Representation of Phonological Alternations in a First and a Second Language: A Preliminary Report

Isabelle Darcy
Indiana University

1. Introduction

The first language word recognition system copes easily with the great variability of spoken language. A particular kind of context-dependent variation is phonological assimilation. In French, for example, there is a process of voicing and devoicing assimilation. For example, *jupe* [ʒy**p**] 'skirt' can be pronounced [ʒy**b**] when followed by an obstruent of opposite voicing, like /g/ in *grise* [**g**ʀiz] 'grey' (voicing assimilation). Conversely, *robe* [ʀo**b**] 'dress' is pronounced [ʀo**p**] when followed by a voiceless obstruent, such as /s/ in *sombre* [**s**ɔ̃bʀ] 'dark' (devoicing assimilation). When the context is a sonorant (/ʀ, l, m, n/), neither voicing nor devoicing occur: *robe rouge* [ʀo**b**]+[ʀuʒ] 'red dress', *jupe noire* [ʒy**p**]+[nwaʀ] 'black skirt'. This process is regressive and systematically occurs across word boundaries, when both words are in the same prosodic domain of the phonological phrase. A possible representation of regressive voicing assimilation in French is an autosegmental spreading and delinking operation (Goldsmith, 1990; Mester & Itô, 1989). Because French has no final devoicing otherwise, it seems that the feature [–voice] can spread from one obstruent to the preceding one, such that French needs to assume a contrastive feature specification of [+voice] and [–voice] for obstruents, and symmetrical spreading of both specifications (Cohn, 1995; Féry, 2003; Wetzels & Mascaró, 2001). Sonorants in French can remain underspecified for the voicing feature (Cho, 1999; Lombardi, 1991; see Darcy, 2006 for a review and typological overview).

Because such modifications alter a word's form across word boundaries and can potentially create ambiguity, they need to be compensated for during the process of word recognition. Previous research established that listeners are able to use abstract phonological knowledge of such alternations in their first language (L1) to compensate for assimilations (Gaskell & Marslen-Wilson, 1996; Darcy et al., *in press;* Darcy & Kügler, 2007). Other interpretations tend to support a universal and early processing capacity involved in compensation for alternations instead of abstract mental representations (Gow, 2002; Mitterer & Blomert, 2003). However, these mechanisms work best if the stimuli presented are gradient rather than categorical (see Darcy & Kügler, 2007). Darcy and Kügler presented results from French listeners hearing categorical occurrences of French voicing and devoicing assimilations. If fully neutralizing, such assimilations are predicted to be challenging for lexical access: reinterpretation of fully, categorically assimilated segments is necessary in order to have successful access to the stored lexical representation. In such a case (for example in the case of [ʀop] where [p] resulting from underlying /b/ is completely voiceless), reinterpretation of a surface [p] into /b/ could depend on prior linguistic knowledge of the alternation in French (voicing/devoicing assimilation), because there are no available auditory or phonetic cues in the stimuli. If on the contrary assimilated segments still carry traces of the underlying sound, like in the case of incomplete or partial assimilation, such reinterpretation can be achieved without specific linguistic knowledge of alternations, using phonetic cues present in the signal (Gow, 2002; Mitterer & Blomert, 2003).

In Darcy and Kügler's study, such phonetic cues have been shown to be absent in acoustic analysis of the stimuli, which were produced naturally. The speaker was instructed and trained to pronounce sentences as naturally as possible, where assimilatory changes were orthographically indicated (e.g. for *robe* 'dress' /ʀob/ in the sentence …[ʀop]+[sal]… 'dirty dress', the sequence <rope sale> was written). The changes made were comparable across conditions. Special attention has been given to not overemphasize the key word pairs. Details of the acoustic analysis are presented in Table

2 below (section 2.4), the same stimuli have been used in the present study. The task used in their study was a word detection task. Listeners were asked to quickly decide if a target word heard in the sentence (e.g. [ʀop]) was the same word that was heard in isolation as the prime word (e.g. /ʀob/). The prime word was always presented in its original unchanged form. In the sentence, crucially, the corresponding target word was embedded in a sentence, and underwent an assimilatory change (e.g. …[ʀop]+[**s**al]… or …[ʀop]+[**ʀ**uʒ]…). The next word's initial ([s] or [ʀ] in the example) was chosen such that it would make the changes on the last consonant of the target word either appropriate (correctly assimilated, '*viable'*), or not appropriate ('*non-viable'* assimilation). Filler items with similar but non assimilatory changes and a "no-assimilation" control condition were included. The rationale was as follows:

- If listeners are able to compensate for assimilation and reinterpret assimilatory modifications to access a lexical representation (reinterpret a [p] into a /b/), they should judge to some extent assimilated targets to be tokens of the prime words.
- If the context matters in their decision, i.e. if they use a phonological knowledge of alternations in order to reinterpret such complete assimilatory changes, they are expected to judge the *correctly assimilated* (*viable* condition) word more often as being a token of the prime than they would for *non-viable assimilated* target words, despite the change being the same ([ʀop] in both conditions).

Listeners interpreted the modified target word as a token of the prime more often in the *viable* assimilation condition (M = 65%) than in the *non-viable* condition (M = 12%; p<.001). Darcy and Kügler (2007) argue that French native speakers have built a detailed mental representation of voicing and devoicing assimilation processes through repeated exposure to those patterns and use this knowledge to compensate for assimilation during the word recognition process.[1] Crucially, Darcy and Kügler uncovered that French listeners compensate differently for voicing vs. devoicing assimilation despite categorical and symmetrical stimuli, further supporting a specific knowledge of the assimilation phenomena in their first language. Listeners compensated for correctly assimilated *voiced* segments much more than for correctly assimilated *devoiced* segments: they recovered /ʒyp/ 'skirt' from [ʒy**b**gʀiz] 'grey skirt' more often (M=80%); they recovered /ʀob/ 'dress' only at chance from [ʀo**p**sal] 'dirty dress' (M=50%, p<.001). The asymmetry observed in compensation parallels an asymmetry reported for production data (Snoeren et al., 2006), where devoicing assimilation occurrences are realized less categorically than voicing assimilation occurrences. This parallel led to the conclusion that an input-driven, detailed knowledge of the processes at work in a language can shape our perception: because listeners are used to fewer categorical occurrences of devoicing assimilation in French, their word recognition system is not optimally prepared to compensate for categorical changes involving a full devoicing. Thus they tend to interpret a fully devoiced segment (ʀo[p]) according to the perceived voicelessness feature as an underlyingly [–voiced] segment in half of all cases (50% compensation). By contrast, the listeners' word recognition system is very effective in compensating for categorical voicing assimilation (80% compensation), presumably because they are used to such cases, which are more frequent than categorical devoicing. We are currently investigating the reasons for such an asymmetry in compensation, as well as the role of lexical knowledge in explaining why the compensation is still as high as 50% (Darcy & Kügler, *in preparation*) in the devoicing cases.

