# The role of phonotactics and pauses in the acquisition of morphosyntactic dependencies

IRENE CANUDAS GRABOLOSA

Submitted in partial fulfillment of the requirements for the degree of

M.Sc. in Brain and Cognition

Department of Information and Communication Technologies
Universitat Pompeu Fabra

Supervisor:
Dr. Luca L. Bonatti

September 2016

upf. Universitat
Pompeu Fabra
Barcelona

## Abstract

A longstanding question in cognitive psychology is how the rules and the words of language are acquired. Peña et al. (2002) argued that learners are endowed with separate acquisition mechanisms dedicated to either extract words from a continuous speech or project generalizations from the structure of such words. To support their claims, Peña et al. (2002) presented results showing that when participants listen to a continuous artificial stream, they can extract items based on long-distance transitional probabilities between syllables, but cannot extract word structure. However when they are presented with even subliminally segmented streams, word structure easily appears.

This thesis has been challenged in several ways. Recently, Frost and Monaghan (2015) argued that two separate mechanisms do not need to be postulated. These authors argued that Peña et al. (2002) results were due to a confounding: their test items presented conflicting information, asking participants to unlearn introduced probability relations induced by the stream with which they were familiarized. When such confounding is removed, they argued, one single mechanism can learn both words and rules. Here, we examined the possible role of phonotactics in Frost and Monaghan's 2015 own material. Four experiments controlling some crucial phonotactic aspects of the material (three with a continuous stream and one with a segmented stream) with Catalan native participants were created.

Unlike what Frost and Monaghan (2015) reported, no learning of abstract regularities was found in the first two experiments, but participants performed above chance when familiarized with a segmented stream, in line with Peña et al. (2002) original results. Our data suggest that non-adjacent structural dependences could not be learned if the segmentation problem was not previously solved (i.e. if non-adjacent regularities were not introduced by segmentation signs). These findings directly contradict the ones from Frost and Monaghan (2015) and suggest a different approach to adjacent and non-adjacent structural regularities.

## Acknowledgements

I would like to thank all the friends who, without asking anything in exchange, took part of all the experiments. In the same way, I'd like to have some words for the ones who did earn something... but used it all in the train ticket to come to the lab. My family was also part of some experiments, and although some were quite afraid of even using the computer, all helped me. I need to mention also my two flatmates, who were the 'naive observers' in making sure the stimuli were well done. Thanks for not throwing me out from home after that. A particular 'thank you' goes to my brother, who was always ready to participate in 'anything you need!'.

I have to acknowledge the help from Lluís de Yzaguirre, who lent me the materials from his doctoral thesis (De Yzaguirre i Maura (1991), managed by the 'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra' (Barcelona)).They were very useful for the second experiment. I need to have a word also for some classmates, like Juana, who found me half of the people who participated in Experiment 2 in a record time. Ana, from the lab group, also helped me in the searching.

Writing that thesis was not easy. I couldn't have done it without my parents' help, in taking care of me, quite a lot, while I was writing.

For all of them, thank you!

# Contents

# 1 Introduction

The way in which infants acquire their mother tongue has been a topic for discussion over the last 50 years. New-born humans have to face with an apparently chaotic language, sometimes more than one. In their process for decoding, they encounter three major problems before assigning meaning to words, according to Jusczyk (2000).

In a first stage, infants need to deal with an extremely variable signal. As speech is produced by people of different sizes and mouth shapes, in different conditions and speaking rate, the differences have an impact on the acoustic characteristics of speech.

Their second problem is segmentation. Natural languages do not contain the equivalence to blank spaces in writing (Cole et al., 1980), therefore, a continuous speech signal has to be first segmented in order to assign, afterwards, a meaning to the segments.

Finally, infants need to understand the relations between items that are present in the language, which constitute the core of grammar. An example of such relations, important for this study, are the so called morphosyntactic dependencies; remote dependencies or a relation between one item and another irrespective of the intermediary elements (Perruchet et al., 2004). A morphosyntactic dependency can be found, in English, between the 3rd person singular and the '-s' in present (*she* speak*s*).

## 1.1 Segmentation

In this study, we will focus on segmentation and the identification of abstract relations between segmented items. Several hypothesis have been advanced to explain about how segmentation, the second problem, is solved. Saffran et al. (1996a) argued that transitional probabilities (TP) between syllables may be fundamental. These statistic computations can be defined as the likelihood that one element predicts another. Consequently, in natural languages, TPs tend to be lower between syllable pairs which straddle word boundaries (Aslin et al., 1998), although they are not equally strong in all languages (Saksida et al., 2016).

Consistent with their hypothesis, Saffran et al. (1996a) found that 8-months-old children were able to segment words from a 2 minutes-speech-continuum on the basis of TPs between syllables.The authors used syllables CV forming trisyllabic words that were put together forming a speech continuum, artificially synthesized. TP within words was 1 (i.e. every time syllable 'A' appeared, it was always followed by syllable 'B'), and between words, 0.33 (1 out of 3 times syllable 'A' appeared, it was followed by 'B'). Infants preferred words to part-words (trisyllabic items spanning over word boundaries) after a very short familiarization, suggesting that they had correctly segmented the stream.

These results have been replicated in adults (Saffran et al., 1996b) and using natural, rather than artificial languages (Pelucchi et al., 2009). Other studies from Saffran et al. (1999) demonstrated that the calculations were part of a general learning mechanism, as statistics could be successfully implemented on stimuli other than language, such as musical tones.

However, transitional probabilities are not the only cue to solve the segmentation problem. Several experiments have shown that transitional probabilities can be affected by prosody, which facilitates or overwrites statistical computations (Saffran et al., 1996b; Monaghan et al., 2007; Shukla et al., 2007; Christophe et al., 2008; Monaghan and Christiansen, 2010; Endress and Hauser, 2010; Ordin and Nespor, 2013). Other computations, such as using high frequency words (Monaghan and Christiansen, 2010), may also play an important role in the segmentation of the speech signal.

Even though transitional probabilities seem to be just one computation in the equation, these calculations are important and may depend on the existence of a universal mechanism.

However, the strength of the mechanism has not been established. One possibility is that it is, indeed, very strong. If it were the case, it would also be possible that the same mechanism could be used to solve part of the third problem, the learning of morphosyntactic relations (Bates and Elman, 1996).

## 1.2 Acquisition of morphosyntactic dependences

In potential contrast with this possibility, Peña et al. (2002) carried out several experiments which suggested that morphosyntactic dependences could not be learned by the same mechanisms as words are segmented. The authors created an artificial stream of trisyllabic words distributed in three families of three items. They followed the structure 'AxC', where the first syllable 'A' predicted the appearance of the last syllable 'C'. Between 'A' and 'C' was, therefore, a non-adjacent TP, a relation corresponding to a simplification of a morphosyntactic dependence: the link between 'A' and 'C' in their design could be be seen either as the detection of a structural dependency between different long-distance items (as in 'she' and '-s' in the example from above) or as the identification of *specific* words defined by that special dependency. To disambiguate between these interpretations, Peña et al. (2002) introduced the notion of 'rule-words'. A rule-word was an item that did not appear during the familiarization but had a structure which was congruent with the generalization of the pattern 'AxC' to the stream.

Peña et al. (2002) discovered that after exposure to a continuous stream participants could not distinguish between part-words (following a structure 'xCA' or 'CAx') and rule-words ('AxC' structure but not appeared before); even though they had no problems in the distinction of words ('AxC' structures that had previously appeared in the stream) against part-words. However, when the stream was even minimally segmented (that is, when they inserted a 25ms pauses between words) participants easily grasped the structure of rule-words. Their conclusion was, therefore, that non-adjacent regularities could be traced even from a continuous stream, but the *structure* induced by such regularities could be identified only if the stream had been previously segmented. That same conclusion was achieved when event related potentials were used to examine subjects' response (Mueller et al., 2008).

Further studies revealed other aspects of the tracing of non-adjacency and structural dependency. Onnis et al. (2004) hypothesized a relation between the variability of the material and learning of the structure. For the authors, the more variable the better the learning, suggesting that material had to be interpreted as encoding a rule in order for non-adjacent relations to be computed.

