

## Cognitive and Neural Mechanisms Sustaining Rule Learning from Speech

Ruth de Diego-Balaguer<sup>1,2,3,4</sup> and Diana Lopez-Barroso<sup>2,3</sup>

<sup>1</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA). Barcelona, Spain

<sup>2</sup> Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona

<sup>3</sup> IDIBELL, L'Hospitalet de Llobregat, Spain

<sup>4</sup> INSERM U955, IM3/ Paris 12, France and Département d'Etudes Cognitives (DEC), Ecole Normale Supérieure, Paris, France

## Abstract

Learners of a new language are faced with the task of extracting from speech the words as well as the rules structuring them. Learners are endowed with the capacity to extract statistical regularities from their environment. Applied to language this capacity allows them to extract words from continuous speech in the absence of other cues. However, it has been proposed that natural languages have an intrinsic cue: the presence of pauses inducing prosodic information. This cue seems to trigger the application of different computational resources that allows the extraction of rules. This review summarizes work indicating the different aspects of executive functions critical in the early stages of language acquisition, when no semantic information is available. Different experiments recording event-related potentials while participants were learning an artificial language with non-adjacent embedded morphological rules indicate a functional dissociation between the brain responses associated with word and rule learning. The results indicate that the presence of salient cues such as prosodic information helps to control attention to bias perception. This bias ignores irrelevant information and attends to the relevant segments containing the rule, shifting from word acquisition to rule extraction. Finally, data from individual differences in connectivity between brain structures related to phonological working memory in addition to data from brain-lesioned patients point to the critical role of the basal ganglia as a coordinator structure between language, working memory and attention through its rich connections with the brain areas responsible for these functions.

## Cognitive and Neural Mechanisms Sustaining Rule Learning from Speech

Decoding language is such a natural process in our native language that it is hard to believe that several complex steps are necessary to reach the final stage of understanding. When listening to a very different and unknown language such as Hindi or Japanese, English speakers, for example, could approximate the actual complexity that a newborn may experience when initially exposed to language. Aside from the acquisition of the phonemes of their specific language, the initially apparent never-ending nonsense stream needs to be parsed into words. Then, these words need to be memorized and associated with meanings. Additionally, rules that structure these words have to be extracted and generalized.

Our understanding of the cognitive and neural mechanisms that permit this achievement is limited. However, we do know that young babies, like other animals, are endowed with mechanisms that can be applied to their surrounding stimuli and, as such, to the language they hear around them. One of these mechanisms is the capacity to capture the statistical structure of the environment, in turn enabling the co-occurrence of events to be learned. The ability to learn that events co-occur has a high adaptive value. Animals can use this ability to run away when a warning signal predicts the appearance of a predator. Humans can also use statistical learning in different domains, such as to segment streams of tones (Abla and Okanoya, 2008; Saffran et al., 1999) or visual patterns (Abla and Okanoya, 2009; Fiser and Aslin, 2001). As for these sequential events, learning their relations is necessary for the acquisition of various aspects of language. Both adults and infants are capable of tracking sequential contingencies between phonemes for the acquisition of phonotactics (Chambers et al., 2003; Onishi et al., 2002), between adjacent syllables for word segmentation (Aslin et al., 1998; Saffran et al., 1996) and between non-adjacent elements for morphosyntactic rules (Gomez, 2002; Peña et al., 2002). Can such an adaptive associative capacity be sufficient to attain all the steps of language acquisition previously listed? Should this mechanism be shared for all of them, are other additional mechanisms necessary for the acquisition of more abstract information, such as grammatical categories or syntactic dependencies?

We do know that after full acquisition, words and rules of language (lexical word forms and their grammatical information) can be functionally dissociated. A number of functional dissociations have been reported in different modalities (perception and production), different languages (Spanish, Swedish, English, German or French, among others) and populations (healthy individuals and patients with lesions in different localizations). Patients display selective difficulties in the retrieval of lexical or grammatical features (De

Diego-Balaguer et al., 2004; Marslen-Wilson and Tyler, 1997; Teichmann et al., 2006). Similarly, dissociable brain networks are reported to be engaged for the retrieval of these two types of information in a variety of languages (Beretta et al., 2003; De Diego-Balaguer et al., 2006; Lehtonen et al., 2009; Tyler et al., 2004; Tyler et al., 2005; Ullman, 2006). Functional distinctions are also manifest in second language acquisition. Generally, second language speakers, even at high levels of proficiency and practice, may not process grammatical rules the way native speakers do, while their vocabulary –in terms of full forms- may be flawlessly acquired (De Diego-Balaguer et al., 2005; Weber-Fox and Neville, 1996; although see also Laufer, 1990; Lemhofer and Dijkstra, 2004 for levels of difficulty according to the similarity across languages for vocabulary acquisition).

Although these distinctions are evident at the final stage of acquisition, for the purpose of this review, it is of interest that this picture is reflected as well in the course of language development. In that sense, it is clear that words carry salient semantic information, whereas the phonological segments carrying rule dependencies do not necessarily have specific meanings on their own. This aspect, however, cannot fully account for the differences in acquisition. Although semantic information is certainly important, pre-linguistic infants have a limited knowledge of semantics. Even at 18 months of age, infants do not know many of the content words of their language, although they know a lot about the most frequent function words and they are deeply sensitive to the sound properties of language (Jusczyk, 1997), including an adequate analysis of phrasal prosody (Christophe et al., 2008). Indeed, even during the acquisition of a second language there is evidence that the absence of semantic information may help to attain native-like processing of some otherwise elusive grammatical rules such case marking in Japanese learned by native speakers of German (Mueller et al., 2008b). As we shall see, studies with artificial languages where no semantic information is available also show dissociations between word and rule acquisition. Hence, the differences observed at the final stage of acquisition may emerge from different cognitive demands in the course of learning before semantic relations with words are fully available. Word and rule extraction from speech may share but also require different processing demands and brain dynamics in the course of acquisition. The present review will focus on the differences in terms of brain functioning and cognitive resources that may underlie the memorization of words and the extraction of the embedded rules in the early stages of language acquisition, when no semantic information is available.

Adjacent Dependencies , Non-adjacent Dependencies and Memory Traces

As previously mentioned, in the early stages of acquisition of a language, babies and adults cannot rely solely on their limited semantic knowledge of the language (Christophe et al., 2008). The use of artificial languages for the study of language acquisition allows one to control for the prior knowledge of infants and adults (although see Finn and Hudson Kam, 2008). Moreover it allows for a better understanding of the weight of the different cues available in the complex stimulation that is present in speech (prosody, statistical information, correlated cues such as coarticulation, etc.). In particular, a simplified language stream made of randomly presented trisyllabic words (e.g. tupiro, bagoli, ...) where embedded dependencies can be introduced (tupiro\_bagoli\_tugoro)<sup>1</sup> has been widely used for the study of speech segmentation (Aslin et al., 1998; Cunillera et al., 2006; McNealy et al., 2006; Saffran et al., 1996) and, with some modifications, to study more abstract rule-learning (De Diego-Balaguer et al., 2007b; Endress and Bonatti, 2007; Peña et al., 2002; Toro et al., 2008). In our studies, we have employed this type of material combined with the comparison of groups of learners that are or are not able to learn the rules of the language, allowing for a perfect matching of the variables under study (rules and words) while avoiding the influence of semantic variables (Figure 1). Although highly simplified compared to real language, this type of material and similar versions have shown convergent results with studies using real language (Pelucchi et al., 2009; Santelmann and Jusczyk, 1998).

Before presenting the studies, let us consider the different types of dependencies that may characterize language learning and describe briefly why there are reasons to believe that not all dependencies in language can be extracted with mechanisms of the same nature.

### *Adjacent Dependencies and Segmentation*

Some of the dependencies present in language can be learned by simple computations also performed by other animals (Table 1). Previous studies using artificial languages have shown that babies as young as 8 months old can exploit statistical relations between adjacent syllables to segment words in a speech stream when no other prosodic or acoustic information about word boundaries are available (Saffran et al., 1996; Aslin et al., 1998). It is also known that animals with no a priori interest in language can also use statistical learning to segment the same linguistic material. This has been reported not only in monkeys with complex vocalizations in their communicative repertoire (cotton-top tamarins) (Hauser et al., 2001), but has also been shown in rats (Toro and Trobalon, 2005) that were passively exposed to the same

material with no contingent reinforcement. Although succeeding in the same task, it is relevant to note that while infants used transitional probabilities between syllables (Aslin et al., 1998) to perform the task, rodents were simply detecting the frequency of the co-occurrence of syllables (Toro and Trobalon, 2005). Transitional probability ( $Prob Y/X = Prob XY / Prob X$ ) is a conditional probability. It considers how often an event occurs associated to another whenever it appears. Thus, it requires keeping track of the different occurrences of the event. In contrast, the frequency of co-occurrence considers only the probability that two events occur together. Thus, this computation is independent of the frequency of each individual event and represents a less complex computational process.

### *Non-adjacent Dependencies and Abstract Rule Learning*

While fairly useful for the acquisition of adjacent dependencies, the frequency of co-occurrence does not seem to be a suitable strategy for the generalization of non-adjacent relations (AXC, where A predicts the appearance of C irrespective of the middle element). While humans are able to segment and discriminate the words presented in an artificial language stream built with nonce words with this structure (Peña et al., 2002), rats tested with this type of material are unable to segment words based on non-adjacent dependencies unless they are pre-segmented with subtle pauses (Toro and Trobalon, 2005) (Table 1). Interestingly, the insertion of these pauses helps humans to learn more abstract relations over the same material. When pauses between words are inserted in the speech stream, humans are not only able to discriminate the words, but are also able to generalize the rules acquired to new words following the same rule (Mueller et al., 2008a; Peña et al., 2002). In other words, when pauses are present, the extraction of these dependencies helps to create categories of syllables as a function of their position (Chemla et al., 2009; Endress and Bonatti, 2007; Mintz, 2003), such as noun and verb categories in real language, and rules can be transferred to completely new material (De Diego-Balaguer et al., 2008). The relevant issue here is that these remote dependencies are more complex dependencies (Newport and Aslin, 2004) than adjacent relations, and they characterize grammatical dependencies such as those observed in morphological relations within words (e.g., unbearable, unbelievable, etc.) and spanning different words (e.g., he is jumping, he is singing, etc.).

The distinction between local (**AB**) and distant relations (**A[AB]B**) has also been used to study their sustaining processing and brain pathways in relation to more abstract syntactic level (Friederici, 2004;

Bahlmann et al., 2006; Opitz and Friederici, 2007; Friederici, 2009). The processing of these distant relations has been claimed to underlie the capacity to understand embedded sentences such as “**The boy** [that the girl saw] **was tall**” characterizing human language. These studies use more complex artificial grammars, with visually presented material and, usually, participants are informed and trained on the rules of the language. Thus, this approach diverts from the one exposed here; however, at some points, converging results will be commented on throughout this review.

