In: J. Cole & J. Hualde (eds.) *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 315-338

Learning the mapping from surface to underlying representations in an artificial language

Sharon Peperkamp and Emmanuel Dupoux

Abstract

When infants acquire their native language they not only extract language-specific segmental categories and the words of their language, they also learn the underlying form of these words. This is difficult because words can have multiple phonetic realizations, according to the phonological context. In a series of artificial language-learning experiments with a phrase-picture matching task, we consider the respective contributions of word meaning and distributional information for the acquisition of underlying representations in the presence of an allophonic rule. We show that on the basis of semantic information, French adults can learn to map voiced and voiceless stops or fricatives onto the same underlying phonemes, whereas in their native language voicing is phonemic in all obstruents. They do not extend this knowledge to novel stops or fricatives, though. In the presence of distributional cues only, learning is much reduced and limited to the words subjects are trained on. We also test if phonological naturalness plays a role in this type of learning, and find that if semantic information is present, French adults can learn to map different segments onto a single underlying phoneme even if the mappings are highly unnatural. We discuss our findings in light of current statistical learning approaches to language acquisition.

1. Introduction

Early language acquisition is a very complex task, because infants have to acquire a great many aspects of their language simultaneously. Yet, they do so with amazing speed. Within the first year of life, infants have been shown to acquire, among others, the segmental categories of their language (Kuhl et al. 1992; Polka and Werker 1994; Werker and Tees 1984), and to start to segment the continuous speech stream into words (Jusczyk and Aslin 1995). There is evidence that for these aspects of language acquisition both adults and infants can exploit algorithms based on a bottom-up acoustical analysis of the input speech (Saffran, Aslin and Newport 1996; Saffran, Newport and Aslin 1996; Maye and Gerken 2000; Maye, Werker and Gerken 2002). However, there is another aspect of early language acquisition that has not yet been investigated: infants have to acquire the underlying forms of words and morphemes. This is difficult, because languages typically display a number of phonological processes that obscure the relationship between underlying representations and their surface phonetic realizations.

Adults have been shown to map surface word forms onto underlying ones in the case of both allophonic (Lahiri and Marslen-Wilson 1991) and non-allophonic (Gaskell and Marslen-Wilson 1996) variation. Given these findings, it is important to examine how infants can learn to infer underlying representations. Consider the case of allophony. Both semantic and distributional information provide evidence as to the presence of allophonic rules. Regarding semantics, the presence of allophony can cause words to have more than one segmental make-up, depending upon the phonological context. For instance, French has an allophonic distinction between [ʁ] and its voiceless variant [χ]; the word *fleur* 'flower' is realized with a final [ʁ] in *fleur jaune* 'yellow flower' but with [χ] in *fleur pourpre* 'purple flower'. Knowledge that *fleu[ʁ]* and *fleu[χ]* have the same meaning thus allows infants to infer that the distinction between [ʁ] and [χ] is allophonic. Alternatively, infants might rely upon distributional information, since segments in allophonic relationships have non-overlapping distributions.[1] For instance, in French, [χ] only occurs next to a voiceless consonant whereas [ʁ] occurs everywhere except next to a voiceless consonant. Hence, computing contextual statistics likewise allows infants to separate allophonic from phonemic distinctions.

Quite another question concerns the type of constraints that guide the acquisition process. Is acquisition based on general learning mechanisms or are linguistic constraints taken into account? Phonologically natural rules share several formal features. For instance, phonological processes fall into several broad categories, such as assimilation, weakening, and strengthening. It is, therefore, reasonable to postulate that a phonological learning algorithm would take these features into account, hence predicting faster and more complete learning for languages that respect them than for languages that do not. In contrast, under general-purpose learning algorithms, one would expect no difference between these two types of languages, as their statistical distributional properties are the same.

A related question concerns the role of natural classes during the acquisition of underlying representations. Phonological processes often concern not just a single pair of segments, but several pairs that constitute a natural class. For instance, the allophonic devoicing of [ʁ] in French extends to the other sonorant consonants. The correct generalization, then, is that voicing is allophonic in sonorants. Saffran and Thiessen (2003) have shown that 9-month old infants are sensitive to natural classes. This raises the question as to whether phonological rules are acquired one by one, or whether generalizations are made within natural classes. More specifically, does knowledge that [ʁ] and [χ] are realizations of a single phoneme /r/

help to infer that, likewise, [l] and [l̥] are realizations of a single phoneme /l/?[2]

Using an artificial language-learning paradigm, we examine both the respective contributions of semantic and distributional information and the role of natural classes for the acquisition of allophonic rules. The participants in our experiments are adults, with whom we assume infants to share a processing architecture. In Experiments 1 and 2, we focus on the distinction between voiced and voiceless obstruents that is phonemic in French, and test if French adults can learn to consider these distinctions to be allophonic. In Experiment 3, we test if French adults more generally can learn to consider segmental distinctions in obstruents to be allophonic, even if the allophonic groupings are highly unnatural. In all experiments, we address the question of natural classes by exposing subjects to a voicing alternation at two places of articulation and testing them on both these places and a novel one.

2. Experiment 1

We created two artificial languages, sharing the same segmental repertoire but not the same set of underlying phonemes. In Language A, voicing is phonemic in stops but allophonic in fricatives, with fricatives being voiced in intervocalic position and voiceless elsewhere. In Language B, voicing is phonemic in fricatives but allophonic in stops, with stops being voiced in intervocalic position and voiceless elsewhere. The underlying obstruent inventories of the two languages are shown in Table 1.