In an interesting study, Newport and Aslin (2004) found that participants could not compute the structure of the stimuli upon syllables but that they were able to do so upon vowels or consonants. [1] From that conclusion it was proposed that the structure of the stimuli could be computed following the Gesthalt principles of similarity. Their hypothesis is supported by the work from Creel et al. (2004) and Endress (2010), who also demonstrated that these regularities can be traced in music, outside linguistic material.

Another important characteristic was stated by Marchetto and Bonatti (2015), who discovered a developmental difference in tracing of the structure and non-adjacent dependence. While twelve-month-old infants showed an adult-like behavior in the computation of structural and non-adjacent dependences, seven-month-old infants failed in non-adjacency tracking, although they were able to trace several statistical relations. Some other studies, like de Diego-Balaguer et al. (2016) tried to establish a relation between development of infants' capacity of attention and the learning of structural dependences.

However, one of the most exhaustive studies on the behavior of non-adjacent and structure tracking was carried by Endress and Bonatti (2007, 2015). Adapting the experimental design from Peña et al. (2002) by changing the length of the familiarization, they found different timebases for adjacent and structural tracing. On the basis of these results and combining them with parallel studies about 'rule learning' (learning the abstract relationship between classes and not between items), the authors proposed that participants were not constrained to non-adjacent statistics (which might occur in a

---

[1] These results have been replicated several times, with the difference that the tracing of structural patterns on consonants has not been reported again. Instead, a growing body of literature has defined the *CV hypothesis*, according to which rules are carried on vowels and segmentation, on consonants (Nespor et al., 2003; Bonatti et al., 2005).

continuous stream in restricted situations) but they could extract a regularity entailing syllable classes.

Both computations, according to Endress and Bonatti (2007), would not be computed in the same way but solved by at least two different mechanisms; the first one for segmenting the stream and the second one for rule generalization. That hypothesis was named 'More than One Mechanism' (MOM). It goes in line with the one from Marcus et al. (1999), who argued that statistical learning operated only at the level of surface statistics, whereas rule learning was doing so at a deeper level, involving abstract patterns.

Their findings are also supported by the work from de Diego-Balaguer et al. (2011), who, by studying brain electrophysiological activity (ERPs), found different activity in the brain depending on whether participants were learning non-adjacent dependences (named 'words') or the structure ('rule-words').

An alternative line or research has been focusing on the explanation of non-adjacent tracking by means of simple statistic computations. Most of the work has centered in the role of pauses and other statistical regularities in Peña et al. (2002). Perruchet et al. (2004) argued that Peña et al. (2002) contained some methodological drawbacks that were creating statistical regularities other than non-adjacent dependences, thus allowing participants to segment using the well-studied adjacent dependences. For Laakso and Calvo (2011), adding pauses to the stream did not allow participants to compute non-adjacent TP but to use the pauses to compute adjacent TPs. A similar conclusion was the one reached by Aslin and Newport (2012), who also advocate for a salience effect, responsible of the success of Peña et al. (2002)'s subjects in tracing non-adjacent dependences. These authors also proposed the context to be playing a role, by making the subjects decide when to apply patterns to several elements that occurred interchangeably and when only to individual elements.

Finally, Vuong et al. (2016) investigated the learning of adjacent, non-adjacent dependences and learning about the structure using a serial reaction time task along with some behavioral tests. While the performance of their subjects in the behavioral tasks was as bad as in previous studies, reaction times were showing that participants could actually trace the structure of the stimuli.

In that line of reinterpretation of Peña et al. (2002), a recent work by Frost and Monaghan (2015) is especially interesting. They proposed an alternative explanation of the results from Peña et al. in which TP would be sufficient to simultaneously account for segmentation generalization from dependencies between words in sequences. According to the authors, the failure of participants to trace non-adjacent structural relations in a continuous stream is due to a violation of the learning structure. Because Peña et al. (2002) test items were creating inserting 'A' or 'C' syllables in the 'x' position, participants would be forced to unlearn the dependency relations for the moved syllable along with the learning of the non-adjacent structural relation (see example in table 1).

| **Familiarization items** | **Words** | **Rule-words** |
|---|---|---|
| PUraKI | PURAKI | PUbeKI |
| BEliGA | BELIGA | BEtaGA |
| TAfoDU | TAFODU | TApuDU |

**Table 1** – Example of the words used in the familiarization and in the test phase in Peña et al. (2002). The words had the same syllables as presented in the familiarization, while rule-words followed the same structure but the second syllable was extracted from the first or the third syllables of the other stimuli.

Following this hypothesis, the 25 ms gaps would be providing an additional cue to relative positions of elements of the language in speech, thus allowing success when they are inserted.

The authors tested their hypothesis with an experiment with three conditions: *segmentation, moved-syllable generalization* and *novel-syllable generalization.* The first condition tested non-adjacency to be sued for segmentation in words, the second one replicated Peña et al. (2002) experiment in a continuous stream and, in the last one the same replication was carried out but with all 'x' syllables being novel syllables which never appeared in the stream. The reasoning was that, by using syllables that had not appeared before, the conflict with the unlearning of the relations of dependency would not appear; therefore it would be possible to generalize the structure to rule-words. The results, which can be seen in figure 1, showed that participants were able to segment and to generalize with novel syllables but they performed about chance when there was a moved-syllable generalization.

If Frost and Monaghan (2015) are right, they would have a strong case against the MOM hypothesis, given the fact that the difference between the failure in rule learning and the success in segmentation when computed on the same familiarization time-frame and material are two is the strongest arguments supporting it. However, there are some reasons to suspect that this is not the case.

Several authors have reported the higher level of difficulty in the acquisition of non-adjacent dependences in comparison to adjacent dependences (Newport and Aslin, 2004; Endress, 2010). This suggests that, as for segmentation, several mechanisms may be playing a role.

Phonology is likely to be among them, as its process of acquisition starts very early,

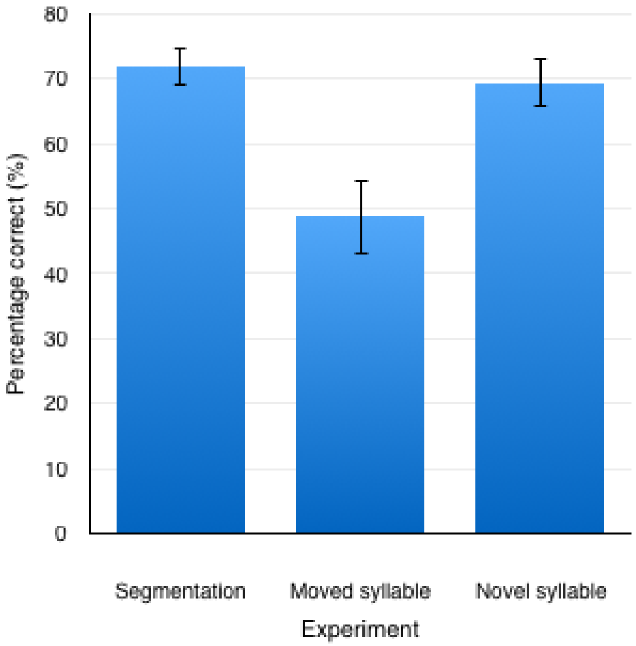


**Figure 1 – Results from the experiments from Frost and Monaghan (2015).** The experiment labelled 'segmentation' was used to test the tracing of non-adjacent dependences in a continuous stream, and the second experiment was used to test the learning of the structural relations, replicating the results from Peña et al. (2002). In the third experiment, in order to avoid the violation of the learning structure, a novel syllable was used in the second position of the rule-words. Error bars indicate standard error.

even in utero (Moon et al., 2010). Because of this fact, it seems reasonable that some phonetic-based mechanism is in place when infants start dealing with different types of dependences and structural regularities.

