Contents lists available at ScienceDirect

Cognition

Attentional effects on rule extraction and consolidation from speech

Diana López-Barroso [a,b,c,d], David Cucurell [a,b], Antoni Rodríguez-Fornells [a,b,e], Ruth de Diego-Balaguer [a,b,e,f,*]

[a] Cognition and Brain Plasticity Unit. Bellvitge Research Biomedical Institute (IDIBELL), Hospitalet de Llobregat, 08097, Spain
[b] Department of Basic Psychology, University of Barcelona, Campus Bellvitge, Hospitalet de Llobregat, 08097, Spain
[c] Cognitive Neurology and Aphasia Unit and Cathedra Foundation Morera and Vallejo of Aphasia, Centro de Investigaciones Médico-Sanitarias, University of Malaga, Malaga, 29071, Spain
[d] Department of Psychobiology and Methodology of Behavioural Sciences, Faculty of Psychology, University of Malaga, Malaga, 29071, Spain
[e] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
[f] Institute of Neurosciences, University of Barcelona, Spain

ARTICLE INFO

ABSTRACT

Incidental learning plays a crucial role in the initial phases of language acquisition. However the knowledge derived from implicit learning, which is based on prediction-based mechanisms, may become explicit. The role that attention plays in the formation of implicit and explicit knowledge of the learned material is unclear. In the present study, we investigated the role that attention plays in the acquisition of non-adjacent rule learning from speech. In addition, we also tested whether the amount of attention during learning changes the representation of the learned material after a 24 h delay containing sleep. For that, we developed an experiment run on two consecutive days consisting on the exposure to an artificial language that contained non-adjacent dependencies (rules) between words whereas different conditions were established to manipulate the amount of attention given to the rules (target and non-target conditions). Furthermore, we used both indirect and direct measures of learning that are more sensitive to implicit and explicit knowledge, respectively. Whereas the indirect measures indicated that learning of the rules occurred regardless of attention, more explicit judgments after learning showed differences in the type of learning reached under the two attention conditions. 24 hours later, indirect measures showed no further improvements during additional language exposure and explicit judgments indicated that only the information more robustly learned in the previous day, was consolidated.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Prediction-based mechanisms appear to play a vital role in the detection of regularities that govern complex situations such as human language. Language contains adjacent and non-adjacent dependencies between elements that should be mastered to, for example, process 'recursion', which is a hallmark of human language (Hauser, Chomsky, & Fitch, 2002). Learning non-adjacent dependencies from language has been claimed to heavily rely on general prediction learning processes (Misyak, Christiansen, & Tomblin, 2010b; Perruchet & Pacton, 2006), which often occurs incidentally, i.e. without the intention to learn ("implicit learning"; Reber, 1967).

Previous research has evaluated non-adjacent rule learning using artificial language learning paradigms (see Gómez, 2002; Peña, Bonatti, Nespor, & Mehler, 2002; Romberg & Saffran, 2013),

in which words or phrases without meaning are built following the structure AXC, establishing that the first element (A) predicts the third one (C), whereas the second element (X) is variable. These artificial paradigms are built as an analogy to what occurs in natural languages (e.g., **he** sleep**s**, **she** run**s**). Statistical learning mechanisms can track these predictive dependencies, extract the existing relationship and allow generalization to new contexts. However, an important question is the degree in which learning in incidental situations relies on attention. Some studies have provided evidence that segmentation of a speech stream into discrete word units can occur incidentally (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). In addition, rule generalization is possible under diverted attention but only as long as learning is based on adjacent dependencies (Toro, Sinnet, & Soto-Faraco, 2011; Toro, Sinnett, & Soto-Faraco, 2005). However, tracking non-adjacent relationships is more complex (Newport & Aslin, 2004). It has been proposed that the only necessary condition to learn adjacent and non-adjacent dependencies is the joint attention for the processing of the two elements in the dependency (Ellis, 2006; Pacton & Perruchet, 2008). In agreement with this view, some experiments

have shown that the presence of pauses in the speech signal (Endress, Nespor, & Mehler, 2009; Peña et al., 2002) and the high variability of irrelevant elements in the sequence (Gómez, 2002) are essential for non-adjacent rule learning. Although these studies did not test the role of attention in the learning process directly, the importance of these features could reside in the enhancement of perceptual salience, guiding attention toward the elements that share the dependency (for a discussion about this topic, see Aslin & Newport, 2012; Romberg & Saffran, 2013).

Electrophysiological data using event-related potentials (ERPs) to track the online learning of non-adjacent rules in an artificial language learning paradigm have also shown that an attention-modulated ERP component, the P2 component, increases as a function of learning, which possibly indexes a change in the locus of the focus of attention during learning (from adjacent to non-adjacent dependencies) (De Diego-Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007) as previously suggested by Gómez and Maye (2005).

Importantly, understanding learning in incidental situations should not be limited to explicit judgments. It is also important to dissociate between the acquisition and storage of new information in relation to the implicit/explicit dimension (Frensch, 1998). During acquisition, knowledge can be initially encoded implicitly, as it often occurs in incidental situations. However, once learned, the invariant features (such as the dependency between non-adjacent elements) are enhanced and can eventually enter consciousness and become more explicit (Cleeremans, 2008). Most studies have not accounted for this distinction because they have only evaluated participants' performances after the learning phase when learning was accomplished (Saffran et al., 1997; Toro et al., 2011). Therefore, no information was available for online implicit learning while manipulating attention. In relation to this point, recent work has clearly shown the importance of introducing online measures in addition to the classical more explicit judgments after learning (Batterink, Reber, Neville, & Paller, 2015; Misyak, Christiansen, & Tomblin, 2010a; Misyak et al., 2010b).

Referring to how representations of the learned information change over time, previous evidence indicates that sleep promotes the lexicalization of new words (Davis, Di Betta, McDonald, & Gaskell, 2009; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). In addition, it promotes the creation of abstract and generalizable representations in rule learning from language (Gómez, Bootzin, & Nadel, 2006; Merkx, Rastle, & Davis, 2011; Tamminen, Davis, Merkx, & Rastle, 2012). Importantly, sleep-related consolidation causes qualitative and quantitative changes in the mental representation of knowledge outside the language domain (for a review: Diekelmann & Born, 2010). Moreover, it plays an important role in promoting the conversion of implicit knowledge into explicit knowledge (Payne, Ellenbogen, Walker, & Stickgold, 2008; Wagner, Gais, Haider, Verleger, & Born, 2004).

