establish distinct units of meaning. In the model presented by Slavova and Soschen (2015a., b.) this process was termed "meaning-encapsulation". It is supposed here that the meaning-encapsulation process is ready to run at birth. The early period of language-acquisition is assumed to be underpinned by the processes of meaning-encapsulation

Two separate regimes within the language-learning process are considered in the basic scheme of the proposed model — the analytic regime, related to language comprehension, and the generative regime, related to speech production or sign-language production (Fig. 3).
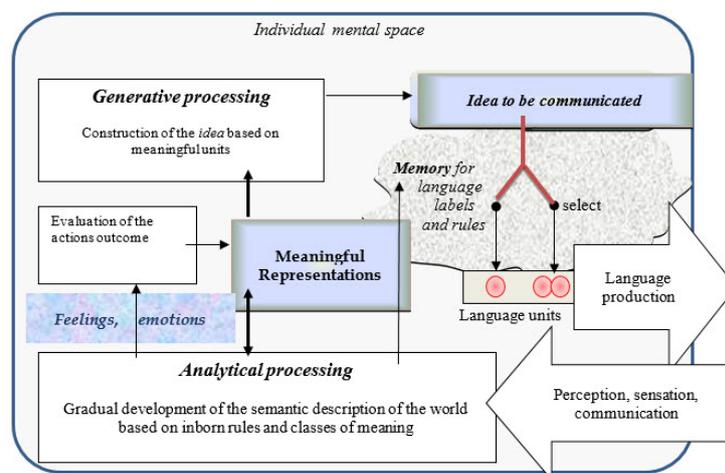

Figure 3. General scheme underlying the proposed model

The analytical processing in language acquisition is concerned with the assigning of meaning to words (or expressions). The acquisition of language-labels and rules and their use is seen as mapping between a child's own meaning-related representations and the labels and rules used in the language-environment (Fig. 3). This process can be presented as blending of the internally "encapsulated" units with the language ingredients. Its outcome is the creation of a lexicalized concept – a capsule with a "name". These names are further involved as words in the generative processing.

It should be taken into account also that children often invent their own labels for the concepts they have formed. These labels (perhaps arbitrary, à la Saussure (1916), perhaps sound-symbolic or to some extent phonetically matching the words used in the language environment) are used by children in communicating, and with persistence (an example is given in Table 2).

Additionally, the term idea is used herein to signify a consciously represented thought, generated by neuronal operations on meaningful units. The process of idea-generation involves creating the assembly of distinguishable meaningful units that express the thought. Recent medical studies suggest that some brain-impairments can harm the regulating of the mechanism of idea-generation. For example, Robinson and colleagues (Robinson et al., 2015) suggest that "When a "brake" to stop message generation mechanism is damaged at the level of conceptual preparation, the speaker will have difficulty stopping new thoughts from being created, generated, and expressed as overt speech.".

Language production commences after the analytic regime has created memory-paths necessary for the retrieval of the labels and rules used in the language. Recent research has shown that children know the meaning of many nouns at the age of 6 months (Bergelson et al., 2012), while the first pronounced words start at 10-13 months. In the speech-generative regime, the child aspires to communicate to others a certain idea, which presupposes its existence already in the child's mental space. One main question addressed here concerns the idea-ingredients, given that the meaningful units are compiled by the analytical processing-regime, the settings for which are assumed to be biologically determined.

When the idea is to be communicated, it has to be converted into expressible language-items, represented in memory. The question to be further addressed concerns what types of conceptualized items are added to children's expressions during the acquisition with regard of their role in the mental representation of the world. As children expressions become more and more reach, the next question is in what proportion and how the growth happens happens as regards of the type of meaningful units.

The analysis that follows relies on the use, during the initial stage of language acquisition, of different parts of speech (POS). POS are seen as offering the building blocks of the language's content, which (by the present hypothesis) has to be bonded to meaningful representations (Fig. 3).

The next section proposes a set of roles that can be distinguished in the mental representation of the world, suggested as basic classes of concepts.

5. MENTAL REPRESENTATION OF THE WORLD – THE MEANING CLASSES HYPOTHESIS

The overall approach in this study looks at meaning as constituting the bedrock of the faculty of language. This is opposite to some linguistic views, which suppose that meaning exists because of language. Obviously, if meaning is a consequence of language, it needs to be explained from where and how language has arisen in order to introduce the meaning that it carries; a scientifically plausible explanation has yet to be put forward, however.

The reasoning in this section relies on the widely accepted psychological model proposed by Lawrence Barsalou (Barsalou 2003), according to which the conceptualizing and in general the mental representing of the world are the undertakings of an Actor, acting in the environment. The hypothesis developed here - termed the "Self-centered model of language faculty" (see Slavova and Soschen 2015 a, b) and derived from Barsalou's model - accords to the concept of Self a central role in developing language faculty. Its authors reasoned that meaningful units of information are created by a substrate of inborn mechanisms that have the task of ensuring the survival of the biological system (the Self) as an "Actor in the environment". The Actor's mental representing of the world can be thought as a system of such meaningful units.

Categorization of meaningful information as distinctive classes is possible if the purpose (i.e., usage) of the information has been determined. The hypothesis put forward here is that the mechanisms that have arisen during evolutionary development have played the role of internal generators of information necessary for the survival of the biological species in the environment. Thus the mental mechanisms responsible for the generation of meaningful units are presumed to build units of importance for the species' survival that are assignable to such classes of information. Each class would have a specific role for representing the environment (the "world") as internal meaning.

To determine the classes, the first step was to deduce the general information-types that the newborn (an autonomous system) should have in order to act adequately with regard to his Self as Actor. Based on this, the environmentally presented "realities" that have to be organized as information units in order to further operate on them are supposed to be related to: 1. Physically negotiable objects in the environment, of significance for its functioning and existence (e.g. energy supply, obstacles, dangers etc.), 2. Their behavior (e.g. actions, states, intents etc.), namely, that which is comparable with the actions and states of the system and of importance for the system's behavior, 3. The manner in which the environment is "organized" and changes (e.g. the spatial and temporal particulars of the significant objects, relative to the system's own functioning in space and time), 4. The qualitative features of the environment (e.g. the same color or form in separate objects) that are of importance for the system's survival in the environment and 5. The quantitative parameters of the environment (e.g. evaluation of proportions between the objects or between groups of objects).