More importantly, phonology has been extensively shown to shape statistical calculations when these are applied to language. Both universal traits such as pitch (Endress and Hauser, 2010) and language-specific characteristics (Yang, 2004; Saksida et al., 2016) are agreed to be crucial for statistics applied to segmentation. Monaghan et al. (2007)'s Phonological-Distributional Coherence Hypothesis, tested in English, Dutch, French and Japanese, even stated a relation between phonological and distributional cues such as when one system is less reliable, the other is stronger.

For morphosyntactic relations, the issue has also been tested. Onnis et al. (2005) studied the importance of phonological cues in non-adjacent structural regularities tracing. Their results are a warn about the possibility to find confounds because of phonotactics of subjects' native language.

Interestingly, Frost and Monaghan (2015) reported a control for 'possible preferences for particular dependencies between syllables not due to the statistical structure of the sequences', by creating eight versions of the language, randomly assigning syllables to 'A' and 'C' roles. However, that did not control for the basic fact of having all 'A' and 'C' syllables starting by a stop.

### 1.3 Aim of the study

Because of the recognized importance that language-dependent mechanisms in tracing of both adjacent and non-adjacent dependences (see section 1.1 and section 1.2), we asked whether the results of Frost and Monaghan (2015), could be biased because of the phonotactics of the subjects's language. By testing this possibility we are also further investigating about the relationship between non-adjacent dependences and phonology.

This relationship is specially important. If it turned out that the extraction of non-adjacent dependences is independent from phonology, then the thesis that universal statistical mechanisms could account for morphosyntactic learning would be reinforced. Indeed, if in the results of Peña et al. (2002), learning occurs after the conflict about unlearning information acquired during the familiarization is removed, then this would also favor a 'one mechanism' approach to language acquisition: one same mechanism would be sufficient to learn both words and rules.

By contrast, if that is not the case, the 'more than one mechanism' hypothesis will gain weight, as it will demonstrate that abstract elements cannot be extracted by statistical computations alone.

In the first experiment we reproduce Frost and Monaghan (2015), using the same stimuli but adapting them to Catalan. The second experiment was prepared with stimuli in which the type-frequency for syllables was controlled, so to test the importance of phonology in these computations. Experiment 3 was meant to control for the effects of familiarization in the first two experiments. Finally, Experiment 4 was done to compare the results from the previous experiments 1-3, in which the familiarization was a continuous artificial stream, with those that can be obtained when the stream contains pauses.

## 2 Experiment 1

Experiment 1 was created as a reproduction of Frost and Monaghan (2015). Methods were as similar as possible to this study, with the necessary adaptations to the Catalan language (substitution of some non-existing sounds and rearrangement of some syllables to prevent the stimuli to be too similar to Catalan).

## 2.1 Participants

Participants were twenty-two Catalan native speakers (8 males, 14 females), age ranging 18- 26 (mean 21.81). All of them spoke Spanish and English, and some a fourth language. They received €5 for participating.

## 2.2 Methods

### 2.2.1 Materials

For comparision with previous results, as well as with those by Frost and Monaghan (2015), we selected stimuli as similar as possible to those by Peña et al. (2002), adapted to Catalan. Stimuli were trisyllabic 'AxC' words where the first syllable 'A' was predicting the appearance of the last syllable C. Familiarization words were grouped in three families, the first and last syllable of them being /pu/ - /ki/, /be/ - /ga/ and /ta/ - /du/. The middle syllable (the 'x' syllable) could be either /li/, /ra/ or /fo/ for the familiarization stream. Thus, all stimuli maintained the same phonological cue as in both Frost and Monaghan (2015) and Peña et al. (2002), in which the first and the last syllable started with a plosive consonant and the middle syllable started with a liquid.

**Familiarization** Participants were exposed to a 10 minutes familiarization stream, composed by concatenating the 'AxC' words from the language. No word from the same family or containing the same middle syllable was immediately repeated. Therefore the transitional probabilities TP between any 'A' and 'x' or between 'x' and 'C' were 0.33, while TPs between 'A' and 'C' were always 1. Because a word from the same family could not follow, the TP between any 'C' and 'A' was 0.5. The stream had a 5 seconds fade in and out to prevent participants from using the onset or offset as a cue.

| **pu_ki** | **be_ga** | **ta_du** |
|---|---|---|
| puliki | beliga | talidu |
| puraki | beraga | taradu |
| pufoki | befoga | tafodu |

**Table 2** – Stimuli 'families' of words presented during the habituation stream

**Test** The test consisted on an auditory two-alternative forced-choice task in which participants had to choose between a word (so following an 'AxC' structure) and a part-word (a trisyllabic part of the habituation stream ignoring word boundary). Words were created in the same way as items from habituation were, but this time 'x' being a novel syllable. The complete list of words and part-words can be found in table 3.

| **words** | **part-words** |
|---|---|
| pumoki | fogane |
| puneki | fokiɲa |
| puɲaki | kitaɲa |
| bemoga | laduɲa |
| benega | lagamo |
| beɲaga | netari |
| tanedu | moberi |
| tamodu | mopuri |
| taɲadu | nedube |

**Table 3** – Stimuli words and part-words presented during the test period

As in Frost and Monaghan (2015), part-words corresponded to strings that had (partially) occurred in the stream but ignoring word boundary. Therefore half of the part-words followed the 'CAN' pattern ('N' standing for 'novel syllable') or an 'NCA' pattern, in which 'x' was substituted by a novel syllable, like app the part-words from Frost and Monaghan (2015). However, unlike them, in our design part-words could also exchange 'A' or 'C' for a novel syllable, thus following a pattern 'xCN' or 'NAx'. The structure of the part-words was the same as the one used by Peña et al. (2002) and Endress and Bonatti (2007).

The speech was synthesized with the MBROLA speech synthesizer (Dutoit et al., 1996) using the Brazilian Portuguese database BR3. The Brazilian Portuguese sounds were used as they were judged to be the most similar to Catalan and they sounded clear and natural enough for experimentation. All selected phonemes also exist in Catalan and phonotactics of the language were taken into account so all the stimuli could be possible words in Catalan. However, for further control, two naive controls double-checked that no word sounded like any existent real word in Catalan. Words had a mean length of 696 ms, each phoneme lasting approximately 116 ms and each syllable, 232 ms. They had a pitch of 200Hz.

### 2.2.2 Procedure

Participants were tested individually in a silent room to avoid distractions. They were informed that they would listen to an imaginary language and then they would have to choose which of the words they would be presented was from the language. The experiment was controlled by a Macbook Pro laptop running Psyscope X (Cohen et al., 1993), and the stimuli delivered through headphones. Subjects had to pass a control test first, an auditory two-alternative forced-choice task, in which they were asked to recognize the syllable [so] by pressing a red or green key on the keyboard. Participants who scored less than 80% correct in this part were excluded later in the analysis. When they finished the control task, the familiarization period started. After its end, the test phase followed.

The total duration of the experiment was about 20 minutes.

Participants were presented 36 test pairs in a pseudo-randomized order to avoid words being always presented in the same position or the comparison of consecutive pairs. They had to choose the item that was more likely to be part of the language they had heard during the habituation period and to guess if they were not sure. Items were presented with a 1.5 s interval, and between the response of the participant and the new trial there was an interval of 2 seconds.

## 2.3 Results

The percentage of the accuracy of the subjects is shown in figure 2. Performance was slightly higher from chance (M= 54.29; SD=9.26; t(21)= 2.17, p= 0.04, two-tailed t-test).

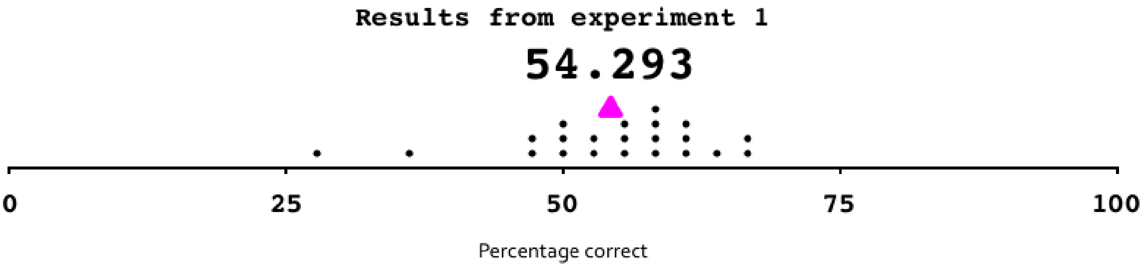


**Figure 2 – Results from Experiment 1.** Means of individual subjects are shown by the black points; while the pink triangle shows the mean of the subjects.