Based on the above-mentioned background, the present study was developed with two different goals. First, our interest was to evaluate whether the amount of attention during the learning of non-adjacent rules affects differently indirect measures of learning and more explicit judgments on the underlying knowledge of the rules. Therefore, we developed a paradigm that allowed us to indirectly evaluate online rule learning in different attentional conditions. An artificial language learning task, in which the participants heard phrases of three artificial words that followed the form of AXC, was implemented with a word-monitoring task that acted as a cover task to manipulate attention. Thus, because learning the underlying dependencies helps to solve the cover task faster, the reaction times (RT) to the cover task provided an indirect online measure of implicit rule learning (Brandon, Terry, Stevens, & Tillmann, 2012; Misyak et al., 2010b). Explicit judgments were also used to assess rule learning by administering a recognition test to the participants after the learning phase. In addition, our secondary objective was to investigate whether attention affects the manner in which rule representations are consolidated leading to different effects in implicit and explicit assessments of this knowledge. Thus, participants' direct and indirect measures were recorded on two consecutive days.

2. Methods

2.1. Participants

Twenty-five students (19 women; mean age: 21.5 SD: 1.9) from the University of Barcelona participated in this study for either 10 euros or course credits. The students were all native Spanish speakers and had no history of auditory problems.

2.2. Materials and procedure

Each participant performed two sessions of the same language learning task (Experiment 1 and 2), separated by 24 hours. Each experiment consisted of a learning and a test phase (see Fig. 1A). For the experiments, 24 CVCV bisyllabic novel words (from now on called "words") were created following Spanish phonotactics. Words were recorded in isolation to avoid intonation cues, in a sound attenuated booth by a female Spanish native speaker. Afterwards they were combined with a sound editor software (Adobe Audition) to form the phrases, taking for each phrase three words from the pool of novel words (e.g., *tagi-male-sira*; Table 1) and inserting a 100 ms interval between words. The average duration of each word was 483.8 ms (± 39.7 ms). The auditory phrases were presented during the learning and test phases, through headphones at a comfortable level and set constant across participants with the Presentation software.

2.2.1. Learning phase

For the learning phase, words were combined to form rule phrases (AXC) and filler (XXX) phrases (Fig. 1B). Following the structure used in previous studies (Gómez, 2002; Gómez & Maye, 2005), *rule phrases* took the form AXC (e.g., **tagi**-male-**sira**, **tagi**-fuse-**sira**, **tagi**-pofi-**sira**) (Table 1), thus establishing that the initial word (A) determined the third word (C) regardless of the middle element (X). Six of the words from the pool were used to build three different AXC rules (i.e. $A_1\_C_1$: tagi_sira; $A_2\_C_2$: jupo_runi; $A_3\_C_3$: pine_ladu). The remaining 18 (i.e. cilu, mego, lofa, tadi, nuso, pume, male, rosu, foli, vidu, supa, pevo, ture, medi, catu, gupe, defa, and nigo) were used as middle words for all A_C structures. Although over the three structures the 18 different words were presented, each structure used only 12 of the 18 X elements. The other 6, different for each structure, were used to test generalization in the recognition phase after learning. *Filler phrases* took the form XXX (e.g., *male-fuse-posi*) and were created by combining the 18 elements that randomly appeared in the middle of the rule phrases (i.e. X element in the AXC phrases) (Table 2). They were combined with the constraint that the same word could not appear twice in the same phrase and each X had the same probability to appear in each position. Each filler phrase was presented only once in the learning phase. These phrases appeared only in the learning phase (see Fig. 1B).

In the *learning phase*, the participants were exposed to 36 rule (see Table 1) and 18 filler phrases (Table 2) that were randomly intermixed. A 100-ms warning tone was used as an arousing signal to prepare the participants for the upcoming presentation of the phrase, which started 400 ms after the tone. Participants performed a word-monitoring task to obtain an indirect measure of learning by means of the reaction times to each phrase presentation. The target word remained printed in the middle of the screen
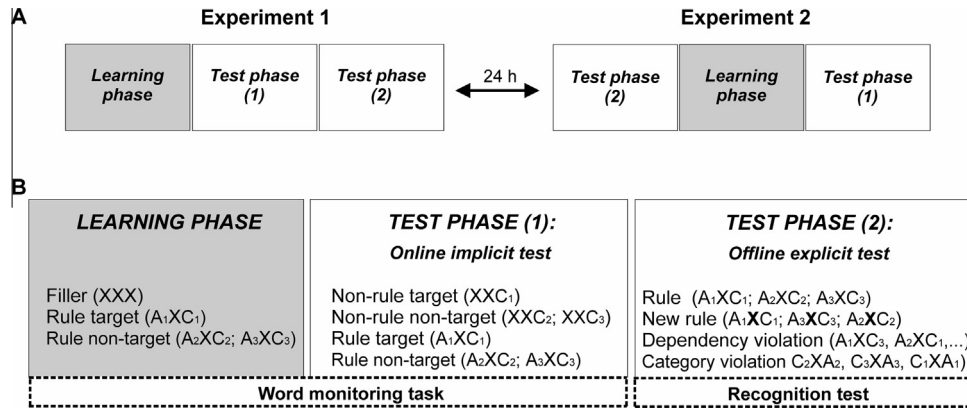
Fig. 1. Design and materials of the study. (A) Illustration of the order of the phases composing Experiment 1 (day 1) and Experiment 2 (day 2). (B) Illustration of the type of phrases that appeared in each phase of the experiments. The learning phase and the online implicit test required a word-monitoring task whereas the offline explicit test took the form of a recognition test.

Table 1
Rule phrases used in the artificial language of the experiment. Words forming the three structures are shaded in grey.

Rule 1			Rule 2			Rule 3		
A	X	C	A	X	C	A	X	C
tagi	male	sira	jupo	male	runi	pine	ture	ladu
tagi	rosu	sira	jupo	rosu	runi	pine	medi	ladu
tagi	foli	sira	jupo	foli	runi	pine	catu	ladu
tagi	vidu	sira	jupo	vidu	runi	pine	gupe	ladu
tagi	supa	sira	jupo	supa	runi	pine	defa	ladu
tagi	pevo	sira	jupo	pevo	runi	pine	nigo	ladu
tagi	cilu	sira	jupo	ture	runi	pine	cilu	ladu
tagi	mego	sira	jupo	medi	runi	pine	mego	ladu
tagi	lofa	sira	jupo	catu	runi	pine	lofa	ladu
tagi	tadi	sira	jupo	gupe	runi	pine	tadi	ladu
tagi	nuso	sira	jupo	defa	runi	pine	nuso	ladu
tagi	pume	sira	jupo	nigo	runi	pine	pume	ladu

Table 2
Filler and non-rule phrases from the artificial language of the experiment. The target-word is shaded in grey. Each participant had always the same target word throughout the experiment.