It should be noted that the reasoning followed takes into account social concepts, science-related concepts etc. as these are presumed of importance for the survival of humans as a species, or, at least, for the survival of humans as they have evolved up until now.

Together with the analysis of the speech data from the corpora, this led to the following classes (Table 3):

- *Entitles,*
- *Relationships,*
- *Circumstances,*
- *Quality and Attribution,*
- *Quantity and Precision and*
- *Others.*

Table 3. Proposed meaning-classes in the model "Actor in the environment". Spread of the POS with examples of annotation.

Meaning-Classes	Examples of annotation, English	Meaning-Classes	Examples of annotation, French
1. Entities		**1. Entities**	
Self	I, my Name, me, my, myself, mine, Baby	Self	je, moi, mon Nom, bébé, mon, ma, mes, mien
Common Noun	n|man, n|work	Comm. Noun	n|cheval, n|pied, n|crasse, n|lapin
Proper Noun	n|Dada, n:prop|Uncle, n:prop|Joe	Proper Noun	n:prop|Papa, n:prop|Raphaël
Pronoun Subj.	pro:sub|he, pro:sub|they	Pronoun Subj.	pro:subj|il, pro:subj|on
Pron. Object.	pro:obj|me, pro:obj|them	Pron. Object.	pro:obj|me, pro:obj|le
Pronouns	pro|it, pro|you	Pronouns	pro|moi, pro|toi, pro|eux
Pron. Reflect.	pro:refl|myself, pro:refl|yourself	Pron. Reflect.	pro:refl|se (+v|garer), pro:refl|se$v|appeler
Pron. Interrog.	pro:wh|what, pro:wh|who	Pron. Interrog.	pro:int|qui, pro:int|quoi
2. Relationships		**2. Relationships**	
Verb - Action	v|go, v|find, v|sit, v|finish	Verb Action	v|marcher, v|connaître, v|dire,
Verb - Modal	mod|can, mod|will, mod|do	Verb - Modal	v:mdl|faire, v:mdl|vouloir, v:mdl|aller
Verb - Auxiliary	aux|be, aux|have, aux|get	Verb - Auxil	v:aux|avoir, v:aux|être
Verb To Be	cop|be		v:aux|être
Participles	part|mix, part|go, part|use	Participles	part|casser; part|boire, part|tomber,

3. Circumstances		3 Circumstances	
Adverbs	adv|out, adv|there, adv|almost, adv|down	Adverbs	adv:place|dehors, adv|d'abord, adv|très, adv|aussi
Prepositions	prep|at, prep|on, prep|with	Prepositions	prep|à, prep|avec, prep|dans, prep|moins
Pron. Demonstr.	pro:dem|that, pro:dem|there	Pron. Dem.	pro:dem|ce, pro:dem|ça,
Conjunctions	conj|but, conj|when, conj|because	Conjunct.	conj|parce pro:rel|que, conj|si
Coordoners	coord|and		conj|et,
Relativisers	rel|what, rel|where	Relativ.	pro:rel|quoi, pro:rel|où; pro:rel|que
4. Quality and Attribution		**5 Quality and Attribution**	
Adjectives	adj|brown, adj|big, adj|good	Adjectives	adj|blanc, adj|petit, adj|beau,
Pron.Possesive	pro:poss:det|my, pro:poss:det|his	Pron. Possess.	det:poss|mon, det:poss|sa
5. Quantity and Precision		**Quantity and Precision**	
Numerals	det:num|four, det:num|million	Numerals	det:num|un; det:num|trois
Quantifiers	qn|more, qn|many, qn|some	Quantifiers	qn|plus, qn|un_peu, qn|plusieurs.
Pron. Indefinite	pro:indef|one, pro:indef|more	Pron. Indef.	det:gen|quelques, det:gen|chaque
Determiners	det|a, det|the, det|this	Determiners	det|le det|un
Post	post|both, post|all; post|else,		
Other		**Other**	
Onomatopoeia	on|beeu, on|bawk, on|ding	Onomatopoeia	on|pin on|pon, on|ham on|ham; on|kof
Interjection	int|da, int|ba, int|da, int|wow !	Interjection	int|wah int|berk int|berk
Communicator	co|please, co|yes, co|no, co|thank_you	Communicat.	co|oui, co|non, co|merci, co|oh, co|héh, co|miam

6. STATISTICAL ANALYSIS OF THE MEANING CLASSES

POS appearing in the data-collections were sorted into the aforenamed meaning-classes, as shown in Table 3. The separation of POS is accomplished by imagining which POS are used in the speech-samples to express each of the classes. The distribution cannot be perfect — a word-form can belong to more than one meaning-class depending on the context. As an example - tout in French is at time an adjective (translated as any, every, entire), an adverb (translated as all, very, in all, all up) a noun (all, whole) and a pronoun (all, any, anything). Here the study relies on the annotation-method applied by the linguists and on their correctness.

The use of the same POS-labels in the two languages can also be a source of errors, but when looking for universals one has to apply common sense in order to find correspondences. For example, the numerals in the French corpus are annotated as nouns as it is following the rules of the French grammar adopted in CHILDES. That has required a retrieval of the numbers in the French corpus and changing the annotation. Some particularities have not been homogenized. For ex-

ample, in the French corpus mien in "le mien" (the mine, m.) is a noun and in "la mienne" (the mine, f.) is an adjective.

The child speech data were statistically treated in respect of these classes. The RUs for each of the meaning-classes were calculated by summing the RUs of their POS-constituents.

As shown in Fig. 4, the paths of use of the meaning-classes during acquisition are very similar in the two languages. (The Sum of RU of the meaning-classes is strictly equal to the SUM of POS RU). The average between-languages correlation for similar meaning-classes is considerably higher than the average POS-to-POS correlations (Table 4).