An univariate ANOVA for family of words (pu_ki, be_ga and ta_fu) showed no statistical significant effects for family, F(2,11)=0.14, p=0.87.

## 2.4 Discussion

Experiment 1 was designed to reproduce the the results obtained by Frost and Monaghan (2015), who introduced a novel syllable in 'x' position in rule-words. As shown from the results (see section 2.3), also our participants preferred words to part-words, although we still do not know why.

However, the rate of success in Catalan participants was quite lower than that of English participants in Frost and Monaghan (2015). An independent-sample t-test was conducted to compare these two results. English participants in Frost and Monaghan (2015) performed better than the Catalan participants in Experiment 1(M=69.3; SD=16), t(38)=3.71, p=0.0007, two-tailed t-test. Two possible reasons for this fact lie on the only two main changes that were introduced for the study: the difference in mother tongue of the subjects (English for Frost and Monaghan (2015) versus Catalan for our subjects) or the fact that the procedure was slightly different, as just one version of the language was created for this replication, while the study from Frost and Monaghan contained 8 different versions.

Although both English and Catalan participants succeed, the sharp difference in results suggests that other processes, beyond statistical tracing of non-adjacent dependences, are at work when the abstract regularity is extracted. As proposed above (see section 1.3) a possibility is that the phonology or phonotactics of the two languages contribute to the result.

Because all stimuli followed the same phonological stop-sonorant-stop pattern, phonological regularities which perhaps have a different impact in the two languages, may have facilitated the learning (as for Onnis et al. (2005)).To address this issue, a second experiment was conducted.

## 3 Experiment 2

In Experiment 2, we controlled for the type-frequency syllables in Catalan. The aim was to get a set of stimuli that, if the plosive-liquid-plosive phonological pattern contributed to extract the rule-words in Experiment 1, favored instead the part-words. If participants then preferred part-words instead of words, it would be an important argument to propose that the results from Frost and Monaghan (2015) were due to phonotactics rather than to a violation of the learning structure. ideally, this material would control for position-frequency (i.e. the amount of times a syllable appears in a position in the real use of the language) of syllables in part-words and words.

However, we had no frequency database available for Catalan to perform these computations. The closest control we could enforce was based on type-frequency (i.e. the amount of times one syllable appeared in all the words from a Catalan dictionary). These frequencies were extracted from the DICSYL corpus, derived from the doctoral thesis of Lluís de Yzaguirre (De Yzaguirre i Maura, 1991), managed by the 'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra' (Barcelona). The obtained frequencies can be observed below in figures 3 and 4, both for relative (i.e. the total amount of times one syllable occurred in a position by the total number of appearances of the syllable) and absolute frequencies (i.e. the total amount of times one syllable occurred in a position by the sum of totals of all the considered syllables). Syllable frequency was found by Messer et al. (2015) to have an effect on infants' verbal memory, in a way that all of them tended to learn better if syllable frequency was observed.

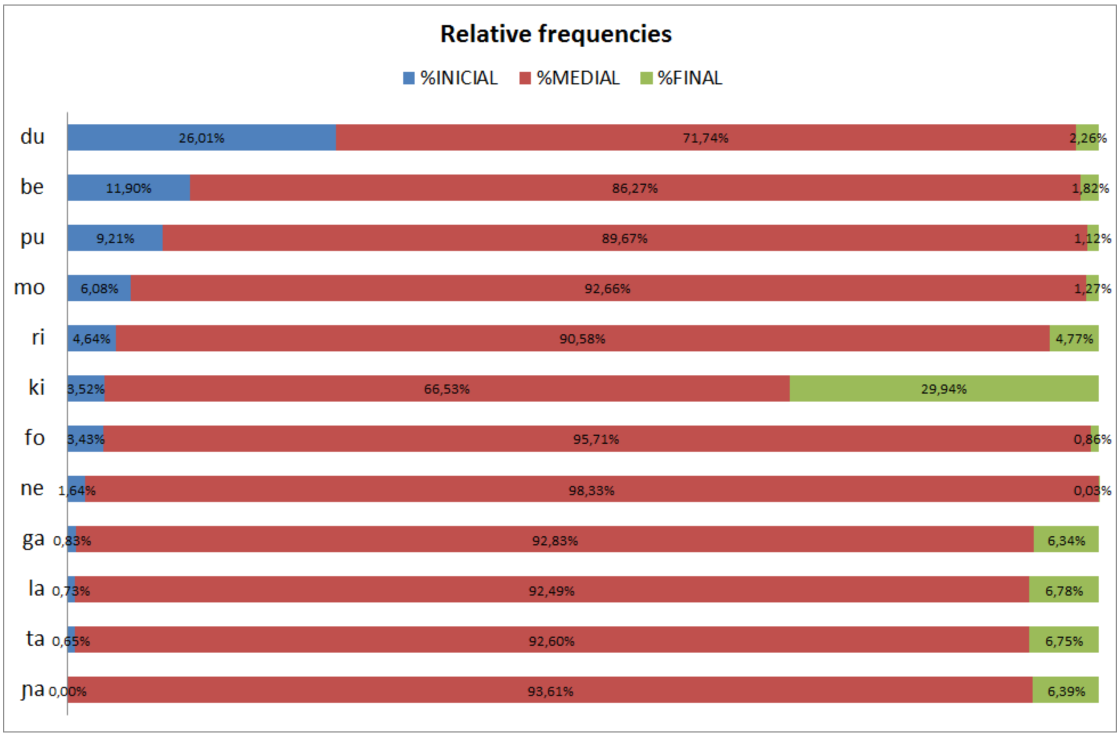


**Figure 3** – Relative frequencies for the observed syllables

## 3.1 Participants

Participants were twenty-three Catalan native speakers (8 males, 15 females), but one male was excluded because of failure on the training part. For the remaining subjects, age ranged 19- 46 (mean 23.08). All of them spoke Spanish and English, and some a fourth language. They received €5 for participating.

## 3.2 Methods

### 3.2.1 Materials

The stimuli had the same structure and syllables as in the first experiment. However, they were created taking into account the type frequency. In order to bias participants

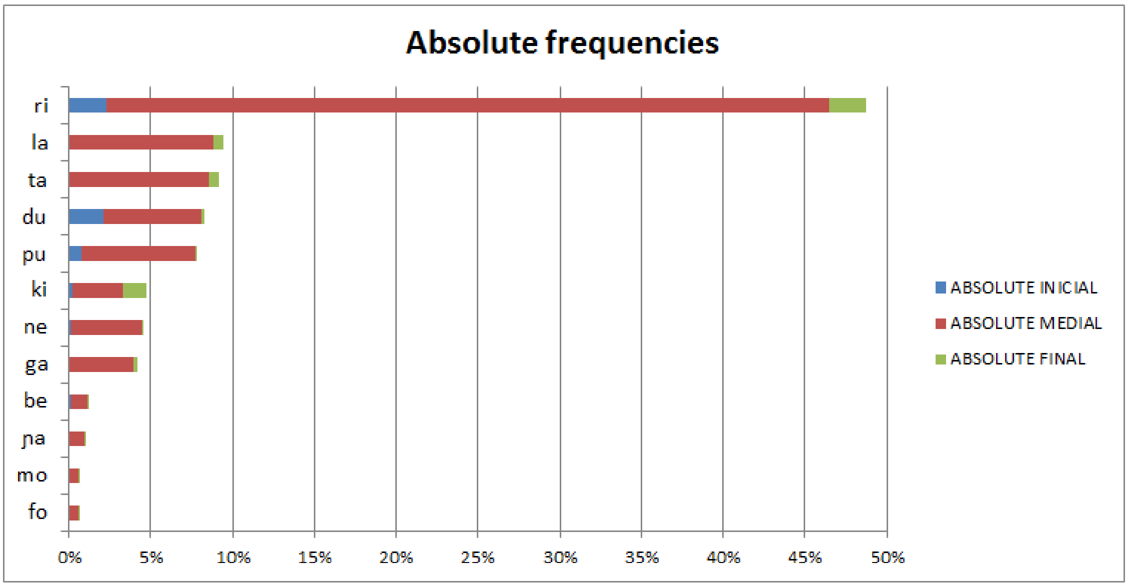


**Figure 4** – Absolute frequencies for the observed syllables

towards 'part-words', these had a better structure than 'words'. That structure was obtained in the following way: the most frequent syllables in first position were used in first position for part-words but never for words and the same was done for most frequent syllables in medial and final position. Stimuli can be found below (tables 4 and 5). Two naive controls were used to check no word was too similar to any real Catalan word.