Filler phrases			Non-rule phrases		
X	X	X	X	X	C
male	ture	cilu	male	cilu	sira
rosu	medi	mego	rosu	mego	sira
foli	catu	lofa	foli	lofa	sira
vidu	gupe	tadi	vidu	tadi	sira
supa	defa	nuso	supa	nuso	sira
pevo	nigo	pume	pevo	pume	sira
ture	cilu	male	ture	male	runi
medi	mego	rosu	medi	rosu	runi
catu	lofa	foli	catu	foli	runi
gupe	tadi	vidu	gupe	vidu	runi
defa	nuso	supa	defa	supa	runi
nigo	pume	pevo	nigo	pevo	runi
cilu	male	ture	cilu	ture	ladu
mego	rosu	medi	mego	medi	ladu
lofa	foli	catu	lofa	catu	ladu
tadi	vidu	gupe	tadi	gupe	ladu
nuso	supa	defa	nuso	defa	ladu
pume	pevo	nigo	pume	nigo	ladu

throughout the task. Participants were instructed to press, for each phrase, the left button of the mouse as soon as the target-word appeared, or the right button if the target was not present in the phrase. The participants were not informed about the presence of rules. They had to respond within 1500 ms after the end of the phrase, otherwise the next trial was delivered.

The target word was always a C word and the three possible C target words were counterbalanced across participants. The target was always the same throughout the experiment for each participant (e.g. C_1). Therefore, from the three AXC structures only the phrases from one structure contained the target word, however, each A could be used to predict the presence or absence of the target (e.g. A_1 fully predicted the presence of the target C_1 while A_2, A_3 fully predicted its absence). Thus, the transitional probability between A and C was 1, between A and X was 0.083, and between X and C or X was 0.05 (i.e., 17 X in filler sentences and 3 Cs in rule sentences). In contrast, if position is considered in the probability computation, an X in the initial position predicted the absence of C. However because X can have 18 values the predictive value of each specific X was only 0.055.

The word-monitoring task allowed us also to manipulate attention. First, because the target-word would always appear in the final position, participants should orient after a few trials their attention to the final word in all phrases. In addition, because only one structure contained the target word (e.g. A_1XC_1), participants should progressively focus their attention to the specific A element (e.g. A_1), and thus the AXC structure predicting the target because all the remaining phrases (filler phrases and rule phrases that did not contain the target-word) led all to "No" responses. Therefore, the rules where the target was present were considered attended rules (target condition), whereas the other two rules were considered unattended rules (non-target condition).

Note that with this design we can also distinguish the prediction of the rule from the preparation of the response. X elements in initial position allow predicting the response but do not predict the specific item that will appear in the last position. A items allow to predict both the response and the rule both in the target and non-target conditions. Because both XXX and AXC non-target conditions require the same "No" response, differences between them in the learning phase can only be explained by facilitation due to rule learning.

2.2.2. Test phase
2.2.2.1. Online implicit test.
After the learning phase, an *online test* was performed with no break or any other indication distinguishing it from the learning phase (Fig. 1A). Participants continued to listen to rule phrases as previously but filler phrases were replaced by non-rule phrases (XXC) (Fig. 1B). Non-rule phrases were like the filler phrases but with a C in the final position (XXC, e.g. *male-fuse-sira*). Thus, the first two elements of the non-rule phrases were randomly assigned with the same constraints as the XXX phrases whereas the third element was always a C element from any of the three rule structures (Table 2). The introduction of non-rule

phrases followed the rationale of the classical serial reaction time tasks. If participants learned the specific dependencies between each initial and final element of the phrase (the AXC structure), participants should respond faster and commit fewer errors in rule phrases than in non-rule phrases. In the rule phrases (AXC) they could still predict the presence or absence of the target (e.g. C_1) based on the initial element (e.g. A_1), whereas in the non-rule phrases (XXC) the presence of the target could not be predicted because X predicts with the same probability the target (i.e. C_1) and also the other non-target Cs (i.e. C_2 and C_3). In contrast to the classical serial reaction time, here learning was not tested introducing violations to the rules in order to spare the predictions of rule phrases for the explicit judgment test after this phase.

2.2.2.2. Offline explicit test. To examine whether more explicit judgments on the rules learned differed as a function of the amount of attention paid during the acquisition, we directly asked the participants to discriminate phrases following the rules (rule phrases and new rule-phrases) from those violating the rules (violation phrases). Participants' explicit judgment of the rules was evaluated using a recognition test (Fig. 1B). The recognition test was administered after the rule-learning task on day 1 (Experiment 1) and at the beginning of day 2 before the rule-learning task (Experiment 2) (Fig. 1A). Presenting the task at the beginning of day 2 allowed us to test for changes in performance in the explicit test from day 1 to day 2 as a function of consolidation, preventing an additional exposure to the rule phrases before the test.

Participants were instructed to press the left mouse button for phrases that belonged to the pre-exposed language or the right button for phrases that did not. There was no maximum time to respond, but the participants were instructed to respond as fast as possible. Each phrase was delivered immediately after the participant's response to the previous phrase.

In order to test rule learning, we built new rule-phrases for each of the three rules. New rule-phrases consisted of each of the three "A_C" structures combined with the 6 words belonging to the X category that never appeared with that specific structure in the learning phase. Therefore these phrases followed the rules but the transitional probability between each element of the structure and this specific X was 0 in the learning phase.

Two types of rule violations were used following Endress and Bonatti (2007): (i) we tested violations of the dependency between the first and third element, in which the first and final word element were placed in the correct order but belonged to different rule structures (e.g., A_1XC_2, A_3XC_1). This measure allows us to see whether the participants learned the specific dependency from a certain A to its corresponding C; (ii) we tested also violations of order positions (category violations), in which the third and first words of a rule phrase were swapped with one another (possible structures: C_1XA_1, C_2XA_2, C_3XA_3; e.g., **sira**-male-**tagi**). This measure allows us to test whether participants learned the positional information of the categories of words A and C. Thirty-six different violation phrases were created: 18 dependency violations and 18 category violations.

Thus, from the whole pool of test phrases, half of them were used for Experiment 1 (day 1) and the other half in Experiment 2 (day 2) to avoid repetition effects. This resulted, for each day, on the presentation of 36 intermixed test phrases consisting in 9 rule phrases, 9 new rule-phrases, 9 dependency violations and 9 category violations.