### *Constraints in the Extraction of Abstract Rules from Non-adjacent Dependencies*

It is crucial to highlight that, despite the greater complexity of non-adjacent relations, statistical learning appears to provide a good tool for segmentation in this context as well (Peña et al., 2002). To be able to abstract and generalize the rule embedded in the language, however, extra information needs to be provided in the speech stream. Indeed, aside from the presence of pauses, other cues in the speech signal have been reported to trigger the appropriate mechanism, thereby enabling a shift in the type of information extracted from these non-adjacent dependencies moving from instance-based memorization to the extraction of more abstract relations. For instance, increasing the variability of irrelevant information (with high variations of the pool of X elements in the AXC dependency) helps the learners to focus on the sources of invariant structure (Gomez, 2002; Gomez and Maye, 2005). In addition, there is evidence showing that pauses are only effective in inducing rule learning if they enhance the salience of syllables carrying the rule, such as presenting these syllables immediately adjacent to the pauses, at the edges of the utterance (Endress and Mehler, 2009a). Moreover, only when the non-adjacent dependency is carried by vowels can the rule be generalized (Toro et al., 2008). When these dependencies are carried by consonants, the specific words are instead segmented and recognized more easily (Bonatti et al., 2005). As summarized in Table 1, constraints are also present in maturational terms, with rule-learning abilities showing a developmental progression. Studies in infants at different ages have shown that there is an early tendency to track adjacent relations that 8-month-old infants (Saffran et al., 1996) can effectively use to segment a speech stream. At 12 months, infants can remember the specific word orders learned and can associate the elements in the dependency to the cues that distinguish them. They are, however, still unable to categorize for generalization (Gomez and Maye, 2005). At 15 months, they start being able to track non-

adjacent dependencies and create categories from these relations (Gomez and Maye, 2005; Gomez and Lakusta, 2004).

In that context, it is relevant to assert that a number of studies have focused on the acquisition of repetition-based structures (Table 2) for the study of abstract rule-learning (ABA structures, where the exact stimuli is repeated at the initial and final position, but also AAB and ABB forms). These rules can also be generalized to new material and even be transferred across modalities (Marcus et al., 2007), indicating that their abstract nature is not confined to the specific material used. Nevertheless, we will not focus on these types of rules in this review because the extraction of repetition-based rules seems to rely on different perceptual primitives than the non-adjacent dependencies previously mentioned (Endress et al., 2009a). That is, the notion of repetition and the ability to classify as a function of this dimension is present in a number of animals other than humans (Murphy et al., 2008; Wallis et al., 2001). In addition, these repetition-based rules can be learned much earlier than AXC relations with no repeated dependencies, as can be observed comparing Table 1 and Table 2. Newborns are able to discriminate between adjacent repetition-based and random sequences (Gervain et al., 2008), and babies as early as 7 months of age (Marcus et al., 1999)(Table 2) are able to acquire (and generalize) both non-adjacent and adjacent repetition-based rules. However, it is not until 15-months of age when they can learn non repetition-based rules on non-adjacent elements (Gomez and Maye, 2005; Gomez and Lakusta, 2004).

Nonetheless, the existence of these constraints for the acquisition of abstract rules compared to segmentation raises a clear question regarding why these constraints are necessary and what their effect is on the cognitive system such that it induces a change in the type of information extracted. Considering the overall picture, therefore, the underlying cognitive mechanisms sustaining rule extraction in the early stages of acquisition are still unclear. One of the recent approaches used to gain insight into these issues has been to attempt to understand what the participants actually learned after exposure to these artificial languages. This was done by varying the items presented in the test phase and the length of exposure to the language (Endress and Mehler, 2009b; Endress and Bonatti, 2007). We have used a complementary approach benefiting from the use of neurophysiological measures, as they allow us (i) to avoid the strategic effects that can arise as a function of the test used because brain responses are recorded during the 'passive' listening to the language; (ii) to follow the online electrophysiological variations as a function of the amount of exposure to the language; and (iii) to relate those responses to the final performance in word and rule learning. We have complemented those results by studying how altering the system affects word and rule

learning either by (iv) interfering with the cognitive components that may sustain learning, or (v) by studying patients with specific lesions in the striatum, a structure of the basal ganglia hypothesized to play an essential role in the rule-learning process.

## Cognitive Differences in Word Learning and Rule Extraction

### *Functional Differences Revealed by Event-Related Potentials*

Adopting this approach, we recorded event-related potentials (ERPs) associated with the presentation of each word in an artificial speech stream, as described in Section 2 (see Figure 1). The ERPs were recorded online throughout a short learning phase (4 minutes) in a group of adult volunteers (De Diego-Balaguer et al., 2007b). Thus, we could observe how their electrophysiological responses varied progressively with increasing exposure and how these related to the final level of acquisition of words and rules (Figure 1). As we wanted to observe exactly what happened under suitable conditions for rule learning, we used languages with the same structure as in Peña et al. (2002), with subtle pauses between words (25 ms).

As exposure progressed, the ERP responses varied, and we observed dissociable brain signatures for word and rule learning. As already reported in speech segmentation tasks with nonsense words (Cunillera et al., 2006; Cunillera et al., 2009; Sanders et al., 2002) and in second language acquisition (McLaughlin et al., 2004), a modulation in the N400 component, a negative going deflection around 400 ms after word onset, with a central scalp distribution, was evident from the second minute of exposure. This modulation correlated only with the performance in the word learning test. In a recent study by Cunillera and collaborators (2009), in which they studied segmentation with adjacent dependencies, the same N400 modulation was present as well from the second minute of exposure. The time-course of the N400 as a function of exposure and its association to word learning was consistent with their results showing significant learning and no variation in the accuracy at tests after one minute of exposure and later. This is despite the fact that segmentation in our study could be performed by simple detection of pauses with no need to apply statistical computations as in the Cunillera et al study (2009). In both the languages with adjacent (Cunillera et al., 2009) and non-adjacent dependencies, word learning achieves an asymptotic

performance after a couple of minutes (De Diego-Balaguer et al., 2009; Endress and Mehler, 2009b Peña et al., 2002).

The second result of the study was more novel and surprising. A second signature appearing progressively in the following minutes of exposure showed a different behavior than the N400. It involved a positive shift around 200 ms (P2 component), and the scalp distribution of the effect was right lateralized in frontal electrodes. Importantly, its amplitude correlated specifically with rule learning performance (Figure 2A). The relationship between this component and rule-learning abilities was confirmed when good and poor learners of the rule were compared (Figure 2B). Although the group of learners was studied as a whole in the previously mentioned studies, this type of artificial language studies shows great inter-individual variability. We took advantage of this fact and compared those participants who were at chance after four minutes of exposure to those with the greatest performance in the rule learning test. Even though the two groups were equal in their word learning capacity, the lexical modulation in the ERPs (N400) appeared earlier in the good learner group (see Figure 2B First minute) and, in this group only, the rule learning component (P2) showed a progressive increase in amplitude throughout the experiment.

Results of the overall group showed that the functional dissociations observed in learning were also present in the ERP recorded in an implicit test and in the explicit classical two-alternative forced-choice task. After four minutes of exposure to the language, participants were not warned that a few violations to the rule (CXA) and new words following the rule that was just acquired ( $A_1A_2C_1$  or  $A_1C_2C_1$  rule words with a middle element that never occurred in that position) would be introduced in the stream while they continued to listen. Both types of test items contained syllables that never occurred in that order in the stream. However, when their ERP response was compared to the actual words present in the language learned, only words that violated the rule induced an early automatic response that appears when a deviant stimulus is detected (a mismatch negativity (MMN) effect). The MMN has also been reported in violations of regular sequences (Carral et al., 2005) and local morphosyntactic agreement violations (Hasting et al., 2007). This modulation was followed by a positive deflection around 600 ms after the onset of the test word, similar to that obtained in grammatical violations (Friederici and Kotz, 2003). In sharp contrast, both in the implicit and the explicit tests, rule words, consistent with the rule showed no ERP differences compared to words that were actually learned, indicating that, after learning, those new words were processed as words of the language.

## *The Role of Attention in the Integration of Multiple Cues*

What does the appearance of this specific P2 component modulation in relation to rule learning tell us about the underlying cognitive mechanism engaged? Functionally, the P2 component description is less clear than the N400 that has been extensively studied and related to lexical processing (Key et al., 2006). Based on the behavior of the P2 modulation in several contexts, however, a tentative explanation can be made that is consistent with the previously outlined constraints required for rule learning. If we reconsider these constraints, they all converge on the possibility that successful extraction of the underlying structural relations requires the presence of cues that could capture the learner's attention. These cues enhance the salience of the syllables carrying the critical rule information either by increasing the stability of these syllables compared to highly variable irrelevant information (Gomez, 2002; Gomez and Maye, 2005) and/or by their consistency with the dependencies (Endress et al., 2005; Peña et al., 2002). Similarly, in our studies, prosodic information (inserted pauses) could act as task-relevant salient information that helps to capture attention resources and focus on the relevant units for rule acquisition (Mueller et al., 2008a). Considering research in other fields, the P2 component has been shown to be modulated by perceptual learning and attention (Hillyard et al., 1973; Reinke et al., 2003).

In a recent proposal, Endress, Nespors and Mehler (2009) suggested that sensitivity to boundaries is a *perceptual primitive* pre-wired in our cognitive system to enable the extraction of grammatical structures. This sensitivity is extremely useful because in natural languages prosodic boundaries always correspond to syntactic boundaries (Christophe et al., 2008). In natural language also, multiple cues converge at the same time and can be exploited in early language acquisition. Because a single cue in isolation is often not fully reliable, the combination of multiple probabilistic cues could facilitate infants' and adults' word and grammatical learning (Christophe et al., 2008; Jusczyk, 1999). The learner, however, needs to be able to use this information in the appropriate way. Prosodic information is always present in natural language; however, rules are not readily extracted by infants from birth. More seems to be needed to abstract and generalize the rules. According to Gomez and Lakusta (2004), two steps seem to be necessary to reach this level of abstraction: first, association between A and C elements; second, categorization of the elements as a function of their association to a particular cue (i.e., number of syllables, repetition and stress pattern).