**Familiarization** Subjects were exposed to a 10 minutes familiarization stream, with exactly the same characteristics as in Experiment 1 (see section 2). However, the stream was formed with the words presented in table 4.

| **la_ga** | **ta_fo** | **ne_mo** |
|---|---|---|
| laɲaga | taɲafo | neɲamo |
| lariga | tarifo | nerimo |
| lapuga | tapufo | nepumo |

**Table 4** – Stimuli 'families' of words presented during the habituation stream

**Test** The test was also created in the same fashion as in Experiment 1. The words that were used are the following:

| lwords | part-words |
|---|---|
| lakiga | pugadu |
| laduga | pufobe |
| labega | fonebe |
| tadufo | ɲamobe |
| tabefo | ɲagaki |
| takifo | kilari |
| nebemo | kitari |
| nedumo | dumola |
| nekimo | duneri |

**Table 5** – Stimuli words and part-words presented during the test part

### 3.2.2 Procedure

The procedure was identical to the one from Experiment 1 (see section 2.2.2).

## 3.3 Results

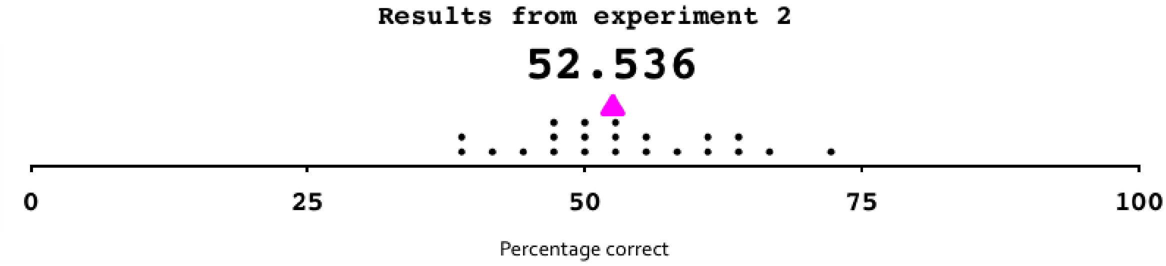


**Figure 5** – **Results from Experiment 2.** Means of individual subjects are shown by the black points; while the pink triangle shows the mean of the subjects.

Results are shown in figure 5. Performance did not differ from chance (M= 52.52; SD= 9.69; t(21)= 1.222, p=0.24 , two-tailed-t-test). A comparison between the results of Experiment 1 and Experiment 2 was performed by means of a two-sample t-test. Experiments did not differ from each others (Difference between means, -1.76; t(42)= -0.62, p=0.54).

A one-way ANOVA was computed to see possible differences for family of words (la_ga, ta_fo and ne_mo). There was no statistical significant effect for this factor (F(2,11)=1.84, p=0.16).

### 3.4 Discussion

Experiment 2 was created to test the role that phonotactics of Catalan could play in the extraction of structural regularities, with the material used in Experiment 1. In experiment 2, whose construction was identical to experiment 1 but with part-words in principle favored the type-frequency of syllables in Catalan, participants did not differ from change. This suggests they did not learn to extract the rules, but answered at random in the test phase. Indeed, we analyzed the words from Experiment 1 using the same principles extracted by the dictionary provided by Lluís de Yzaguirre. For this criterion, they showed to be 'better-formed' than part-words, therefore expected to be preferred if type positional frequency plays a role in the extraction of acceptable patterns. This contrasted with the words of Experiment 2.

Even though these results seem to confirm the importance of phonetics of the mother tongue on learning of structural patterns in an artificial language, the weak effect was unpredicted. Stimuli were prepared to favor part-words; hence participants were expected to prefer part-words over words, which was not the case. Moreover, results from Experiment 1 and 2 did not differ statistically. We would have predicted a stronger effect, although the very weak success in Experiment 1 was making it difficult to obtain it. Overall, considering the random response found in Experiment 2 and the weak effect of Experiment 1 the results suggest that in Catalan the role of the phonological pattern plosive-liquid-plosive is very small. Based on these results, the role of phonotactics on extracting regularities seems to be weaker than the hypothesized.

We should not loose from sight, however, that we also found a difference between the results in Frost and Monaghan (2015) and the other experiments performed here (see section 2.4), there must be some fact that can account for it.

Experiment 2 was constructed under the hypothesis that increasing positional type-frequency in their constriction would create a bias for part-words. This manipulation was only the best we could do given the Catalan dictionaries available. However, the source of the material had two important problems.

First, token-frequency (or the real use of a syllable in a corpus and in spoken language) was not contemplated. Although there is a relation between type and token frequencies, they hold differences (Berg, 2014). Potentially, type-frequency is not so relevant for Catalan speakers.

Second, items were constructed taking into account the absolute position of syllables (i.e. being an initial, medial or final syllable) but not their *relative* place (i.e. following or preceding a specific syllable). Further research, which necessarily goes beyond that point is necessary to address these factors.

A third possibility, which we could control within the limits of the current work, concerns baseline preference for the experimental items. Perhaps, despite our principles of construction, there was no baseline preference for part-words, independently of the effect of the familiarization. To explore this possibility, we constructed Experiment 3.

# 4 Experiment 3

Because the results of Experiment 1 and 2 did not differ, although the latter was not different from chance and the former was, we asked whether the results could be affected by some baseline preference. Therefore, in Experiment 3 we tested the test couples of Experiment 1 and 2 without familiarization, so to see if the participants of the study were learning something in the basis of the familiarization, or just reflecting the preference for the specific items we chose in the test phases, presumably for some prior bias induced by their general language experience.

## 4.1 Participants

Participants were forty-four Catalan native speakers (18 males, 26 females), age ranging 15- 77 (mean 39.54). All of them spoke Spanish and most of them a third and fourth language.

## 4.2 Methods

### 4.2.1 Materials

The stimuli were the ones used in the test part of Experiment 1 (see table 3) and Experiment 2 (see table 5).

### 4.2.2 Procedure

Participants were split in two groups, one exposed to stimuli from Experiment 1 (we will call the part of the experiment Experiment 3.1). The second group (Experiment 3.2) was exposed to the stimuli of Experiment 2. No group was exposed to any familiarization. Otherwise, the methods and materials were identical to those of Experiments 1 and 2.

## 4.3 Results

The percentages of the subject's responses in accuracy are shown in figures 6 and 7.

In Experiment 3.1 participants were not different from chance (M=50.37; SD=9.16; t(21)= 0.19, p=0.85 ,two-tailed-t-test). Also, their performance did not differ from that of participants in Experiment 1 (difference between means= 3.91; t(42)= 1.41, p= 0.17, 2-sample t-test).