A strange behavior is displayed by the class of "Others", in that its use in the two languages is negatively correlated. It is not clear from the data why the use of this class displays so different a statistical picture in the two languages. However, some reasoning is proposed in the next section.
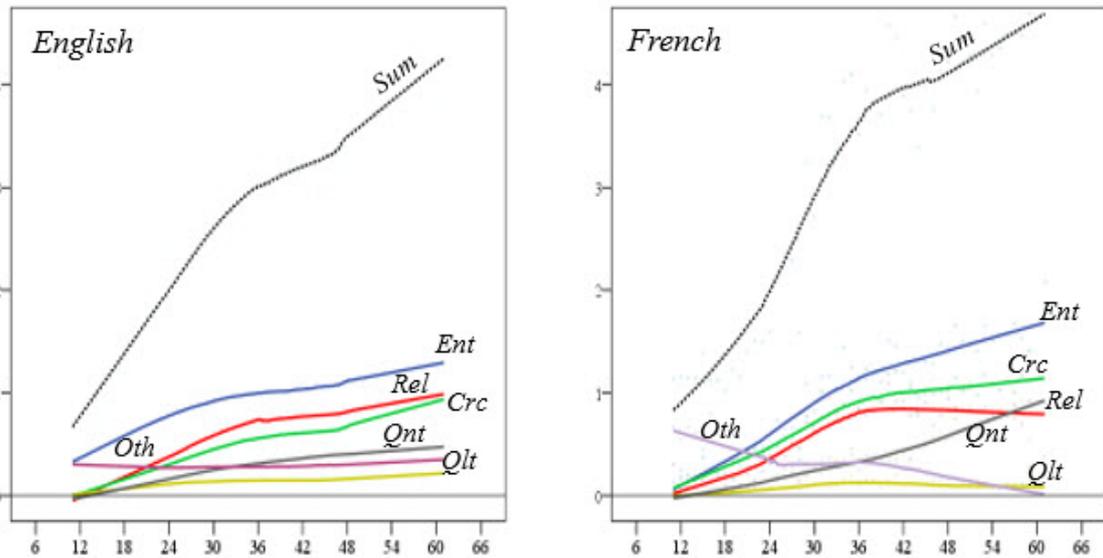
Figure 4. Meaning classes' use with advancing age : Ent – Entities, Rel – Relationships, Crc – Circumstances, Qlt – Quality, Qnt – Quantity, Oth – Others, Sum of all classes (Sum of POS RU).

Table 4. Correlations of the use of classes between English and French

Meaning class	Correlation
Entities	,830
Relationships	,857
Circumstances	,886
Quality and Attribution	,460
Quantity and Precision	,842
Others (expressive and communication)	-,121

The conclusion at this point is: when children's expressions are classified to the proposed classes of meaning, the statistical pictures that describe the two acquisition processes are very similar, as seen from the plot in Fig. 4 and from the correlations.

7. DATA OBSERVATION

The proposed meaning-classes are viewed here as language replications of the semantic roles that humans mentally construct from their interaction with the environment. If the classes have this function, they would be detectable in children's speech from the first stage of language production. Observation of the data-collections confirmed their use from its outset, i.e., at 10-14 months; examples of the first use of each class in the two languages are given in Appendix C.

A commonly accepted fact in the field of child language-acquisition is that the ability to learn arbitrary associations between words and objects develops until about 14 months of age (e.g. Werker et al., 1998). Brain studies (e.g. Friedrich and Friederici, 2005) also suggest that the processes underlying semantic integration are already developed at the age 14 months. The analysis of the data shows that at the age of 14 months all the classes of the proposed set are used by the two language groups taken as a whole.

The investigation of the English data collection confirmed statistically that there is difference in the acquisition process reflecting children's individual abilities (Atanasov at al., 2016). In the data examined here, for the age-group of 10-13 months there are dialogues of 21 children, 14 acquiring English and 7 French. The dialogues of the 13-months old English-acquiring children (9 different children) contain all the classes except Quantity and Precision and of the French-acquiring group of children (4 children) contain all the classes. The dialogues of 14-months-old English-acquiring (15 children) contain already all the classes. At 16 months the use of all classes is already intensive for the two language-samples (fig. 4). At 16 months, 2 (of the 4 recorded) English and 2 (of the 4 recorded) French acquiring children used all the classes within the confines of single dialogues. Two of the "classes-incomplete" dialogues belong to children who are recorded at younger age, which allowed seeing that these children have used the "missing" classes in their dialogues at 14 and 15 months of age.

The conclusion that can be derived is that the two language-collections' samples support the primary character of the proposed meaning classes.

There are questions concerning the behavior of the classes, however, which have to be clarified. As discussed in the previous section, the classes' acquisition displays a very similar, smoothly growing development, except for the class of Others. The class comprises Onomatopoeias, Interjections and Communicators and its use in both languages is initially high and tends to decrees over the time (Fig 4). The detailed plot of the development over the time of its components is given in Fig. 5.

Analyzing results obtained in specialized domains offers some reasons concerning the behavior and the differences accounted statistically for the use of the class of Others. A body of research supports Imai and Kita's sound symbolism bootstrapping hypothesis (Imai and Kita 2014) stating that sound iconicity facilitates language learning in general (e.g. Assano et al., 2015). For example, studies of adults' and children's language-learning have shown that non-Japanese speakers learn easer sound symbolic Japanese adjectives (Lockwood et al., 2016) and verbs (Imai et al., 2008). The results obtained by Fenson and colleagues (Fenson et al., 1994) regarding English and Spanish children's language-acquisition showed that the earliest-acquired words were those judged as being most iconic, where Onomatopoeias and Interjections were rated as being highest in iconicity.
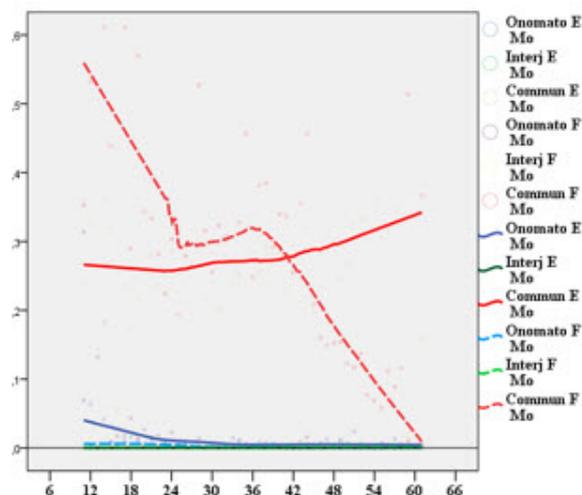
Figure 5. The class of Others.
Correlations : Communicators: -0,415 Onomatopoeias: 0,352, Interjections: -0,084,.

