Possible contributions of simulation and robotics to understanding human language acquisition

Peter Ford Dominey

Sequential Cognition and Language Group
Institut des Sciences Cognitives - CNRS
<dominey@isc.cnrs.fr>

Résumé/Abstract

The goal of the research presented here is to establish a developmental trajectory of neurophysiologically grounded models that simulate successively more complex aspects of human language acquisition based on a minimum of pre-wired language-specific functionality, that is compatible with observations of perceptual and language capabilities in human development.

1. *Introduction*

One of the most interesting open questions and debates in modern cognitive science concerns the nature of the mechanisms responsible for human language acquisition. One of the principal axis along which theories diverge concerns the degree to which innate language-specific mechanisms are required. Related to this is the question concerning the role of learning as for the individual acquires her native language. As outlined in more detail below, one theoretical position holds that man is genetically endowed with a highly language-specific capability with a pre-specified universal grammar, and that learning consists in setting the parameters of this UG to accommodate the target native language (Chomsky 1995). One of the central tenets of this position is based around the "poverty of the stimulus" argument which holds that the linguistic data to which the child is exposed is highly impoverished with respect to the complexity of the target grammar, thus eliminating the possibility that acquisition could be accomplished by a general learning mechanism. In this perspective, syntax is the central organizer of language, from which semantic structure is derived. An alternative theoretical position holds that the language-specific component is significantly reduced, and that learning exploits more general capabilities, and plays a much more significant role (Langacker, Feldman, Tomasello etc). One of the central tenets of this approach is that the problem of language acquisition requires that we take the structure of meaning and the communicative function of language seriously, as this structure highly influences the structure of language. Indeed, in this context we can consider the possibility that perhaps it is the structure of meaning and communicative functions that drives the structure of syntax, rather than the opposite.

One method to test such a theory is to construct a model based on that theory, and then to confront the model with experimental conditions to determine the extent to which the model can account for (and predict) behavioral and neurophysiological results. In this context, the objective of the current research program is to use emerging technology in neural network simulation, computer vision, speech technology and robotics in order to construct robotic systems that permit the testing of theories of language development. In general, the advantage of this type of « constructionist » approach, is that by actually attempting to build a system that can acquire language (or a « miniature » language), one is forced to take seriously a number of issues. In particular, if language acquisition is defined as the construction of the relations between sentences and meaning, then the issue of meaning must be carefully considered. Likewise, the domain called « language » is immense, and so we must also seriously consider which issues to address.

Methodological issues

A fundamental issue that must be addressed in this context, is that language is a highly complex object of study, that interacts with essentially all aspects of higher sensory and cognitive functions in an intricate manner. In this context three important points should be made.

(1) One should make it clear that one is aware of this complexity of language, including an at least partial enumeration of the key issues. From a functional perspective of language as the interface between sentences and meaning, these issues would include classical elements of lexical and phrasal semantics including thematic role assignment (syntactic comprehension more generally), focus and informational content, anaphoric reference, time-mode and aspect, etc.

(2) One should then clearly identify which of these particular aspects will be addressed, including the performance measure based on documented human experimental result.

(3) Finally, one should clearly motivate these particular aspects, and justify that the choice is interesting, informative and scaleable.

This type of modeling approach can serve two related goals : (1) to understand how the real system might function, and (2) to produce a non-human system that can display interesting language performance/behavior.

In the first case, validation of the model will consist in demonstrating that it can account for psycholinguistic data, while the second would consist in demonstrating that the system displays interesting/robust language behavior. It is also possible to attempt to pursue both of these goals, with the idea that if we understand how the real system works, then we can use this understanding to build high performance artificial systems.

2. Theoretical approach : defining the problem

Feldman et al. (1990) posed the problem of « miniature » language acquisition based on <sentence, image> pairs as a « touchstone » for cognitive science. In this task, an artificial system is confronted with a reduced version of the problem of language acquisition faced by the child, that involves both the extraction of meaning from the image, and the mapping of the paired sentence onto this meaning.

Extraction of meaning

In this developmental context, Mandler (1999) suggested that the infant begins to construct meaning from the scene based on the extraction of perceptual primitives. From simple representations such as contact, support, attachment (Talmy 1988) the infant could construct progressively more elaborate representations of visuospatial meaning. Thus, the physical event « collision » is a form of the perceptual primitive « contact ». Kotovsky & Baillargeon (1998) observed that at 6 months, infants demonstrate sensitivity to the parameters of objects involved in a collision, and the resulting effect on the collision, suggesting indeed that infants can represent contact as an event predicate with agent and patient arguments.

Siskind (2001) has demonstrated that force dynamic primitives of contact, support, attachment can be extracted from video event sequences and used to recognize events including pick-up, put-down, and stack based on their characterization in an event logic. The use of these intermediate representations renders the system robust to variability in motion and view parameters. Most importantly, Siskind demonstrated that the lexical semantics for a number of verbs could be established by automatic image processing.

Sentence to meaning mapping

Once meaning is extracted from the scene, the significant problem of mapping sentences to meanings remains. The nativist perspective on this problem holds that the <sentence, meaning> data to which the child is exposed is highly indeterminate, and underspecifies the mapping to be learned. This « poverty of the stimulus » is a central argument for the existence of a genetically specified universal grammar, such that language acquisition consists of configuring the UG for the appropriate target language (Chomsky 1995). In this framework, once a given parameter is set, its use should apply to new constructions in a generalized, generative manner.

An alternative functionalist perspective holds that learning plays a much more central role in language acquisition. The infant develops an inventory of grammatical constructions as mappings from form to meaning (Goldberg 1995). These constructions are initially rather fixed and specific, and later become generalized into a more abstract compositional form employed by the adult (Tomasello 1999). In this context, construction of the relation between perceptual and cognitive representations and grammatical form plays a central role in learning language (e.g. Feldman et al. 1990, 1996; Langacker 1991; Mandler 1999; Talmy 1998).

These issues of learnability and innateness have provided a rich motivation for simulation studies that have taken a number of different forms. Elman (1990) demonstrated that recurrent networks are sensitive to predictable structure in grammatical sequences. Subsequent studies of grammar induction demonstrate how syntactic structure can be recovered from sentences (e.g. Stolcke & Omohundro 1994). From the « grounding of language in meaning » perspective (e.g. Feldman et al. 1990, 1996; Langacker 1991; Goldberg 1995), Chang & Maia (2001) exploited the relations between action representation and simple verb frames in a construction grammar approach. In effort to consider more complex grammatical forms, Miikkulainen (1996) demonstrated a system that learned the mapping between relative phrase constructions and multiple event representations, based on the use of a stack for maintaining state information during the processing of the next embedded clause in a recursive manner.

In a more generalized approach, Dominey (2002) exploited the regularity that sentence to meaning mapping is encoded in all languages by word order and grammatical marking (bound or free) (Bates et al. 1982). That model was based on the functional neurophysiology of cognitive sequence and language processing and an associated neural network model that has been demonstrated to simulate interesting aspects of infant (Dominey & Ramus 2000) and adult language processing (Dominey et al. 2003).