When recognizing words in a second language, non-native listeners are faced with a challenging task. Because of L1 influences in L2 perception, L2 categories can be interpreted according to distinctions used in L1 (for reviews see Sebastian-Gallés, 2005; Strange, 1995). These phoneme confusions can obscure lexical distinctions when no difference can be perceived, for example between

[1] It is still not clearly understood when this knowledge about alternations comes into play during the word recognition process, whether at the lexical or at a sublexical level. Similarly, the format for those mental representations of alternations is not clearly defined. It is not known if they are separated from lexical entries and stored at a different level or if they are incorporated into each lexical entry. The present experiments cannot distinguish between those options (see general discussion).

[neta] 'clean' and [nɛta] 'granddaughter' for Spanish learners of Catalan (Pallier, Colomé & Sebastian-Gallés, 2001). Such lack of distinction at the perceptual level could result in listeners' encoding of only one lexical representation for both forms; at least, Pallier et al.'s pattern of results show that their listeners are not accessing two different lexical representations, either because they don't have two, or because they do not perceive the [e]-[ɛ] contrast during the online task they used. These less detailed representations in turn cause an increase in word activation (Sebastian-Gallés, Echeverrìa & Bosch, 2005) and competition, as the listeners' native vocabulary can also be activated while hearing the L2 (Marian & Spivey 2003; Spivey & Marian, 1999; Weber & Cutler, 2004). Taken together, these results suggest that less precise prelexical processing increases word competition, leading to less efficient lexical access (but see Ju & Luce, 2004).

To date little research has been done on second language processing of phonological alternations. Recently, Darcy, Peperkamp and Dupoux (2007) provided evidence that highly proficient late learners had acquired the ability to perceptually compensate for the assimilatory pattern of their second language (L2), suggesting that they had developed phonological knowledge of alternations for the L2 that is comparable to their knowledge of alternations for the L1. In a similar way, Weber and Cutler (2006) reported that proficient second language learners can achieve mastery of phonotactic sequencing in a second language, which they can use in word segmentation and recognition. However, L2 listeners in their study were not successful in inhibiting unwanted interference from their L1 phonotactic knowledge. In the Darcy, Peperkamp and Dupoux study (2007), a paradigm similar to the Darcy and Kügler (2007) study was used to compare a native assimilation process in French (voicing assimilation) with a non-native process (place assimilation). In English the situation is reversed: place assimilation is part of the phonology, whereas voicing assimilation does not apply in the same form as it does in French. Two sets of sentences were constructed in each French and English, where both voicing and place assimilations modified target words differently. For example, in the English set, for the prime *wet* /wɛt/, targets words and contexts were *we*[**p**]+[**p**]*ants* in the viable assimilation condition, and *we*[**p**]+[**s**]*ocks* in the non-viable assimilation condition.[2] For voicing assimilation, the following were constructed: for the prime *black* /blæk/, the viable assimilation condition was the target *bla*[**g**]+[**g**]*love* and the non-viable one was *bla*[**g**]+[**r**]*ug*. The main results of the study showed that French listeners compensated for voicing assimilation in French in the *viable* condition (like in the Darcy and Kügler study), but not in the *non-viable* one. Crucially, they did not compensate for place assimilation, regardless of the change being theoretically a *viable* or *non-viable* assimilation. English listeners displayed a parallel pattern: they compensated for place assimilation in English sentences, in the *viable* but not in the *non-viable* condition; they didn't compensate as much for voicing assimilation, regardless of condition. These results support the claim that compensation is language-specific. Therefore, the ability to compensate for assimilation seems to rely on the presence of a mental representation for the alternation, established through exposure to the language. When tested on their second language (where American English learners of French listened to French sentences and vice versa), compensation patterns were different according to proficiency in the second language. Beginning learners who spent less than two years abroad showed the same compensation pattern in both the first and second language. French learners of English accordingly compensated for voicing assimilation in both French and English sentences, but much less for place assimilation. On the other hand, less proficient American English learners of French compensated for place assimilation changes in both English and French sentences, and only very little for voicing assimilation in both languages. However, according to the results, highly proficient American English learners of French had mastered the ability to compensate for assimilated words similarly to French native listeners in French sentences, and used this knowledge to recognize words. Contrary to Weber and Cutler's results, they also seemed to successfully stop compensating for a process that did not exist in their L2 (place assimilation), even if it did exist in their L1. Given the aforementioned differences between native listeners and learners with respect to less detailed mental representations, the question then arises as to whether L2 learners also represent phonological knowledge of alternations in the same way as native listeners do. The

[2] Here, viable and non-viable are defined according to place assimilation: the trigger context for place assimilation is a labial or velar consonant (/p, b, m, k, g, ŋ/). Alveolar consonants (/t, d, n/) do not propagate the place feature onto the preceding consonant.

present study more thoroughly explores the representation of such knowledge in a first and a second language, using evidence from voicing/devoicing assimilation in French.

Based on Darcy and Kügler's results (2007), this study asks whether second language learners of French would build appropriate representations for voicing and devoicing assimilation, thus showing a similar sensitivity to voicing and devoicing differences in French. American English learners of French have been tested on the same task and with the same stimuli as the native speakers of French. The results for the native speakers of French are those already reported in Darcy and Kügler (2007). The paper is organized as follows. I will first present a description of how the stimuli were constructed and of the experimental setup, followed by an acoustic analysis of the experimental sentences that were used in the experiments that followed. Section 3 introduces a perceptual categorization experiment that verifies how word final consonants are identified when words are not followed by any context. This experiment is preliminary to both word detection experiments in context. In section 4 and 5, the results of the two word detection experiments will be reported for native speakers first, and then for second language learners. The paper concludes with a discussion of the results, in section 6.