Hence, the passive capture of attention by salient cues appears to be insufficient for rule abstraction and generalization. Although in our previous study, all participants were exposed to the same materials with pauses, the P2 modulation appeared only in the good learner group. This result, along with that of our follow-up work, indicated that the P2 effect associated with rule learning was endogenously generated<sup>4</sup> (de Diego-Balaguer et al., 2007a). Thus, this effect did not simply reflect the external capture of attention by the pauses. It is more likely to correspond to the necessary shift of attention to the relevant cues to be integrated. In this study (de Diego-Balaguer et al., 2007a), we recorded ERP responses when participants were exposed to randomly presented syllables, thus preventing any possible word or rule learning. Streams could be either continuous or with pauses every three syllables, as in the previously described artificial language. Comparing the ERP recorded during the learning phase, the two random conditions allowed us to observe the effect of the exogenous capture of attention by the presence of pauses. In contrast to the previous results, this exogenous effect led to no variation whatsoever in the P2 component. Instead, as can be observed in Figure 2 (Left), the N1, an early negativity enhanced 100 ms after each syllable presentation, was evident in the continuous stream. This effect indicated an increased attention to each syllable onset when no pauses could be used to segment the speech stream. Both the N1 and the previously mentioned N400 ERP components have been proposed as speech segmentation indexes. The N1, however, appears to be sensitive to onset perception, whereas the N400, described in a previous study, appears to tap the identification of recently segmented words (Abla et al., 2008; Cunillera et al., 2009; Sanders et al., 2002). The equivalent comparison with language streams (non-random) was also accompanied here by the same N1 enhanced in the continuous artificial language. In this case, however, the N1 enhancement appeared only for the first syllable (Figure 3, Right). This result suggests that the perceptual onset was located at the initial syllable of the trisyllabic chunk perceived as a segmented word as learning progressed.

In contrast to the results obtained in the random condition, the artificial language with pauses did induce the expected P2 modulation (Figure 3, Right), replicating our previous study (De Diego-Balaguer et al., 2007b). This modulation was consistent with the proposal by Peña et al. (2002) and with our behavioral results showing that pauses enhanced word learning and enabled rule-extraction, whereas the absence of pauses in the continuous language prevented the extraction of rules even though words could still be segmented. Indeed, it has been claimed that pauses are necessary for rule extraction (Peña et al., 2002) because they trigger the mechanism for rule-extraction by rendering the artificial language more similar to

natural speech. Nevertheless, in our study some of the individuals participating in the study were still able to extract the rules even in the continuous condition, although no prosodic information was present. Significant learning was also obtained in a grammaticality judgement task in a similar study by Mueller et al. (2008a). We therefore compared the good learners of the rule with those participants that were not able to learn. Here again, even in the absence of pauses, the P2 enhancement was present for the good learner group. This result supports the assumption of an endogenous source of this modulation because clearly, in this case, no external prosodic cues were present that could attract the learners' attention. This result rather points to the relation of the P2 effect to the necessary endogenous shift of attention from the whole word to the relevant segments carrying the rule.

The entire pattern of results indicates that, in addition, this shift can be favored by several relevant convergent cues. In the stream with pauses, prosodic cues and the systematic relation between initial and final syllables compared to the variable middle syllable favored rule learning and enhanced word learning. This shift requires the learner to both selectively attend to the syllables carrying the structural information and ignore the variable middle syllable. Indeed, in agreement with this idea, previous learning studies have found a P2 enhancement when sounds were actively ignored (Ceponiene et al., 2005; Sheehan et al., 2005). Sheehan et al. (2005: p. 552) interpreted the amplitude of the P2 component in their study as "an index of inhibitory process that is strengthened when a repeated sound proves to have no relevance". The results of the P2 component modulation agree with the idea of Gomez and Lakusta (2004) regarding the critical role of attention to group together the elements of the input carrying the regularities while ignoring irrelevant information. In this way, the different cues can be viewed as pointers in the input that can be used by the learner to shift the focus of attention from the whole word level to the relevant information to be clustered for the extraction of the rule.

The role of attention in biasing language learning processes is in line with more general proposals (Echols and Newport, 1992; Gleitman and Wanner, 1982) and with the classical notion of the role of prosodic information as a crucial cue for word learning and grammatical acquisition (Gleitman and Wanner, 1982; Morgan et al., 1987). From their perspective, infants might be attentionally biased to extract salient elements from the language stream while ignoring others to simplify the word-learning endeavor. On similar grounds, N. C. Ellis (2008) proposed that, particularly for the acquisition of grammatical relations, attention is tuned to enhance the perception of the relevant information. Interestingly, according to this author, this tuning affects second language acquisition. While similar structural patterns may help learning through

transfer from the first to the second language, interference may arise when the new structural pattern requires a differential allocation of attention.

## Coordination of Different Brain Regions in Word and Rule Learning

### *Long-range Synchronisation between Distant Brain Regions in Early Language Acquisition*

The evidence presented thus far indicates that control of attention plays a critical role in the early stages of language acquisition. More precisely, rule learning seems to influence perceptual processing of the prosodic cues by shaping endogenous attention. This attentional influence on the processing of sensory information is reflected in the P2 modulation (Nan et al., 2006; Nan et al., 2009). Snyder et al. (2006) showed that the amplitude of the P2 auditory evoked-response was positively correlated with the perceptual segregation of a single continuous stream of tones in two separate streams. In that case, the effect could be driven by the physical characteristics of the stimulation, because the segregation pops out when a threshold in the gap between tones is reached. The enhancement of this component, however, is also observed when the salient stimuli are specifically relevant cues for the task (Luck and Hillyard, 1994), hence, when endogenous attention is more clearly engaged. Similarly, in language processing, the P2 modulation has also been observed when prosodic cues were relevant for segmentation in a pause detection task (Mattys et al., 2005) and in an artificial language task when the stress indicated initial word boundaries (Cunillera et al., 2006; Cunillera et al., 2008).

We obtained additional evidence for the involvement of endogenous attention and how this mechanism induces the coordination between distant areas of the brain by analyzing the oscillatory patterns appearing in the electrophysiological recordings (De Diego-Balaguer et al., 2007b). Large-scale coordination between frontoparietal and sensory cortices (e.g., auditory areas) enabling control of attention (Corbetta and Shulman, 2002) is reflected in the oscillatory synchronization between neuronal groups in distant cortical areas (Buzsaki and Draguhn, 2004; Engel et al., 2001; Fries, 2005; Siegel et al., 2008). A specially suited technique to discern endogenous and exogenous processing interactions through attention is, therefore, time-frequency analysis of electrophysiological recordings (Buschman and Miller, 2007; Engel et al., 2001). Oscillatory patterns have been shown to vary as a function of learning (Gruber et al., 2001; Miltner et al., 1999; Popescu et al., 2009) and provide complementary

information from that obtained from ERPs (Buiatti et al., 2009; Hagoort et al., 2004). Oscillatory patterns refer to the fact that neuronal assemblies can synchronize at different frequency bands. For the purposes of our research, it is interesting that these frequencies have been related to different cognitive processes, particularly to different aspects of executive control (Laufs et al., 2003; Makarov et al., 2005), including working memory and attention.

In our investigation (De Diego-Balaguer et al., 2009), the oscillatory variations during the initial exposure to the artificial language with embedded rules and inserted pauses clearly showed different patterns of oscillatory activity in the group of participants focused on word memorization and those who extracted the embedded rule (Figure 2C). Both groups showed progressive synchronization (phase coherence) between the activities in distant regions of the scalp. This synchronization, however, occurred at different frequency ranges in the two groups. Only the group that was able to generalize the rule showed increased synchronization between frontal, temporal and parietal regions in the gamma-band range (30-80 Hz)(Figure 2C, bottom). Interregional coherence in this frequency band does not seem to be dependent on sensory characteristics of the stimulus presentation (Siegel et al., 2008). This long-range synchronization reflects the endogenous control of attention in the good learners of the rule. In contrast, poor rule learners who focused on whole-word memorization displayed greater synchrony between frontal and parietal regions in the theta band (4-8 Hz)(Figure 2C, top). Theta synchrony between these regions has been reported in human studies involving periods of information retention and has been attributed to a common mechanism of neural interaction that sustains working memory functions (Sarnthein et al., 1998). In our study, the effect in the poor learner group could reflect the incremental neural efficiency of the memory-matching process between the segmented pieces held in working memory and the incoming stimulation, progressively enhancing their memory traces (Perruchet and Vinter, 2002). This pattern is also in agreement with the implication of working memory in first and second language word acquisition in children and adults (Baddeley, 2007).

#### *Coordination and Integration through the Basal Ganglia*

As we have seen, working memory and control of attention appear to play an important role in word and rule learning in the early stages of language acquisition. The basal ganglia structures occupy a privileged position to hold an integrative function between the different pathways related to

these functions. The basal ganglia refer to a set of grey matter subcortical structures localized deep in the brain that are connected to several areas throughout the brain in parallel loops holding different motor and cognitive functions (Alexander et al., 1986; Middleton and Strick, 2000). There are several reasons to believe that the basal ganglia and, particularly, the striatum (one of the primary components of the basal ganglia), may play a prominent role in the course of language acquisition.

There is evidence indicating that the striatum is implicated in attention (Couette et al., 2008; Georgiou-Karistianis et al., 2007; Karnath et al., 2005) as well as in storing and rehearsal in verbal working memory (Chang et al., 2007). The striatum acts as a funnel that receives inputs from different neocortical areas responsible for these cognitive functions and sends its outputs back to the cortex, thereby forming different functional loops (Middleton and Strick, 2000). However, the cortico-striatal parallel circuits, classically described as functionally segregated, have been shown more recently to have a great deal of interaction (Haber, 2003; Haber and Calzavara, 2009; Yin and Knowlton, 2006). There is anatomical overlap between the areas of the striatum connecting to regions of the prefrontal cortex related to executive functions and those related to language processing (Figure 5). This includes the premotor cortex, which is engaged in the rehearsal component of phonological working memory, and those connecting to Broca's area, which are associated with syntactic processing (Leh et al., 2007; Lehericy et al., 2004). This overlap in the projections may be particularly important for language learning (Rodriguez-Fornells et al., 2009), and could uphold the coordination between attention, working memory and language. This reinforces the idea that this structure may play a key role in integrating inputs from different networks.

Moreover, work outside the language domain, indicates that these subcortical structures play a clear role in sequence and category learning (Casey et al., 2006; Seger and Cincotta, 2006; Seger, 2006). Because language is sequential by nature, a preponderant subcortical implication has been proposed for different aspects of language processing (Lieberman, 2000). Indeed, studies from brain damaged patients and brain imaging indicate that the striatum may be involved in those aspects of rule application that require non-automatized behavior (Friederici and Kotz, 2003; Friederici et al., 2003; Longworth et al., 2005; Teichmann et al., 2005), such as in syntactic integration under ambiguity resolution or in the presence of syntactic violations (Friederici and Kotz, 2003; Friederici et al., 2003; Wahl et al., 2008). Because rules in languages are, by definition, not yet automatized in the earliest

phases of language acquisition, subcortical structures are likely to play a predominant role during the learning process.