In this context, one objective of the research presented in this paper is first, to demonstrate that a neurophysiologically motivated model of sequence learning can account for a significant body of psycholinguistic data, and then to demonstrate that when combined with a system for extracting meaning from visual scenes, an extended version of this model can account for a number of observations of developmental language capabilities.

Some challenges in phrasal semantics

Before proceeding with the model descriptions and simulations, it is important to establish the framework within phrasal semantics in which we will operate. As noted above, our approach relies on the presence of a certain degree of isomorphism between conceptual structures and the linguistic structures that communicate them.

Jackendoff has recently proposed a three tiered framework for describing phrasal semantics, i.e. the manner in which sentence structure communicates meaning beyond that of the sum of the lexical elements. In this context Jackendoff notes that for simple compositional structure based on argument satisfaction, modification, and lambda extraction and variable binding, there is a close correspondence between the configurations of lexical items in syntax and conceptual structure (Jackendoff 2002, 387). However, what is referred to as enriched composition such as reference transfer depicted in the sentence *The ham sandwich over in the corner wants more coffee* manifests situations in which this iconicity is claimed to break down. For our current purposes, the objective from a developmental perspective, is to explore how far the infant can go with the « close correspondence » hypothesis.

In Jackendoff's framework, phrasal semantics is organized into descriptive, referential and information/focus tiers. The descriptive tier comprises thematic role assignment and associated argument satisfaction. In this context we will demonstrate how grammatical constructions indexed by word order and grammatical marking fulfill these functional criteria. The referential tier includes co-reference and modal structure. In this context we will demonstrate how the grammatical construction framework allows for resolution of pronoun reference and reflexive verb argument assignments. The information topic/focus tier third tier includes representation of pragmatic focus that can also be addressed in the grammatical construction framework as described below. Considering the following sentences : *John pushed the block*, and *The bock was pushed by John*, one can see that these are not perfectly equivalent statements, due to the focus component. Interestingly this indicates that the representation of meaning *requires* a discourse variable to indicate whether agent, object or recipient is in focus. In the succeeding sections these different issues will be addressed.

3. Thematic role assignment by the ATRN model

Returning to the problem posed by Feldman, our task is to develop a system that can learn the structural mapping between sentences and their meanings. Functionally, we will pose this in terms of thematic role assignment, the process by which one can determine from a sentence « who did what to whom ». More specifically, for a given action, we consider that there is an agent and object and optional recipient, and that these roles should be assigned based on a decoding of the sentence. How does the decoding proceed ?

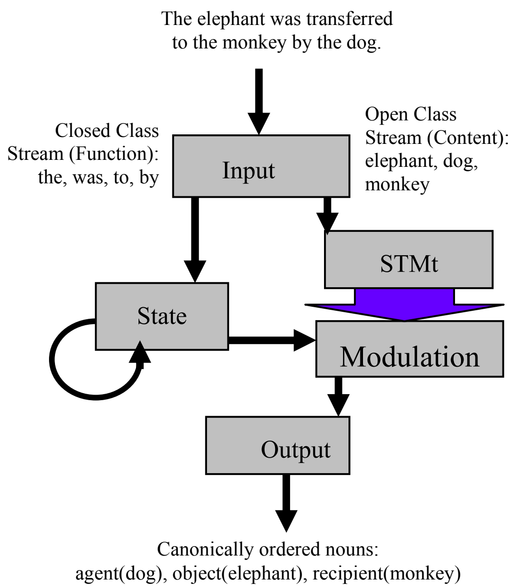

Figure 1. Thematic Role Assignment in the Abstract Temporal Recurrent Network (ATRN) Sequence Processing Model. The recurrent State system provides sensitivity to temporal and serial structure, forming the core of the Temporal Recurrent Network. The Short Term Memory and Modulation components provide for the representation of abstract sequences of variables, guided by the TRN, forming the core of the Abstract Recurrent Network.

Bates et al. (1982) proposed the powerful universal property that across languages, this meaning structure is encoded regularities including word order, grammatical marking (morphology and function words), and prosody. Exploiting this theory, we mapped the problem of thematic role assignment onto the model architecture in Figure 1. The task of the model after presented with an input sentence, word by word, is to respond by producing the agent, object and recipient of the action, in that (canonical) order. As words in a sentence are presented to the model, closed class words are directed to the recurrent State network that encodes their identity and order of arrival. Open class words are directed to a short term memory from which they can be retrieved in an arbitrary order based on input from the State system. Sentence to meaning mapping, or thematic role assignment, corresponds for each sentence to the retrieval of the open class elements from STM in the canonical order. Thus, each sentence type (dative, dative passive etc.) corresponds to a particular re-ordering of elements in STM, based on the coding of grammatical structure by closed class elements in the recurrent network, unique to each sentence type. This corresponds to the concept of grammatical construction (Goldberg 1995).

Based on this mapping of thematic role assignment onto the model, we are able to demonstrate that the system can learn the mappings for the nine sentence types of the Caplan et al. (1985) syntactic comprehension task. It is of interest that the model was initially developed to describe neural activity in the primate prefrontal cortex during sensorimotor sequence learning (Dominey, Arbib & Joseph 1995) and the learning of abstract sequencing rules in humans (Dominey et al. 1998). This provides support for the argument that the underlying processes are not wholly language specific.

Data the model accounts for

The validation of such a model takes place by determining to what extent the model accounts for (and predicts) data in its performance domain. Here we will consider performance of the model in terms of psycholinguistic results in infants and adults ; performance of agrammatic patients in language and sequencing tasks, and human event related potential (ERP) results.

Infant performance

From the developmental perspective, a series of experiments with children ranging in age from a few days to 8 months have demonstrated their sensitivity to serial, temporal and abstract structure of sound sequences in the context of language acquisition. Saffran et al. (1996) presented sound sequences made up of three-syllable nonsense words to 8 month old infants. They demonstrated that after only two minutes of exposure the infants had learned the statistical regularity that syllables within words follow one another with higher frequency than syllables that span word boundaries. We simulated this behavior with syllables presented as sequence elements to the model, and demonstrated significantly longer response times to syllables that spanning word boundaries than those within words, indicating that the model was also sensitive to these statistical regularities. Indeed this was no surprise, as this was precisely the kind of sequence learning task the model was developed for.

In a related study, Nazzi et al. (1998) presented new born infants with speech that had been low pass filtered, and demonstrated that after 10 minutes of exposure the children acquired sufficient knowledge of the prosodic structure of the target language to allow them to discriminate it from languages in different rhythm classes. To simulate these experiments we recoded sentences as consonant vowel sequences in order to capture the prosodic structure. When trained on these temporally structured sequences the TRN (temporal recurrent network) was also able to discriminate between sentences originating from different rhythm classes (Dominey & Ramus 2000). Finally, Marcus et al. (1999) demonstrated that after a two minute exposure to sound sequences such as *la-di-di, wo-fe-fe* 7 month old infants acquired significant knowledge of the underlying abstract ABB structure and could use this in discriminating sequences made up of new syllables. This significantly differed from the Saffran et al. (1996) results because the knowledge was abstract and rule based and could be applied to sequence elements that the infants had not previously been exposed to.