2. Methods and Materials

2.1 Stimuli

Sixteen target words were selected. They were all monosyllabic nouns, with a C(C)VC structure. All targets ended in a final obstruent that was voiced for half of the items (e.g. *robe* /ʀob/; the devoicing set, n=8), and voiceless for the other half (e.g. *lac* /lak/; the voicing set, n=8). They were matched in average frequency (devoicing: 3553; voicing: 6122, $t(7) = -0.9$, $p > .1$) according to the Brulex Corpus (frequency per 100 millions, from Content, Mousty & Radeau, 1990). Sixteen matched non-words (i.e. 'voiced' or 'devoiced' counterparts of the targets), indicated by an *, were constructed by switching the voicing feature of the final obstruents (e.g., devoicing: *robe* /ʀob/ 'dress' – *ʀop /ʀop/, or voicing: *lac* /lak/ 'lake' - *lag /lag/).

Each target, including its changed counterpart (e.g., [ʀob] and *[ʀop]), was associated with a triplet of context words. In this case, context words were always adjectives since the standard noun phrase has the shape 'determiner + noun + adjective' in French. Each adjective in a triplet corresponded to one of the experimental conditions: *viable change, non-viable change,* and *no-change*: Both *viable* and *non-viable change* conditions associated with the changed counterpart of the target (e.g. *[ʀop]). Only the *no-change* condition associated with the original target ([ʀob]).

For the *viable* change condition, the adjective's initial consonant was an obstruent agreeing in voicing feature with the target counterpart (e.g., *ro*[**p**]+[**s**]*ale*). The adjectives in the *non-viable* change and *no-change* conditions both started with a neutral consonant which was not involved in the relevant assimilation process (e.g., for non-viable change: *ro*[**p**]+[**n**]*oire*; for no-change: *ro*[**b**]+[**ʀ**]*ouge*). For both devoicing and voicing equally, this neutral consonant was always a sonorant (nasals and liquids, as well as the standard French uvular trill [ʀ]), that does not trigger assimilation in French. In all three conditions the association (changed)target+adjective always yielded a legal consonant cluster in French and did not contain any violation of voicing assimilation[3]. Table 1 provides an overview of all three conditions with examples.

Finally, three sentence frames were constructed for each target. A sentence frame consisted of a sentence beginning and sentence ending, where each of the three (changed) target+adjective combinations could be inserted and resulted in a plausible sentence. For example, for the pairs *ro[ps]ale* 'dirty *dress' (*viable*), *ro[pn]oire* 'black *dress' (*non-viable*), and *ro[bʀ]ouge* 'red dress' (*no-change*), one of the 3 sentences would be *Elle a mis sa* ___ ___ *aujourd'hui*. 'She put on her ___ ___ today.' Globally, the sentence frames were matched in number of words and position of the insertion slots across both sets. No violation of voicing agreement occurred in the frames either. The target+adjective combinations were selected such as the target did not predict the adjective in a strong way. For example, the target *chèque* 'check' would not have been followed by *bancaire* 'bank-issued',

[3] Half of the items were geminated in each set. The speaker produced all geminates as a single long consonant, without release in between.

because both words are often associated together in the sequence *chèque bancaire* 'bank-issued check'. Combining the three conditions with the three sentence frames gave rise to nine actual sentences associated with each target. This resulted in a total of 144 sentences.

Table 1. Overview of context conditions and changes for French stimuli.

	Voicing type			
Condition	Devoicing (D)		Voicing (V)	
	target "robe" [ʀɔb] 'dress'		target "lac" [lak] 'lake'	
appropriate change (viable context)	1a	ro[**p**+s]ale 'dirty dress'	2a	la[**g**+ʒ]elé 'frozen lake'
non appropriate change (non-viable context)	1b	ro[**p**+n]oire 'black dress'	2b	la[**g**+n]ordique 'nordic lake'
no-change	1c	ro[b+ʀ]ouge 'red dress'	2c	la[k+l]impide 'clear lake'

Thirty additional sentences were constructed similarly to the experimental sentences (the same kind of change on targets involving one feature, and the same proportion of identical and changed words), and served as practice (N=18), or distractors (N=12). Modifications involved voicing, manner and place contrasts at the end or beginning of target words, in order to steer participants' attention to the precise form of words (e.g. target *cube* [kyb] 'cube', distractor sentence containing **gube* [gyb]). Crucially, these distractor/training sentences did not contain any case of assimilation in either viable or non-viable context, so that the feedback provided here (during the training only) was unambiguous and could not influence participants' later responses on experimental sentences.

The 144 experimental, 12 distractor and 18 training sentences were recorded by the author, a female native speaker of French, together with another similar set of sentences containing alternation in place of articulation. The speaker familiarized herself intensively with the stimuli before recordings, in order to maximize the naturalness for all conditions, especially for the non-words that occur in the non-viable context (e.g. [ʀop]+[nwaʀ]). Intonation and speech rate were kept similar in all three sentences for the same target word. All sentences were recorded naturally several times. Sentence triplets that best matched with respect to "naturalness", "similar intonation", and "speech rate" were then selected and acoustically analyzed. They were digitized at 16kHz and 16bits on an OROSAU22 sound board, and edited using the sound preparation software programs CoolEdit and Praat. The onset of the carrier word and the onset of the following adjective were marked with digital labels.

2.2 Experimental setup for the word detection task

A typical experimental trial had the following form: first, a **prime** (i.e., a target word in citation form in isolation) was presented auditorily, the sentence followed after 500 ms. In the sentence, the target word occurred in its original form ([ʀob]) or in its changed form ([ʀop]). Listeners were asked to quickly decide if a target word heard in the sentence (e.g. [ʀop]) was the same word as the prime word heard in isolation (e.g. /ʀob/)[4]. The prime word was always presented in its original unchanged form. Crucially, the corresponding target word was embedded in the sentence, and underwent an assimilatory change (e.g. … [ʀop]+[**s**al]… or …[ʀop]+[nwaʀ]…) in two of the three conditions. In order to avoid a pure acoustic/phonetic matching between prime (e.g. /ʀob/) and target in the following sentence (e.g. …[ʀob]+[ʀuʒ]…), it has proven useful to present the prime word in a different voice.