Laing, C. E. (2017) used an approach based on picture-mapping task and reported an advantage for onomatopoeia in the mapping word-to-semantic-item in broader perceptual sense. The extended analysis of the results related to these phenomena, proposed by Laing, states that Onomatopoeia probably constitutes the most obvious and common form of iconicity, but ideophones (e.g., glisten, jingle) and mimetics (found in Japanese) also contribute to iconicity. The author states that the extent of this contribution varies across languages.

Concerning the statistical picture obtained in the present study, in the light of these cited results it is hypothesized here that Onomatopoeias, known to be dominant in infants' early lexicons, are initially used to name Entities, Relationships etc., which explains the decrease of their use over the course of time, in both languages. It can be supposed that the use of onomatopoeias is language and culture dependent, which may explain the observed differences of the two languages.

As may be seen from the plot in Fig. 5, the negative between-languages correlation observed for the Others class is due mainly to the dissimilar use of Communicators (yes, no, thank you, hi!, oui, merci, salut! etc.) and Interjections (see Table 3). It should be noted that the p-value for the correlation of Communicators is 0.17, so, formally, the correlation represents an unreliable result. In fact, observation of the data shows that the intensity of use of Communicators displays a big dispersion over the dialogues. Unsurprisingly, the use of Communicators is dialogue-dependent. This dependency reflects the influence of the context occasioning the dialogue on its content. It is plausible, too, that children's individual habits have an impact on their use of Communicators. In all cases, in the English dialogues the level of use of Communicators is approximately constant, whereas the French data contain dialogues in which the use of Communicators drops drastically over the course of time.

The plots in Fig. 6 show the distances found after multidimensional scaling for the set of POS RU in the two languages (for the entire period of language acquisition investigated). The plots suggest that use of POS advances en bloc (with the exception of Nouns and Verbs) in both languages. Only the use of Communicators displays a markedly distant point, suggesting their separate role in language expression.
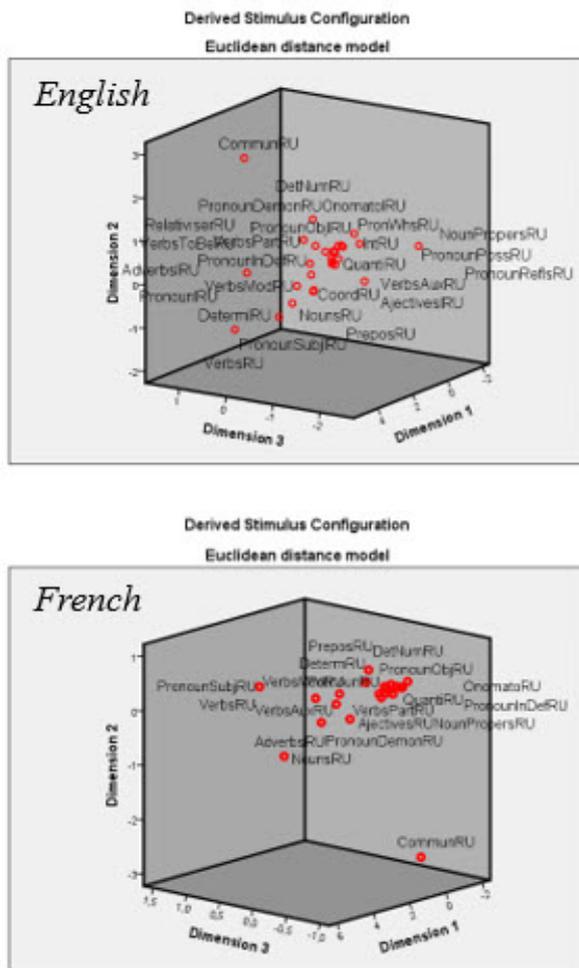
Figure 6. Euclidian distance model after multidimensional scaling of the POS RU

In French, the use of the Others class decreases as language-acquisition progresses (Fig. 4). One may suppose that the decrease is offset by the use of alternative POS (as is found regarding Onomatopoeias), or perhaps by other means of communicating.
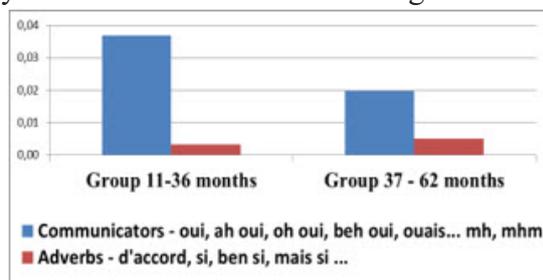
Figure 7. Example - use of some affirmative communicators and adverbs in French

Fig 7, for example, depicts the analyzed samples regarding use of affirmative communicators in French, together with other means often used by French-speakers for expressing them (where "si" is a quite specific manner of expressing a double negation - i.e., expressing

approval in contradiction to a negative statement just made by the other speaker in the exchange).

Speech communication in free dialogues is an act. The internal dispositions, intents, and emotional states implied by the speaker are not transferred in a uniformly faithful manner by the pronounced words alone. Even when taking into account that the prosody of the announcement conveys a lot of information, the accompanying signs and reactions such as gestures and gaze make part of the communication. This behavioral aspect of language communication is culture dependent. As an example, the communication within a group of speakers of Italian looks different to that within a group of speakers of Dutch. It seems reasonable to propose that the difference observed in the statistical result is due to the influence of the adult's language and communication practices in the two cultures.