Interestingly, this is precisely the kind of task that the ARN was developed for. We thus demonstrated that indeed, after training with sequences adhering to a given ABA or ABB rule, the model was capable of discriminating between sequences made up of new elements that did or did not adhere to the learned rule. The intuition that manipulation of these abstract rule-based structures will have applications in language processing will be developed below.

While it was quite interesting and encouraging that the model could capture these infant psycholinguistic behavioral results, the real challenge was to determine whether the model could account for more « practical » language related behavior. In this context, it has been demonstrated that new born infants are capable of performing the lexical categorization of English function *vs* content words based on their auditory properties (Shi et al. 2000). Given the likelihood that this early discrimination capability is based at least in part on sensitivity to prosodic structure, we attempted to determine if the model could perform this lexical categorization based on the fundamental frequency of the auditory speech signal. Indeed, we demonstrated that the F0 structure of French and English contains information that allows a successful lexical categorization based on the presence of F0 peaks more predominantly in content than function words, and that the TRN is sensitive to this structure (Blanc et al. 2003). These results suggest a functional relation between temporal sequence processing and language acquisition *via* lexical categorization (see related papers in Morgan & Demuth 1996). In this context, we demonstrated that increasing the time constants of neurons in the recurrent network of the TRN resulted in a failure in lexical categorization as well as in the processing of temporal sequences of short duration auditory elements as observed by Tallal et al. (1998) and Wright et al. (1997), similar to the behavioral deficits observed by these authors in children with specific language impairment (SLI). These results indicate the potential importance of functional sensitivity to temporal structure for early lexical categorization.

They also indicate that the presence of a perturbation of this capability, can result in a performance profile corresponding to a subgroup of SLI.

« Adult » performance

These developmental results lead to the question of how they can be extended to provide adult-like processing. As described above, we reformulated syntactic comprehension (thematic role assignment) in terms of the selection of abstract mappings selected by configurations of closed class elements in the recurrent network. This allowed the simulation of adult performance as described above, and in detail in Dominey et al. (2003).

This indicates that the model is sensitive to grammatical structure as coded by grammatical markers, and word order both of which contribute to the specification of thematic role assignment (Bates et. al. 1982) corresponding to the descriptive tier of phrasal semantics (Jackendoff 2002). Another dimension in which communicative content can be encoded is in the modulation of the prosodic structure of sentences to communicate different attitudes (e.g. surprise, certitude, resignation etc.)

This prosodic information is principally carried by the fundamental frequency of the speech signal (F0) (see Blanc & Dominey 2003 for a review). As we demonstrated that the temporal recurrent network (TRN) was sensitive to F0 structure for lexical categorization, it appeared likely that this sensitivity would be functional in the prosodic attitude discrimination domain as well. We thus trained the model with sentences spoken in six different prosodic attitudes, based on the study from Morlec et al. (2001), and then tested its discrimination capability with new sentences, demonstrating performance at the level of human subjects (Blanc & Dominey 2003), extending the model's performance into the Information tier (Jackendoff 2002).

Aphasic/impaired behavior

A further crucial test of such a model is its ability to explain and predict performance in conditions of functional impairment. Cerebral lesions in and around Broca's area in the left peri-sylvian region result in agrammatic aphasia in which the ability to use purely syntactic or grammatical cues (e.g. function words and grammatical morphology) is severely impaired (Caplan et al. 1985). From the perspective of the model, impaired processing of function words in the recurrent network would yield agrammatic aphasia, but would also impair the ability to manipulate non-linguistic transformations as in the rule ABC-BAC.

While the model could reproduce observed aphasic performance in linguistic tasks, the more interesting possibility was that it predicted that these linguistic impairments should be correlated with impairments in non-linguistic rule based sequencing tasks since the two are realized — for the model — by the same mechanism. Thus, Figure 2A illustrates the striking correlation between performance on the Caplan task, and performance on recognizing sequences that adhere or not to the abstract structure ABC-BAC in aphasic patients (Lelekov et al. 2000, Dominey et al. 2003). This correlation suggests a shared underlying mechanism.

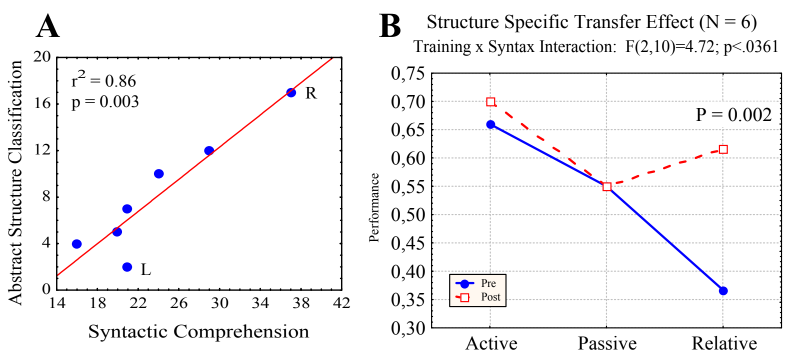

Figure 2A. Correlated impairments in syntactic comprehension and in non-linguistic sequence transformation processing. *Figure 2B*. Training on a "relativised" non-linguistic sequences yields improved performance in comprehension of relativised sentences.

To further investigate this possibility, we predicted, based on the model, that training in the sequence processing domain should transfer to improved performance in the language processing domain. We thus considered that comprehension of the relative phrase *the cat(1) that the dog(2) chased(3)* would require recovery of the canonical order *the dog(2) chased(3) the cat(1*), structurally corresponding to the rule 123-231. We trained 6 agrammatic patients on this non-linguistic sequencing rule over 10 weekly sessions, and then tested their performance on sentence processing. Figure 2B illustrates their pre- and post-reeducation performance. We see that there is no change for the active and passive sentence types, while there is a significant improvement for the relative sentences (Hoen et al. 2003). This further supports the hypothesis that a common underlying neurophysiological system is responsible.

Human neurophysiology

If a common neurophysiological mechanism was responsible for the transformation processing in thematic role assignment and for our non-linguistic tasks, we would expect to see neurophysiological evidence of this. In this context, we thus set out to reproduce two characteristic language-related ERP (event related potential) profiles in our non-linguistic sequencing tasks. Lelekov et al. (2000) compared ERP profiles in response to errors in sequential order *vs* errors in rule-based structure. The rule-based structural errors yielded a positive wave centered around 600ms with a spatial topography quite similar to that of the language-related P600, typically evoked in response to structural language violations (Lelekov, Dominey & Garcia-Larrea 2000). Similarly, the left anterior negativity (LAN) is associated with syntactic structural processing as triggered by grammatical function words (Brown et al. 1999). We thus observed that the cerebral processing of function symbols, that indicated the appropriate transformation in abstract sequences, resulted in a left anterior negativity remarkably similar to the LAN associated with function word processing (Hoen & Dominey 2000).