Table 1. Underlying obstruent inventories of Languages A and B

	Language A						Language B					
	voiceless			voiced			voiceless			voiced		
stops	/p/	/t/	/k/	/b/	/d/	/g/	/p/	/t/	/k/			
fricatives	/f/	/s/	/ʃ/				/f/	/s/	/ʃ/	/v/	/z/	/ʒ/

Subjects were exposed to short phrases accompanied by referential pictures, in one of the two languages. These phrases were of the type determiner + noun, where the determiner was either *nel*, meaning ‘two’ or *ra*, meaning ‘three’, and the noun started with either a stop or a fricative followed by a vowel. Nouns starting with a labial, palatal, or velar obstruent appeared with both determiners, whereas nouns starting with a dental obstruent appeared with only one of the determiners. Crucially, the determiner *ra* but not *nel* created the context for the intervocalic voicing rule; words starting with a non-dental stop or a fricative thus appeared in two phonetic forms in Language B and Language A, respectively, depending on the determiner. Hence, subjects received evidence as to the phonemic or allophonic status of voicing in stops and fricatives except the dental ones.

After the exposure phase, a word-picture matching paradigm was used to test if subjects had learned the allophonic distributions and therefore treated voicing distinctions differently in stops and fricatives for purposes of word identification. That is, subjects were first presented with a phrase-picture pairing, for instance *nel pama* with a picture of two bears, and then heard a minimally different phrase, *ra bama*, which they had to match to either a picture of three bears or a picture of three tokens of a novel object. In this paradigm, the crucial measure is to what extent subjects choose the same object despite the change in the initial consonant of the noun. We predicted that subjects exposed to Language A would do so more often if the noun started with a fricative than with a stop, whereas, conversely, subjects exposed to Language B would do so more often if the noun started with a stop than with a fricative. Moreover, if learning generalizes within natural classes, subjects should treat all stops and fricatives differently, even the dental ones, for which the status of voicing could not be inferred from the exposure.

2.1. Materials

For the exposure phase, 12 disyllabic items for Language A were selected and 12 matched items for Language B. In each language, half of the items started with a stop and half with a fricative. Initial stops were either voiced or voiceless in Language A but always voiceless in Language B. Conversely, initial fricatives were either voiced or voiceless in Language B but always voiceless in Language A. In all items, non-initial stops and fricatives were voiced if they occurred in between vowels and voiceless otherwise, thus respecting the phonotactics of both languages. All items were pseudo-words in French.

For both languages, twenty-four short phrases were then constructed by prefixing the pseudo-words *nel* and *ra* to each one of the 12 items. In the phrases with *ra* belonging to Language A, the initial voiceless fricatives were replaced by their voiced counterparts, thus respecting the allophonic voicing rule for fricatives of Language A. Likewise, in the phrases with *ra* belonging to Language B, the initial voiceless stops were replaced by their voiced counterparts, thus respecting the allophonic voicing rule for stops of Language B.

For the test phase, 48 novel disyllabic pseudo-words were equally selected, 24 starting with a voiceless stop and 24 with a voiceless fricative. Forty-eight pairs of short test phrases were then created by prefixing *nel* and *ra* to each one of these words; for the phrases with *ra*, the initial voiceless consonants were replaced by their voiced counterparts. In addition, another 24 novel disyllabic pseudo-words were selected, to be used in filler trials, 6 starting with a voiceless stop, 6 with a voiceless fricative, and 12 with a nasal or a liquid. For these words, only one short phrase was created, in half of the cases by prefixing *nel* and in the other half by prefixing *ra*. As before, initial voiceless obstruents were replaced by their voiced counterpart in the phrases with *ra*. Finally, for a short training phase, another six disyllabic pseudo-words starting with a nasal or a liquid consonant were selected. On the basis of these items, ten phrases were created by prefixing *nel* to one of them, *ra* to another one, and both *nel* and *ra* to the remaining four. In all 78 words, non-initial stops and fricatives were voiced if they occurred in between vowels and voiceless otherwise. Hence, all phrases were legal in both languages.

All phrases were recorded by a female native speaker of French. They were digitized at 22050 Hz and 16 bits and stored on a computer disk.

In addition to the phrases, 192 pictures of objects, animals, and body parts were selected.

2.2. Procedure

The experiment consisted of a training phase, an exposure phase, and a test phase. Half of the subjects were exposed to Language A and half to Language B.

At the beginning of the experiment, subjects were told that they would be studying short phrases of the type ‘two dolls’ or ‘three boats’ in an unknown language, in which *nel* means ‘two’ and *ra* means ‘three’, and that they were to memorize the words in this language.

During the training phase, subjects were familiarized with the experimental task. They were first shown a picture of two or three identical objects. A short phrase describing the picture was presented orally at the same time. Two pictures - one of which was showing the same object as the first picture - were then shown simultaneously, one on the left-hand side and one on the right-hand side of the computer screen, and a new phrase was presented. If the first phrase started with *nel*, the second one started with *ra*, and vice versa. The subjects’ task was to indicate which one of the two pictures was described by the new phrase. There were six trials. In the first two, only the determiner changed; for instance, after being exposed to [nɛlmule] with a picture of two brushes, subjects were tested on [ramule], where they had to choose between a picture of three brushes and one of three stars (the former one being the correct response). In the last four trials, the noun changed as well, and this change was either very large ([nɛlrobil] – [ralaro]) or it involved only the first consonant ([nɛlregap] – [ramegap]). For these four trials, the correct response was constituted by the picture showing the new object. During the training phase, the nouns always started with a nasal or a liquid consonant. Subjects replied by pressing either key [1] (for the picture on the left-hand side) or key [2] (for the picture on the right-hand side) on a computer keyboard, and they received feedback as to whether their responses were correct or incorrect.