Re-addressing the questions of concept formation and language production after undertaking her huge analysis, Laing (2017) discusses the role of iconicity in early language development as follows: "...These 2 words [dog and ball, found to be among the first 10 most frequently used by small children] are among the 3 least iconic of the 10 words overall, and dog is both the least iconic and least systematic of the 2 words. In these two cases, therefore, the motivation behind their early acquisition cannot be driven by iconicity. "

Perhaps the reason for children's frequent production of such words could reside in the inborn necessity to mentally represent the related semantic types.

8. DISCUSSION

The idea underlying this study is not novel in linguistics. The overall approach can be seen as a statistical investigation towards the "semantic bootstrapping hypothesis" proposed by Steven Pinker (Pinker, 1987). Pinker supposes the existence of a "semantic inductive basis" that helps children in the acquisition of language rules by means of "syntax-semantic pairing". The content of the inductive basis proposed by Pinker comprises categories such as "name of person or thing", "action or change of state", "attribute" and "spatial relation, path or direction". Further, Pinker's work shows how these categories can lead to the acquisition of syntactic rules.

The set of classes proposed here, inferred from the Actor in the environment mod-

el, came to be quite similar to the categories proposed by Pinker. The results of the present study support the hypothesis that syntactic rules are based on semantic determinants and suggests, too, that this basis is common for all humans.

This implies that during the course of evolution, over a large time-span, the development of languages has been dictated by the development of the mechanism for mental representing of the world.

Let us present language development (and acquisition) as depending on two complementary factors: the one being how necessary it is to encode some item of information in order to communicate it (in terms of its importance for the continued communal existence of humankind), and the other being how feasible it is to do so. One may suppose that, in evolutionary terms, the necessity to communicate in order to survive has influenced the development of the mental capacities required.

Communicating information by means of spoken language necessitates first conceptualizing it and then according to it a phonetic content (label). This raises the question as to abilities necessary for conceptualization and those regarding the phonological encoding.

In the languages examined, Communicators and Interjections are mostly expressed with short syllables comprising simple phonetic content, easily memorized and pronounced. These lexemes are expressing internally generated, affective reactions to information that has been processed (immediately or in the past), and express products generated by limbic system processes. They can be seen as speech-expressions of internal information flows that encode intrinsic characteristics of the Actor. This explains their intensive use in the initial period of language production (see Fig. 5).

From the standpoint of conveyed meaning, Ccommunicators are the most complex representatives of the speech as they serve to communicate agreement or disagreement, intents, internal dispositions, etc., evaluating the overall conceptualized situation. Indeed, Communicators serve to summarize and convey, in a single word, both the Actor's overall conception of a situation and his or her immediate stance or inclinations in reaction to it. As stated in the Wikipedia article on "Yes and No", "They are sometimes classified as a part of speech in their own right: sentence words or word sentences."

Onomatopoeias, most likely, are used intensively at first because their sound-symbolic nature facilitates the word-to-concept mapping. From the standpoint of mental representing of concepts, their use is equivalent to the use of Entities, Relationships, Circumstances, Quality and Quantity.

The problematic is related to mechanisms ensuring the concept-creation and their dependence on brain resources. It has been argued by Fennell and Werker (Fennell and Werker, 2003) that 14 month old children's failure in associative word-learning situations is due to a processing overload (which, however, does not incapacitate their discriminating of the words' phonetic detail).

Children's speech demands the use of so-far extant concepts, so their volume in the speech reflects the processing charge which the mental system has allowed at that given age. The plot of the use of classes (Fig. 4) shows that the line of the Sum of classes develops in a very similar way for the two language groups of children (correlation 0.87). This suggests that some resource underlying the conceptualization abilities is used quite similarly by the two language groups.

The statistical picture displayed entails several questions. One is why the classes participate with different weights within the meaning construction. The reason for this should be related to the processing charge that they demand.

A second question is related to the proportions of meaning-classes – in English the intensity of use develops in the order Entities – Relationships – Circumstances (Fig. 4) and in French the order of intensity is Entities – Circumstances – Relationships. These classes serve to mentally re-describe an Event. The mental image of an Event consists (in general) of Entities, Relationships and Circumstances. The proportion of their use to express an Event can be language dependent. The Sums of the RU of discussed three classes develop in very similar way (correlation till 36 months - 0.905). That suggests that the mental process treats the two schemes with an equal effort.

All these questions necessitate establishing a model depicting the complexity of the mental processing associated with different meaning-classes that can explain the reported statistical observation.

9. CONCLUSION

Theories and studies in the field of child language-acquisition have increasingly concentrated on the relation between language

units and semantic representations. Despite the huge amount of brain studies investigating the reactions to semantic stimuli, semantic confusion, word-to-concept mapping and other aspects of semantics, there is still no explanation from where the semantic representations come and what primary role do they have. In other words, why do they exist?

The present study proposes a model which, as first step, portrays the general biological foundation for the existence of semantic representations as information substance. The model posits that children are born equipped for the role of Self-actor in the environment and supposes that inborn information-treatment mechanisms organize information into general meaning-related classes that have the role of ensuring the Actor's survival in the environment.

The classes-hypothesis is tested by analyzing data from child language-acquisition of two languages. When children's speech is considered in terms of use of these classes, the similarity between the two language acquisition processes is important. An essential statistical observation is that children use representatives of all these classes from the onset of language production - an indication that the proposed classes reflect inborn mechanisms for mental representing of the world.

The presented result and reasoning, as it often happens in science, give rise to several novel questions. The most important of them is related to the processing load which, following the data is different for the different classes.