[4] Reaction times have been collected and will be reported where appropriate. However, they are difficult to compare across both experiments, because they have been collected differently (for "yes" responses only in one experiment, for both in the other). Similarly, they are difficult to compare them across conditions, given that different conditions elicited different numbers of "yes" responses. They were relatively slow on average, and do not interact significantly with condition.

An acoustic matching might bias listeners to respond to subtle low level differences between the prime and the target, rather than to phonological differences that would require that the context to be taken into account. A male voice was therefore used for the prime words.

In order to counterbalance all sentences, three experimental lists were defined. In each list, all three conditions were present for each target word, but in different sentence frames. The sentence frames were rotated across the three lists, so that across the experimental lists all three conditions appeared in all three sentence frames according to a Latin Square design. Each list contained three blocks of experimental sentences. Each participant heard three repetitions of all targets (one per block), each in a different condition (viable, non-viable and no-change). In order to avoid sentence repetition, each condition was embedded in one of the three sentences that were rotated across the three lists. The 144 experimental sentences used in this study were originally combined with another 144 trials that were similar in shape but that were targeting a different type of assimilation of place of articulation. For the purposes of this study, place assimilation sentences were considered distractor sentences and are not analyzed further.

2.3 Acoustic Analyses

Only the experimental sentences used in the following perception experiments were analyzed, not the training/distractor sentences or prime words. In order to validate the full neutralization of the voicing contrast in the stimuli, for both change conditions in the sequence (changed)target + adjective ($C\underline{V_1C_1+C_2V_2}C$), assimilated word-final consonants must be completely devoiced or voiced to the same extent as their underlying voiceless or voiced equivalent in the other set. In other words, changed targets in the “devoicing set” (*[ʀop] which is the changed counterpart of the target /ʀob/) should be similar in voicing to the targets in the “voicing set” (e.g., [p] should have the same degree of voicelessness as the underlyingly voiceless [k] in /lak/). The changed (assimilated) targets of one set will therefore be compared to the unchanged targets in the other set. Similarly, vowel duration must not give cues about the underlying voicing of the following consonant (Mack 1982) in a way that would interfere with the word detection pattern in different conditions.

For each word pair (changed)target + adjective (144 in total), digital labels were set manually at zero crossings. The intervals were set to define the vowel in the target word (V1), the final consonant (C1), the initial consonant of the context word (C2) and the vowel of the context word (V2). Within the C1 interval, the following sub-intervals were defined: *voiced portion of closure* (vc), *closure duration* (cd), and *transition to the following consonant* (f).[5] The following critical values were determined:

a) Vowel length in V_1 position
b) Closure duration for the C_1+C_2 sequence
c) Duration of voicing into closure for C_1

The duration of (c) indicates the portion of the closure where some glottal activity persists. Voicing into closure for voiced obstruents comprises between 80% and 100% of the closure duration of C_1, while for unvoiced obstruents it is not necessarily zero; it can be around 30% of the C_1 closure duration (Snoeren et al. 2006).

2.4 Results

The following comparisons were made: Duration of intervals was compared across condition (*viable, non-viable* and *no-change*) and/or voicing type (*voicing* vs. *devoicing*). ANOVAs were conducted for V_1, C_1, C_2 and V_2 intervals separately. Only critical results are reported.

[5] Only the crucial values are reported here. It is important to note that the articulatory transition (f) cannot be equated to a release.

Analyses of V_1 vowel durations revealed the following facts. The mean duration of V_1 before an underlyingly voiced segment is slightly longer than before an underlyingly voiceless one, and this is irrespective of the actual surface voicing of the consonant (see Table 2). Even if the difference is very small (13 ms on average), as compared to, e.g., English voicing-induced vowel length differences,[6] it is still significant (V_1 before voiced segments: 75ms; V_1 before voiceless segments: 62 ms; $F(1,138)=7.5$, $p<.007$). There is no interaction with condition, however ($p>.7$). This suggests that the vowel length of V_1 varies according to the underlying voicing of the upcoming consonant, but not according to whether this consonant will be actually voiced or devoiced, and irrespective of the condition (viable or non-viable assimilation). An ANOVA on vowel length restricted to both change conditions (viable vs. non-viable) declaring the factor "voicing type" and "condition" did not find any significant difference. There is no main effect of "voicing type" (V vs. D) across viable and non-viable conditions (V: 63ms; D: 73ms, $p>.05$), no effect of condition ($p>.8$) and no interaction between both ($p>.8$). This leads to the conclusion that vowel length is very unlikely to be responsible for a potential difference observed between these two conditions in later word detection experiments.

A voicing degree measure is computed from the duration of voicing into closure as a ratio of C_1 duration. Basis of comparison for each assimilatory change is the non assimilated (*no-change*) underlying form (*no D*, *no V*). An underlyingly voiceless consonant (*no V*) following a vowel is voiced for 13% of its closure duration; underlyingly voiced segments (*no D*) display clear voicing for the most part of C_1-closure (87%), compared to Snoeren et al. (2006). Stimuli might induce some biases in the upcoming word detection task if they show that the degree of assimilation (in terms of surface voicing) is weaker in viable *devoicing* than *voicing*. While the acoustic analyses of voicing degree alone do not provide a robust measure of a strength difference, a look at Table 2 does not prompt the impression that the degree of assimilation is weaker in devoicing than in voicing stimuli. Indeed, voicing degree in changed conditions (assimilated) almost perfectly equates the corresponding value for the underlying counterpart.