During the exposure phase, subjects listened to 20 of the 24 prepared phrases in either Language A or Language B, accompanied by their referential pictures, as shown in Table 2. Note, firstly, that for each lexical item, either the phrase with *nel* or the phrase with *ra* is identical in Language A and Language B; secondly, that each obstruent occurs an equal number of times word-initially in Language A and Language B; and, thirdly, that for items starting with a dental obstruent, only one of the two phrases occurs, such that subjects did not receive evidence as to the status of voicing in dental obstruents.

Table 2. Phrases in Languages A and B used in exposure phase of Experiment 1

	Language A: allophonic fricative voicing		Language B: allophonic stop voicing	
'rabbit'	nɛl pemuʃ	ra pemuʃ	nɛl pemuʃ	ra bemuʃ
'flower'	nɛl bovi	ra bovi	nɛl povi	ra bovi
'apple'	nɛl kelaf	ra kelaf	nɛl kelaf	ra gelaf
'fork'	nɛl ginel	ra ginel	nɛl kinel	ra ginel
'hat'	nɛl timu		nɛl timu	
'tree'		ra daru		ra daru
'cat'	nɛl foʒam	ra voʒam	nɛl foʒam	ra foʒam
'nose'	nɛl fulek	ra vulek	nɛl vulek	ra vulek
'bottle'	nɛl ʃagip	ra ʒagip	nɛl ʃagip	ra ʃagip
'house'	nɛl ʃubo	ra ʒubo	nɛl ʒubo	ra ʒubo
'balloon'	nɛl sano		nɛl sano	
'snail'		ra zelum		ra zelum

Phrases in which the noun started with a labial, palatal, or velar obstruent were presented 16 times each, and those in which the noun started with a dental obstruent were presented 8 times each.[3] All phrases were presented in a semi-random order with an SOA of 3000ms, for a total of around twenty minutes. The pictures appeared on the screen 500ms prior to the presentation of the corresponding phrase and were presented for the entire 3000ms. Subjects could take a short brake halfway through the exposure. They were reminded that *nel* and *ra* meant 'two' and 'three', respectively, and they were asked to try to memorize as many of the words as possible in the language.

The test phase consisted of two parts and was identical for all subjects. In the first part, subjects were tested on their lexical representations of the words they had learned during the exposure phase, and during the second part, they were tested on their lexical representations of novel items. Within each part, trials were presented in a random order, and subjects did not receive feedback.

The first part of the test phase consisted of 12 test trials (one for each lexical item of the exposure) and 6 filler trials. In the test trials, subjects were presented with the phrase-picture pairings that were identical in both languages and that they had been exposed to, and tested on corresponding phrases in which the determiner and the voicing value of the initial obstruent of the noun were changed. For instance, they would first hear [nɛlpemuʃ] and then be tested on [rabemuʃ]. As in the training phase, their task was to indicate which one of two pictures described the new phrase, where one of the pictures showed the same object as before, and the other one showed a novel object. According to the language of exposure and the initial consonant of the noun, a trial belonged to either the Phonemic or the Allophonic condition. That is, for subjects exposed to Language A (with allophonic fricative voicing), trials with a stop-initial noun belonged to the Phonemic condition and trials with a fricative-initial noun to the Allophonic condition; for subjects exposed to Language B (with allophonic stop voicing), trials with a stop-initial noun belonged to the Allophonic condition and trials with a fricative-initial noun to the Phonemic condition. Trials in which the noun started with a labial, palatal, or velar obstruent belonged to the Trained condition; during exposure, these nouns had been presented both in the context of *nel* and in the context of *ra*. Trials in which the noun started with a dental obstruent belonged to the Untrained condition; during exposure, these nouns had been presented either in the context of *nel* or in the context of *ra* (see Table 2). For Trained items, trials in the Phonemic condition contained a test phrase that had not been part of the exposure, and trials in the Allophonic condition contained a test phrase that had been part of the exposure. For Untrained items, all test phrases were new with respect to the exposure. Half of the test phrases started with *nel* and the other half with *ra*. Likewise, in half of the trials, the same object appeared on the left-hand side and in the other half it appeared or the right-hand side of the screen. As for the filler trials, there were three trials with nouns that were part of the exposure and three trials with novel nouns, starting with a nasal or a liquid consonant. In all filler trials, the test phrase was completely identical to the one presented before.

The second part of the test consisted of 48 different test trials and 24 different filler trials. During this part, all phrase-picture pairings were novel. In the test trials, half of the nouns started with a stop (⅓ labial, ⅓ dental, ⅓ velar) and the other half with a fricative (⅓ labial, ⅓ dental, ⅓ palatal). In phrases with *nel* the initial obstruent was voiceless, and in phrases with *ra* it was voiced. In these trials, subjects were first presented with a phrase-picture pairing and then tested on a corresponding phrase in which both the determiner and the voicing value of the noun-initial obstruent were changed. Their task was again to indicate which one of two pictures described the new phrase, where one of the pictures showed the same object, and the other one showed a new object. As before, according to the language of exposure and the initial consonant of the noun (stop or fricative), a trial belonged to either the Phonemic or the Allophonic condition, and according to the place of articulation of the initial consonant (labial/palatal/velar or dental), a trial belonged to either the Trained or the Untrained condition. In the filler trials, subjects were presented with phrase-picture pairings and tested on the identical phrases. Half of the nouns in the filler trials started with a stop or a fricative (voiceless in phrases with *nel* and voiced in phrases with *ra*), and the other half with a nasal or a liquid consonant. In both test and filler trials, the same object appeared on the left-hand side in half of the cases and on the right-hand side in the other half. The 72 trials of this second test phase were presented in two blocks, the order of which was counterbalanced across subjects.