Conflict of interests
Author declares no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex, 63*, 196-205. https://doi.org/10.1016/j.cortex.2014.08.025

Atanasov, D., Slavova, V., & Andonov, F. (2016, July). A Statistical Study of First Language Acquisition: No Gender Differences in the Use of Parts of Speech. *In Proc. of the 12th Annual International Conference on Computer Science and Education in Computer Science (CSECS 2016)* (pp. 1-4). https://www.researchgate.net/profile/Velina_Slavova/publication/315729185_A_STATISTICAL_STUDY_OF_FIRST_LANGUAGE_ACQUISITION_NO_GENDER_DIFFERECES_IN_THE_USE_OF_PARTS_OF_SPEECH/links/58dfa453a6fdcc41bf920578/A-STATISTICAL-STUDY-OF-FIRST-LANGUAGE-ACQUISITION-NO-GENDER-DIFFERECES-IN-THE-USE-OF-PARTS-OF-SPEECH.pdf

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and cognitive processes, 18*(5-6), 513-562. http://dx.doi.org/10.1080/01690960344000026

Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of child language, 27*(03), 521-559. https://www.cambridge.org/core/journals/journal-of-child-language/article/early-development-of-nouns-and-verbs-in-french-exploring-the-interface-between-lexicon-and-grammar/854B3A644E6542A31487A7B2343828ED

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258. doi:10.1073/pnas.1113380109 http://www.pnas.org/content/109/9/3253.full

CHILDES - the child language component of the Talk Bank system for sharing and studying conversational interactions. http://childes.talkbank.org/

De Saussure, F. (1916). *Cours de linguistique générale,* publié par Ch. Bally et A. Sechehaye avec la collaboration de A. Riedlinger. Paris: Payot. https://books.google.bg/books?hl=en&lr=&id=wmLQflL01Y4C&oi=fnd&pg=PA10&dq=De+Saussure,+F.+(1916).+Cours+de+linguistique+g%C3%A9n%C3%A9rale,+&ots=nqYpoyz0Rh&sig=V6Si7Ckl8YMEFfTHqBnHKTwThBU&redir_esc=y#v=onepage&q=De%20Saussure%2C%20F.%20(1916).%20Cours%20de%20linguistique%20g%C3%A9n%C3%A9rale%2C&f=false

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech, 46*(2-3), 245-264. doi:10.1177/00238309030460020901

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development,* i-185. doi:10.2307/1166093

Friedrich, M., & Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neu-*

roreport, 16(6), 653-656. http://journals.lww.com/neuroreport/Abstract/2005/04250/Lexical_priming_and_semantic_integration_reflected.28.aspx

Frishberg, N. (1987). Home sign. *Gallaudet encyclopedia of deaf people and deafness, 3,* 128-131. https://scholar.google.bg/citations?view_op=view_citation&hl=en&user=Sw1utaAAAAAJ&citation_for_view=Sw1utaAAAAAJ:Y0pCki6q_DkC

Glezer, L. S., Kim, J., Rule, J., Jiang, X., & Riesenhuber, M. (2015). Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *Journal of Neuroscience, 35*(12), 4965-4972. https://doi.org/10.1523/JNEUROSCI.4031-14.2015

Hausser, Roland. 1989. Principles of computational morphology. Pittsburgh, PA, Carnegie Mellon University, Laboratory for Computational Linguistic, *Technical report.* The application: https://pdfs.semanticscholar.org/7dbb/02fdb2f14cf3f701e76fbf3165661197ceba.pdf

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). (2010). *Action meets word: How children learn verbs.* Oxford University Press. https://books.google.bg/books?hl=en&lr=&id=McEVDAAAQBAJ&oi=fnd&pg=PR9&dq=Hirsh-Pasek+and+Golinkoff+(2006),+%E2%80%9CAction+meets+word:+How+children+learn+verbs.%E2%80%9D&ots=KI19ZZAWI4&sig=NBgr3jPCWuIFoS2VoyTuWzvtSgU&redir_esc=y#v=onepage&q=Hirsh-Pasek%20and%20Golinkoff%20(2006)%2C%20%E2%80%9CAction%20meets%20word%3A%20How%20children%20learn%20verbs.%E2%80%9D&f=false

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65. https://doi.org/10.1016/j.cognition.2008.07.015

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B, 369*(1651). doi:10.1098/rstb.2013.0298

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology, 161*, 32-45. https://doi.org/10.1016/j.jecp.2017.03.017

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(8), 1274. http://dx.doi.org/10.1037/xlm0000235

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of child language, 12*(02), 271-295. https://doi.org/10.1017/S0305000900006449

MacWhinney, B. (2012). T*he CHILDES Project. Tools for analyzing talk-Electronic edition.* Part 1. The CHAT transcription format. August 6, 2012. https://pdfs.semanticscholar.org/7dbb/02fdb2f14cf3f701e76fbf3165661197ceba.pdf

Moran P. & J. Tommerdahl (2011). A case study of linguistic isolation and questions about subsequent language support and educational provision in the United Kingdom. In Patricia Sutcliff, William J. Sullivan & Arle Lommel (eds.), The Linguistic Association of Canada and the United States, *LACUS Forum 36: Mechanisms of Linguistic Behavior,* 229-240. Houston, TX: LACUS, 01.05.17: http://www.lacus.org/volumes/36/216_moran_p.pdf

Pinker, S. (1987). The bootstrapping problem in language acquisition. *Mechanisms of language acquisition,* 399-441. https://goo.gl/USrGZM

Robinson, G. A., Butterworth, B., & Cipolotti, L. (2015). "My Mind Is Doing It All": No "Brake" to Stop Speech Generation in Jargon Aphasia. *Cognitive and Behavioral Neurology, 28*(4), 229-241. doi:10.1097/WNN.0000000000000080

Slavova, V. & A. Soschen. (2015 a.). On mental representations: Language structure and meaning revised, *International Journal Information theories & applications 2*(4), 316-325. http://www.foibg.com/ijita/vol22/ijita22-04-p02.pdf

Slavova, V. & A. Soschen. (2015 b.) Syntactic operations - modelling language faculty, *International Journal Information theories & applications 2*(4), 326-337. http://www.foibg.com/ijita/vol22/ijita22-04-p03.pdf

Slavova, V. (2016, July). Data Collection for Studying Language Acquisition. *In Proc. of the 12th Annual International Conference on Computer Science and Education in Computer Science (CSECS 2016)* (pp. 1-4). https://www.researchgate.net/profile/Velina_Slavova/publication/316406783_DATA_COLLECTION_FOR_STUDYING_LANGUAGE_ACQUISITION/links/58fc6cb2aca2723d79d89506/DATA-COLLECTION-FOR-STUDYING-LANGUAGE-ACQUISITION.pdf

Torigoe, T., & Takei, W. (2002). A descriptive analysis of pointing and oral movements in a home sign system. *Sign Language Studies, 2*(3), 281-295. doi:10.1353/sls.2002.0013

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental psychology, 34*(6), 1289. http://dx.doi.org/10.1037/0012-1649.34.6.1289

Appendix A

List of the corpora in CHILDES data repository used in this study.