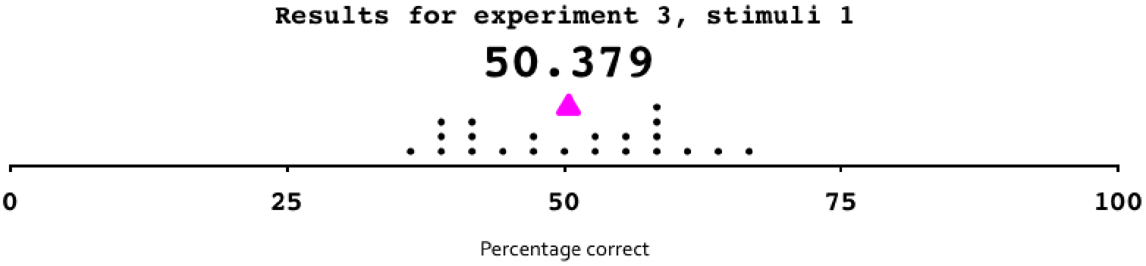


**Figure 6 – Results from Experiment 3.1, stimuli from Experiment 1** Means of individual subjects are shown by the black points; while the pink triangle shows the mean of the subjects.

To control for the effect of family of words, a univariate ANOVA was performed, with family of words as a factor (see the families in table 3). Results revealed a significant statistical effect for family (F(2,11)=16.82,p=0.0001). Post-hoc analysis, with a Sheffé test, showed a rejection of family be_ga (36.36% correct) over pu_ki (59.84% correct), dif=0.23, serr=0.04, p=0.00001; and over ta_du (54.2% correct), dif=0.19, s.err=0.04, p=0.00009; but there was no significant statistical difference between pu_ki and ta_du (dif=-0.05, s.err=0.04, p=0.51).

Also in Experiment 2.3 participants were no different from chance(M=54.80; SD=15.81; t(21)= 1.42, p=0.17 ,two-tailed-t-test). Nor did their performance differ from that of Experiment 2 (difference between means= -2.27; t(42)= -0.57, p= 0.57, 2-sample t-test).

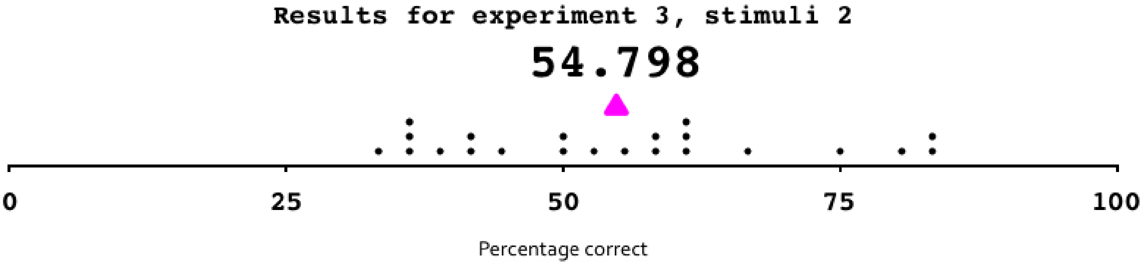


**Figure 7 – Results from Experiment 3.2, stimuli from Experiment 2.** Means of individual subjects are shown by the black points; while the pink triangle shows the mean of the subjects.

An univariate ANOVA for family of words (la_ga, ta_fo and ne_mo) was conducted. Results demonstrated significant effects for family, F(2,11)=7.86, p=0.0004. Posterior analysis (Scheffé test) demonstrated a preference for la_ga (64.4% correct) over ta_fo(51.9% correct), dif=-0.12, s.err=0.043,p=0.01; and over ne_mo (48.1% correct), dif=-0.16, s.err=0.043, p=0.0008. No statistical difference was found between the last two families (dif=-0.04, s.err=0.04, p=0.67).

## 4.4 Discussion

The aim of this 'double' experiment was to establish a baseline to check the effect of familiarization in experiments 1 and 2. We found that the results for Experiment 1 and 2 did not differ from the baselines established here. That is, we found no learning due to the familiarization occurred in any of the two experiments.

In other words, according to performance of subjects in Experiment 3, statistical computations were not helpful for participants to learn the non-adjacent structural rules.

There was, in fact, no learning, contrary to the thesis advanced by Frost and Monaghan (2015). Perhaps, the only effect of familiarization was to create a normalization of the responses. As reported, preference for words in Experiment 3 depended on the family of words. For Experiment 3.1, two families ( pu_ki and ta_du ) for Experiment 1 and one for Experiment 2 (la_ga) were clearly preferred over the others. That was not the case for Experiment 1 or 2. That familiarization only induces normalization of preferences, though is a far cry from the thesis that one single mechanism can account for all language learning.

Our results show that the stimuli from Experiment 2 did not have the expected bias, but a baseline preference level of about 56% for words against part-words. This was not the effect that we tried to trigger, and it confirms the importance of the two problems proposed in the discussion of the previous experiment (see section 3.4); most likely without taking into account the token-frequency nor the relative position of the syllables, it is difficult to direct participants' preferences for artificial tokens.

It is to be noticed that age proved to be an important indicator for success in 3.2: the older the subjects, the more prone to prefer part-words (exact calculations can be found in appendix A.3). Older participants were indeed sensitive to positional type-frequency in the familiarization as expected. Because the used source for the stimuli was created about 25 years ago (De Yzaguirre i Maura, 1991), a third possibility to partially explain the lack of induced preferences in Experiment 2 is that the databases must be relevant to the tested age.

Until now, our experiments found very minor evidence for learning of structural generalizations. We have only used continuous familiarizations, which, as per Frost's thesis, would be sufficient to induce learning of linguistic generalizations. How would Catalan participants behave if they were exposed to a segmented familiarization? If our Catalan participants are simply bad learners of rules in artificial languages, then we should also find poor evidence of learning in these conditions. If instead what counts for the acquisition of structural generalizations is the nature of the input, as Peña et al. (2002) argued, then we should find clear evidence of abstract learning in spite of the failures found until now. Experiment 4 tests these alternative options.

# 5 Experiment 4

## 5.1 Participants

Participants were twenty-three Catalan native speakers (10 males, 13 females), but one female was excluded due failure during training. For the remaining twenty-two, their age ranged 19-34 (mean 21.59). All of them spoke Spanish, most of them also English and some a fourth language. They received €5 for participating.

## 5.2 Methods

### 5.2.1 Materials

The stimuli that were used for this experiment were the same as the ones used in Experiment 2 (see section 3.2.1) [2]. However, in the familiarization stream, a 25 ms gap was inserted between words, following the same procedure as in Peña et al. (2002) and Endress and Bonatti (2007, 2015). According to these authors, these gaps are not noticeable for the participants.

### 5.2.2 Procedure

The followed procedure was the same as in Experiment 2 (see section 3.2.2). The length of the familiarization and the number of trials of the test remained unchanged, the only difference were the pauses in the familiarization stream. rule-

## 5.3 Results

The percentage of the accuracy of the subjects is shown in figure 8. Performance was robustly higher than chance (M=76.97; SD=10.01; t(21)= 26.00, p=0.0001,two-tailed-t-test). These results were higher than the baseline preferences tested in Experiment 3 (difference between means= 14.86; t(42)= 3.06; p=0.0045, 2-sample t-test). They were also higher than the results of Experiment 2, obtained after a continuous familiarization (difference between means= -18.81; t(42)= -5.48; p=0.0001, 2-sample t-test).

---

[2] Even though the material design was proved to fail in its principal function, stimuli 2 were still carefully designed stimuli that seemed to adapt better to Catalan than stimuli from Experiment 1 did (which were the adapted stimuli from Frost and Monaghan (2015).

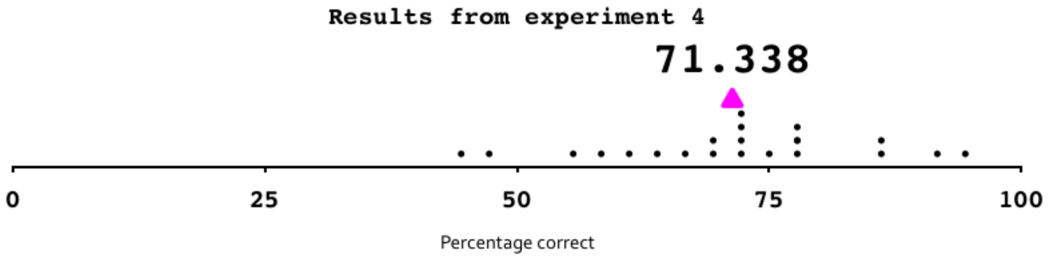


**Figure 8 – Results from Experiment 4.** Means of individual subjects are shown by the black points; while the pink triangle shows the mean of the subjects.