The entire experiment lasted around 40 minutes.

2.3. Subjects

Twelve native speakers of French with normal hearing and normal or corrected-to-normal vision participated in the experiment.

2.4. Results and discussion

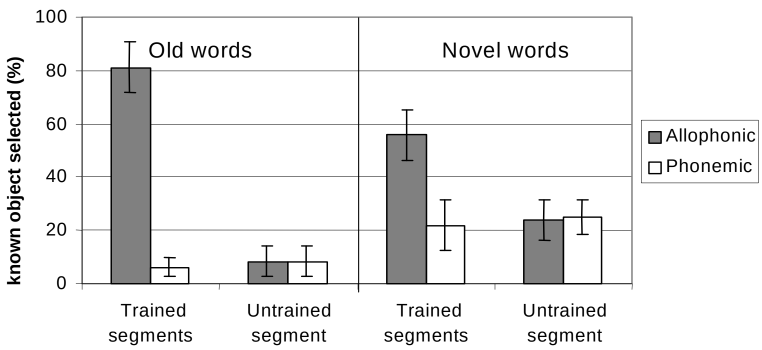

Figure 1. Percentage of pairings of the novel phrase to the known object for the test trials in the two test phases.

The results of the test phases with the old and the novel words were subject to separate ANOVAs.

First, the results of the test phase with the old words were subject to an ANOVA with between-subject factor Language (A vs. B) and within-subject factors Contrast (Phonemic vs. Allophonic) and Segment (Trained vs. Untrained). There was a significant main effect of Contrast (F(1,10) =

47.6, $p<.0001$), due to the fact that the percentage of pairings with the known object was higher in the Allophonic than in the Phonemic condition, and a significant main effect of Segment ($F(1,10) = 27.3$, $p<.0001$), due to the fact that the percentage of pairings with the known object was higher in the Trained than in the Untrained condition. Furthermore, the interaction between these two factors was significant ($F(1,10) = 35.2$, $p<.0001$), reflecting the fact that there was an effect of Contrast in the Trained ($F(1,10) = 54$, $p<.0001$) but not in the Untrained condition. No other main effect or interaction reached significance.

The results of the test phase with the novel words were subject to an ANOVA with between-subject factors Language (A vs. B) and Order (counterbalancing factor, Block1 first vs. Block2 first) and within-subject factors Contrast (Phonemic vs. Allophonic) and Segment (Trained or Untrained). There was a significant main effect of Contrast ($F(1,10) = 8.6$, $p<.016$), due to the fact that the percentage of pairings with the known object was higher in the Allophonic than in the Phonemic condition, and a marginally significant effect of Segment ($F(1,10) = 4.4$, $p<.064$), due to the fact that the percentage of pairings with the known object was higher in the Trained than in the Untrained condition. Moreover, the interaction between these two factors was significant ($F(1,10) = 6.0$, $p<.035$), reflecting the fact that there was an effect of Contrast in the Trained ($F(1,10) = 8.7$, $p<.015$) but not in the Untrained condition. No other main effect or interaction reached significance.

These results show that subjects treated voicing distinctions differently, depending on whether in the language of exposure they were phonemic or allophonic. Specifically, they ignored a voicing difference on labial, palatal, and velar obstruents more often in the Allophonic condition than in the Phonemic condition for purposes of word identification. Importantly, they did so not only for the items to which they had been exposed, but also for novel items. This, then, is evidence that French adults can learn to map voiced and voiceless stops or fricatives onto the same underlying phonemes after twenty minutes of exposure. By contrast, subjects failed to generalize the acquired knowledge towards dental obstruents, in both known and novel items. This suggests that they learned to ignore the voicing distinction in two pairs of segments without making any inference about the remaining pair belonging to the same natural class.

In the next experiment, we test whether subjects can learn to create abstract phoneme categories on the basis of distributional information only, that is, in the absence of word meanings.

3. Experiment 2

This experiment was identical to Experiment 1, with one modification: during the exposure phase, for each lexical item a referential picture was shown with either *nel*-phrases or *ra*-phrases but not both. There was therefore no semantic information revealing whether a pair of phrases referred to the same entities or not.

3.1. Materials and procedure

The materials and procedure were as in Experiment 1, except that for the exposure phase, we replaced the referential pictures of half of the phrases containing a lexical item starting with a labial, palatal, or velar obstruent by a picture showing either two (for phrases with *nel*) or three (for phrases with *ra*) question marks.[4] For half of these items, this was done for the phrase with *nel*, and for the other half for the phrase with *ra*. Table 3 shows the materials for the exposure phase; the phrases in italic are accompanied by the question marks. Note that the phrases that are shown with their referential picture are always identical in the two languages.

Table 3. Phrases in Languages A and B used for the exposure phase of Experiment 2. The phrases shown in italics are not accompanied by a referential picture

	Language A: allophonic fricative voicing		Language B: allophonic stop voicing	
‘rabbit’	nɛl pemuʃ	*ra pemuʃ*	nɛl pemuʃ	*ra bemuʃ*
‘flower’	*nɛl bovi*	ra bovi	*nɛl povi*	ra bovi
‘apple’	nɛl kelaf	*ra kelaf*	nɛl kelaf	*ra gelaf*
‘fork’	*nɛl ginel*	ra ginel	*nɛl kinel*	ra ginel
‘hat’	nɛl timu		nɛl timu	
‘tree’		ra daru		ra daru
‘cat’	nɛl foʒam	*ra voʒam*	nɛl foʒam	*ra foʒam*
‘nose’	*nɛl fulek*	ra vulek	*nɛl vulek*	ra vulek
‘bottle’	nɛl ʃagip	*ra ʒagip*	nɛl ʃagip	*ra ʃagip*
‘house’	*nɛl ʃubo*	ra ʒubo	*nɛl ʒubo*	ra ʒubo
‘balloon’	nɛl sano		nɛl sano	
‘snail’		ra zelum		ra zelum

3.2. Subjects

Twenty-four native speakers of French with normal hearing and normal or corrected-to-normal vision participated in the experiment. None of them had participated in Experiment 1.