3. Results

A Kolmogorov-Smirnov test revealed that reaction times followed a normal distribution in both Experiment 1 ($D(98) = 0.057$, $p > 0.1$) and Experiment 2 ($D(92) = 0.075$, $p > 0.1$) (Fig. S1), therefore raw reaction times with no transformation were

introduced in the analyses. Since we had a gender bias (19 females, 6 males in the sample), we initially introduced the gender factor in all the analyses (i.e. Learning and Test phases). Since the gender factor was never significant and did not interact with any of the other factors, the analyses reported here do not include this factor. In both experiments, accuracy in the word-monitoring task was almost at ceiling during learning [> 90%]. No differences between error rates in the different conditions were observed. Raw data can be accessed in the Supplementary data.

3.1. Experiment 1: Attention effects on learning (Day 1)
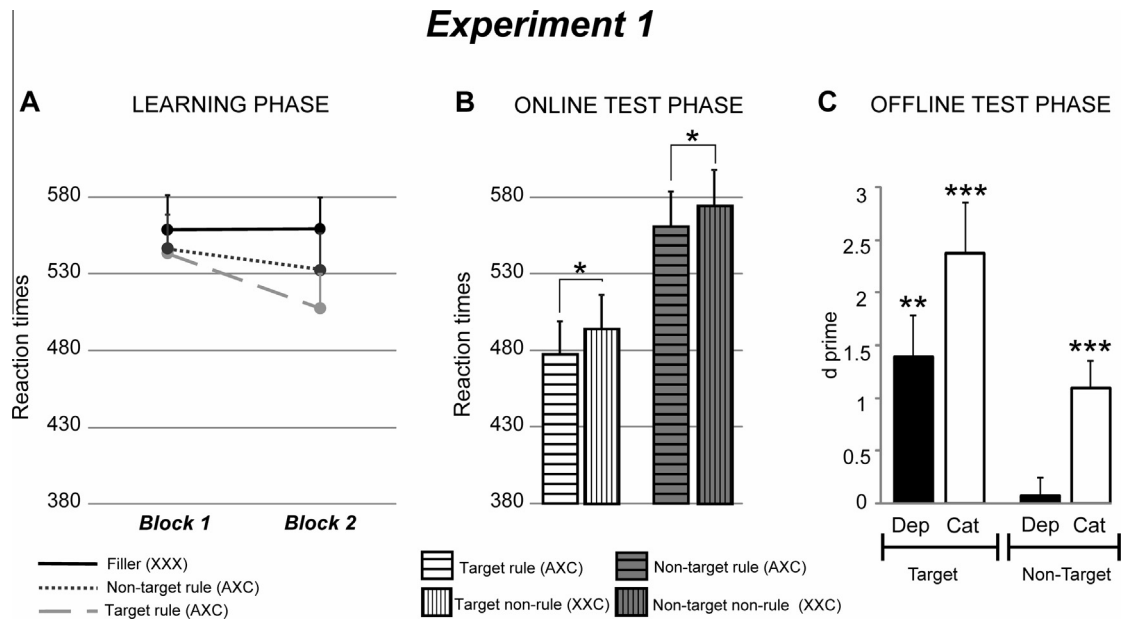
3.1.1. Learning phase

Reaction times were calculated from the onset of the third element (C) in the phrase. Reaction times from incorrect responses were not included in the analysis. Outliers, considered when reaction time was 2 standard deviations above and below the mean of each trial (e.g., the mean reaction time of the first AXC trial for all the participants, the mean reaction time of the second AXC trial for all the participants, and so on), were also removed (3.6% of all trials).

In order to analyze the learning effect, the learning phase was divided into two blocks, each block containing the mean reaction times of the first and second halves of the phase (block 1 and block 2 of the learning phase). A 3 (Type of phrase: target rule vs. non-target rule vs. filler) × 2 (block: 1 vs. 2) within-subject factors repeated-measures ANOVA was performed. We found a significant main effect of type of phrase ($F(2,42) = 3.35$, $p < 0.05$, $\eta_p^2 = 0.13$) and a significant type of phrase × block interaction ($F(2,42) = 4.11$, $p < 0.05$, $\eta_p^2 = 0.16$) (see Fig. 2A). Pairwise t-test comparison indicated that in block 1, all conditions had equivalent reaction times (all $ps > 0.3$). This occurred despite rule phrases containing the target being the only phrases corresponding to "yes" responses, and these are usually faster than "no" responses. In contrast, in block 2 a learning effect was observed. Participants were faster in the rule phrases compared to the filler phrases for both the target ($t(21) = 3.7$, $p = 0.001$) and non-target conditions ($t(23) = 2.6$, $p = 0.014$). No significant difference was found between the target and the non-target phrases ($p = 0.14$). These analyses showed that during the course of the learning phase, participants were able to take advantage of the predictability in the rule phrases. Importantly, that occurred both in the target and in the non-target phrases.

3.1.2. Online implicit test

As we did for the learning phase, mean reaction times were calculated for the first and second halves of the trials from the online implicit test, resulting in two blocks (block 1 and block 2 of the online implicit test) (see Fig. 2B). A 2 (rule: rule vs. non-rule) × 2 (attention: target vs. non-target) × 2 (block: 1 vs. 2) within-subjects repeated-measures ANOVA was performed. This analysis revealed that participants were faster in rule phrases than in non-rule phrases (main effect of rule ($F(1,23) = 4.9$, $p = 0.036$, $\eta_p^2 = 0.17$) and this occurred irrespective of the amount of attention (none of the interactions were significant; all $p > 0.1$). No significant effect of block was found ($p = 0.53$). Target conditions were faster than non-target conditions (main effect of attention: $F(1,23) = 40.7$, $p = 0.0001$, $\eta_p^2 = 0.63$). This result is of little interest since in this comparison target conditions correspond to "yes" responses that tend to be executed faster than "no" responses.

Then, we also proceeded to the analyses of the errors in the detection of the target. Since the number of errors was very low, we collapsed block 1 and 2 to perform the analysis. A 2 (rule: rule vs. non-rule) × 2 (attention: target vs. non-target) within-subject repeated-measured ANOVA revealed a main effect of rule ($F(1,23) = 7.62$, $p = 0.011$, $\eta_p^2 = 0.24$) indicating that participants committed

Fig. 2. Experiment 1. (A) Learning phase: reaction times from blocks 1 and 2. (B) Online test phase: average of the reaction times from blocks 1 and 2 in the online implicit test. (C) Offline test phase: participant's discrimination indexes (d prime) of the rules in relation to category and dependency violations. Dep: dependency violations; Cat: category violations. The bars represent SEM: standard error of the mean. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

more errors in the non-rule condition for both the target and the non-target conditions (Fig. S2). The main effect of attention was marginal (p = 0.056).