In order to test the importance of striatal lesions in language acquisition, we performed a cross-sectional study with a group of Huntington's disease patients at different stages of the disease (De Diego-Balaguer et al., 2008). Huntington's disease (HD) is an inherited neurodegenerative disorder with primary neuronal dysfunction and death in the striatum (caudate and putamen) at early stages of the disease (Myers et al., 1988; Peschanski et al., 1995; Vonsattel et al., 1985). At later stages of the disease, cortical degeneration is also present (Douaud et al., 2006; Douaud et al., 2009). Patients and control participants matched in age and educational background to the symptomatic group were tested for word learning and rule extraction with the same classical two-alternative forced choice task (Figure 1) after four minutes of exposure to the artificial language with pauses and embedded rules. Patients at the earliest stages of Huntington's disease, when neural degeneration is mostly confined to the striatum, showed a poor rule learning ability only, whereas both word and rule learning abilities were compromised at later stages of Huntington's disease, when cortical degeneration may also have been present (Douaud et al., 2006; Rosas et al., 2003). We observed that language impairment progressed along with the disease, showing a correlation with striatal atrophy. A stronger effect of Huntington's disease progression and striatal degeneration was observed for rule-based acquisition than for word learning abilities when we matched the Huntington's disease patients and the healthy control participants on their word learning ability.

Interestingly, more subtle abnormalities were observed in pre-HD individuals who, despite their preclinical status, have been reported to have striatal dysfunction (Lawrence et al., 1998; Thieben et al., 2002). Although the pre-HD population showed a spared performance in the learning tests, the rule they acquired was less abstract than that of the control group. The healthy controls benefitted from the acquisition of the rule in the first artificial language. They displayed greater performance in the second artificial language they learned showing a capacity to transfer their acquired structural knowledge from one language to a completely new one with rules organized in the same manner (AXC dependencies). Huntington's disease patients, even at a preclinical stage, did not show this transfer capacity. They were able to learn and generalize the rule to new words but only in the specific language acquired each time from scratch. Transfer is an intrinsic characteristic of procedural knowledge, and data from outside the language domain has shown that the striatum appears to be a

key structure for these transfer abilities (Dahlin et al., 2008). Finally, it is worth noting that the results of the neuropsychological evaluation of the patients supported the role of working memory and attention during early language acquisition. In the whole group of patients, the scores from the rule learning test correlated with selective attention, working memory tests and with tests that required a sequencing ability in a classical neuropsychological evaluation. In contrast, the largest correlation for word learning arose with a test engaging episodic memory. These results emphasize the coordinating and integrative role of the basal ganglia necessary during early language acquisition and its importance, particularly, in rule-learning.

These results, pointing to the importance of subcortical pathways in language acquisition, converge with our findings from a study using diffusion tensor imaging (DTI) (Lopez-Barroso et al., submitted). This technique has emerged as a new neuroimaging tool that has been very useful in the fields of learning and development research. It is sensitive to microstructural changes in white matter based on the properties of water diffusion. The diffusion of water molecules is forced into certain *anisotropic* (i.e., preferential) directions (e.g., following axon membranes and myelin sheaths) (Le Bihan D., 2006). Because the main direction of water molecules is aligned to the direction of the dominant axonal fibres, the technique allows the extraction of information about both the orientation of the fibres connecting different areas and about how well these areas are connected. Indeed, recent studies have shown that individual differences in the success of different cognitive tasks are related to variations in white matter connectivity between specific brain areas (Boorman et al., 2007; Charlton et al., 2009; Fuentemilla et al., 2009; Gold et al., 2007), and that extensive learning induces changes in diffusivity measures (Bengtsson et al., 2005; Scholz et al., 2009). Thus, the use of the DTI method could provide information about the connections between those regions and their relation to individual differences in language acquisition.

In our study (Lopez-Barroso et al., submitted), interesting patterns of connectivity were found correlated with segmentation and rule learning (Figure 3). The left external capsule was related to individual differences in the performance of both language acquisition tasks under working memory constraints (Figure 4C). This region, between the claustrum and the insula, is crossed by fibres connecting the superior temporal region and the premotor cortex and nearby inferior frontal region (Charlton et al., 2009; Schmahmann et al., 2007) (Figure 5). It has been proposed as an alternative pathway to the dorsal longitudinal fascicle (Friederici, 2009; Hickok and Poeppel, 2004) in connecting

those temporal and frontal areas. A recent DTI study (Charlton et al., 2009) reports a correlation between the performance in two basic working memory tasks (i.e., the Digit Span Backward Task and a Word-Number Sequencing Task) and individual differences in –among others- the medial temporo-frontal pathway passing through the external capsule. The task used in this study was not focused on the phonological working memory role in language acquisition, but rather on pure individual differences in working memory tasks. However, it is noteworthy that the same cluster in the external capsule was highlighted in our study. It is of interest to note that the external capsule also represents a corticostriatal projection system that connects several regions of the cerebral cortex (e.g., premotor and supplementary motor areas) to the basal ganglia (Schmahmann and Pandya, 2006) (Figure 5). These areas have also been related to speech production and working memory rehearsal (Awh et al., 1995; Chang et al., 2007).

#### Developmental and Clinical Implications

Despite the use of artificial languages, allowing us to control for previous knowledge and semantic influence, all the studies reviewed here were performed on adults, making them more comparable to second than to first language acquisition. Studies in infants with similar paradigms are critical to understand the developmental progression of these abilities and the possible differences, in the role of cognitive functions described, for infants compared to adult language acquisition. In light of the available evidence, some similarities may exist between first and second language acquisition. The importance of being able to exploit the available information given by prosodic information is underscored in the study by Sabish et al. (2009), where deficits in the processing of prosodic information are identified in children with specific language impairment (SLI). This disorder is primarily characterized by difficulties in phonological and grammatical acquisition. On the other hand, fronto-striatal abnormalities have been documented in attention-deficit hyperactivity disorder (ADHD) (Doyle et al., 2005; Williams et al., 2000). Along with the attentional deficits in these children, this disorder often co-occurs with SLI and with speech and language disabilities in general (Williams et al., 2000).

Finally, it is worth pointing out that if control of attention was as necessary during first language acquisition as it seems to be during second language acquisition in adults, then the developmental progression observed in infants from adjacent segmentation to non-adjacent rule

learning could be rooted in the characteristics of the progression of cognitive development and brain maturation. On the one hand, attentional abilities, such as selectively attending to relevant stimuli while successfully ignoring irrelevant stimuli and the ability to shift attention quickly and effectively, improves progressively throughout childhood (Neville, 2006). In parallel, brain maturation displays impressive changes during the first and second year after birth. Myelination and brain maturation happens at different rates in different regions, thereby constraining the development of cognitive functions (Casey et al., 2000; Uylings, 2006). Concerning the pathways commented on in this review, although subcortical structures seem to be in place from birth (Uylings, 2006; Casey et al., 2004), prefrontal regions and their fronto-striatal connections seem to develop progressively over the course of childhood. In that sense, it is relevant to signal that whereas synaptogenesis peaks at the age of 3 months in auditory cortex, the peak appears at 15 months of age for the middle frontal gyri that are responsible for the development of executive function (Huttenlocher, 1997). Until cognitive and brain maturation allows for the mastering of a rudimentary control of attention, infants may not be able to shift from their default segmentation and word learning strategy. At the cognitive level, this idea has already been proposed by Gomez and Maye (2005). They reported a developmental progression from a default tendency to detect adjacent relations in 12-month-old infants to the extraction of more abstract non-adjacent relations at 18 months of age. Their explanation for these results was framed in terms of progressive development of processing capacity (Newport, 1988). From this point of view, maturation of this attentional shifting mechanism may be necessary in order for infants to detect the structural information of speech. The limitations of younger children may allow for a perceptual simplification of the complex speech stream. This would lead to an earlier acquisition of whole words, followed by a shift to the relevant cues and their integration for the acquisition of more distant information when other capacities, such as control of attention, are developed. This idea is consistent with later maturation of the brain structures responsible for the control of attention.

## Summary

The studies reviewed here applied a two-fold approach: 1) they used artificial languages, which allowed for the control of the cues available in speech in the absence of semantic information (i.e., statistical relations between syllables, presence of prosodic information by the insertion of

pauses); and 2) the combination of different methodologies, which complemented the behavioral measures of learning. These methodologies included the study of brain lesioned patients, measures of structural (DTI) and functional connectivity (phase coherence) and electrophysiological measures that enabled the recording of changes in brain activity during learning. The overall results indicate that different aspects of executive functions, specifically attention and working memory, play important roles in the early stages of language acquisition.

First, we have seen that the brain responses associated to word and rule learning can be clearly dissociated. Functionally dissociable electrophysiological modulations have been observed for pure segmentation, progressive memorization of segmented words and rule extraction from speech. More precisely, we have seen that the presence of prosodic cues modifies the brain responses associated with the treatment of the speech stream. While the absence of cues induces an increase of attention to syllable onsets (N1 modulation), the presence of pauses exogenously captures the learner's attention to word boundaries. This detection enhances segmentation, leading to progressive memorization and focus of attention in the whole segmented words (N400 modulation). It is at this point when control of attention has to be put in place to shift from this word learning procedure. The prosodic cue has then to be used to extract the embedded rule and categorize the elements of the dependency, thereby promoting grammatical acquisition. We have observed this change to be reflected in a different, endogenously-generated electrophysiological response (P2 modulation) that is associated with rule learning performance. This effect is consistent with the engagement of a control of attention to shift to the specific elements carrying the structural information while ignoring irrelevant elements. In terms of brain dynamics, we have observed that this mechanism is implemented by the synchronization between frontal and parietal activity via gamma-band coherence, a frequency band associated with attentional influences. Although the specific anatomical source of this effect is still lacking, this control is likely to be performed by a fronto-parietal network exerting cognitive control over sensory areas (Corbetta and Shulman, 2002).

Finally, we have also observed that the coordination between the different functions engaged in language acquisition (i.e., attentional control, language processing and working memory) could be in part performed by the basal ganglia. Whereas the brain regions responsible for speech processing and verbal working memory are in part overlapping (Jacquemot and Scott, 2006) and connected through ventral and dorsal pathways (Hickok and Poeppel, 2007; Marslen-Wilson and Tyler, 2007), the

basal ganglia share functional specializations and structural connections with these brain regions (Figure 5). In that sense, we have reviewed data showing how lesions to these structures in Huntington's disease patients induce difficulties in language acquisition, particularly for rule extraction and the capacity to abstract and transfer the rule when it is acquired. In a similar direction we have also reported how individual differences in segmentation and rule extraction under working memory restrictions were related to the integrity of the fronto-temporal connections passing through subcortical white matter via the left external capsule.