3.3. Results and discussion

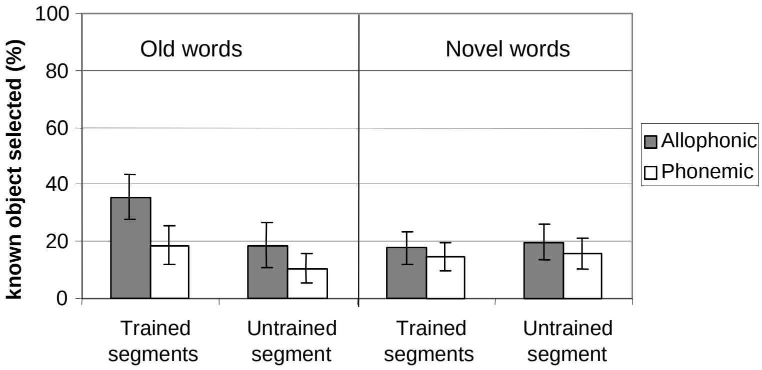

Figure 2. The percentage of pairings of the novel phrase to the known object for the test trials in the two test phases.

The results of the two test phases were analyzed as in Experiment 1.

The ANOVA of the results of the test phase with the old words revealed a main effect of Contrast ($F(1,22) = 8.02$, $p<.02$), due to the fact that the percentage of pairings with the known object was higher in the Allophonic than in the Phonemic condition, and a main effect of Segment ($F(1,10) = 9.8$, $p<.006$), due to the fact that the percentage of pairings with the known object was higher in the Trained than in the Untrained condition. No other main effect or interaction reached significance. In particular, the interaction between Contrast and Segment was not significant ($F<1$). Restricted analyses revealed a significant effect of Contrast in the Trained condition ($F(1,22) = 6.7$, $p<.018$), but not in the Untrained condition.

The ANOVA of the results of the test phase with the novel words revealed an interaction between Contrast and Language ($F(1,22) = 5.0$, $p<.037$). No other main effect or interaction reached significance. Restricted analyses revealed a significant interaction between the factors Contrast and Language in both the Trained ($F(1,22) = 4.8$, $p<.05$) and the Untrained ($F(1,22) = 3.5$, $p<.08$) condition. As in the global ANOVA, no other main effect or interaction reached significance.

These results show that as in Experiment 1, subjects treated voicing distinctions differently, depending on whether in the language of exposure they were phonemic or allophonic. Specifically, they ignored a voicing difference on labial, palatal, and velar obstruents more often in the Allophonic condition than in the Phonemic condition for purposes of word identification. The effects were numerically much smaller than in Experiment 1, as the predominant response in all cases was to choose the unknown picture. Furthermore, the difference between phonemic and allophonic distinctions yielded significant effects only for the items to which the participants had been exposed, and not for novel words. Hence, removing relevant semantic information had the effect of severely weakening, although not eliminating, the acquisition of the difference between phonemic and allophonic distinctions. As in Experiment 1, this acquisition failed to generalize towards dental obstruents, in both trained and novel words.

In the last experiment, we test whether subjects can learn unnatural allophonic rules.

4. Experiment 3

We created two new artificial languages, C and D, sharing the same segmental repertoire as Languages A and B but not the same set of underlying phonemes. In Language C, all obstruents are phonemic except [f], [t] and [ʒ], which are allophones of /g/, /s/, and /p/, respectively, occurring in intervocalic position. In Language D, likewise, all obstruents are phonemic except [s], [k] and [b], which are intervocalic allophones of /d/, /v/, and /ʃ/, respectively. Note that these allophonic groupings are highly unnatural. They involve at least two changes among the dimensions of voicing, manner, and place, that are unrelated to the conditioning context (intervocalic position) and that differ from one another within the same language (for instance, in Language C, the voiced velar stop /g/ has the voiceless labial fricative [f] as an allophone, whereas the voiceless labial stop /p/ has the voiced palatal fricative [ʒ] as an allophone).

The underlying obstruent inventories of the two languages are shown in Table 4.

Table 4. Underlying obstruent inventories of Languages C and D

	Language C						Language D					
	voiceless			voiced			voiceless			voiced		
stops	/p/		/k/	/b/	/d/	/g/	/p/	/t/			/d/	/g/
fricatives		/s/	/ʃ/	/v/	/z/		/f/		/ʃ/	/v/	/z/	/ʒ/

4.1. Materials and procedure

Two new sets of materials for the exposure phase and the test phase were created, following the logic of those in Experiment 1, but based on the phoneme categories of the new languages. Table 5 shows the materials for the exposure phase. As in the previous experiments, word-medial and word-final obstruents are legal in both languages.