Table 2**:** Vowel duration and voicing degree according to assimilation type (V=Voicing, D=Devoicing), underlying and surface voicing.

assim. type (condition)	underlying	surface	mean duration V_1 (ms) / (SD)	voicing degree (SE)
viable D	/b/	**[p]**	72 (35)	15 % (01)
viable V	/p/	[b]	63 (13)	94 % (03)
non-viable D	/b/	**[p]**	73 (37)	13 % (02)
non-viable V	/p/	[b]	63 (11)	89 % (02)
no D	/b/	[b]	79 (39)	87 % (04)
no V	/p/	**[p]**	61 (11)	13 % (01)

In other words, Table 2 shows that an underlying voiceless segment (e.g., boldface underlying [p]: *no V*) is not "more voiceless" than an assimilated devoiced segment (boldface surface [p]: *non-viable D* and *viable D*), and vice-versa. Comparisons across conditions were performed on the basis of the surface voicing. Neither the analysis of surface voiced conditions ($p>.2$) nor that of surface unvoiced conditions ($p>.5$) revealed any effect of "condition", suggesting that a given surface voicing is similar in all three context conditions (viable, non-viable and no-change).

Another issue would be the degree of assimilation being weaker in *viable* than in *non-viable* conditions. An ANOVA has been conducted on voicing degree including the variables "condition" (restricted to change conditions: viable and non-viable) and "underlying voicing" (voiced, unvoiced). The analysis revealed that "condition" has no effect on the voicing degree ($p>.1$). Not surprisingly, there is a highly significant main effect of "underlying voicing" on the voicing degree ($F(1,92)=1057.1$, $p<.0001$), but no interaction of the two variables ($p>.5$).

[6] For example, Mack (1982) presents duration differences of 150 ms or more; Crowther & Mann (1992) obtained values of 80 ms on average.

In conclusion, acoustic analyses confirm that stimuli are categorical with respect to voicing degree; they do present occurrences of "complete assimilation" (neutralization) where assimilation was intended. Furthermore, statistical analyses support the conclusion that no subtle acoustic difference across conditions would strongly bias a word detection response.

3. Perceptual Categorization: Identification

3.1 Procedure and participants

All target words were excised out of the carrier sentences used for the following word detection experiment, and presented in isolation in a forced-choice categorization task (e.g., [ʀop] excised out of *Elle a mis sa [ʀop] noire 'she put on her black dress'*]. Extra care was taken in selecting the cutting point at the end of target words where no auditory cue about the next segment was perceptible (in general, at the boundary between closure duration (cd) and transition (f) within the C_1 interval). In the case of voiced or voiceless geminates (for the viable condition, half of stimuli contained a geminate), the cutting point was set to roughly half of the closure duration between C_1 and C_2.

Eighteen French native speakers who had not participated in any other similar experiment were recruited for this experiment. They were students in the Parisian area at the time of testing. Due to experimenter error, their exact age range could not be retrieved. They had only limited and late experience with English, none of them spoke another second language fluently. They were tested individually in a quiet room. None of them reported any history of hearing impairment. Participants were told that they would hear a list of words (like [ʀob]) or non-words (like [ʀop]) pronounced quite rapidly, their task being to identify the final sound of each token. They had to tick the consonant they heard on a response sheet during the 3000 ms separating each token, choosing between the original consonant and the assimilated one. For the word *robe [ʀob]* 'dress' for example, the choice was between [b] (unchanged) and [p] (assimilated). A free cell allowed them to report any better matching sound, if needed. Responses are measured as "congruent response", i.e. consistent with underlying voicing.

3.2 Results

One item in the devoicing set containing an affricate (*badge* 'name tag') turned out to induce high error rates (mean error: 30%), and was excluded from analyses and from the subsequent experiment. No participant was excluded.

Analyses of variance restricted to both change conditions (viable and non-viable) revealed that devoicing produced significantly ($p=.012$) more "congruent" judgments (3.9%) than did voicing items (1.6%); this is due to the presence of voiceless geminate closures, which are most difficult to identify. Globally, final consonants of target words, when presented without their following context, were perceived mostly categorically as voiced or voiceless (as shown in the last column), reflecting very closely the surface voicing and the acoustic properties of the stimuli.

***Table 3*: Mean percentages (and standard error) of "responses congruent with underlying voicing" according to assimilation type, underlying and surface voicing (N = 18). The last column is a transformation of the congruent responses into a percentage of stimuli "perceived as [voiced]".

assimilation type	underlying voicing	% congruent	SE	Surface voicing	% as [+vd]
viable D	/b/	7.1	1.3	[p]	7.1
viable V	/p/	1.8	0.6	[b]	98.1
non-viable D	/b/	0.8	0.4	[p]	0.8
non-viable V	/p/	1.3	0.6	[b]	98.6
no D	/b/	97.8	0.5	[b]	97.8
no V	/p/	98.3	0.4	[p]	1.6

This suggests that no traces of the underlying voiced or voiceless counterpart, for example minimal cues following from coarticulation, that might induce unexpected perceptual biases, are retained in the assimilated conditions. If participants then display a different interpretation of the same consonants, once target words are presented with their context (see following experiment), the source of this re-interpretation might be the context itself.

4. Word detection in context: First language results

4.1 Procedure and participants

Compensation for assimilation can be assessed using a word detection task: first, a prime (i.e., a target word in citation form) is presented auditorily (male voice), the sentence follows after 500 ms (female voice). Participants were instructed to press "yes" as soon as possible when they thought that the prime presented was the same (i.e., had the same *form*) as the target in the sentence, and "no" otherwise. The critical measure is the detection rate in each condition. Reaction times were collected for "yes" responses only. The experimental setup is as described above in section 2.2.

Eighteen native speakers of French (11 women and 7 men) who did not take part in the previous experiment participated in return for a small payment. They all grew up monolingually, and had only limited and late experience with English. They reported no other fluent L2. They were tested individually in a quiet room. They ranged in age from 19 to 28 years and were students in the Parisian area at the time of testing. None of them reported any history of hearing impairment. They were randomly assigned to one of the three experimental lists.

4.2 Results

For these and all the following results, the same criterion for item and participant exclusion as in Darcy et al. (2007) applies: items causing more than 50% false alarms or misses in the native speakers' results on L1 are dropped. The error rate for each participant is based on the performance for the no-change and non-viable change conditions alone. Participants who made more than 37.5% errors globally on these two conditions were considered as failing to perform the word detection task for this contrast and replaced.