3.1.3. Offline explicit test

Discrimination indexes were calculated by transforming the percentage of correct responses into d prime scores (d') to control for response bias (MacMillan & Creelman, 2005). To determine the d' for each subject and condition, hits ("Yes" responses for correct items: rule and new-rule test items) and false alarms ("Yes" responses for incorrect items: dependency and category violations) were calculated. In order to better understand the type of knowledge acquired from the pre-exposed structures in target and non-target conditions, we calculated two different d'. First, we measured the ability of discriminating rule items from violations of dependency. To do that, we used as false alarms for the target conditions, the proportion of times that participants say "yes" to a violation of the dependency of the target rule condition (where the first A element was from the target rule); for the non-target condition, the false alarms were calculated as the proportion of times that participants say "yes" to a violation of the dependency of the non-target rule conditions (where the first A element was part of one of the two non-target rules). Second, we calculated the ability of discriminating rule items from violations of category items. For that, we used as false alarms for the target conditions, the proportion of times that participants responded "yes" to a violation of the category of the target rule (an item that started with a C element from the target rule); for the non-target condition, the false alarms were calculated as the proportion of times that participants responded "yes" to a violation of the category of the non-target rule conditions (an item that started with the C element of the non-target rules). Positive d' values indicate good discrimination and learning, values of d' close to 0 indicate no discrimination (see Fig. 2C).

The d' values were entered into a repeated-measures ANOVA with 2 (attention: target vs. non-target) × 2 (violation: dependency vs. category) within-subject factors. A main effect of attention (F(1,24) = 11.5, p = 0.002, $\eta_p^2 = 0.32$) and a main effect of violation (F(1,24) = 12.83, p = 0.002, $\eta_p^2 = 0.34$) were observed. No

significant interaction was found (p > 0.9). Nevertheless, in the target condition, post-hoc one-sample t-test against 0 confirmed that participants succeeded discriminating rule phrases from phrases with dependency violations (t(24) = 3.5, p = 0.002) and from phrases with category violations (t(24) = 5.01, p = 0.0001). In contrast, in the non-target condition, participants were able to discriminate the rule phrases only from phrases with violations of category (t(24) = 4.22, p = 0.0001). Their discrimination ability from phrases with a violation of dependency was not significant (p > 0.6). Therefore, while in the target condition participants were able to learn not only the position of the A and C category but also the specific dependency between them, this latter information was not learned from the rules without the target.

3.2. Experiment 2: Effect of attention on consolidation

3.2.1. Offline explicit test (day 2)

In order to look at the consolidation effects of the more explicit knowledge of the rules, the same ANOVA analysis that was applied to the offline explicit test in experiment 1, was performed on this same phase in experiment 2, which was carried out after a delay containing sleep. This test was administered at the beginning of the session in experiment 2 (see Fig. 1). The analysis revealed a main effect of violation (F(1,22) = 25.2, p = 0.0001, $\eta_p^2 = 0.53$) (Fig. 3A). Neither the main effect of attention nor the interactions between these factors were significant (both p > 0.1). Post-hoc one-sample t-test against 0 revealed that on day 2, participants were able to discriminate rule phrases from category violations in both the target (t(22) = 4.22, p = 0.0001) and the non-target conditions (t(22) = 3.09, p = 0.005). However they were not able to distinguish the rule phrases from phrases containing dependency violations neither in the target nor in the non-target conditions (both p > 0.3). In order to test the differences in consolidation we included the factor day into the ANOVA (day 1 vs day 2) and performed a t-test between the discrimination indexes in day 1 and day 2 for each condition. The interaction between day, attention and violation was marginally significant (F(1,22) = 3.42, p = 0.078, $\eta_p^2 = 0.13$). The decrease in performance observed from day 1 to 2 for the discrimination of dependency violations in target
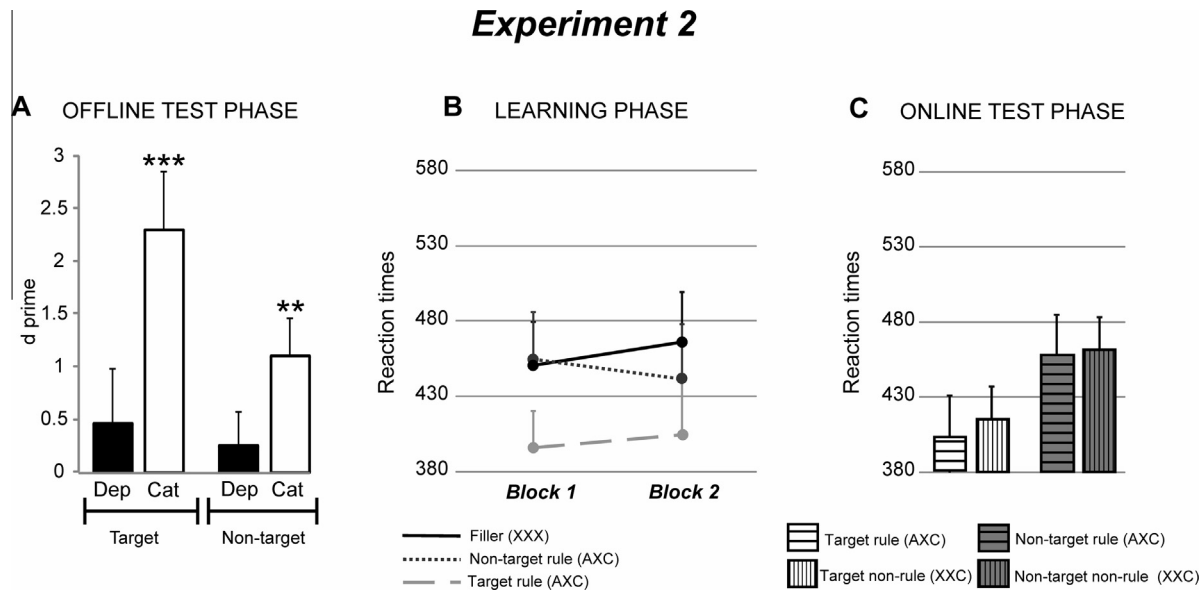
Experiment 2

Fig. 3. Experiment 2. (A) Offline test phase: participant's discrimination indexes (d prime) of the rules in relation to category and dependency violations. (B) Learning phase: reaction times from blocks 1 and 2. (C) Online test phase: average of the reaction times from the blocks 1 and 2. (C) Dep: dependency violations; Cat: category violations. Bars represent standard error of the mean. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

conditions was also marginally significant ($t(22) = 1.92$, $p = 0.068$). The remaining comparisons were not significant (all $p > 0.5$) (Fig. 3A).