Table 5. Phrases in Languages A and B used in exposure phase of Experiment 3

	Language C: /p/ → [ʒ], /g/ → [f], (/z/ → [t]) } / V_V		Language D: /ʃ/ → [b], /v/ → [k], (/d/ → [s]) } / V_V	
'rabbit'	nɛl pemuʃ	ra ʒemuʃ	nɛl pemuʃ	ra pemuʃ
'flower'	nɛl bomi	ra bomi	nɛl ʃomi	ra bomi
'apple'	nɛl kela	ra kela	nɛl vela	ra kela
'fork'	nɛl girel	ra firel	nɛl girel	ra girel
'balloon'	nɛl doba		nɛl doba	
'tree'		ra tirur		ra tirur
'cat'	nɛl goʒa	ra foʒa	nɛl foʒa	ra foʒa
'nose'	nɛl vusen	ra vusen	nɛl vusen	ra kusen
'bottle'	nɛl ʃanip	ra ʃanip	nɛl ʃanip	ra banip
'house'	nɛl puko	ra ʒuko	nɛl ʒuko	ra ʒuko
'hat'	nɛl zifu		nɛl zifu	
'snail'		ra setum		ra setum

The presence of the four single phrases *nel doba*, *nel zifu*, *ra tirur* and *ra setum* ensures that the exposure is comparable to that in Experiment 1, but is otherwise uninteresting for the purposes of the present experiment. Indeed, the alternations /z/ → [t] (Language C) and /d/ → [s] (Language D) are impossible to infer, since they do not represent a generalization within a natural class of one of the alternations that are present overtly. Hence, we expect that during the test phase, subjects treat both the [z-t] and the [d-s] distinctions in the Untrained condition as phonemic; our classification of these distinctions as phonemic or allophonic in the two languages is arbitrary.

The procedure was as in Experiment 1.

4.2. Subjects

Twelve native speakers of French with normal hearing and normal or corrected-to-normal vision participated in the experiment. None of them had participated in Experiments 1 or 2.

4.3. Results and discussion

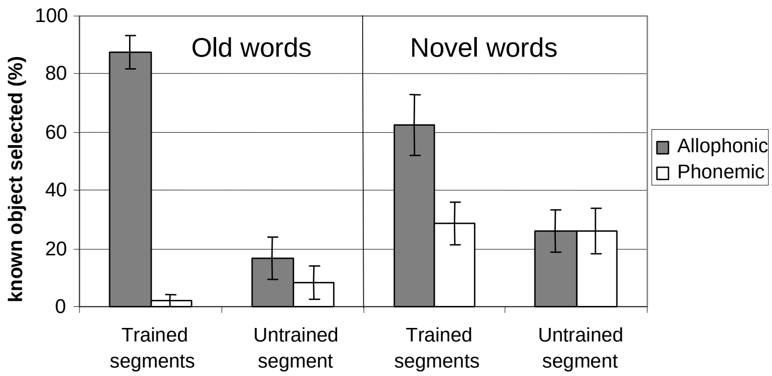

Figure 3. Percentage of pairings of the novel phrase to the known object for the test trials in the two test phases.

The results of the two test phases were analyzed as in Experiments 1 and 2.

The ANOVA of the results of the test phase with the old words revealed a main effect of Contrast ($F(1,10) = 139$, $p<.0001$), due to the fact that the percentage of pairings with the known object was higher in the Allophonic than in the Phonemic condition, and a main effect of Segment ($F(1,10) = 30.6$, $p<.0001$), due to the fact that the percentage of pairings with the known object was higher in the Trained than in the Untrained condition. Furthermore, the interaction between these two factors was significant ($F(1,10) = 47.2$, $p<.0001$), reflecting the fact that there was an effect of Contrast in the Trained ($F(1,10) = 159$, $p<.0001$) but not in the Untrained condition. Finally, the interaction between Language and Contrast was significant ($F(1,10) = 11.6$, $p<.008$), reflecting the fact that the effect of Contrast was stronger in Language A ($F(1,5) = 145$, $p<.0001$) than in Language B ($F(1,5) = 29.1$, $p<.004$). No other main effect or interaction reached significance.

The ANOVA of the results of the test phase with the novel words revealed a main effect of Contrast ($F(1,10) = 6.0$, $p<.035$), due to the fact that the percentage of pairings with the known object was higher in the Allophonic than in the Phonemic condition, and a main effect of Segment ($F(1,10) = 13.1$, $p<.006$), due to the fact that the percentage of pairings with the known object was higher in the Trained than in the Untrained condition. Moreover, the interaction between these two factors was significant ($F(1,10) = 5.7$, $p<.039$), reflecting the fact that there was an effect of Contrast in the Trained ($F(1,10) = 7.6$, $p<.021$) but not in the Untrained condition. No other main effect or interaction reached significance.

These results show that as in Experiments 1 and 2, subjects treated voicing distinctions differently, depending on whether in the language of exposure they were phonemic or allophonic. Specifically, they ignored word-initial segmental differences more often in the Allophonic condition than in the Phonemic condition for purposes of word identification, and they did so not only for the items to which they had been exposed but also for novel items starting with the same segments. This, then, is evidence that French adults can learn to map segments onto underlying phonemes on the basis of semantic information even if this is the result of unnatural allophonic rules. As expected, subjects treated both untrained distinctions as phonemic, both in the items to which they had been exposed and in new ones. Indeed, the untrained distinctions being phonologically unrelated to the trained ones, there was no evidence that one of them would be allophonic.

5. General discussion

We used an artificial language learning paradigm to evaluate how semantic and distributional information can be used to acquire the difference between phonemic and allophonic distinctions. In our paradigm, two counterbalanced artificial languages contained the same sets of surface segments (all phonemic in French, the subjects' native language) resulting from different underlying phoneme inventories. This design ensures that our results are due to the phonological properties of the two languages, rather than to some inherent phonetic properties of the segments being used.