The devoicing item *badge* 'name tag' has been removed from the analyses (together with its counterpart). No participant has been excluded. Results are presented in Fig. 1. They clearly show that compensation is higher for both voicing and devoicing when the context is appropriate for assimilation (viable change condition: M=65%; voicing M=79%, devoicing M=51%). For both, the percentage of "yes" answers is low in the non-viable context (M=12%) and highest when no change occurred (M=96%). Detection rate ("yes answers") is used as the dependent measure in an ANOVA with "voicing" (voicing or devoicing) as within subject (F_1; respectively between item, F_2) factor, and "condition" (viable, non-viable or no-change) as within subject and item factor. There is a main effect of condition showing that detection rates are dependant on the assimilation being viable or not ($F_1[2,34]=231.8$, $p<.0001$; $F_2[2,26]=204.0$, $p<.0001$). There is also a main effect of voicing, indicating that listeners behave differently as a function of voicing or devoicing ($F_1[1,17]=23.4$, $p<.0001$; $F_2[1,13]=11.3$, $p<.005$). This is however only due to the pattern of results in the viable condition, as the following analyses show. Considering each condition separately, detection rates are higher for voicing as compared to devoicing ($F_1[1,17]=18.6$, $p<.0001$; $F_2[1,13]=10.4$, $p<.007$) in the viable condition only; they are equal in both other conditions (p>. 1, participants & items). Mean reaction times by subjects range from 523 ms to 2104 ms (mean RT for n=18: 1358 ms). Analyses of reaction times and detection values did not reveal any interaction of *RT* with the factor *voicing/devoicing* ($F[1,17]<1$).

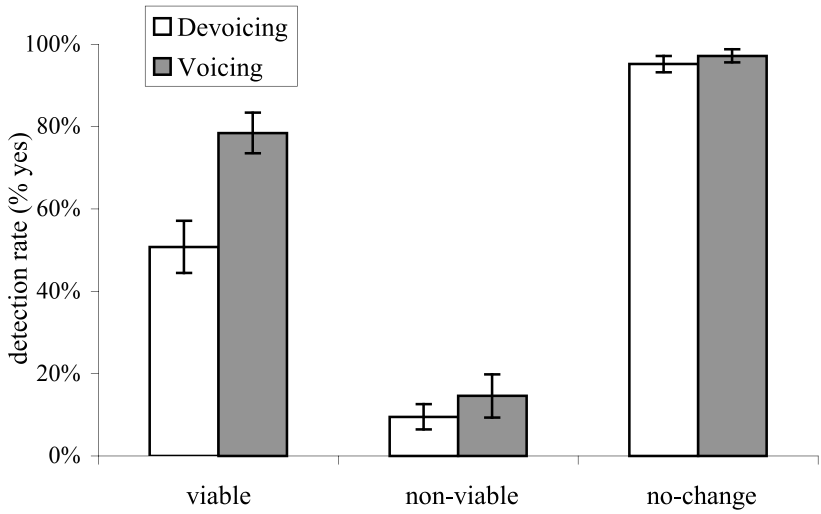

***Figure 1*: Compensation results (% yes) in each assimilation type according to underlying voicing, N=18 (from Darcy & Kügler, 2007)

This experiment confirms that French speakers strongly compensate for voicing assimilation in the viable condition. However, compensation is asymmetrical in this condition only: the percentage of "yes" answers is higher for voicing items (e.g. /ʒyp/ 'skirt') than for devoicing items (e.g. /ʀob/ 'dress'). This asymmetry cannot be explained by the higher "congruent with underlying voicing" judgment rate for devoicing items in the categorization experiment (section 3), which would lead here to higher detection responses in this experiment. Rather, listeners here are interpreting half of all devoicing assimilation occurrences as being different from the heard prime word (only 50% "yes"). Interestingly, this asymmetry parallels the one observed in naturally produced assimilation, where devoicing is systematically less categorical (only 6% are) than voicing assimilation (52% are categorical) (Snoeren et al. 2006). However, this asymmetry does not parallel the actual shape of the stimuli heard by listeners in this experiment, which were comparable in terms of the degree of voicing or devoicing. When hearing categorical assimilations, as in the present experiment, voicing is compensated for more (79% vs. 51%, p<.0001) than devoicing, whose compensation is less efficient. Darcy and Kügler (2007) suggested that this effect might occur because categorical occurrences of devoicing challenge the recognition system, whereas categorical occurrences of voicing do not. This seems to indicate a different representation for each kind of alternation (Darcy & Kügler 2007). Indeed, there is no reason to expect that categorical occurrences would be less efficiently compensated for, if the word recognition system uses a specific abstract knowledge of voicing and devoicing assimilation to compensate for appropriate changes.

One possibility is that listeners represent both alternations in a different way, targeting different input types. We are currently investigating this issue in greater detail (Darcy & Kügler, *in preparation*).

Thus far, combined results of acoustic analyses, categorization, and word detection have shown that a sound, categorized clearly without context as voiced or unvoiced, is given the other value in appropriate contexts, inducing compensation for voicing assimilation. The correct reinterpretation depends on prior linguistic knowledge of the voicing alternation in French, rather than on auditory or phonetic cues present in the stimuli. Because native listeners of French are accustomed to incomplete occurrences of devoicing assimilation in French (Snoeren et al. 2006), their word recognition system is not optimally prepared to compensate for categorical changes. In contrast, French listeners' word recognition system is very effective in compensating for categorical voicing assimilation. Do L2 learners represent phonological knowledge in the same way as native listeners? I turn now to the bilingual results, where American English listeners have been presented with the same stimuli and tasks.

5. Word detection in context: Second language results

5.1 Stimuli and procedure

In this experiment, the same stimuli, procedure and experimental setup as in the previous experiment have been used.