3.2.2. Learning phase (day 2)

We looked at the pattern of learning progression after a delay containing sleep by performing the same ANOVA as the one performed for the learning phase on experiment 1. We found a main effect of type of phrase (filler vs. target rule vs. non-target rule) ($F(2,42) = 17.47$, $p = 0.0001$, $\eta_p^2 = 0.45$) (Fig. 3B). Neither block nor type of phrase × block interaction were significant (all $p > 0.3$), indicating that no further learning was observed. This effect was driven by overall faster reaction times in target rule phrases than in the other two (non-target rule and fillers) that did not differ from each other. Post-hoc *t*-test comparison between type of phrase collapsing by block, revealed that target rule phrases were faster than both filler ($t(22) = -5.5$, $p = 0.0001$ and non-target rule phrases ($t(22) = -6.2$, $p = 0.0001$), whereas reaction times from filler and non-target rule phrases did not differ ($p = 0.4$).

3.2.3. Online implicit test (day 2)

The same ANOVA performed for the online test of experiment 1 was applied to day 2 (Fig. 3C). It revealed a significant main effect of attention, which again may relate to the difference between "yes" and "no" responses ($F(1,21) = 16.9$, $p = 0.0001$, $\eta_p^2 = 0.44$), and a main effect of block ($F(1,21) = 4.51$, $p = 0.046$, $\eta_p^2 = 0.17$), where responses were faster on block 2 compared to block 1. The main effect of rule and the interactions were not significant on day 2 (all $p > 0.4$).

As we did for the experiment 1, we proceeded to the analysis of the errors in the detection of the target. After collapsing the block 1 and 2, a 2 (rule: rule vs. non-rule) × 2 (attention: target vs. non-target) within-subject repeated-measured ANOVA revealed a main effect of rule ($F(1,22) = 6.3$, $p = 0.02$, $\eta_p^2 = 0.22$). In addition, in contrast to experiment 1, we found here a rule × attention interaction (Fig. S3). A paired-samples *t*-test revealed that participants committed more errors only in the non-rule condition that carried the target word: target rule vs. target non-rule ($t(22) = 2.5$, $p = 0.02$); non-target rule vs. non-target non-rule ($t(22) = 0.6$, $p = 0.53$).

4. Discussion

In the present study, we sought to investigate how the amount of attention paid during the learning phase (1) affects the indirect implicit measures and more direct explicit judgments of the non-adjacent rule knowledge; and (2) whether it affects to the form in which rules are consolidated. Our online implicit measures during the learning phase revealed learning both for the target and the non-target rules, suggesting that learning occurred irrespective of the amount of attention. However when we directly asked for more explicit judgments in the recognition test we found differences in the type of knowledge acquired. Participants were able to learn both the positional information of the word categories and the specific dependency of the target rule ("attended") but only the former for the non-target ("unattended") rules. Nevertheless, only the knowledge on the category information turned to be long lasting 24 hours later in both target and non-target conditions.

Previous studies on language learning with artificial languages (Gómez, 2002; Peña et al., 2002; Saffran, Aslin, & Newport, 1996; Saffran et al., 1997; Toro et al., 2005) have used direct measures of learning administered after the learning phase. It is however notable that the authors used 2-alternative forced choice tests that are more sensitive to influences from implicit memory (Voss, Baym, & Paller, 2008; Voss & Paller, 2009). More direct measures using recognition tests have proved to be less sensitive to capture the underlying knowledge in this type of language learning paradigms (Kabdebon, Pena, Buiatti, & Dehaene-lambertz, 2015). Directly asking about one's knowledge of the rules has been considered to evaluate conscious and explicit knowledge (Dienes & Perner, 1999). However, knowledge may not be accessible through explicit measures in all the cases (Reber, 1967). For instance, prediction-based cognitive mechanisms can track the sequences (e.g. the non-adjacent relations in this study) and extract knowledge of the dependencies between the elements (as it occurs in serial reaction time tasks). Then, the knowledge regarding the exposed sequence is more sensitive to indirect (i.e. through measurement of reaction times) than direct tests, in which the participants are openly questioned about the sequence (Cleeremans, Destrebecqz, & Boyer, 1998; Jiménez, Méndez, & Cleeremans,

1996). The measurement of the reaction times during the cover word-monitoring task offered us the possibility to get an index that is more likely to reflect implicit rule knowledge. Although some explicit knowledge may influence reaction times, this learning measure was inherently more sensitive to implicit knowledge because the participants were not directly requested to inform about the rule and were merely performing a cover task in which performance could only improve if the rules were learned.

Explicit judgments were also collected after learning, using a recognition test. In this test, the participants were directly asked to judge whether a sentence belonged to the language previously learned. Although the answer to this direct question may also be influenced by implicit knowledge, the answer requires the participants' conscious access to the acquired information and therefore it is more likely to reflect a measure of explicit memory in comparison with the online indirect test. There are other measures that can be used to complement this type of direct test in order to measure explicit representations. For example, a recent study has used confidence ratings about the answers in a 2-alternative-forced choice in which they asked about the learned adjacent and non-adjacent dependencies structures in order to measure the explicit access to the knowledge (Romberg & Saffran, 2013). Our results add to this and other previous research (Batterink et al., 2015; Misyak et al., 2010a) showing that the combination of online measures and also different offline tests tapping more or less explicit knowledge increase our sensitivity to better comprehend the type of underlying knowledge acquired by the participants.

Our results show that participants were able to exploit the dependency between the elements forming the rules to perform faster the cover task, leading also to increased errors when the last word of the dependency appeared unexpectedly. Importantly, this occurred in both the target and in the non-target conditions. The ability of acquiring useful information about the underlying structure of the material incidentally has been previously shown in the visual domain (Chun & Jiang, 1998). Consistent with our results, serial reaction time tasks measuring implicit learning have shown that learning occurs even when attention is engaged in a secondary counting task (Frensch, 1998; Jimenez & Mendez, 1999). However, it has also been proposed that a minimal level of attention is required, even for implicit learning, to capture the relationship between non-adjacent elements (such as the elements presented in our task) (Pacton & Perruchet, 2008). Regarding this last point, it is worth mentioning that our manipulation does not imply that attention was completely removed in the non-target conditions. In fact, it is likely that initially, before learning the specific dependency (A-C), the participants rapidly noticed that the target was always located in the final position (in the C, given an AXC sequence). This prior learning might have then guided attention to that location (regardless of whether the sequence contained the target or not) and represented minimal attentional involvement to the non-target sequences that might have been sufficient to implicitly learn the rules. The participants then would have learned that a stimulus (i.e. the A element of the AXC rule) helped predict the appearance of the target (i.e. C), as it occurs in perceptual learning experiments (Seitz & Watanabe, 2008; Watanabe, Náñez, & Sasaki, 2001). The participants then might have focused on the elements of the target rules and ignored the others (non-target rules). Hence, the representation of the rule containing the target became more explicit and therefore more sensitive to the explicit test, whereas the non-target rules remained more implicit (which required less attention).