Corpus in CHILDES:	Number of Dialogues included in this study
English Belfast Corpus	10
English Bernstein-Ratner Corpus	5
English Bliss Corpus	2
English Bloom73 Corpus	6
English Braunwald Corpus	111
English Brent Corpus	24
English Brown Corpus	31
English Clark Corpus	1
English Cornell Corpus	11
English Demetras1 Corpus	25
English Feldman Corpus	19
English Fletcher Corpus	10
English Gleason Corpus	48
English Hall Corpus	2
English Higginson Corpus	4
English HSLLD Corpus	86
English MacWhinney Corpus	22
English NewEngland Corpus	19
English Peters Corpus	8
English Post Corpus	20
English Rollins Corpus	24
English Sachs Corpus	4
English Snow Corpus	4
English Suppes Corpus	1
English VanHouten Corpus	15
English Warren Corpus	4
English Weist Corpus	10
English-USA Bates Corpus	99
Eng-USA Soderstrom Corpus	1
French Champaud Corpus	32
French Geneva Corpus	15
French Hammelrath Corpus	224
French Hunkeler Corpus	22
French Leveille Corpus	34
French Lyon Corpus	207
French MTLN Corpus	299
French Paris Corpus	52
French York Corpus	23
Phonbank English Providence Corpus	1

Appendix B: Presentation of the Data. Example: English — the beginning of a dialogue.

a) As presented in CHILDES, the data source. Window of the CLAN interface.

b) Tagged and additionally annotated records of the same dialogue, as stored locally in the constructed database

c) Child speech, after extraction from the same of dialogue, stored in the local database table "Children speech" (in 3th Normal form).

Appendix C. First use of the proposed meaning classes

Examples for the period 09-19 months in English and 11-20 months in French.
There are no occurrences of use of the classes observed in the corpora before the earliest month, shown in the listed 10 examples for each class.

Months	Dialogue	Speech English	POS
		Entities	
09	ale09.br	Mama .	n:prop|Mama .
09	mor09.br	Dada .	n:prop|Dada .
09	mor09.br	egg .	n|egg .
10	mir10.br	Ma &=noise .	n:prop|Ma .
11	may11.hg	shoes .	n|shoe-PL .
11	may11.hg	baby .	n|baby .
11	may11.hg	(ba)nana .	n|banana .
11	dil11.br	no , Dillon &=noise .	co|no cm|cm n:prop|Dillon . (Self)
12	rol12.ro	ca(r) .	n|car .
12	rol12.ro	me me .	pro:obj|me pro:obj|me .
		Relationships	
11	mog11.br	go &=noise !	v|go !
12	rat12.ro	<what's> [/] what's that ?	pro:wh|what~cop|be&3S pro:dem|that ?
12	chg12.ro	pulls on hat .	v|pull-3S prep|on n|hat .
13	mrg13.br	done .	part|do&PASTP .
14	lin14.ne	draw .	v|draw .
15	ali15.br	&dæ sit .	v|sit .
16	wil16.pr	‹tired [?]› .	part|tire-PASTP .
16	ali16.bl	climb .	v|climb .
16	ali16.bl	gone .	part|go&PASTP .
17	lah17.bw	eating .	part|eat-PRESP .
		Circumstances	
12	rat12.ro	<what's> [/] what's that ?	pro:wh|what~cop|be&3S pro:dem|that ?
13	bry13.ne	yyy up .	adv|up .
14	mir14.br	out .	adv|out .
14	nao14.sa	dere [: there] .	adv|there .
15	taz15.bw	out Dee baba [= bottle] .	prep|out n:prop|Dee n|baby .
15	taz15.bw	Mommy out .	n:prop|Mommy adv|out .
16	ali16.bl	away .	adv|away .
16	ali16.bl	there Mama .	adv|there n:prop|Mama .
17	lae17.bw	outside [= actually says side] .	adv|outside .
17	lae17.bw	down .	adv|down .
		Circumstances	
12	rat12.ro	<what's> [/] what's that ?	pro:wh|what~cop|be&3S pro:dem|that ?
13	bry13.ne	yyy up .	adv|up .
14	mir14.br	out .	adv|out .
14	nao14.sa	dere [: there] .	adv|there .
15	taz15.bw	out Dee baba [= bottle] .	prep|out n:prop|Dee n|baby .
15	taz15.bw	Mommy out .	n:prop|Mommy adv|out .
16	ali16.bl	away .	adv|away .
16	ali16.bl	there Mama .	adv|there n:prop|Mama .
17	lae17.bw	outside [= actually says side] .	adv|outside .
17	lae17.bw	down .	adv|down .
		Quality and Attribution	
10	mir10.br	yummy [?] &=noise .	adj|yum&dn-Y .
13	mrg13.br	hot .	adj|hot .
14	nor14.ne	orange .	adj|orange .
14	mir14.br	big .	adj|big .
15	taz15.bw	my me Dada .	pro:poss:det|my pro:obj|me n:prop|Dada .
16	stf16.pe	gross .	adj|gross .
16	ali16.bl	dirty .	adj|dirt&dn-Y .
17	lae17.bw	I want my bottle .	pro:sub|I v|want pro:poss:det|my n|bottle .
17	lag17.bw	a@z:sc <my bike> [?] .	unk|a pro:poss:det|my n|bike .
18	ger18.cl	he sleepy .	pro:sub|he adj|sleep&dn-Y .
		Quantity and Precision	
14	nor14.ne	the woof .	det|the on|woof .
14	nor14.ne	that duck .	det|that n|duck .
15	ali15.br	a mommy .	det|a n|mommy .
16	ali16.bl	more .	pro:indef|more .
17	lah17.bw	xxx have that one . [+ PI]	v|have det|that pro:indef|one .
18	lae18.bw	<eat all> [?] .	v|eat pro:indef|all .
18	ger18.cl	two , I .	det:num|two cm|cm pro:sub|I .
19	lah19.bw	xxx this one [?] [>] . [+ PI]	det|this pro:indef|one .
19	laf19.bw	<six [?] egg [* 0s]> [<] .	det:num|six n|egg .