Experiment 1 shows that when semantic information regarding the identity of lexical items is provided during exposure, a robust separation of phonemic and allophonic voicing distinctions is acquired. That is, whereas subjects judged that a change in voicing of the initial obstruent must correspond to a change in lexical item for the consonants in which voicing was phonemic in the language of exposure, the change in voicing was taken to be lexically irrelevant for the obstruents in which voicing was allophonic. Not only did the subjects apply this distinction to the words they had been exposed to (and hence had explicitly learned during the twenty minutes of exposure), but they also generalized it to novel words starting with the same segments. This is evidence that they have learned that in the language of exposure, voicing is phonemic and hence carrier of semantic information in one set of segments whereas it is allophonic and hence irrelevant for word identification in another set.

These results are in accordance with much recent work with artificial language learning paradigms, showing that both adults and infants can learn various properties of a language's sound structure such as segments (Maye and Gerken 2000; Maye, Werker, and Gerken 2002), phonotactics (Onishi, Chambers, and Fisher 2002; Chambers, Onishi, and Fisher 2003; Saffran and Thiessen 2003) and word segmentation (Saffran, Aslin, and

Newport 1996; Saffran, Newport, and Aslin 1996) from brief auditory exposure. All of the above-mentioned studies, though, concern surface phonological properties. The present study provides evidence that adults can likewise learn the mapping between surface segments and underlying phonemes. That is, the subjects in our experiments learned that some segments are surface manifestations of a single underlying phoneme and hence that the distinction between them can be ignored for purposes of word recognition.

We also tested in Experiment 1 whether the acquisition of the difference between phonemic and allophonic distinctions extends to novel segments belonging to the same natural classes. In particular, the exposure phase provided evidence for the status of voicing in labial, palatal, and velar obstruents, but not for its status in dental obstruents. If the regularity that subjects have learned is represented in terms of distinctive features or natural classes, one would expect that, depending on the language of exposure, either [t-d] or [s-z] would be considered a likely allophonic distinction. This was not what we found, though: subjects considered both distinctions as phonemic. In other words, at least in this paradigm, the difference between phonemic and allophonic distinctions seems to be learned on a segment-by-segment basis. This is a bit surprising, as the phonological system of natural languages is typically organized around natural classes.[5]

In Experiment 2, we examined whether the difference between phonemic and allophonic distinctions can be learned in the absence of semantic information, on the basis of distributional cues only. In order to test this, we removed the referential pictures of one of the two phrases associated to each lexical item. There was therefore no semantic information revealing whether a pair of phrases contained the same noun or a different one. The only information that participants could use was distributional: that is, in the case of allophonic voicing, the two word forms had complementary distributions, with one form occurring only after the determiner *nel* and the other one only after *ra*. We found that the difference between phonemic and allophonic distinctions was learned only for the lexical items that were part of the exposure phase; for the novel items, the subjects considered all distinctions to be phonemic. The performance on the lexical items that were part of the exposure indicates that the distributional cues separating phonemic from allophonic distinctions has some influence. The absence of an effect on novel lexical items, by contrast, suggests that the subjects did not learn the phonological system of the language, with voicing being phonemic in one class of obstruents and allophonic in the other class. Of course, it could be that the effect on novel items is too small to be measured in this experiment given that the effect on the items that were part of the exposure was already small to begin with. In order for the distributional cues to be more informative, it is probably necessary to increase the amount of exposure and the range of contexts in which the different obstruents appear.

In Experiment 3, we returned to the paradigm with full semantic information and examined whether the difference between phonemic and allophonic distinctions can be learned if the allophonic groupings are highly unnatural. In order to test this, we created two new languages in which the allophonic groupings are each the result of a different rule that is unnatural in terms of both its conditioning context and the phonological distance between the two related segments. Yet, despite the unnaturalness of the language to which they were exposed, subjects successfully learned to associate word forms that are allophonic variants of one another, and generalized this performance to novel items starting with the same segments. The results of Experiment 3 are as strong as that of Experiment 1, which used the same natural rule applying to different segments within a natural class. This, then, shows that within the present paradigm, subjects can learn arbitrary mappings from surface segments onto underlying phonemes. This conclusion is compatible with the finding in Experiment 1 that subjects learn allophonic distinctions on a one-by-one basis, and do not generalize their acquisition to other segments within the same natural class. We propose four alternative interpretations of these results.

First, it is possible that the artificial language learning paradigm and/or the phrase-picture matching task does not induce subjects to use their linguistic resources. Although this possibility is always to be kept in mind, there is no reason to suspect that our paradigm is less linguistic in nature than the ones used previously to test the acquisition of word segmentation, segmental categories, and phonotactics. Quite the opposite could be claimed, since our paradigm contains word-meaning pairings that are presented within short phrases. Furthermore, our phrase-picture matching task is modeled on the one used to test lexical knowledge in young children (see, for instance, Swingley and Aslin 2000).

The second possibility is that the paradigm is fine, but that our adult subjects use metalinguistic rather than linguistic abilities during the task. Specifically, they might have relied on an orthographic code and learned a substitution rule based on letters, not phonetic segments. Since letters do not reflect the phonological structure of the corresponding segments, this would explain why they did not show any effect of phonological naturalness. Testing preschool children with the same paradigm will allow us to test this hypothesis. Moreover, it is quite possible that the acquisition of a novel phonological system is more difficult than that of the native language. Although most artificial language learning experiments have found parallel results in infants and adults (Saffran, Aslin, and Newport 1996; Saffran, Newport, and Aslin 1996; Maye and Gerken 2000; Maye, Werker, and Gerken 2002; Onishi, Chambers, and Fisher 2002; Chambers, Onishi, and Fisher 2003), there are some discrepancies too. In particular, whereas Maye and Gerken (2001) found that adults do not generalize the acquisition of a non-native segmental distinction to another pair of segments belonging to the same natural class, Maye and Weiss (2003) obtained such generalization in infants. This finding is particularly interesting in the light of the failure of our adult subjects to generalize the acquisition of an allophonic voicing distinction from two pairs of segments to a third one within the same natural class. It thus appears to be important to carry out the same type of experimental research with infants.