4. Model 2 : mapping sentences to meaning

The results described thus far provide preliminary evidence that there is some functional overlap between abstract transformation processing, and the mapping of grammatical structure to meaning. The semantics of the model described above are quite impoverished, however, thus placing severe limitations on the types of structure mapping that can be investigated. In this section we present a modified model that retains the operating principles of the first model, while introducing a richer semantics.

In order to describe the structure mapping model, we first summarize the essential properties of the ATRN (Abstract Temporal Recurrent Network). Given an input sequence of open and closed class elements, the ATRN should return open class elements in their canonical order, i.e. the (possibly) re-ordered sequence of open class elements corresponding to the learned rule that is associated with the particular configuration of closed class elements in the sequence. The closed class configuration is coded in the recurrent network of the TRN that guides the output ordering via selection of memory contents from the ARN. This allows the model to perform abstract sequencing tasks as well as thematic role assignment. The underlying assumptions of the model corresponding quite closely with those of the « cue competition » model of Bates et al. (1982). These authors proposed that across languages, the mapping of phrasal structure to meaning is encoded or specified by cues including word order, grammatical morphology and function words, and prosody. During acquisition, these cues compete such that the most reliably informative cues are detected and exploited.

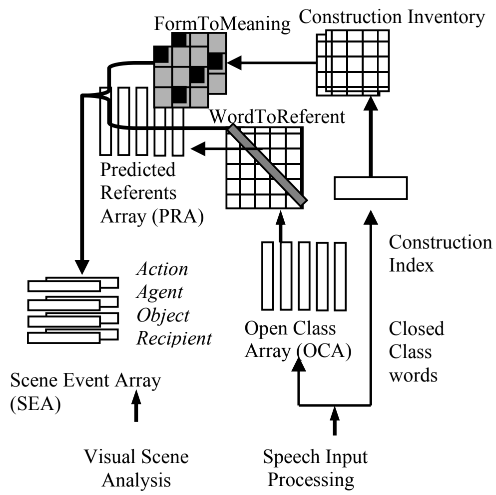

Figure 3. Sentence to Meaning Mapping Model. See text for functional details.

Functional model overview

Words in sentences, and elements in the scene are coded as single bits in respective 25-element vectors. On input, open class words populate the Open Class Array (OCA), and closed class words populate the Construction Index. On the meaning side, meaning is encoded as event (agent, object, recipient) in the Scene Event Array (SEA). In the initial simulations these meanings were encoded by hand, and in later simulations were subsequently extracted directly from Visual Scene Analysis described below. Learning occurs at the lexical and the phrasal semantics levels. For lexical semantics, words in OCA are translated to Predicted Referents via the WordToReferent mapping to populate the Predicted Referents Array (PRA). These WordToReferent mappings are initially learned by associating each word in the input sentence with each referent in the scene. Based on cross-situational statistics a word will co-occur more often with its referent than with other referents, so the learning is effective (see Siskind 1996). This crude learning method is later replaced by syntactic bootstrapping described below. For learning phrasal semantics, once a minimum of lexical semantics is in place the FormToMeaning mapping of PRA elements onto their roles in the Scene Event Array (SEA) for a given sentence type is extracted, and stored in the ConstructionInventory, an associative memory that uses ConstructionIndex as an index for storage and retrieval. ConstructionIndex encodes the closed class words that uniquely characterize each sentence type. Once a grammatical construction is thus learned it can be used for all new sentences of that construction type in a robust form of systematic generalization.

Initial conditions

Before embarking on this exercise, we must clearly establish what are the initial pre-wired processing capabilities available to the language learner. From the linguistic perspective it has been established that by the 16-19 month period that concerns us, children are capable of segmenting words from continuous speech. Already at 9 months, infants are capable of using statistical regularities of artificially generated sound sequences to detect the boundaries of words after only 2 minutes of exposure (Saffran et al. 1996), and by the time they reach the 16-19 month period, they have a well developed word segmentation capability (see Pinker 1987, Jusczyk 1997 for reviews). In addition to this early segmentation capability, it also appears that infants are able to use auditory cues in the speech signal in order discriminate between the major lexical categories of closed class *vs* open class words (Shi et al. 1999). This early discrimination between the closed class of grammatical morphemes and function words (*the, a, to, by, from*, etc.) vs. the open class of nouns, verbs, adjectives, adverbs etc. relies not only on acoustic form differences, but also on different statistical distributions and additional cues to which the infants are sensitive, and forms a crucial component of the initiation of syntactic processing (see Morgan & Demuth 1996 for an extensive treatment of this issue). The result of this early, discrimination capacity applied to the child's target language will then be expressed in adulthood. Indeed, in adults, extensive data from event related potentials, brain imagery and psycholinguist studies indicate that these lexical categories are treated by distinct and dissociated neurophysiological processing streams (e.g. Friederici 1985, Osterhout 1997, Pulvermüller 1995, Brown, Hagoort, & ter Keurs 1999). So word segmentation and the capability to discriminate open vs. closed class lexical categories are capabilities that can be considered in place at or before 16 months of age.

If language acquisition is the learning of a mapping between sentences and meanings, then the infant must also have some pre-linguistic capacity for representing meaning. From this non-linguistic perspective, already at 6 months of age, children are capable of processing causal events with agents, objects and actions and using these « naive physics » representations to understand simple action scenarios that involve goal-directed reaching for objects (Woodward 1998). Similarly, infants in this same age range display rather sophisticated knowledge of the physical properties of objects that allows them to « parse » and understand dynamic scenes with multiple objects (Carey & Xu 2000). This implies the existence of conceptual representations that can be instantiated by non-linguistic (e.g. visual) perceptual input prior to the development of language. These conceptual representations will form the framework upon which the mapping between linguistic and conceptual structure can be built. This approach does not exclude the possibility that the conceptual representation capability will become more sophisticated in parallel with linguistic development (see Bowerman & Levinson 2001 for a survey of the issue). It does require, however, that at least a primitive conceptualization capability that can deal with multiple-agent events exists in a pre-linguistic state, a position that may still be open to debate by formal linguists (see Crain & Lillo-Martin 1999 for the formal linguistics perspective).

Identification of the preceding initial conditions attempted to define constraints on the system itself, but we must also take into account the constraints on the environment in which the system is to learn. In particular, we have identified requirements on the initial conceptual and linguistic processing capabilities. We must now have some assurance that indeed, the environment is such that the scenes to be conceptualized and the accompanying sentences will have some relation. Specifically, to some extent we must assume that language input describes ongoing events that the infant can perceive (see Hirsh-Pasek & Golinkoff 1996, Pinker 1987). Having identified these initial state processing capabilities, we can now describe the model that will embody these conditions.

Accounting for developmental data

Part of the goal of this work was to consider how simulation studies of language acquisition can become more realistic by the introduction of conceptual grounding. Conceptual grounding is the use of an internal or « conceptualized » representation of the perceived external world of the language learner. In essence, the benefit of this approach is that this structured internal representation provides an additional dimension of structural regularity that is highly correlated with the linguistic input, and will thus have a substantial influence on the efficacy of learning. Several aspects of this modeling approach are discussed in the following paragraphs (see Dominey 2000).