Overall, we underline the role of executive functions in language learning, particularly, the control of attention in the acquisition of language rules. We should nevertheless clarify that the involvement of executive function by no means implies that all the processes described here have to be explicit and volitional. Attentional tuning is not necessarily explicit and there is evidence indicating that shifts in endogenous attention can be performed automatically (Ranzini et al., 2009).

## Author Note

Ruth de Diego-Balaguer, Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona; IDIBELL, L'Hospitalet de Llobregat, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA). Barcelona, Spain; INSERM U955, IM3/ Paris 12, France and Département d'Etudes Cognitives (DEC), Ecole Normale Supérieure, Paris, France; Diana Lopez-Barroso, Departament de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona and IDIBELL, L'Hospitalet de Llobregat, Spain.

This work was supported by a pre-doctoral Grant (2009FI\_B 00138) from the Catalan government to DLB and a Grant from the Spanish Ministry of Science to RDB (MICINN, PSI2008-3885). We would like to thank A.-C. Bachoud-Lévi, E. Camara and L. Fuentemilla, T. Cunillera and T.F. Münte, co-authors of the reviewed work, and I. Nogué and D. Cucurell for their help preparing and running some of the experiments reported. We also thank A. Rodriguez-Fornells, J.M. Toro, M. Gullberg and P. Indefrey for their valuable comments and suggestions to the manuscript. Correspondence concerning this article should be addressed to: Ruth de Diego-Balaguer, Dept. de Psicologia Bàsica, Facultat de Psicologia, Universitat de Barcelona, Pg. Vall d'Hebron 171, 08035 Barcelona (Spain). Internet: [ruth.dediego@icrea.cat](mailto:ruth.dediego@icrea.cat)

## Notes

<sup>1</sup> Dependencies between elements will be highlighted in bold throughout the document.

<sup>2</sup> D : Determinant, N : Noun, V : Verb, NP : Noun Phrase, Adj : Adjective

<sup>3</sup> According to Nespor (*Nespor, M., Pena, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. Lingue e Linguaggio, ii(2), 201-227*) vowels and consonants have different roles in language processing, with vowels conveying grammatical variations while consonants more necessary for lexical identification. From this view the learning system is bias to exploit structural generalisations from vowels while transitional probabilities to segment words are more readily extracted over consonants. Bonnati et al (2005) and Toro et al (2008) interpret their results within this perspective.

<sup>4</sup> Exogenous attention is automatic and triggered by external cues (i.e., a sudden flash), while endogenous attention is classically considered as controlled and volitional (Posner, 1980). It is, however, important to keep in mind that endogenous attention can also be highly automatic when symbolic cues have an internalised association, for example, with specific locations (i.e., an arrow pointing to the right triggers a reflexive endogenous attentional shift to the right) (Ranzini et al., 2009; Posner, 1980), while still keeping a similar brain network and brain responses as volitional shifts of attention.

## References

- Abla,D., Katahira,K., and Okanoya,K. (2008). On-line Assessment of Statistical Learning by Event-related Potentials. *J. Cogn Neurosci.* *20*, 952-964.
- Abla,D. and Okanoya,K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia* *46*, 2787-2795.
- Abla,D. and Okanoya,K. (2009). Visual statistical learning of shape sequences: an ERP study. *Neurosci. Res.* *64*, 185-190.
- Alexander,G.E., DeLong,M.R., and Strick,P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* *9*, 357-381.
- Aslin,R.N., Saffran,J.R., and Newport,E.L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* *9*, 321-324.
- Awh,E., Smith,E.E., and Jonides,J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. *Ann. N. Y. Acad. Sci.* *769*, 97-117.
- Baddeley,A.D. (2007). Why do we need phonological loop? In *Working memory, thought and action*, A.D.Baddeley, ed. (Oxford: Oxford University Press), pp. 15-33.
- Bahlmann,J., Gunter,T.C., and Friederici,A.D. (2006). Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience* *18*, 1829-1842.
- Bengtsson,S.L., Nagy,Z., Skare,S., Forsman,L., Forssberg,H., and Ullen,F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* *8*, 1148-1150.
- Beretta,A., Campbell,C., Carr,T.H., Huang,J., Schmitt,L.M., Christianson,K., and Cao,Y. (2003). An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain Lang* *85*, 67-92.
- Bonatti,L.L., Pena,M., Nespors,M., and Mehler,J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* *16*, 451-459.
- Boorman,E.D., O'Shea,J., Sebastian,C., Rushworth,M.F., and Johansen-Berg,H. (2007). Individual differences in white-matter microstructure reflect variation in functional connectivity during choice. *Curr. Biol.* *17*, 1426-1431.

- Buiatti, M., Pena, M., and Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44, 509-519.
- Buschman, T.J. and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860-1862.
- Buzsaki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926-1929.
- Carral, V., Corral, M.J., and Escera, C. (2005). Auditory event-related potentials as a function of abstract change magnitude. *Neuroreport* 16, 301-305.
- Casey, B.J., Amso, D., and Davidson, M.C. (2006). Learning about learning and development with modern imaging technology. In *Processes of change in brain and cognitive development*, Y. Munakata and M.H. Johnson, eds. Oxford U. Press), pp. 513-533.
- Casey, B.J., Davidson, M.C., Hara, Y., Thomas, K.M., Martinez, A., Galvan, A., Halperin, J.M., Rodriguez-Aranda, C.E., and Tottenham, N. (2004). Early development of subcortical regions involved in non-cued attention switching. *Dev. Sci.* 7, 534-542.
- Casey, B.J., Giedd, J.N., and Thomas, K.M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol. Psychol.* 54, 241-257.
- Ceponiene, R., Alku, P., Westerfield, M., Torki, M., and Townsend, J. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology* 42, 391-406.
- Chambers, K.E., Onishi, K.H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87, B69-B77.
- Chang, C., Crottaz-Herbette, S., and Menon, V. (2007). Temporal dynamics of basal ganglia response and connectivity during verbal working memory. *Neuroimage* 34, 1253-1269.
- Charlton, R.A., Barrick, T.R., Lawes, I.N., Markus, H.S., and Morris, R.G. (2009). White matter pathways associated with working memory in normal aging. *Cortex*.
- Chemla, E., Mintz, T.H., Bernal, S., and Christophe, A. (2009). Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Dev. Sci.* 12, 396-406.
- Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Lang Speech* 51, 61-75.
- Corbetta, M. and Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201-215.

- Couette,M., Bachoud-Levi,A.C., Brugieres,P., Sieroff,E., and Bartolomeo,P. (2008). Orienting of spatial attention in Huntington's Disease. *Neuropsychologia* 46, 1391-1400.
- Cowan,N. (1995). Attention and memory: an integrated framework.
- Cunillera,T., Camara,E., Toro,J.M., Marco-Pallares,J., Sebastian-Galles,N., Ortiz,H., Pujol,J., and Rodriguez-Fornells,A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*.
- Cunillera,T., Gomila,A., and Rodriguez-Fornells,A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC. Neurosci.* 9, 23.
- Cunillera,T., Toro,J.M., Sebastian-Galles,N., and Rodriguez-Fornells,A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Res.* 1123, 168-178.
- Dahlin,E., Neely,A.S., Larsson,A., Backman,L., and Nyberg,L. (2008). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510-1512.
- De Diego-Balaguer, R., André, M., Rodriguez-Fornells, A., and Bachoud-Levi, A. C. Decoding speech to extract the rules of language. 2007a. In Proc. of the 15th Meeting of the European Society of Cognitive Psychology, Marseille, France
- De Diego-Balaguer,R., Costa,A., Sebastian-Galles,N., Juncadella,M., and Caramazza,A. (2004). Regular and irregular morphology and its relationship with agrammatism: evidence from two Spanish-Catalan bilinguals. *Brain Lang* 91, 212-222.
- De Diego-Balaguer,R., Couette,M., Dolbeau,G., Durr,A., Youssov,K., and Bachoud-Levi,A.C. (2008). Striatal degeneration impairs language learning: evidence from Huntington's disease. *Brain* 131, 2870-2881.
- De Diego-Balaguer,R., Fuentemilla,L., and Rodriguez-Fornells,A. (2009). Brain dynamics sustaining rapid rule extraction from speech. submitted.
- De Diego-Balaguer,R., Rodriguez-Fornells,A., Rotte,M., Bahlmann,J., Heinze,H.J., and Munte,T.F. (2006). Neural circuits subserving the retrieval of stems and grammatical features in regular and irregular verbs. *Hum. Brain Mapp.* 27, 874-888.
- De Diego-Balaguer,R., Sebastian-Galles,N., Diaz,B., and Rodriguez-Fornells,A. (2005). Morphological processing in early bilinguals: an ERP study of regular and irregular verb processing. *Brain Res. Cogn Brain Res.* 25, 312-327.

- De Diego-Balaguer,R., Toro,J.M., Rodriguez-Fornells,A., and Bachoud-Levi,A.C. (2007b). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS. ONE.* 2, e1175.
- Douaud,G., Behrens,T.E., Poupon,C., Cointepas,Y., Jbabdi,S., Gaura,V., Golestani,N., Krystkowiak,P., Verny,C., Damier,P., Bachoud-Levi,A.C., Hantraye,P., and Remy,P. (2009). In vivo evidence for the selective subcortical degeneration in Huntington's disease. *Neuroimage* 46, 958-966.
- Douaud,G., Gaura,V., Ribeiro,M.J., Lethimonnier,F., Maroy,R., Verny,C., Krystkowiak,P., Damier,P., Bachoud-Levi,A.C., Hantraye,P., and Remy,P. (2006). Distribution of grey matter atrophy in Huntington's disease patients: a combined ROI-based and voxel-based morphometric study. *Neuroimage* 32, 1562-1575.
- Doyle,A.E., Willcutt,E.G., Seidman,L.J., Biederman,J., Chouinard,V.A., Silva,J., and Faraone,S.V. (2005). Attention-deficit/hyperactivity disorder endophenotypes. *Biol. Psychiatry* 57, 1324-1335.
- Echols,C.H. and Newport,E.L. (1992). The role of stress and position in determining first words. *Language Acquisition* 2, 189-220.
- Ellis,N. (2008). Usage-based and Form-focused Language Acquisition: The Associative Learning of Constructions, Learned-Attention and the Limited L2 Endstate. In *Handbook of Cognitive Linguistics and Second Language Acquisition*, Lawrence Erlbaum).
- Endress,A.D. and Bonatti,L.L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105, 247-299.
- Endress,A.D. and Mehler,J. (2009a). Primitive computations in speech processing. *Q. J. Exp. Psychol.* (Colchester. ) 1-23.
- Endress,A.D. and Mehler,J. (2009b). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, 60. 353-369.
- Endress,A.D., Nespors,M., and Mehler,J. (2009). Perceptual and memory constraints on language acquisition. *Trends Cogn Sci.* 13, 348-353.
- Endress,A.D., Scholl,B.J., and Mehler,J. (2005). The role of salience in the extraction of algebraic rules. *J. Exp. Psychol. Gen.* 134, 406-419.
- Engel,A.K., Fries,P., and Singer,W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704-716.
- Finn,A.S. and Hudson Kam,C.L. (2008). The curse of knowledge: first language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition* 108, 477-499.