A one-way ANOVA was computed to see possible differences for family of words (la_ga, ta_fo and ne_mo). There was no effect (F(2,11)=1.87, p=0.15).

### 5.4 Discussion

Experiment number 4 was intended to check whether participants could learn non-adjacent abstract rulesafter familiarization with a minimally segmented stream. Results show that participants learned the rules and that they were able to perform better than they did in Experiment 2 and 3.

Furthermore, familiarization caused normalization. No preference for family of words was observed, as happened in Experiment 2 (but not in Experiment 3).

These results confirm the ability to learn non-adjacent rules from segmented streams while showing that learning from unsegmented streams seems much more difficult (in this case, it was not even found).

## 6 Conclusions

Which mechanisms are in role to understand morphosyntactic structures in the language? Since 2002 there has been an important debate whether these non-adjacent structural rules can be learned using the same underlying mechanisms as used in the cracking the speech signal. Researchers (Peña et al., 2002; Newport and Aslin, 2004; Endress and Bonatti, 2007; Balaguer et al., 2007; Mueller et al., 2008; Marchetto and Bonatti, 2015; Endress and Bonatti, 2015) have found evidences that the behavior of the mechanism used in tracking structural dependences differs from statistical computations in a variety of fields (age of acquisition, input needed or timebase).

Not everybody agrees in that conclusion, as there is a growing body of literature defending the possibility to learn such structural relations using statistical computations (Laakso and Calvo, 2011; Aslin and Newport, 2012; Vuong et al., 2016).

One of these last claims in favor of using the same mechanism for segmentation and morphosyntactic learning is the one from Frost and Monaghan (2015). According to the authors, it is possible to learn structural dependences in the same conditions as for non-adjacent dependences. For them, the failure of the participants from Peña et al. (2002) was due to requiring them to learn the generalization of non-adjacency while unlearning a dependency relation for the moved syllable (see section 1.2). Because the impossibility to use the same timebase and material in rule-learning and segmentation are two of the most important arguments in support of the MOM hypothesis; the results of Frost and Monaghan (2015) are a real challenge for considering a need of two mechanisms.

In this study, the claims from Frost and Monaghan (2015) were examined and used to test the possible role of phonology and mother tongue in the solving of non-adjacent statistics.

Four experiments were conducted, to test several crucial phonotactic aspects of the material were created. Two of them used a continuous stream, a third did not have familiarization as it was used as a baseline for 1 and 2 and the forth had a segmented stream familiarization. In all of them, we tested Catalan native speakers.

The percentages of correct responses of all the experiments are summarized in figure 9, where also the results from Frost and Monaghan (2015) are presented.

According to these results, it seems that non-adjacent dependences could not be learned if the segmentation problem was not previously solved (i.e. if non-adjacent structural regularities were not introduced by pauses). These findings directly contradict the ones from Frost and Monaghan (2015), who achieved positive results with the same procedure used in experiments 1 and 2.

Because non-adjacent structural dependences were not learned using the same material that could be used in a segmentation task (that is a continuous stream of words), it seems that the approach to structural regularities is different from the computation of adjacent regularities.

The nature of such mechanisms is not that clear. However, as stated in the introduction (see section 1.2) several studies have found that same impossibility to use the same material for statistical learning and tracing structural dependences (Peña et al., 2002; Newport and Aslin, 2004; Endress and Bonatti, 2007; Mueller et al., 2008; Marchetto and Bonatti, 2015; Endress and Bonatti, 2015). Actually, the nature of the stimuli that needs to be used resembles the material used for the learning of some other computations, such as center-embedded structures, that need to be marked by prosodic cues to be exploited (Mueller et al., 2010); or the detection of repetitions in the structure (Endress et al., 2005), which has been shown to use a Gestalt-like operation that has been related to the tracing of non-adjacency structural dependency (Aslin and Newport, 2012).

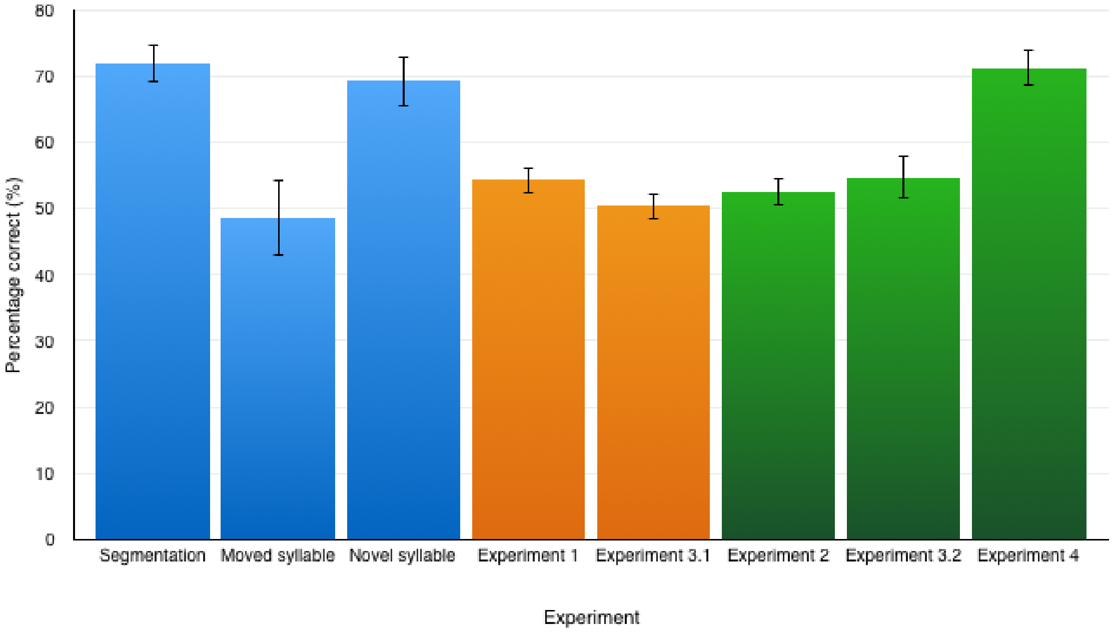


**Figure 9 – Results from all the experiments along with the results from Frost and Monaghan (2015).** The results from Frost and Monaghan are shown in blue, results using stimuli from Experiment 1 are shown in orange and the ones in which stimuli from Experiment 2 were used are in green. Error bars indicate the standard error.

It is not completely true, however, that nothing was learned in experiments 1 and 2. The results from these experiments, compared with those from Experiment 3, seem to show a normalization of the preferences. This is an interesting result, as it demonstrates some learning. However, it would be interesting to see whether this phenomenon is not due to external factors, such as age. Because of the difference of ages in Experiment 1, 2 and 4 compared with the third, it is not impossible that this had some importance in the responses, like it had for Experiment 3.2, thus creating a false effect of normalization.

By comparing the results from these experiments and the ones from Frost and Monaghan (2015) it is seen that phonology is important for the learning of structural delendences. However, that fact could not be proven entirely, as we tested participants in type-frequency, which was not ideal. For a good reaffirmation of the results found in this study, further work would be necessary to see the reaction of participants with a lower baseline preference for the stimuli when they are presented both in continuous and segmented streams. Going back to Frost and Monaghan (2015), it would be also interesting to check whether their participants' performance depended on the baseline preference based on linguistic knowledge, like it was the case in our experiments 1 and

2. If their results were achieved independently of the preference baseline, the role of mother tongue, and by extension, phonology, would be undiminishable to explain such differences. On the other way, similar results would confirm the work presented here.