Importance of synergistic grounding for learning and generalization

In ideal conditions free of ambiguity and noise, learning the mapping between language and meaning would be relatively straightforward. This is not the case however, and learning systems must thus take advantage of all possible sources of regularity, particularly in initiating the acquisition process. In particular, knowledge of word meaning can aid in learning syntactic regularities, which in turn can aid in refining word meaning. From this perspective, we observed that in the face of referential noise or variability in the input, the standard statistical acquisition of word meaning yielded quite weak results. However, when this weak but present semantic information was allowed to interact in a synergy with the then weak but present syntactic structure the situation changed dramatically. Indeed, we observed that there is a true and unavoidable benefit from the synergistic interaction of syntactic and semantic knowledge. Weak initial knowledge of word meaning allowed the learner to establish an initial mapping between structure in the sentence and structure in the world (or, indeed in the conceptualized scene). In turn, this knowledge of structural correspondences could then be used to refine and accelerate the learning of word meanings *via* syntactic bootstrapping by drastically reducing the search in the mapping of words to their referents, as syntax identifies the referent. This refined semantic representation then contributes to an improved syntactic representation in a truly synergistic coupling. The performance benefit of this synergistic bootstrapping was also made clear in Experiment 4 which examined how acquired knowledge of syntactic structure would affect the learning of new verbs. Indeed, as syntactic knowledge became successively acquired with exposure to training, new verbs were more rapidly acquired. This demonstrates how knowledge of syntactic structure simplifies the learning of new words.

From a simulation perspective, it is reassuring that this synergistic grounding or bootstrapping in the model is due to a central choice in how the learning architecture works. Indeed, without this mechanism the learning is too weak and cannot proceed. Thus, in implementing a system that must learn, we in a sense rediscover the properties (i.e. synergistic bootstrapping) of the learning system that we simulate.

Effects of concreteness and lexical category on learning rate

Models or theories of language acquisition should not only achieve some level of performance, but should also explain behavioral observations about changes in the capabilities of language learners during the course of acquisition. For children learning English, it is a highly reliable observation that during the early stages of learning there is an asymmetrical vocabulary distribution that heavily favors nouns over verbs (Bates et al. 1995). A recent paper by Gillette et al. (1999) reviews two theoretical explanations, and experimental data that may tip the balance. One position holds that nouns are learned first because they are conceptually simpler (they describe single objects) than verbs (that describe relations between multiple objects), thus placing the burden of explanation on a conceptual system that is initially not prepared for representing verbs. An alternative position explains the noun verb gradient by two related mechanisms. The first has to do with the concreteness of word to world mappings. Concrete nouns (and particularly those that are among the most frequent in child vocabularies) correspond to concrete, easily observed objects in the child's environment. The high observability factor contributes to the ease in early acquisition based on simple associative word-to-world mapping. According to this view, the noun preference in early vocabularies is more related to an effect of concreteness than of lexical category per se. Indeed, highly concrete verbs are present in the earliest vocabularies.

Once this initial vocabulary of concrete terms has been established, it provides the grounded basis for the subsequent acquisition of more abstract terms, including verbs. The presence of known nouns in the context of an unknown verb provides the required scaffolding of a clause level syntax that allows even an abstract verb to now be correctly associated with the appropriate aspect of the scene. In this framework then, the noun verb discontinuity is more related to concreteness than to lexical category, and the development of a concrete set of nouns provides a syntactic — thematic mapping that allows the subsequent acquisition of more abstract verbs. In other words there is no need for explaining late verbs in terms of lack of an appropriate conceptual representation at the outset. What is missing is the syntactic representation that must first be built from the scaffolding of concrete nouns (Gillette et al. 1999).

The simulation results in Experiments 1 and 2 support this explanation of the vocabulary development. In particular, Experiment 2 revealed that differences in learning over time were related to concreteness and that lexical category per-se for nouns and verbs did not influence learning rate independent of concreteness. A prediction that issues both from the theoretical position defended by Gleitman and colleagues (Gillette et al. 1999), and supported by the model, is that for human languages in which verbs are associated with more concrete aspects of scenes than are nouns, one should observe an increase in the proportion of verbs in the initial vocabularies. Mandarin appears to be such a language, and indeed, recent evidence from studies of Mandarin children indicate that this prediction is born out in the data (Snedeker & Li 2001).

5. A first step towards addressing a « usage based » acquisition of grammatical constructions in a robotic system

Tomasello (2003) proposes a « usage based » model of language acquisition that differs from the current theoretical trends towards generativist and connectionist approaches by three principal points. First, the model is functionalist, based explicitly on the expression and comprehension of communicative intentions and real world meaning. Second the model is construction based, focusing on utterances and their meanings rather than isolated words and morphemes as the fundamental units of language acquisition. Finally the model is usage based, such that as concrete utterances are stored and used, pattern-finding processes are at work extracting increasingly abstract grammatical structure to allow generalization to a full fledged generative capability.

The model that we described in the previous section, joined with a perceptual scene analysis capability described below addresses important aspects of Tomasello's three points, and provides an interesting point of departure for implementing the usage based model. From the functionalist perspective, the model is centered around the grounding of grammatical structure in meaning, and in extracting meaning from the perceptual environment as described below. From the construction perspective, this grounding of grammatical structure in meaning is explicitly achieved via grammatical constructions that define the mapping from open class elements to scene referents. Finally, from the usage-based perspective, through the accumulation of experience with sentences with common grammatical structures, pattern-finding operators extract these grammatical structures that are stored and retrieved based on the construction specific pattern of open and closed class words *à la* Bates and MacWhinney. The rest of this section outlines the functional characteristics of this system.

Extraction of meaning

As noted by Feldman and Miikkulainen, the complexity of meaning places serious limitations on the associated complexity of grammatical representations. Indeed we will consider that to a certain extent, it is the structure of meaning that will drive the construction of grammatical representations. This approach in a sense displaces the burden of complexity from syntax onto semantics. In the current section we outline how meaning can be extracted from visual scenes based on well specified perceptual primitives.

Our goal in this endeavor is to demonstrate that these capabilities, including perceptual scene analysis, can be implemented with a minimum of built in capability. In this context, we look to the human developmental literature for clues on how children develop their ability to extract meaning. As noted, in the domain of perceptual primitives, physical contact is highly salient quite early in development (e.g. Kotovsky & Baillargeon 1998). Analyzing a set of physical events including *touch*, *push*, *take*, and *give*, it became clear that all of these events could be described in terms of contact. An example of how the event *give* is decomposed into contacts is illustrated in Figure 4B.