- Fiser, J. and Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science* 12, 499-504.
- Floel, A., de Vries, M.H., Scholz, J., Breitenstein, C., and Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *Neuroimage*. 47, 1974-1981.
- Frey, S., Campbell, J.S., Pike, G.B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435-11444.
- Friederici, A.D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn Sci* 8, 245-247.
- Friederici, A.D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci.* 13, 175-181.
- Friederici, A.D. and Kotz, S.A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *Neuroimage.* 20 *Suppl 1*, S8-17.
- Friederici, A.D., Kotz, S.A., Werheid, K., Hein, G., and von Cramon, D.Y. (2003). Syntactic comprehension in Parkinson's disease: investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology.* 17, 133-142.
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci.* 9, 474-480.
- Fuentemilla, L., Camara, E., Munte, T.F., Kramer, U.M., Cunillera, T., Marco-Pallares, J., Tempelmann, C., and Rodriguez-Fornells, A. (2009). Individual differences in true and false memory retrieval are related to white matter brain microstructure. *J. Neurosci.* 29, 8698-8703.
- Georgiou-Karistianis, N., Sritharan, A., Farrow, M., Cunnington, R., Stout, J., Bradshaw, J., Churchyard, A., Brawn, T.L., Chua, P., Chiu, E., Thiruvady, D., and Egan, G. (2007). Increased cortical recruitment in Huntington's disease using a Simon task. *Neuropsychologia* 45, 1791-1800.
- Gervain, J., Macagno, F., Cogoi, S., Pena, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. U. S. A* 105, 14222-14227.
- Gleitman, L. and Wanner, E. (1982). The state of the state of the art. In *Language acquisition: The state of the art*, E. Wanner and L. Gleitman, eds. (Cambridge, U. K.: Cambridge University Press), pp. 3-48.
- Gold, B.T., Powell, D.K., Xuan, L., Jiang, Y., and Hardy, P.A. (2007). Speed of lexical decision correlates with diffusion anisotropy in left parietal and frontal white matter: evidence from diffusion tensor imaging. *Neuropsychologia* 45, 2439-2446.

- Gomez,R. and Maye,J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy* 7, 183-206.
- Gomez,R.L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431-436.
- Gomez,R.L. and Lakusta,L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science* 7, 567-580.
- Gruber,T., Keil,A., and Muller,M.M. (2001). Modulation of induced gamma band responses and phase synchrony in a paired associate learning task in the human EEG. *Neurosci. Lett.* 316, 29-32.
- Haber,S.N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317-330.
- Haber,S.N. and Calzavara,R. (2009). The cortico-basal ganglia integrative network: the role of the thalamus. *Brain Res. Bull.* 78, 69-74.
- Hagoort,P., Hald,L., Bastiaansen,M., and Petersson,K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438-441.
- Hasting,A.S., Kotz,S.A., and Friederici,A.D. (2007). Setting the stage for automatic syntax processing: the mismatch negativity as an indicator of syntactic priming. *J. Cogn Neurosci.* 19, 386-400.
- Hauser,M.D., Newport,E.L., and Aslin,R.N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78, B53-B64.
- Hauser,M.D., Weiss,D., and Marcus,G. (2002). Rule learning by cotton-top tamarins. *Cognition* 86, B15-B22.
- Hickok,G. and Poeppel,D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393-402.
- Hillyard,S.A., Hink,R.F., Schwent,V.L., and Picton,T.W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177-180.
- Huttenlocher,P.R. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology* 387, 167-178.
- Jacquemot,C. and Scott,S.K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci.* 10, 480-486.
- Jusczyk,P.W. (1997). *The discovery of spoken language.* (Cambridge, MA: MIT Press).
- Jusczyk,P.W. (1999). How infants begin to extract words from speech. *Trends Cogn Sci.* 3, 323-328.

- Karnath,H.O., Zopf,R., Johannsen,L., Fruhmann,B.M., Nagele,T., and Klose,U. (2005). Normalized perfusion MRI to identify common areas of dysfunction: patients with basal ganglia neglect. *Brain* 128, 2462-2469.
- Laufer,B. (1990). 'Sequence' and 'order' in the development of L2 lexis. *Applied Linguistics* 11, 281-296.
- Lawrence,A.D., Weeks,R.A., Brooks,D.J., Andrews,T.C., Watkins,L.H., Harding,A.E., Robbins,T.W., and Sahakian,B.J. (1998). The relationship between striatal dopamine receptor binding and cognitive performance in Huntington's disease. *Brain* 121 ( Pt 7), 1343-1355.
- Le Bihan D. (2006). From Brownian motion to mind imaging: diffusion MRI. *Bull. Acad. Natl. Med.* 190, 1605-1627.
- Leh,S.E., Ptito,A., Chakravarty,M.M., and Strafella,A.P. (2007). Fronto-striatal connections in the human brain: a probabilistic diffusion tractography study. *Neurosci. Lett.* 419, 113-118.
- Lehericy,S., Ducros,M., Van de Moortele,P.F., Francois,C., Thivard,L., Poupon,C., Swindale,N., Ugurbil,K., and Kim,D.S. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Ann. Neurol.* 55, 522-529.
- Lehtonen,M., Vorobyev,V.A., Soveri,A., Hugdahl,K., Tuokkola,T., and Laine,M. (2009). Language-specific activations in the brain: Evidence from inflectional processing in bilinguals. *Journal of Neurolinguistics* 22, 495-513.
- Lemhofer,K. and Dijkstra,T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Mem.Cognit.* 32, 533-550.
- Lieberman,P. (2000). *Human language and our reptilian brain: the subcortical bases of speech, syntax and thought.* (Cambridge, MA: Harvard University Press).
- Longworth,C.E., Keenan,S.E., Barker,R.A., Marslen-Wilson,W.D., and Tyler,L.K. (2005). The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. *Brain* 128, 584-596.
- Lopez-Barroso,D., De Diego-Balaguer,R., Cunillera,T., Camara,E., and Rodriguez-Fornells,A. (submitted). Role of the working memory phonological loop subcomponents in rule-based learning and segmentation of language: a behavioural and DTI neuroimaging study.
- Luck,S.J. and Hillyard,S.A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1000-1014.
- Makris,N., Papadimitriou,G.M., Kaiser,J.R., Sorg,S., Kennedy,D.N., and Pandya,D.N. (2009). Delineation of the middle longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb. Cortex* 19, 777-785.
- Marcus,G.F., Fernandes,K.J., and Johnson,S.P. (2007). Infant rule learning facilitated by speech. *Psychol. Sci.* 18, 387-391.

- Marcus,G.F., Vijayan,S., Rao,S.B., and Vishton,P.M. (1999). Rule learning by seven-month-old infants. *Science* 283, 77-80.
- Marslen-Wilson,W.D. and Tyler,L.K. (1997). Dissociating types of mental computation. *Nature* 387, 592-594.
- Marslen-Wilson,W.D. and Tyler,L.K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 362, 823-836.
- Mattys,S.L., Pleydell-Pearce,C.W., Melhorn,J.F., and Whitecross,S.E. (2005). Detecting silent pauses in speech: a new tool for measuring on-line lexical and semantic processing. *Psychol. Sci.* 16, 958-964.
- McLaughlin,J., Osterhout,L., and Kim,A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703-704.
- McNealy,K., Mazziotta,J.C., and Dapretto,M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J. Neurosci.* 26, 7629-7639.
- Middleton,F.A. and Strick,P.L. (2000). Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain Cogn* 42, 183-200.
- Miltner,W.H., Braun,C., Arnold,M., Witte,H., and Taub,E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature* 397, 434-436.
- Mintz,T.H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91-117.
- Morgan,J.L., Meier,R.P., and Newport,E.L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cogn Psychol.* 19, 498-550.
- Mueller,J.L., Bahlmann,J., and Friederici,A.D. (2008a). The role of pause cues in language learning: the emergence of event-related potentials related to sequence processing. *J. Cogn Neurosci.* 20, 892-905.
- Mueller,J.L., Girgsdies,S., and Friederici,A.D. (2008b). The impact of semantic-free second-language training on ERPs during case processing. *Neurosci. Lett.* 443, 77-81.
- Murphy,R.A., Mondragon,E., and Murphy,V.A. (2008). Rule learning by rats. *Science* 319, 1849-1851.
- Myers,R.H., Vonsattel,J.P., Stevens,T.J., Cupples,L.A., Richardson,E.P., Martin,J.B., and Bird,E.D. (1988). Clinical and neuropathologic assessment of severity in Huntington's disease. *Neurology* 38, 341-347.
- Nan,Y., Knosche,T.R., and Friederici,A.D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res.* 1094, 179-191.

- Nan, Y., Knosche, T.R., and Friederici, A.D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biol. Psychol.* *82*, 70-81.
- Neville, H. (2006). Different profiles of plasticity within human cognition. In *Processes of change in brain and cognitive development*, Y. Munakata and M.H. Johnson, eds. Oxford U. Press), pp. 287-314.
- Newport, E.L. (1988). Constraints in learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Language Sciences* *10*, 147-172.
- Newport, E.L. and Aslin, R.N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* *48*, 127-162.
- Newport, E.L., Hauser, M.D., Spaepen, G., and Aslin, R.N. (2004). Learning at a distance - II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology* *49*, 85-117.
- Onishi, K.H., Chambers, K.E., and Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition* *83*, B13-B23.
- Onnis, L., Waterfall, H.R., and Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition* *109*, 423-430.
- Opitz, B. and Friederici, A.D. (2007). Neural basis of processing sequential and hierarchical syntactic structures. *Hum. Brain Mapp.* *28*, 585-592.
- Peña, M., Bonatti, L.L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* *298*, 604-607.
- Pelucchi, B., Hay, J.F., and Saffran, J.R. (2009). Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition* *113*, 244-247.
- Perruchet, P. and Vinter, A. (2002). The self-organizing consciousness. *Behav. Brain Sci.* *25*, 297-330.
- Peschanski, M., Cesaro, P., and Hantraye, P. (1995). Rationale for intrastriatal grafting of striatal neuroblasts in patients with Huntington's disease. *Neuroscience* *68*, 273-285.
- Popescu, A.T., Popa, D., and Pare, D. (2009). Coherent gamma oscillations couple the amygdala and striatum during learning. *Nat. Neurosci.*
- Posner, M.I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* *32*, 3-25.
- Ranzini, M., Dehaene, S., Piazza, M., and Hubbard, E.M. (2009). Neural mechanisms of attentional shifts due to irrelevant spatial and numerical cues. *Neuropsychologia* *47*, 2615-2624.