5.2 Participants

Twenty-nine American English speakers (10 men), late learners of French with long (n=12, >2 years) or short (n=17, <2 years) exposure have been tested; three of them (long exposure) had to be excluded directly from any further analyses because their status as late learners of French could not be guaranteed given their knowledge of other languages or a mixed language background. This left a total of 26 late L2 learners of French (long exposure: n=9; short exposure: n=17). They received a small monetary compensation for taking part. Length of exposure equals length of residence in France, and has been determined through detailed language background questionnaires. None of the included participants grew up bilingually nor had any early contact with French. None of them reported being fluent in another L2. Their mean age is 28.9 (SD=11) years. Their average age of first exposure to French is 12.5 years (SD=4.4), and their average length of residence is 3.7 years (SD=7). The difference in sample size is due to difficulties in finding American English speakers with long exposure to French.

5.3 Results

Five participants did not reach the inclusion criterion and were excluded from analyses, leaving a group of 21 participants (long exposure: n=8; short exposure: n=13). Results are presented for each group (according to the length of exposure) in Fig. 2 and 3. For simplicity, I will term the long-exposure group “advanced” and the short-exposure group “beginners”. However, this should not be misleading: both groups were highly proficient in L2 and could understand without any problems all the sentences presented in the experiment.

Detection rate is used as the dependent measure in ANOVAs by subjects and items, declaring the factors “exposure” (long vs. short), “voicing” (voicing vs. devoicing) and “condition” (viable, non-viable and no-change). A main effect of condition was visible by subject and by item ($F_1[2,38]=173.5$, $p<.0001$; $F_2[2,26]=110.9$, $p<.0001$), as well as an effect of exposure (marginal by subjects but significant by items [$F_1[1,19]=3.9$, $p<.061$; $F_2[1,13]=9.5$, $p<.009$]) – as well as a strong interaction between the two ($F_1[2,38]=8.6$, $p<.001$; $F_2[2,26]=14.8$, $p<.0001$). This interaction suggests that advanced and beginners behave differently on the various conditions as a result of exposure. There was no effect of voicing across both groups (both $p>.1$), indicating a similar detection rate for both voicing and devoicing across conditions. Mean reaction times by subjects range between 1273 ms and 1996 ms (mean RT for n=21: 1695 ms). Analyses of reaction times and detection values revealed only a marginal interaction of *RT* with the factor *voicing/devoicing* (M=1753 ms for voicing; M=1636 ms for devoicing, $F[1,20]=5.9$, $p<.05$).

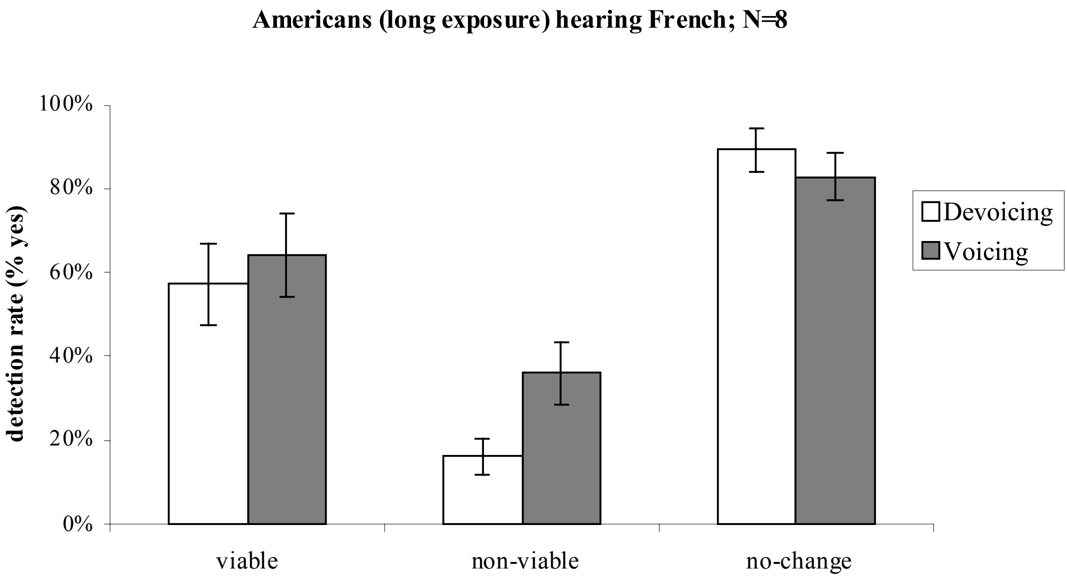

***Figure 2*:** Compensation results (% yes) in each assimilation type according to underlying voicing, for the advanced learners (N=8).

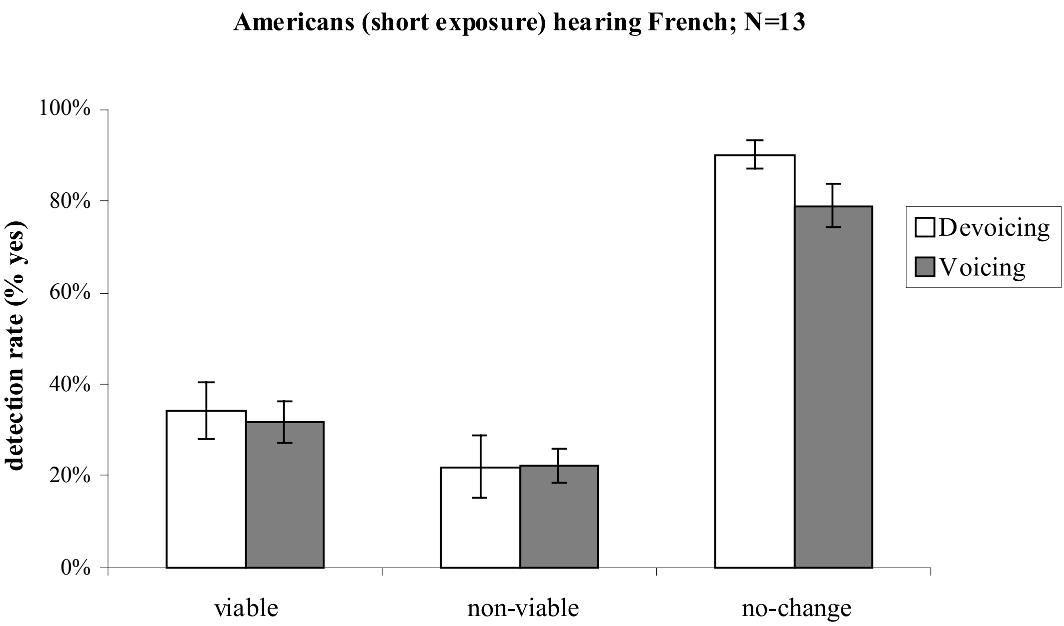

***Figure 3*:** Compensation results (% yes) in each assimilation type according to underlying voicing, for the beginner-level learners (N=13).