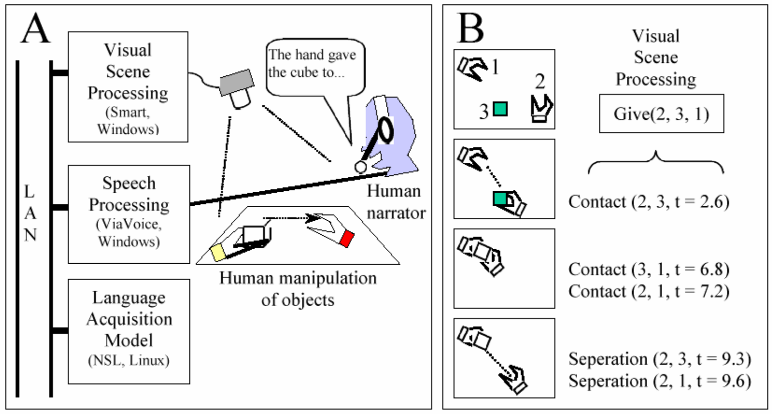

Figure 4. A. « Usage-Based » language acquisition architecture. *B*. Extraction of event description for give based on a pattern of contacts between objects.

We thus developed a system in which visual scenes are analyzed by a commuter vision system that recognizes and tracks objects based on their color (Smart - PanLab), and detects contact between the recognized objects. Based on the temporal sequences of contacts, the system can discriminate between different categories of events including *touch, push, take* and *give* (Dominey 2003). Figure 4B illustrates how the lexical semantics of the verb *give* can be defined in terms of the primitive « contact ». For a given video sequence the visual scene analysis generates the corresponding event description in the format *event(agent, object, recipient)*.

Single event labeling

Events are defined in terms of contacts between elements. A contact is defined in terms of the time at which it occurred, the agent, object, and duration of the contact. The agent is determined as the element that had a larger relative velocity towards the other element involved in the contact. Based on these parameters of contact, scene events are recognized as follows :

Touch(agent, object) : A single contact, in which (a) the duration of the contact is inferior to *touch_duration* (1.5 seconds), and (b) the *object* is not displaced during the duration of the contact.

Push(agent, object) : Similar to touch, with a greater contact duration, superior or equal to *touch_duration* and inferior to *take_duration* (5 sec), and object displacement.

Take(agent, object) : A single contact in which (a) the duration of contact is superior or equal to *take_duration*, (b) the object is displaced during the contact, and (c) the agent and object remain in contact.

Take(agent, object, source): Multiple contacts, as the agent takes the object from the source. Same as Take(agent, object), and for the optional second contact between agent and source (a) the duration of the contact is inferior to *take_duration*, and (b) the agent and source do not remain in contact. Finally, contact between the object and source is broken during the event.

Give(agent, object, recipient) : Multiple contacts as agent takes object, then initiates contact between object and recipient.

These event labeling templates form the basis for a template matching algorithm that labels events based on the contact list, similar to the spanning interval and event logic of Siskind (2001).

Complex « Hierarchical » Events : The events described above are simple in the sense that there have no hierarchical structure. This imposes serious limitations on the syntactic complexity of the corresponding sentences (Feldman et al. 1996, Miikkulainen 1996). The sentence *The block that pushed the moon was touched by the triangle* illustrates a complex event that exemplifies this issue. The corresponding compound event will be recognized and represented as a pair of temporally successive simple event descriptions, in this case : *push(block, moon)*, and *touch(triangle, block)*. The « block » serves as the link that connects these two simple events in order to form a complex hierarchical event.

We tested the system using ~300 sentence, scene pairs that were obtained by a human experimenter who manipulated blocks in the visual field of our scene analysis system, and simultaneously narrated his actions in the context of the setup presented in Figure 4B. This corpus of training data consisted of 10 different types of grammatical constructions (identified below), and their corresponding meanings.

Hirsh-Pasek & Golinkoff (1996) indicate that children use knowledge of word meaning to acquire a fixed SVO template around 18 months, then expand this to non-canonical sentence forms around 24+ months. Tomasello (1999) indicates that fixed grammatical constructions will be used initially, and that these will then provide the basis for the development of more generalized constructions (Goldberg 1995). The following experiments attempt to follow this type of developmental progression. Training results in changes in the associative WordToReferent mappings encoding the lexicon, and changes in the ConstructionInventory encoding the form to meaning mappings, indexed by the ConstructionIndex.

A. Learning of active forms for simple events

(1) Active : The block pushed the triangle.

(2) Dative : The block gave the triangle to the moon.

For this experiment, 17 scene/sentence pairs were generated that employed the 5 different events, and narrations in the active voice, corresponding to the grammatical forms 1 and 2. The model was trained for 32 passes through the 17 scene/sentence pairs for a total of 544 scene/sentence pairs. During the first 200 scene/sentence pair trials, no syntactic bootstrapping was used. This was necessary in order to avoid the random effect of syntactic knowledge on semantic learning in the initial learning stages. The trained system displayed error free performance for all 17 sentences, and generalization to new sentences that had not previously been tested.

B. Passive forms

This experiment examined learning active and passive grammatical forms, employing grammatical forms 1-4. Word meanings were used from Experiment A, so only the structural FormToMeaning mappings were learned.

(3) Passive : The triangle was pushed by the block.

(4) Dative Passive : The moon was given to the triangle by the block.

Seventeen new scene/sentence pairs were generated with active and passive grammatical forms for the narration. Within 3 training passes through the 17 sentences (51 scene/sentence pairs), error free performance was achieved, with confirmation of error free generalization to new untrained sentences of these types. The rapid learning indicates the importance of lexicon in establishing the form to meaning mapping for the grammatical constructions.

C. Relative forms for complex events

Here we consider complex scenes narrated by relative clause sentences. Eleven complex scene/sentence pairs were generated with narration corresponding to the grammatical forms indicated in 5 – 10 :

(5) The block that pushed the triangle touched the moon.

(6) The block pushed the triangle that touched the moon.

(7) The block that pushed the triangle was touched by the moon.

(8) The block pushed the triangle that was touched the moon.

(9) The block that was pushed by the triangle touched the moon.

(10) The block was pushed by the triangle that touched the moon.

After presentation of 88 scene/sentence pairs, the model performed without error for these 6 grammatical forms, and displayed error-free generalization to new sentences that had not been used during the training for all six grammatical forms.

D. Combined test with and without lexicon

A total of 27 scene/sentence pairs, used in Experiments B and C, were employed that exercised the ensemble of grammatical forms 1 – 10 using the learned WordToReferent mappings. After exposure to 162 scene/sentence pairs the model performed and generalized without error. When this combined test was performed without the pre-learned lexical mappings in WordToReferent, the system failed to converge, illustrating the advantage of following the developmental progression from lexicon to simple to complex grammatical structure.

E. Some scaling issues

A small lexicon (n<25) and construction inventory (n=10) are used, as the objective was to demonstrate the integrated system and grammatical structure learning capability. Based on the independent word and construction representations, and their synergistic interaction, the architecture scales well. The model is being tested with a larger lexicon, and has learned over 40 grammatical constructions. Importantly, the system should extend to all languages in which sentence to meaning mapping is encoded by word order and/or grammatical marking (Bates et al. 1982). In the current study, deliberate human event production yielded essentially perfect recognition, though the learning model is relatively robust to elevated scene error rates as documented in Dominey (2000).