The fact that the focus of attention changes as learning occurs is related to, for example, Jiang and Chun's (2001) proposal of a bidirectional relationship between attention and previous experience. Thus, once a minimum degree of learning has occurred, the focus of attention is internally guided toward elements involved in that learning. Simultaneously, what is learned from that point on is affected by the amount of attention allocated to it. This scenario can be obviously observed in first language acquisition during early infancy. In this sense, child-directed speech, with its prosodic exaggerations, may guide attention to the important components of the speech signal that simplify the acquisition of words and rules from fluent speech (Dominey & Dodane, 2004). Once some learning has occurred, attention may be guided internally based on previous experience and does not have to rely only on salient external cues (De Diego-Balaguer, Martinez-Alvarez, & Pons, 2016).

One set of experiments that did not directly address the role of attention in learning provided evidence suggesting that attention may promote non-adjacent rule learning. In this sense, the importance of some cues presented in the stream could lie, at least partially, in their ability to attract attention. For example, Peña et al. (2002) showed that the extraction and generalization of AXC rules from fluent speech was only possible when short pauses were inserted between the AXC words during the training phase. Pauses, because of their natural salience, may automatically capture attention (De Diego-Balaguer et al., 2007, 2016) and help segmentation. Attention is then free to be allocated to the first and last elements of the segmented words to learn rule dependency (A-C) because given a sequence, attention tends to be allocated to elements from the start and end (Endress et al., 2009). On the other hand, Gómez (2002) showed that given a stream formed by successive AXC units, the degree in which the rules are learned and generalized depends on the variability of the variable element (X). Hence, in this case again, the system learns to ignore the variable elements and focus attention on the stable ones, helping learning of the non-adjacent rule.

Previous studies taking offline measures with 2-alternative forced choice tasks have addressed the role of attention in speech segmentation, which is a related and important aspect of language acquisition (Saffran et al., 1997; Toro et al., 2005). Saffran et al. (1997) administered a speech segmentation task to two groups of participants, children and adults, engaged in a cover task (i.e. a computer coloring program). Word learning occurred incidentally in both groups, although the performance when a concurrent task was administered was moderate compared to performance without interference (Aslin, Saffran, & Newport, 1998). The authors concluded that learning occurred even in the absence of focused attention to the language input. However, Toro et al. (2005), who also used the identical speech segmentation task but with three different attentional manipulations, concluded that some degree of attention was necessary to attain a certain level of word segmentation. In addition, Toro et al. (2011) studied the role of attention in rule generalization using a repetitive-based rule (AAB or ABA) that was different from the non-repetitive rule used in our study. The authors observed that attention had a different role in rule generalization depending on the underlying structure of the rules (attention was required to generalize non-adjacent structures but not adjacent structures). Because learning and generalizing non-adjacent structures is more difficult than learning adjacent structures (Newport & Aslin, 2004), the results suggest that the importance of attention in learning also depends on the degree of difficulty of the given task (see also Jimenez and Mendez (1999) for non-linguistic material). In our study it is interesting to notice that learning of the categories, that is the positional information of each word category (i.e. A initial, C final) was learned both in target and non-target conditions when an explicit judgment was required in the recognition test. In contrast, the specific dependency information required attention to be correctly recognized despite the online test indicated that participants learned the specific dependencies of the rules. This suggests that the amount of attention provided during the learning phase affects the way that information is stored and accessible. Our results

contribute evidence to how attention is necessary for the explicit access of the knowledge of non-adjacent rules, whose difficulty may be greater than for adjacent elements (Saffran et al., 1997; Toro et al., 2005) or repetitive-based rules (Toro et al., 2011).

In relation to consolidation, previous studies have observed that sleep qualitatively affects the consolidation of rules. Studies using serial reaction time tasks observed that the process of conversion of implicit to explicit knowledge was supported by sleep (Cleeremans, 2008; Wagner et al., 2004; Wilhelm et al., 2013). In the language domain, Gómez et al. (2006) exposed 15-month-old infants to an artificial language formed by three-word sentences following the structure AXC. The authors reported that only infants that napped after the learning session were able to generalize the learned structure to new material, thus suggesting that sleep promoted the formation of more abstract representations of the rules. Similarly, Tamminen et al. (2012) observed that when adults learned new affixes with an associate meaning embedded with existing words (e.g., buildnule, climbnule), the generalization of those affixes to new words only appeared after a sleep period (see also Merkx et al., 2011).

In our case, irrespective of attention, performance did not improve after a 24 h delay containing sleep. Nevertheless, it is interesting to notice that category learning, that is the positional information of each word category (i.e. A initial, C final), was more robust than learning of the dependencies, since it was maintained after sleep in both the target and the non-target conditions. In contrast, the specific dependency information required attention to be learned on day 1 and was forgotten after sleep. Because the learning of the specific dependencies was modest even in day 1, it could be the case that only more robustly learned information was consolidated (Wilhelm, Metzkow-Mészàros, Knapp, & Born, 2012). As an alternative or additional factor, it could also be that because the creation of categories is a more abstract learning, it was more prone to be consolidated after a short exposure. This result is similar to what Gómez et al. (2006) and Hupbach, Gomez, Bootzin, and Nadel (2009) reported in infants. The authors reported that only infants that napped after the learning session retained the abstract relation between the initial and final elements of the rule, thus they were capable of generalizing this abstract relation to similar but new stimuli, suggesting that sleep promoted the formation of abstract representations of the acquired knowledge. Even if our results are in the same direction (i.e. participants remembered the positional information but not the specific dependencies after a period of 24 hours containing sleep), we did not test participants with completely new material to assess if they were able to transfer the abstract knowledge to a completely new language. In addition, it would have been interesting to study the consolidation effect in the implicit measure, however the design of our study did not allow us to observe it because on the day 2, participants were tested with the offline explicit test before the word monitoring/language learning task (see Fig. 1A), thus the participants were previously exposed to the material that day. Nevertheless we did observe that after a 24 h delay containing sleep, reaction times continued to decrease overall with further exposure to the language. This improvement was not specific to the rules. Indeed no differences in reaction times between rule and non-rule conditions were observed in the online test and differences in error rates were only present in the target condition. This loss of learning compared to what we observed on day 1 was consistent with the explicit measures previously commented where only knowledge of positional information of the word categories appeared to be preserved. Because participants did not sleep immediately after learning, the interference received before the day 2 testing could have prevented the consolidation effects to arise (Hupbach et al., 2009; Talamini, Nieuwenhuis, Takashima, & Jensen, 2008). However, this would not explain why learning effects were not present

on day 2. Alternatively, the salience of the category violations in the offline explicit test on day 1 may have biased the information consolidated for day 2. This may have also biased attention to focus only on positional information in the online task that followed the offline explicit test on day 2. Despite the same attentional manipulation and same artificial language was used on day 2, we observed sensitivity to the dependencies only in the analysis of errors and only for the target condition on day 2 (Fig. S3). No differences were observed for the online measures of learning and test (Fig. 3B and C).