The third possibility is that adults do use their linguistic abilities, but that these abilities do not take phonological naturalness into account. In particular, it could be the case that acquisition processes in speech perception are not constrained by phonological naturalness, which might be tied to constraints on the articulatory system only. In light of this hypothesis, our current work carries out the same set of experiments with a production rather than a perception task.

Finally, it is possible that even though we introduced an allophonic rule in our artificial language, the participants interpreted it as another type of rule, specifically, a morphophonological one. Our languages were designed such that word-medially and word-finally, voiced and unvoiced obstruents occurred only in intervocalic and non-intervocalic position, respectively. Hence, the phonotactics of allophonic intervocalic voicing were respected throughout the phrases. However, given that only word-initially - i.e. after the determiners - voiced and voiceless obstruents actively alternated, the possibility of a morphophonological interpretation was not excluded. Such an interpretation could explain why we did not find any effect of phonological naturalness; indeed, morphophonological processes often escape phonological naturalness. In order to make sure that our paradigm taps purely allophonic rule learning, we would have to expose subjects to occurrences of the alternation(s) in other contexts than at the border between determiner and noun.

To conclude, we have established an experimental paradigm to study the acquisition of underlying phonological representations. Our results clearly show that adults can acquire novel allophonic distinctions during a short exposure. The nature of the mechanism involved in our experiments, its functioning in infants, and its relationship with other learning mechanisms that are responsible for native language acquisition, remain to be further elucidated.

Acknowledgements
This research was supported by a grant from the Centre National de la Recherche Scientifique (‘Cognition et Traitement de l’Information’). We would like to thank Diogo Almeida and Katrin Skoruppa for their research assistance, and Jennifer Cole, José Hualde, and an anonymous reviewer for comments and questions.

Notes

1. This is an oversimplification, in that the phonetic properties of segments are defined along continuous acoustic and articulatory parameters. Consequently, there often is a certain amount of overlap in the distributions of segments in allophonic distinctions, but these distributions are crucially bimodal.
2. We use the term phoneme to refer to the abstract units that underlying representations are made of. Likewise, we use the term allophone to refer to surface segments that constitute non-default realizations of phonemes. A distinction between two segments is said to be phonemic if the segments are realizations of two different phonemes and allophonic if they are realizations of a single phoneme.
3. Dental-initial nouns are only part of the exposure to provide evidence that dental obstruents occur in the language and can be word-initial; this allows us to test if learning generalizes towards the dental place of articulation. Given that there is nothing to be learned regarding these nouns during the exposure, they do not need to occur as often as the other nouns.
4. We chose this design rather than one in which there were no referential pictures at all, since subjects might lose interest in the exposure if it does not contain pictures. We thus stayed as close as possible to the design of the previous experiment.
5. The lack of generalization towards the dental place of articulation is not due to a potential special status of dentals. Indeed, in a further experiment, not reported on here, the trained distinctions involved labial and dental obstruents and the untrained ones palatal/velar obstruents; in this experiment, subjects likewise treated both of the untrained voicing distinctions as phonemic.

References

Chambers, Kyle, Kristine Onishi, and Cynthia Fisher
2003 Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87: B69–B77.

Gaskell, Gareth, and William Marslen-Wilson
1996 Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22: 144-58.

Jusczyk, Peter, and Richard Aslin
1995 Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* 29: 1-23.

Kuhl, Patricia, K. Williams, Francisco Lacerda, Kenneth Stevens, and Björn Lindblom
1992 Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606-608.

Lahiri, Aditi, and William Marslen-Wilson
1991 The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38: 245-94.

Maye, Jessica, and LouAnn Gerken
2000 Learning phonemes without minimal pairs. In *Proceedings of the 24th Annual Boston University Conference on Language Development. Volume 2*, S. Howel, S. Fisch, and T. Keith-Lucas (eds.), 522-533. Sommerville, MA: Cascadilla Press.
Maye, Jessica, and LouAnn Gerken
2001 Learning phonemes: How far can the input take us? In *Proceedings of the 27th Annual Boston University Conference on Language Development. Volume 2*, Anna Do, Laura Domínguez, and Aimée Johansen (eds.), 480-490. Sommerville, MA: Cascadilla Press.
Maye, Jessica, and Daniel Weiss
2003 Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In *Proceedings of the 27th Annual Boston University Conference on Language Development. Volume 2*, Barbara Beachley, Amanda Brown, and Frances Conlin (eds.), 508-518. Sommerville, MA: Cascadilla Press.
Maye, Jessica, Janet Werker, and LouAnn Gerken
2002 Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101-B111.
Onishi, Kristine, Kyle Chambers, and Cynthia Fisher
2002 Learning phonotactic constraints from brief auditory exposure. *Cognition* 83: B13-B23.
Polka, Linda, and Janet Werker
1994 Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421-435.
Saffran, Jenny, Richard Aslin, and Elissa Newport
1996 Statistical learning by 8-month-old infants. *Science* 274: 1926-1928.
Saffran, Jenny, Elissa Newport, and Richard Aslin
1996 Word segmentation: the role of distributional cues. *Journal of Memory and Language* 35: 606-621.
Saffran, Jenny, and Erik Thiessen
2003 Pattern induction by infant language learners. *Developmental Psychology* 39: 484-494.
Swingley, Daniel, and Richard Aslin
2000 Spoken word recognition and lexical representation in very young children. *Cognition* 76: 147-166.
Werker, Janet, and Richard Tees
1984 Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49-63.