Months	Dialogue	Speech French	POS
		Entities	
11	jli11.Pa	yyy maman .	n|maman&f .
11	jli11.Pa	papa .	n|papa&m .
12	mrc12.Ly	bébé !	n|bébé&m !
12	mra12.Ly	des ‹cubes [= ! gémit]› .	prep|de&les n|cube&m-PL .
12	tmc12.Ly	chat .	n|chat&m .
12	mra12.Ly	‹(...) oh› les pingouins .	co|oh det|les&pl n|pingouin&m-PL .
12	mra12.Ly	un ‹vélo [= ! gémit]› .	det|un&m&sg n|vélo&m .
13	ans13.Ly	c'est bien Ana . (N.B. Self)	pro:dem|ce$v:aux|être&PRES&3s n|bien&m n:prop|Ana . (N.B. Self)
13	tmd13.Ly	‹gâteau [/]› gâteau .	n|gâteau&m .
13	ani13.Ly	ramasse un lego .	v|ramasser-PRES&SUB&13s det|un&m&sg n|lego&m=toy .
		Relationships	
11	tmb11.Ly	+‹ caché .	part|cacher-PP&m .
12	mra12.Ly	tu lis !	pro:subj|tu v|lire-PASS&PRES&12s !
12	nad12.Ly	+‹ ‹a [= ! rit]› .	v:aux|avoir&PRES&3s .
12	mra12.Ly	+‹ veux dormir là .	v:mdl|vouloir&PRES&12s v|dormir-INF adv:place|là .
14	mrb14.Ly	çà tourne .	adv|çà v|tourner-PRES&SUB&13s .
14	ana14.Ly	eh regardes .	co|eh v|regarder-PRES&SUB&2s .
14	mrb14.Ly	me ‹voir [?]› .	pro:obj|me v|voir&INF .
15	tmb15.Ly	est là .	v:aux|être&PRES&3s adv:place|là .
16	tmd16.Ly	0il est où> [?] doudou .	pro:subj|il v:aux|être&PRES&3s pro:rel|où n|doudou&m=blankie .
16	tmd16.Ly	a pas .	v:aux|avoir&PRES&3s adv:neg|pas .
		Circumstances	
11	jli11.Pa	ça ?	pro:dem|ça ?
12	tmc12.Ly	encore ah .	adv|encore co|ah .
12	mra12.Ly	+‹ ‹non ,› tout_de_suite .	co|non=no cm|cm adv|tout_de_suite .
13	tmc13.Ly	et c(el)ui-là ?	conj|et pro:dem|celui-là ?
14	mrb14.Ly	‹ça [//]› ça tourne yyy .	pro:dem|ça v|tourner-PRES&SUB&13s .
15	anz15.Ly	et là .	conj|et adv:place|là .
15	ans15.Ly	et ‹voilà [?]› .	conj|et adv:place|voilà .
16	tmc16.Ly	‹à côté› [?] .	prep|à n|côté&m .
17	mra17.Ly	ah ‹dedans [?]› .	co|ah adv:place|dedans .
17	ana17.Ly	à moi .	prep|à pro|moi&sg .
		Quality and Attribution	
12	nah12.Ly	+‹ ‹ma [?]› maman .	det:poss|ma&sg n|maman&f .
12	tmd12.Ly	grand .	adj|grand&m .
15	ann15.Ly	bleu [?] .	adj|bleu&m !
16	tma16.Ly	fermé [?] .	adj|fermé&m .
16	tmc16.Ly	xxx rigolo .	adj|rigolo&m .
17	tmb17.Ly	sa tête .	det:poss|sa&f&sg n|tête&f .
17	ane17.Ly	pas gentille [?] .	adv:neg|pas adj|gentille&f .
18	tmg18.Ly	rouge [?] .	adj|rouge .
18	tme18.Ly	est lourd[?] .	v:aux|être&PRES&3s adj|lourd&m .
19	mrd19.Ly	yyy mon sac sac .	det:poss|mon&m&sg n|sac&m n|sac&m
		Quantity and Precision	
12	mra12.Ly	et les deux .	conj|et det|les&pl num|deux .
12	mra12.Ly	parce que ‹il [/]› ‹il [/]› ‹il [/]› il a lit un petit peu .	conj|parce pro:rel|que pro:subj|il v:aux|avoir&PRES&3s n|lit&m det|un&m&sg adj|petit&m n|peu&m .
13	mra13.Ly	‹(.) la la la [?]› .	det|la&f&sg det|la&f&sg det|la&f&sg .
14	mrc14.Ly	le chat .	det|le&m&sg n|chat&m .
15	jli15.Pa	elle va aller dans le yyy .	pro:subj|elle v:mdl|aller&PRES&3s v:mdl|aller&INF prep|dans det|le&m&sg .
16	tmb16.Ly	une abeille .	det|une&f&sg n|abeille&f=bee .
17	ana17.Ly	c'est le feutre .	pro:dem|ce$v:aux|être&PRES&3s det|le&m&sg n|feutre&m .
18	tme18.Ly	‹deux trois› [?] .	num|deux&m num|trois&m .
19	jli19.Pa	c'est un pain !	pro:dem|ce$v:aux|être&PRES&3s det|un&m&sg n|pain&m !
20	ant20.Ly	et un carré pour Ana petit (N.B. Self)	conj|et det|un&m&sg adj|carré&m n|pour&m n:prop|Ana adj|petit&m .

N.B. The youngest children in the French group are 11 months old.