This study did not entirely provide evidence for a very strong role of phonology. Instead, we showed that no learning of structural regularities was achieved with statistical computations. However, tracing of non-adjacent structural dependences is not simple. Even though the results of this study advocate for a different mechanism than segmentation to explain this operation, they do not want to deny the importance of statistical calculations, which at seem to affect the performance of the subjects by creating a normalization of their responses (but see above). It is highly possible, therefore, that the learning of such regularities in real language uses all the mechanisms to which infants have access, like transitional probabilities, rule-learning, phonotactics or already-known elements.

# A Appendix: Further analysis of the results

## A.1 Experiment 1

In order to determine whether the response data from subjects in Experiment 1 was in a normal distribution, the Kolmogorov Test was used (t=0.145; p=0.74).

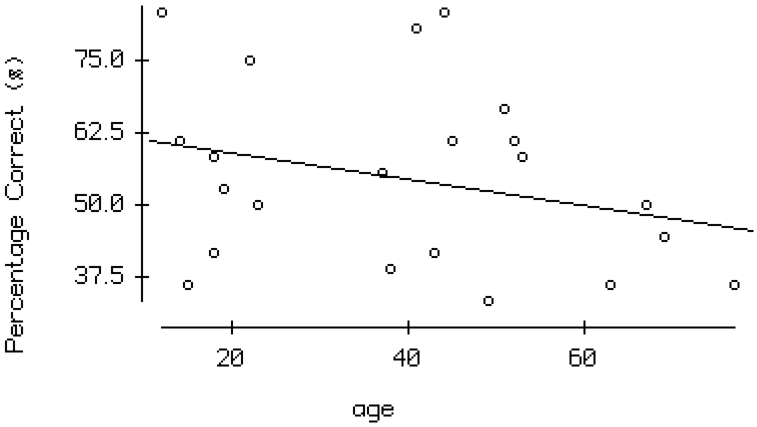
## A.2 Experiment 2

The Kolmogorov Test was used to determine the response data from subjects was in normal distribution (t=0.116; p=0.91).

## A.3 Experiment 3

The Kolmogorov Test was administered to determine that the data response data from subjects from both sets of stimuli was in normal distribution (stimuli from Experiment 1: t=0.117; p-value=0.92; stimuli from Experiment 2: t=0.147; p=0.72).

We run a regression analysis for age and response correct for experiment 3.2. The analysis demonstrated a weak regression between percentage correct and age (see figure 10l Percentage correct descended with age.



| Source | Sums of Squares | df | Mean Square | F-ratio |
|---|---|---|---|---|
| Regression | 15197.3 | 1 | 15197.3 | 69.1 |
| Residual | 173793 | 790 | 219.991 | |

| Variable | Coefficient | SE(Coeff) | t-ratio | p-value |
|---|---|---|---|---|
| Intercept | 63.7984 | 1.204 | 53.0 | 0.0001 |
| age | -0.227598 | 0.0274 | -8.31 | 0.0001 |

**Figure 10** – Regression between percentage correct and age. R squared = 8.0%; R squared (adjusted) = 7.9%; s = 14.83 with 792 - 2 = 790 degrees of freedom. Percentage correct seems to descend with age.

## A.4 Experiment 4

The normal distribution of the response data from subjects was ensured through the Kolmogorov Test (t=0.126; p=0.872).

1.25

# Bibliography

Aslin, R. N. and Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current directions in psychological science*, 21(3):170–176.

Aslin, R. N. and Newport, E. L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language learning*, 64(s2):86–105.

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

Balaguer, R. D. D., Toro, J. M., Rodriguez-Fornells, A., and Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One*, 2(11):e1175.

Bates, E. and Elman, J. (1996). Learning rediscovered. *Science*, 274(5294):1849.

Batterink, L. J., Reber, P. J., Neville, H. J., and Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of memory and language*, 83:62–78.

Berg, T. (2014). On the relationship between type and token frequency. *Journal of Quantitative Linguistics*, 21(3):199–222.

Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations the role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6):451–459.

Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and speech*, 51(1-2):61–75.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2):257–271.

Cole, R. A., Jakimik, J., and Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America*, 67(4):1323–1332.

Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3):356–371.

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, memory, and cognition*, 30(5):1119.

de Diego-Balaguer, R., Fuentemilla, L., and Rodriguez-Fornells, A. (2011). Brain dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience*, 23(10):3105–3120.

de Diego-Balaguer, R., Martinez-Alvarez, A., and Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in psychology*, 7.

De Yzaguirre i Maura, L. (1991). *Estructura sil·làbica del català central.* Departament de Filologia catalana Universitat de Barcelona.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Van der Vrecken, O. (1996). The

MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on Spoken Language Processing*, volume 3, pages 1393–1396. IEEE.

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta psychologica*, 135(2):182–190.

Endress, A. D. and Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105(2):247–299.

Endress, A. D. and Bonatti, L. L. (2015). Words, rules, and mechanisms of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(1):19–35.

Endress, A. D. and Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, 61(2):177–199.

Endress, A. D., Scholl, B. J., and Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, 134(3):406.

Frost, R. L. and Monaghan, P. (2015). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147:70–74.

Gonzalez-Gomez, N. and Nazzi, T. (2015). Constraints on statistical computations at 10 months of age: the use of phonological features. *Developmental science*, 18(6):864–876.

Jusczyk, P. W. (2000). *The discovery of spoken language*. MIT press.

Laakso, A. and Calvo, P. (2011). How many mechanisms are needed to analyze speech? a connectionist simulation of structural rule learning in artificial language acquisition. *Cognitive science*, 35(7):1243–1281.

Lany, J. and Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19(12):1247–1252.

Lany, J., Gómez, R. L., and Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31(3):481–507.

Marchetto, E. and Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of child language*, 42(04):873–902.

Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398):77–80.

Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., and Leseman, P. P. (2015). Growth of verbal short-term memory of nonwords varying in phonotactic probability: A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language*, 84:24–36.

Monaghan, P. and Christiansen, M. H. (2010). Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *Journal of child language*, 37(03):545–564.

Monaghan, P., Christiansen, M. H., and Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive psychology*, 55(4):259–305.

Moon, C., Lagercrantz, H., and Kuhl, P. K. (2010). Phonetic learning in utero. *The*

*Journal of the Acoustical Society of America*, 127(3):2017–2017.

Mueller, J. L., Bahlmann, J., and Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*, 20(5):892–905.

Mueller, J. L., Bahlmann, J., and Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive science*, 34(2):338–349.

Nespor, M., Peña, M., and Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, 2(2):203–230.

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance i. statistical learning of non-adjacent dependencies. *Cognitive psychology*, 48(2):127–162.

Onnis, L., Monaghan, P., Christiansen, M. H., and Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th annual conference of the Cognitive Science Society*, pages 1047–1052. Lawrence Erlbaum Mahwah, NJ.

Onnis, L., Monaghan, P., Richmond, K., and Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53(2):225–237.

Ordin, M. and Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning*, 63(4):800–834.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child development*, 80(3):674–685.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593):604–607.

Perruchet, P., Tyler, M. D., Galland, N., and Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133(4):573.

Romberg, A. R. and Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):906–914.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44(4):493–515.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, 274(5294):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(1):27–52.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of memory and language*, 35(4):606–621.

Saffran, J. R. and Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, 4(2):273–284.

Saksida, A., Langus, A., and Nespor, M. (2016). Co-occurrence statistics as a language-dependent cue for speech segmentation. *Developmental science*.

Sandoval, M. and Gomez, R. L. (2013). The development of nonadjacent dependency learning in natural and artificial languages. *Wiley Interdisciplinary Reviews: Cognitive*

*Science*, 4(5):511–522.

Shukla, M., Nespor, M., and Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54(1):1–32.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008). Finding words and rules in a speech stream functional differences between vowels and consonants. *Psychological Science*, 19(2):137–144.

Vuong, L. C., Meyer, A. S., and Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning*, 66(1):8–30.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in cognitive sciences*, 8(10):451–456.