Grammatical constructions in a robotic system : conclusion

The goal of the current study was to identify minimal event recognition and form-to-meaning mapping capabilities that could be integrated into a coherent system that performs at the level of a human infant in the first years of development when the construction inventory is being built up, and to do so in a manner consistent with the usage-based model of language acquisition proposed by Tomasello (2003). The current study demonstrates (1) that the perceptual primitive of contact (available to infants at 5 months), can be used to perform event description in a manner that is similar to but significantly simpler than Siskind (2001), (2) that a novel implementation of principles from construction grammar can be used to map sentence form to these meanings together in an integrated system, (3) that relative clauses can be processed in a manner that is similar to, but requires less specific machinery (e.g. no stack) than that in Miikkulainen (1996), and finally (4) that the resulting system displays robust acquisition behavior that reproduces certain observations from developmental studies with very modest “innate” language specificity.

6. General discussion

Independent of the theoretical stance taken, language acquisition takes place in a perceptual and behavioral context that cannot be ignored. The objective of this review has been to provide examples from our work that indicate how relatively simple learning mechanisms can exploit regularities in the acoustic signal and in the perceptual world to achieve certain significant language acquisition functions. This includes processing of the acoustic signal to perform initial lexical categorization, and then use this knowledge in learning the mapping from grammatical structures onto meaning. Likewise we demonstrate that meaning can be extracted from the perceptual scene by simple perceptually based mechanisms. The resulting studies indicate that interesting (though clearly not all) aspects of lexical and phrasal semantics can be accounted for in this context.

In this context, several important universals of language and semantic structure have emerged. Bates and MacWhinney (et al. 1982) indicated some time ago that in all languages, phrasal semantics is encoded by combinations of a small set of parameters, particularly word order, grammatical marking and prosody. This is an extremely potent universal in that it captures an immense typological variability (across languages) with a highly compact set of variables. In this sense it is of immense interest from a modeling perspective. In developing the functional mapping from this hypothesis onto a system for language acquisition, our neural network models that are sensitive to serial order, temporal (prosodic) structure, and abstract mapping structure provide the appropriate vehicle. Indeed, we initially demonstrated that the combined temporal abstract recurrent network ATRN behaved like human infants in tests of sensitivity to serial, temporal and abstract structure of sound sequences and language (Dominey and Ramus 2000).

In the ATRN framework we subsequently characterized thematic role assignment (a principal aspect of phrasal semantics) as the mapping of open class elements onto their semantic referents, guided by the configuration of function words that uniquely characterize each sentence type. We demonstrated that the required lexical categorization (function vs. content) could be achieved via the temporal recurrent network (TRN) and its sensitivity to the prosodic features that distinguish these categories (Blanc, Dodane and Dominey 2003). The structure mapping was initially simulated in the combined temporal abstract recurrent network ATRN (Dominey et al. 2003). While this model allowed the generation and testing of a number of predictions concerning normal and aphasic behavior, and the underlying neurophysiology, its architecture posed limits on the complexity of meaning that we could investigate.

We thus developed a sentence to meaning mapping model (Dominey 2000, 2003) that allows a much richer representation of meaning, including the representation of complex hierarchical structures corresponding to the meaning of relative phrases. The rich structure of meaning provided a solid foundation onto which grammatical structure could be mapped. Jackendoff (2002) has indicated indeed that semantics should be considered as an independently generative system that is linked to but does not require syntax. I have taken this argument a step further suggesting that the generative structure of semantics was the driving force behind the development of generative syntax (Dominey in press). This in a sense displaces the burden of explanation from syntax to semantics. The adult conceptual system is clearly rich and generative, and as suggested by Mandler (1996) infants can construct meaning representations from perceptual primitives and thus develop progressively complex conceptual representations. Dominey (2003) demonstrates that with the simple perception of physical contact, complex hierarchical event representations can be constructed.

The generative program analyses syntax as a recursive formal system, independent of the parallel conceptual structure (Chomsky 1995), and based on arguments of the poverty of the stimulus places immense stock in an innate universal grammar. Connectionist systems posit radically less innate structure (Elman 1990) but do not clearly bridge the gap between syntactic structure and meaning. Tomasello (2003) has proposed the usage-based model of language acquisition as an alternative. This functionalist model is construction (e.g. Goldberg 1995) oriented, and relies on the usage based extraction of successively abstract structural patterns. In this spirit, and from the simulation perspective, recent efforts seek to posit relatively simple mechanisms for extracting meaning, and for building the structure of language around this meaning (Chang and Maia 2001, Dominey 2003, Feldman et al. 1996, Steels 2001).

Clearly, there is an important interaction between innate mechanisms, and structure in the environment via learning. The open issue today concerns the allocation of function to these resources. The current research joins in the argument that there is enormous structure in the environment, including the communicative behavior of adult and infant. This reduces requirements on a genetic universal grammar, reallocating these requirements to more generalized learning mechanisms for the extraction and mapping of grammatical and conceptual structure onto one another.

Acknowledgements : Supported by the Human Frontiers Science Program, the ACI *Action and Causality* Project (Ann Reboul PI), and the Eurocores *OMLL* Project.

Bibliography

BATES E., DALE P.S. & THAL D. (1995), « Individual differences and their implications for theories of language development », In FLETCHER P. & MACWHINNEY B. (eds), *Handbook of Child Language*, Oxford, Basil Blackwell, 96-151.

BATES E., MCNEW S., MACWHINNEY B., DEVESCOVI A. & SMITH S. (1982), « Functional constraints on sentence processing : A cross linguistic study », *Cognition* 11, 245-299.

BLANC J.M. & DOMINEY P.F. (2003), « Discrimination of Prosodic Attitudes by a Temporal Recurrent Network », *Cognitive Brain Research* 17, 693-699.

BLANC J.M., DODANE C. & DOMINEY P.F. (2003), « Temporal processing for syntax acquisition : A Simulation Study », *Proceedings of the 25th Ann Conf. Cog. Sci. Soc.*, MIT, Cambridge (Mass.).

BOWERMAN M. & LEVINSON S.C. (2001), *Language Acquisition and Conceptual Development*, Cambridge, Cambridge University Press.

BROWN C.M., HAGOORT P. & TER KEURS M. (1999), « Electrophysiological signatures of visual lexical processing : Open- and closed-class words », *Journal of Cognitive Neuroscience*, 11 : 3, 261-281

CAPLAN D., BAKER C. & DEHAUT F. (1985), « Syntactic determinants of sentence comprehension in aphasia », *Cognition* 21, 117-175.

CAREY S. & XY F. (2001), « Infant's knowledge of objects : beyond object files and object tracking », *Cognition* 80, 179-213.

CHANG N.C. & MAIA T.V. (2001), « Grounded learning of grammatical constructions », *AAAI Spring Symposium On Learning Grounded Representations*, Stanford.

CHOMSKY N. (1995), *The Minimalist Program*, Cambridge (Mass.), The MIT Press.

CRAIN S. & LILLO-MARTIN D. (1999), *An Introduction to Linguistic Theory and Language Acquisition*, Malden (Mass.), Blackwell.

DOMINEY P.F. & RAMUS F. (2000), « Neural network processing of natural language : I. Sensitivity to serial, temporal and abstract structure of language in the infant », *Language and Cognitive Processes* 15 : 1, 87-127.