- Reinke, K.S., He, Y., Wang, C., and Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Research* 17, 781-791.
- Rodriguez-Fornells, A., Cunillera, T., Mestres-Misse, A., and De Diego-Balaguer, R. (2009) Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 364, 3711-3735
- Rosas, H.D., Koroshetz, W.J., Chen, Y.I., Skeuse, C., Vangel, M., Cudkovicz, M.E., Caplan, K., Marek, K., Seidman, L.J., Makris, N., Jenkins, B.G., and Goldstein, J.M. (2003). Evidence for more widespread cerebral pathology in early HD: an MRI-based morphometric analysis. *Neurology* 60, 1615-1620.
- Sabisch, B., Hahne, A., Glass, E., von Suchodoletz, W., and Friederici, A.D. (2009). Children with specific language impairment: The role of prosodic processes in explaining difficulties in processing syntactic information. *Brain Res.*
- Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926-1928.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., and Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27-52.
- Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, 69, 105–134.
- Sanders, L.D., Newport, E.L., and Neville, H.J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5, 700-703.
- Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G.L., and von, S.A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proc. Natl. Acad. Sci. U. S. A* 95, 7092-7096.
- Saur, D., Kreher, B.W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci U. S. A* 105, 18035-18040.
- Schmahmann, J.D. and Pandya, D.N. (2006). *Fibers Pathways of the Brain.*, J.D.Schmahmann and D.N.Pandya, eds. (Oxford: Oxford University Press).
- Scholz, J., Klein, M.C., Behrens, T.E., and Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nat. Neurosci.* 12, 1370-1371.
- Seger, C.A. (2006). The basal ganglia in human learning. *Neuroscientist*. 12, 285-290.

- Seger,C.A. and Cincotta,C.M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb. Cortex* 16, 1546-1555.
- Sheehan,K.A., McArthur,G.M., and Bishop,D.V. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Brain Res. Cogn Brain Res.* 25, 547-553.
- Siegel,M., Donner,T.H., Oostenveld,R., Fries,P., and Engel,A.K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron* 60, 709-719.
- Snyder,J.S., Alain,C., and Picton,T.W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J. Cogn Neurosci.* 18, 1-13.
- Teichmann,M., Dupoux,E., Kouider,S., and Bachoud-Lévi,A.-C. (2006). The role of the striatum in processing language rules: evidence from word perception in Huntington's disease. *J. Cogn Neurosci.*
- Teichmann,M., Dupoux,E., Kouider,S., Brugieres,P., Boisse,M.F., Baudic,S., Cesaro,P., Peschanski,M., and Bachoud-Levi,A.C. (2005). The role of the striatum in rule application: the model of Huntington's disease at early stage. *Brain* 128, 1155-1167.
- Thieben,M.J., Duggins,A.J., Good,C.D., Gomes,L., Mahant,N., Richards,F., McCusker,E., and Frackowiak,R.S. (2002). The distribution of structural neuropathology in pre-clinical Huntington's disease. *Brain* 125, 1815-1828.
- Toro,J.M., Nespors,M., Mehler,J., and Bonatti,L.L. (2008). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychol. Sci.* 19, 137-144.
- Toro,J.M. and Trobalon,J.B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics* 67, 867-875.
- Tyler,L.K., Bright,P., Fletcher,P., and Stamatakis,E.A. (2004). Neural processing of nouns and verbs: the role of inflectional morphology. *Neuropsychologia* 42, 512-523.
- Tyler,L.K., Stamatakis,E.A., Post,B., Randall,B., and Marslen-Wilson,W. (2005). Temporal and frontal systems in speech comprehension: an fMRI study of past tense processing. *Neuropsychologia* 43, 1963-1974.
- Ullman,M.T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex* 42, 480-485.
- Uylings,H.B.M. (2006). Development of the human cortex and the concept of "critical" or "sensitive" periods. In *The Cognitive Neuroscience of Second Language Acquisition*, M.Gullberg and P.Indefrey, eds. (Oxford: Blackwell Publishing), pp. 59-90.
- Vonsattel,J.P., Myers,R.H., Stevens,T.J., Ferrante,R.J., Bird,E.D., and Richardson,E.P., Jr. (1985). Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* 44, 559-577.

- Wahl, M., Marzinzik, F., Friederici, A.D., Hahne, A., Kupsch, A., Schneider, G.H., Saddy, D., Curio, G., and Klostermann, F. (2008). The human thalamus processes syntactic and semantic language violations. *Neuron* 59, 695-707.
- Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953-956.
- Weber-Fox, C.M. and Neville, H.J. (1996). Maturation constraints on functional specialization for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8, 231-256.
- Williams, D., Stott, C.M., Goodyer, I.M., and Sahakian, B.J. (2000). Specific language impairment with or without hyperactivity: neuropsychological evidence for frontostriatal dysfunction. *Dev. Med. Child Neurol.* 42, 368-375.
- Yin, H.H. and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464-476.

Table 1.

Summary of relevant studies where positive learning of adjacent and non-adjacent non repetition-based learning has been reported. Only results on infants and non-human animal studies are included. The population, the type of dependency studied (Adjacent and Non-adjacent dependency) and examples of the learning and test material used are described.

<i>Non Repetition-based learning</i>				
Sample	Dependencies	Material		References
		Training streams	Test	
Monkeys (Cotton-top tamarins)	Adjacent ( <u>segmentation</u> )	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku C <sub>2</sub> B <sub>1</sub> A <sub>1</sub> : tibida B <sub>1</sub> B <sub>2</sub> C <sub>3</sub> : dapiku	(Hauser et al., 2001)
Rats	Adjacent ( <u>segmentation</u> )	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku C <sub>2</sub> B <sub>1</sub> A <sub>1</sub> : tibida B <sub>1</sub> B <sub>2</sub> C <sub>3</sub> : dapiku	(Toro and Trobalon, 2005)
Monkeys (Cotton-top tamarins)	Non-Adjacent ( <u>segmentation</u> )	A <sub>1</sub> x C <sub>1</sub> : pokiga ( <u>syllable dependency</u> ) A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : pagute ( <u>vowel dependency</u> )	A <sub>1</sub> x C <sub>1</sub> : pokiga C <sub>2</sub> A <sub>1</sub> x: bupoki A <sub>1</sub> x C <sub>1</sub> : pagute C <sub>2</sub> A <sub>1</sub> x : baepagu	(Newport et al. 2004)
8 month-old infants	Adjacent ( <u>segmentation</u> )	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku	A <sub>1</sub> B <sub>1</sub> C <sub>1</sub> : bidaku C <sub>2</sub> B <sub>1</sub> A <sub>1</sub> : tibida B <sub>1</sub> B <sub>2</sub> C <sub>3</sub> : dapiku	(Saffran et al., 1996; Aslin et al., 1998)
15 month-old infants	Non-Adjacent ( <u>generalization</u> )	A <sub>1</sub> x C <sub>1</sub> : pel-kicey-jic (x = 18 values)	A <sub>1</sub> x C <sub>1</sub> : pel-wadim-rud A <sub>1</sub> x C <sub>2</sub> : pel-kicey-rud	(Gomez and Maye, 2005)
17 and 18 month-old infants	Non-Adjacent ( <u>generalization</u> )	A <sub>1</sub> x C <sub>1</sub> : pel-kicey-jic (x = 24 values)	A <sub>1</sub> x C <sub>1</sub> : pel-wadim-rud A <sub>1</sub> x C <sub>2</sub> : pel-kicey-rud	(Gomez, 2002)

Table 2.

Summary of relevant studies where positive learning of adjacent and non-adjacent repetition-based learning has been reported. Only results on infants and non-human animal studies are included. The population, the type of dependency studied (Adjacent and Non-adjacent dependency) and examples of the learning and test material used are described.

<i>Repetition-based learning</i>				
Sample	Dependencies	Material		References
		Training	Test	
Newborns	Adjacent	Different brain responses between ABB (e.g. <i>mubaba</i> ) and ABC (e.g. <i>mubage</i> ) during exposition		(Gervain et al., 2008)
7 months-old infants	Adjacent & Non-Adjacent ( <u>generalization</u> )	ABB: gatiti	ABA: wofewo	(Marcus et al., 1999)
		ABA: gatiga	ABB: wofefe ABA: wofewo	
Monkeys (Cotton-top tamarins)	Adjacent ( <u>generalization</u> )	ABB: gatiti	AAB: wowofe	(Hauser et al., 2002)
		AAB: gagati	ABB: wofefe	
Rats	Non-Adjacent ( <u>generalization</u> )	ABA auditory patterns	DCD auditory patterns	(Murphy et al., 2008)

## Figure Captions

Figure 1. Schematic illustration of the materials used in the studies reviewed. Left: illustration of the language streams used in the learning phase. Each word was composed of three syllables; the first and the last of syllable were placed following a dependency rule. Words (i.e., *patem*) and rules (i.e., the structure *pa\_\_mi*) could be acquired from the same material. One of the possible structures that can be learned is highlighted in bold. The “\_” represents the 25 ms pause between words. Right: examples of test items used in the two-alternative forced choice test performed after learning to evaluate word and rule acquisition. After a learning phase lasting for four minutes, a behavioral test was administered. Half of the streams were tested for word acquisition and half for rule-learning using a two-alternative forced-choice test.

Figure 2. Modulation of the ERP components as a function of learning and time of exposure. A. Correlation between the mean amplitude of the P2 component at Fz in the third minute of learning (at the 120-220 ms time-window) and performance on the rule-learning test. B. ERP averages of the language conditions for each group at a frontal location (Fz), showing the evolution of the differences between groups over the learning period (first, second and third minute). C. Differences between Good and Poor Learners for each spectral band (theta, alpha and gamma) at blocks 1, 3, 5, 7, 9, 11, 13 and 15 including 50 consecutive word stimuli each. Dotted lines indicate transition between first to second minute, second to third minute and third to fourth minute of exposure to the language. Significantly (Mann-Wilcoxon rank-sum test,  $P < 0.05$ ) higher coherence for Good over Poor learners is plotted in red lines while significantly (Mann-Wilcoxon rank-sum test,  $P < 0.05$ ) greater inter-electrode coherence for Poor over Good Learners is represented with blue lines. \*  $P < 0.05$ , corrected for multiple comparisons (\*\* indicates the peak of the effect).