Summarizing, this study used an artificial language task, which allowed the online implicit measurement of learning of non-adjacent rules acquired throughout the session in addition to explicit offline learning measures. The present results showed that attention modulates the knowledge acquired from learning. Whereas incidental learning of the rules could be observed regardless of the amount of attention, the offline more explicit measures suggest that the structural information acquired from the rule exposure depends on the amount of attention paid during the acquisition. After a night of sleep, at least with the online and offline tests we have used in this study, we cannot conclude that the amount of attention affects differentially the consolidation processes.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.03.016.

References

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*(3), 170–176. http://dx.doi.org/10.1177/0963721412436806.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. a. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83*, 62–78. http://dx.doi.org/10.1016/j.jml.2015.04.004.

Brandon, M., Terry, J., Stevens, C. J., & Tillmann, B. (2012). Incidental learning of temporal structures conforming to a metrical framework. *Frontiers in Psychology, 3*(August), 1–10. http://dx.doi.org/10.3389/fpsyg.2012.00294.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71. http://dx.doi.org/10.1006/cogp.1998.0681.

Cleeremans, A. (2008). Consciousness: The radical plasticity thesis. In R. Banerjee & B. K. Chakrabarti (Eds.), *Models of brain and mind: Physical, computational and psychological approaches. Progress in brain research* (pp. 19–33). Amsterdam: Elsevier.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2*(10), 406–416. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21227256>.

Davis, M. H., Di Betta, A. M., McDonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803–820.

De Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology, 7*(February), 1–15. http://dx.doi.org/10.3389/fpsyg.2016.00044.

De Diego-Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule

extraction from speech. *PLoS ONE, 2*(11), e1175. http://dx.doi.org/10.1371/journal.pone.0001175.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114–126. http://dx.doi.org/10.1038/nrn2762.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *The Behavioral and Brain Sciences, 22*(5), 735–755. Discussion 755–808. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11301570>.

Dominey, P. F., & Dodane, C. (2004). Indeterminacy in language acquisition: The role of child directed speech and joint attention. *Journal of Neurolinguistics, 17*(2–3), 121–145. http://dx.doi.org/10.1016/S0911-6044(03)00056-3.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164–194. http://dx.doi.org/10.1093/applin/aml015.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*(2), 247–299. http://dx.doi.org/10.1016/j.cognition.2006.09.010.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13*(8), 348–353. http://dx.doi.org/10.1016/j.tics.2009.05.005.

Frensch, P. A. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47–104). Thousand Oaks, CA: Sage Publications.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*(8), 670–674. http://dx.doi.org/10.1111/j.1467-9280.2006.01764.x.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science (New York, N.Y.), 298*(5598), 1569–1579. http://dx.doi.org/10.1126/science.298.5598.1569.

Hupbach, A., Gomez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science, 12*(6), 1007–1012.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 54*(4), 1105–1124. http://dx.doi.org/10.1080/713756001.

Jimenez, L., & Mendez, C. (1999). Which attention is needed for implicit sequence learning ? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25*(1), 236–259.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 22*(4), 948–969.

Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*. http://dx.doi.org/10.1016/j.bandl.2015.03.005.

MacMillan, N., & Creelman, C. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Elbaum Associated, Publishers.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology, 64*(6), 1200–1220. http://dx.doi.org/10.1080/17470218.2010.538211.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010a). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*(September), 1–9. http://dx.doi.org/10.3389/fpsyg.2010.00031.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010b). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2*(1), 138–153. http://dx.doi.org/10.1111/j.1756-8765.2009.01072.x.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162. http://dx.doi.org/10.1016/S0010-0285(03)00128-2.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*(1), 80–96. http://dx.doi.org/10.1037/0278-7393.34.1.80.

Payne, J. D., Ellenbogen, J. M., Walker, M. P., & Stickgold, R. (2008). The role of sleep in memory consolidation. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 663–685). New York: Elsevier.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*, 604–607.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233–238. http://dx.doi.org/10.1016/j.tics.2006.03.006.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science, 37*, 1290–1320. http://dx.doi.org/10.1111/cogs.12050.

Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*(2), 101–106.

Seitz, A. R., & Watanabe, T. (2008). Is task-irrelevant learning really task-irrelevant? *PLoS ONE, 3*(11), e3792. http://dx.doi.org/10.1371/journal.pone.0003792.

Talamini, L. M., Nieuwenhuis, I. L. C., Takashima, A., & Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory, 15*, 233–237. http://dx.doi.org/10.1101/lm.771608.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125*(1), 107–112. http://dx.doi.org/10.1016/j.cognition.2012.06.014.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*(43), 14356–14360. http://dx.doi.org/10.1523/JNEUROSCI.3028-10.2010.

Toro, J. M., Sinnet, S., & Soto-Faraco, S. (2011). Generalizing linguistic structures under high attention demands. *Journal of Experimental Psychology. Learning, Memory, and Cognition 37*(2), 493–501.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*, B25–B34. http://dx.doi.org/10.1016/j.cognition.2005.01.006.

Voss, J. L., Baym, C. L., & Paller, K. A. (2008). *Accurate forced-choice recognition without awareness of memory retrieval*, 454–459. http://dx.doi.org/10.1101/lm.971208.and.

Voss, J. L., & Paller, K. a. (2009). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience, 12*(3), 349–355. http://dx.doi.org/10.1038/nn.2260.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature, 427*(6972), 352–355. http://dx.doi.org/10.1038/nature02223.

Watanabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*(6858), 844–848. http://dx.doi.org/10.1038/35101601.

Wilhelm, I., Metzkow-Mészàros, M., Knapp, S., & Born, J. (2012). Sleep-dependent consolidation of procedural motor memories in children and adults: The pre-sleep level of performance matters. *Developmental Science, 15*(4), 506–515. http://dx.doi.org/10.1111/j.1467-7687.2012.01146.x.

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience, 16*(4), 391–393. http://dx.doi.org/10.1038/nn.3343.