DOMINEY P.F. (2000), « Conceptual grounding in simulation studies of language acquisition », *Evolution of Communication* 4 : 1, 57-85.

DOMINEY P.F. (2002), « A model of learning syntactic comprehension for natural and artificial grammars », in WITRUK, FRIEDERICI & LACHMANN (eds), *Basic Functions of Language, Reading and Reading Disability*, Dordrecht, Kluwer, 61-77.

DOMINEY P.F. (2003), « Miniature Language Acquisition from Narrated Video Events », *Proceedings of the 25th Ann Conf. Cog. Sci. Soc.*, MIT, Cambridge (Mass.).

DOMINEY P.F. (2004), « A conceptuocentric shift in the characterization of language », Review of Jackendoff : *Foundations of Language*, *Behavioral Brain Sciences*, in press.

DOMINEY P.F., HOEN M., LELEKOV T. & BLANC J.M. (2003), « Neurological basis of language in sequential cognition : Evidence from simulation, aphasia and ERP studies », *Brain and Language,* 86, 207-225.

ELMAN J. (1990), « Finding structure in time », *Cognitive Science* 14, 179-211.

FELDMAN J., LAKOFF G., BAILEY D., NARAYANAN S., REGIER T. & STOLCKE A. (1996), « L0: The first five years », *Artificial Intelligence Review* 10, 103-129.

FELDMAN J.A., LAKOFF G., STOLCKE A,. WEBER S.H. (1990), « Miniature language acquisition : A touchstone for cognitive science », *Proceedings of the 12th Annual Conference of the Cognitive Science Society,* MIT, Cambridge (Mass.), 686-693.

FRIEDERICI A.D. (1985), « Levels of processing and vocabulary types : evidence from on-line comprehension in normals and agrammatics », *Cognition* 19, 133-166.

GILLETTE J., GLEITMAN H., GLEITMAN L. & LEDERER A. (1999), « Human simulation of vocabulary learning », *Cognition* 73, 135-151.

GOLDBERG A. (1995), *Constructions*, Chicago, The University of Chicago Press.

HIRSH-PASEK K. & GOLINKOF R.M. (1996), *The Origins of Grammar : Evidence from Early Language Comprehension*, Cambridge (Mass.), The MIT Press.

JACKENDOFF R.J. (2002*), Foundations of Language*, Oxford, Oxford University Press.

JUSCZYK P.W. (1997), *The Discovery of Spoken Language*, Cambridge (Mass.), The MIT Press MIT Press.

KOTOVSKY L. & BAILLARGEON R. (1998), « The development of calibration-based reasoning about collision events in young infants », *Cognition* 67, 311-351.

LANGACKER R. (1991), *Foundations of Cognitive Grammar. Practical Applications, Volume 2*, Stanford, Stanford University Press.

MANDLER J. (1999), « Preverbal representations and language », in BLOOM P., PETERSON M.A., NADEL L. & GARRETT M.F. (eds), *Language and Space*, Cambridge (Mass.), The MIT Press, 365-384.

MARCUS G.F., VIJAYAN S., BANDI RAO S. & VISHTON P.M. (1999), « Rule learning by seven-month-old infants », *Science* 283, 77-80.

MIIKKULAINEN R. (1996), « Subsymbolic case-role analysis of sentences with embedded clauses », *Cognitive Science* 20, 47-73.

MORGAN J.L. & DEMUTH K. (1996), *Signal to syntax : boostrapping from speech to grammar in early acquisition*, Mahwah (NJ), Lawrence Erlbaum.

MORGAN J.L., SHI R. & ALLOPENNA P. (1996), « Perceptual bases of rudimentary grammatical categories », in MORGAN J.L. & DEMUTH K. (eds), *Signal to Syntax*, Mahwah (NJ), Lawrence Erlbaum, 263-286.

MORLEC Y., BAILLY G.L. & AUBERGE V. (2001), « Generating prosodic attitudes in French : Data, model, and evaluation », *Speech Communication* 33, 357-371.

NAZZI T., BERTONCINI J. & MEHLER J. (1998), « Language discrimination by newborns : Towards an understanding of the role of rhythm »*, Journal of Experimental Psychology, Human Perception & Performance* 24 : 3, 1-11.

OSTERHOUT L. (1997), « On the brain response to syntactic anomalies : Manipulation of word position and word class reveal individual differences », *Brain and Language* 59, 494-522.

PINKER S. (1987), « The bootstrapping problem in language acquisition », in MACWHINNEY B. (ed.), *Mechanisms of Language Acquisition*, Hillsdale (NJ), Lawrence Erlbaum.

PULVERMÜLLER F. (1995), « Agrammatism : behavioral description and neurobiological explanation », *Journal of Cognitive Neuroscience* 7 : 2, 165-181.

ROY D. & PENTLAND A. (2002), « Learning Words from Sights and Sounds : A Computational Model », *Cognitive Science* 26 : 1, 113-146.

SAFFRAN J.R., ASLIN R.N. & NEWPORT E.L. (1996), « Statistical learning by 8-month-old infants », *Science* 274, 1926-1928.

SHI R., WERKER J.F. & MORGAN J.L. (1999), « Newborn infants' sensitivity to perceptual cues to lexical and grammatical words », *Cognition* 72 : 2, B11-B21.

SISKIND J.M. (1996), « A computational study of cross-situational techniques for learning word-to-meaning mappings », *Cognition* 61, 39-91.

SISKIND J.M. (2001), « Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic », *Journal of AI Research* 15, 31-90.

SNEDEKER J. & Li P. (2000), « The limits of observation : Can the situations in which words occur account for cross-linguistic variation in vocabulary composition ? », *7th International Symposium on Chinese Languages and Linguistics*.

STEELS L. (2001), « Language Games for Autonomous Robots », *IEEE Intelligent Systems* 16 : 5, New York, IEEE Press, 16-22.

STOLCKE A. & OMOHUNDRO S.M. (1994), « Inducing probablistic grammars by Bayseian model merging », in *Grammatical Inference and Applications : Proceedings of the 2nd International Colloquim on Grammatical Inference*, Springer Verlag.

TALLAL P., MERZENICH M., MILLER S. & JENKINS W. (1998), « Language learning impairments : Integrating basic science, technology and remediation », *Experimental Brain Research* 123, 210-219.

TALMY L. (1988), « Force dynamics in language and cognition », *Cognitive Science* 10 : 2, 117-149.

TOMASELLO M. (1999), « The item-based nature of children's early syntactic development », *Trends in Cognitive Science* 4 : 4, 156-163.

TOMASELLO M. (2003), *Constructing a Language : A Usage-Based Theory of Language Acquisition*, Cambridge (Mass.), Harvard University Press.

WOODWARD A.L. (1998), « Infants selectively encode the goal of an actor's reach », *Cognition* 69, 1-34.

WRIGHT B.A., LOMBARDINO L.J., KING W..M, PURANIK C.S., LEONARD C.M. & MERZENICH M.M. (1997), « Deficits in auditory temporal and spectral resolution in language-impaired children » *Nature* 387,176-178.