Figure 3. Grand Average ERPs for each trisyllabic presentation at central (Cz) electrode location for language and random streams in the continuous condition and the condition with pauses. The location of the plotted electrode is highlighted with a red dot. The dotted line indicates the onset of each syllable.

Figure 4. Material and diffusion tensor imaging results from the working memory study. A. Structure of the two types of languages used to induce rule learning and word segmentation. The diagram illustrates the phonological working memory components and how in the working memory constrained condition (i.e. the

articulatory suppression condition [ASC, in orange], the only information that enters straight into the store is the relevant information, while the rehearsal is not available since participants were required to utter repeatedly the syllable “bla”. B. Structure of the test items for each condition. Participants were asked to choose between pairs of words after the learning phase. In the Rule learning condition, pairs of rule-words, and non-words were presented. In the segmentation condition, pairs of words and part-words (combining the final syllable of one word and the two first syllables of another word, or the last two syllables of one word and the first one of another word) were presented. C. Left: Significantly white matter correlated clusters between the FA values and the performance in ASC rendered on the FA mean image showing the left external capsule ( $p < .01$ ,  $n = 60$  voxels). Right: Scatter-plot of the correlation between the mean FA value for each participant at the peak of the left external capsule cluster [MNI coordinates  $x = -27$ ,  $y = 9$ ,  $z = 0$ ] and ASC in the language learning conditions (segmentation and rule together).

Figure 5. Areas related to language processing, control of attention and working memory and pathways connecting them. The diagram illustrates the two major dorsal and ventral pathways connecting frontal, parietal and temporal areas and concomitant connections of these areas with the basal ganglia. These connections project to the anterior part of the striatum (the head of the caudate and the anterior putamen) that are framed in the diagram.

Figure 1.

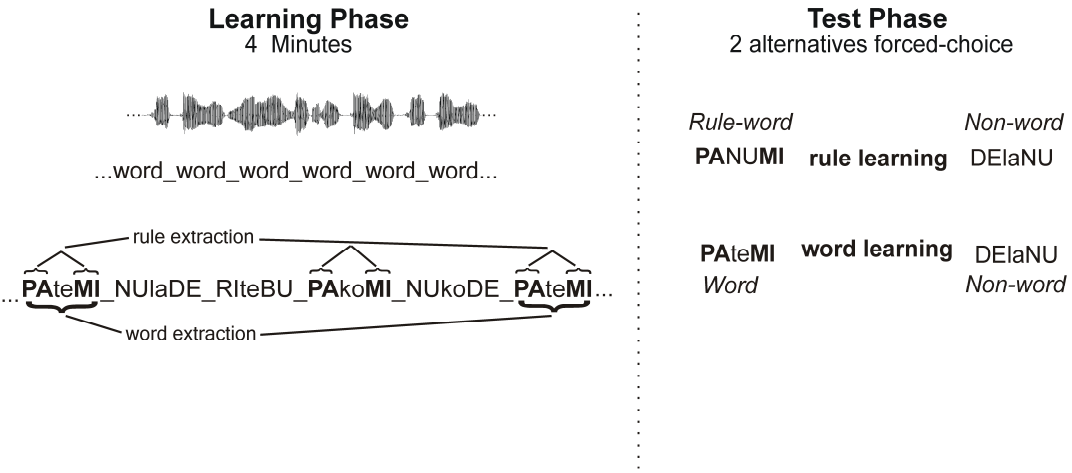


Figure 2.

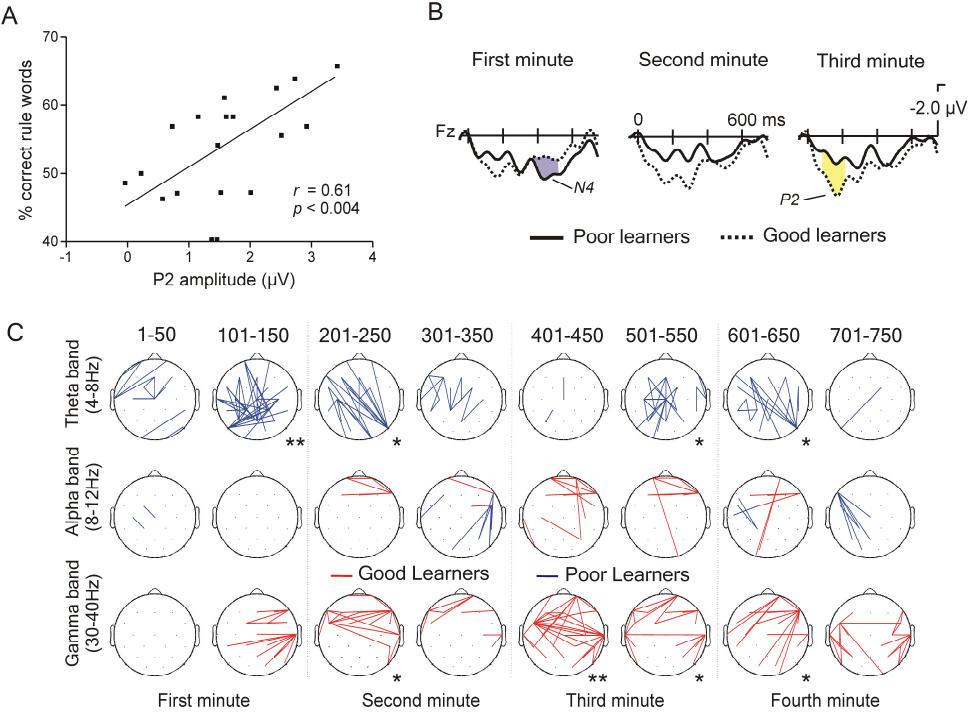


Figure 3.

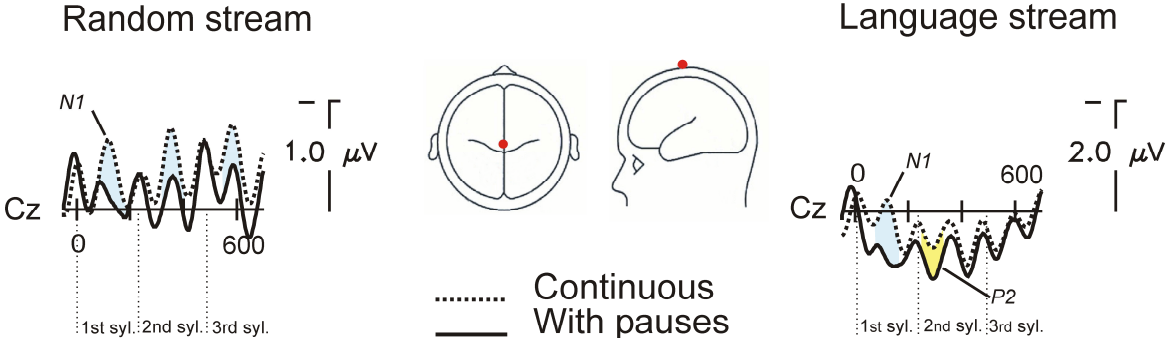


Figure 4.

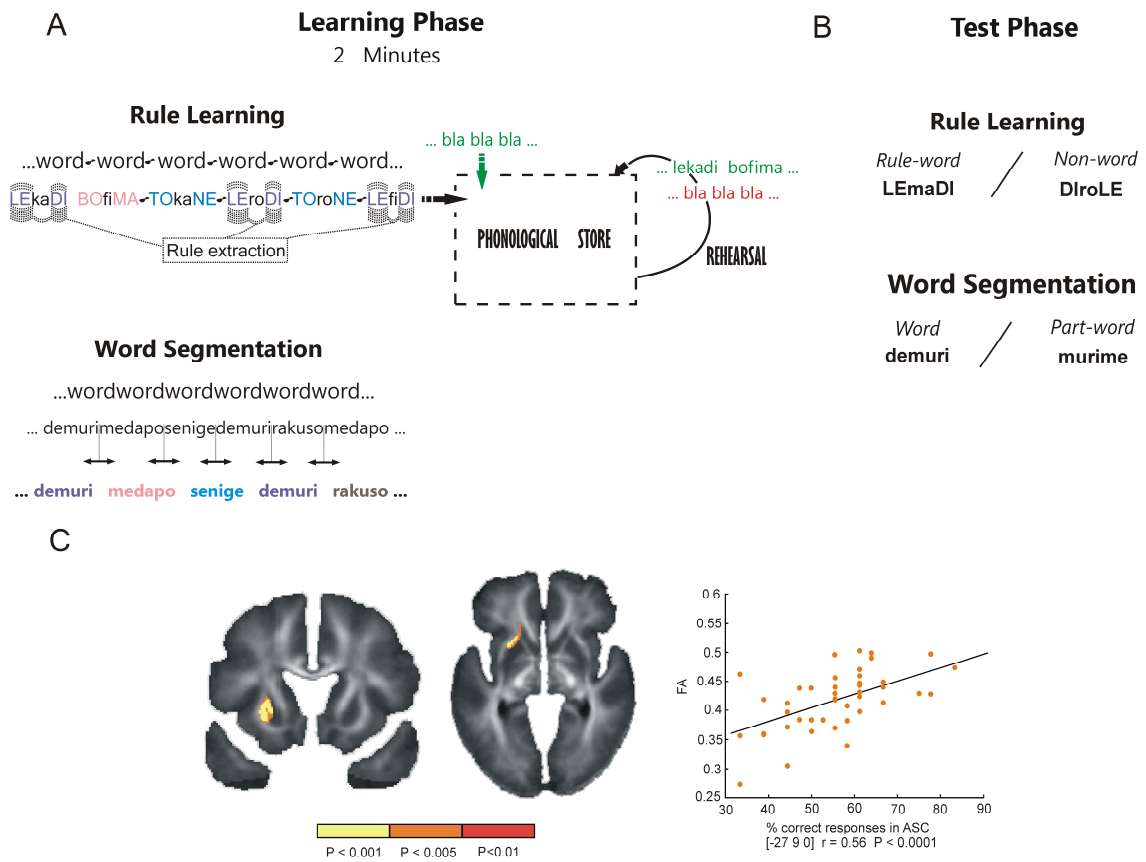


Figure 5.