When compared directly on the viable condition, both advanced and beginners show a significant difference in compensation. An ANOVA on detection rate, restricted to the viable condition, declared the factor "exposure" and was significant by subject and item ($F_1[1,19]=11.0$, $p<.004$; $F_2[1,14]=22.0$, $p<.0001$). This supports the conclusion that beginners do not compensate for voicing/devoicing assimilation to the same extent as do advanced learners. Looking at the specific difference between voicing and devoicing in each group, an ANOVA declaring "voicing" (voicing vs. devoicing) and "condition" as within subject and between item factors revealed neither effect of voicing (both p >.1) nor any interaction with condition (both p >.1) in either group (advanced and beginner).

Comparisons between the learners and the French listener group yielded the following results. Overall, the advanced learner group does not differ significantly from the French native listeners, whereas the beginning learners do. A by subject and by item ANOVA on detection rates declaring the factor "L1" (French vs. American English) fails to produce significant L1 effects for the advanced group ($F_1[1,24]=0.002$, $p>.1$; $F_2[1,14]=0.005$, $p>.1$). In contrast, the same comparison with the beginner group yields a significant main effect of L1 ($F_1[1,29]=7.8$, $p<.009$; $F_2[1,14]=14.8$, $p<.002$). The same pattern remains visible when the analysis is restricted to the viable condition. A final comparison on detection rates, restricted to the viable condition, between the learner groups and the French native listeners includes the factors „voicing" (V/D) and „L1".There is a main effect of voicing ($F_1[1,37]=9.9$, $p<.003$; marginal by items: $F_2[1,13]=3.22$, $p<.09$) and of L1 ($F_1[1,37]=9.3$, $p<.004$; $F_2[1,13]=28.1$, $p<.0001$), as well as an interaction between both ($F_1[1,37]=9.6$, $p<.004$; $F_2[1,13]=10.1$, $p<.007$), suggesting that when hearing French, American English learners and French native listeners indeed do differ with respect to the voicing / devoicing pattern.

A closer look at individual data reveals the following interesting facts that might be linked to the asymmetry observed in the non-viable condition for advanced learners: Out of eight participants in the advanced group, four show a pattern of compensation that is similar in shape to native listeners of French (but only one shows a pattern similar in the amount of compensation with 50% for devoicing to 80% for voicing). They compensate less for devoicing (M=46%) than for voicing (M=81%) on the viable condition. For comparison, only one out of 18 French native listeners showed a higher compensation for devoicing than for voicing, both being high (85% vs. 65%). The four other participants show a flatter pattern on this condition, which tends to go in the other direction: they compensate more for devoicing (M=67%) than for voicing (M=46%). This asymmetrical pattern mimics the behavior of the native American English listeners when hearing their native language (see Darcy et al. 2007), even though the raw amount of compensation is higher in the present study (50% on average). Interestingly, the group of four that behave French-like on the viable condition maintains this compensation asymmetry also on the non-viable condition: they compensate on average at 10% for devoicing but as much as 46% for voicing. However, this high percentage might be due in particular to one individual (75% compensation for voicing, but 22% for devoicing). But it is interesting to see that these are the same learners. The less French-like four participants, by contrast, have a very flat pattern in this condition (21% vs. 25%). It is tempting to see in this pattern of results two different compensation strategies developed by learners; one group (French-like) seems to over generalize their compensation for voicing (which, as a reminder, is not present in their L1 phonology) to the non-viable context condition.

6. General Discussion

In this study, results have been reported about the way second language listeners recognize assimilated words and process assimilated sounds in a second language, as compared to native listeners of that language. After having validated the clarity and the symmetry of the stimuli through acoustic analyses and a perceptual identification task, I have established that native French listeners systematically compensate more for voicing assimilation than for devoicing assimilation despite similarly categorical stimuli.

In section 5, two groups of second language listeners have heard the same stimuli. The results for the group of advanced learners parallel those of native listeners, whereas the results for the beginners' group are different from the native listeners. Overall, advanced learners have acquired the pattern of alternation for French voicing assimilation, as shown by their pattern of compensation, whereas less advanced learners did not yet do so. This has been shown previously (Darcy et al., 2007). The new and more critical results for the present study reside in the difference between voicing and devoicing items visible in French native listeners. No group of second language learners evidenced any significant difference between voicing and devoicing items on the critical viable condition.

For French, when hearing categorical assimilations (which are less common in the case of devoicing assimilation), voicing is compensated for much more strongly than devoicing (78% vs. 51%, $p<.0001$), but in the viable condition only (when assimilation is appropriate). It is therefore proposed that this compensation pattern reflects a different representation for each alternation. Recall that categorical voicing assimilation is very common in naturally produced assimilation. More than half of them fall into the category “full assimilation” (Snoeren et al. 2006). By contrast, only 6% of naturally produced devoicing assimilation fall into this category. This means that the overwhelming majority of devoicing occurrences are not categorical, but gradient. Therefore, when it comes to mentally representing this alternation within phonological knowledge, one possibility would be that the voicing alternation, but not the devoicing alternation, is part of phonological knowledge. If this is the case, for voicing, compensation is expected to be high in the viable but not the non-viable condition. The predictions for devoicing are less clear-cut. If no devoicing alternation is represented abstractly, the straightforward prediction would be to expect no or very little compensation, as is the case for the non -viable condition. However, word detection rate for devoicing items in the viable conditions is as high as 50%. While it does not reject this first hypothesis, this relatively high but random compensation rate calls for an explanation. A different view would require the inclusion of phonetic gradiency within the representation. This would result in a less effective but still present compensation for categorical items. A direct prediction of it would be to see a compensation pattern for devoicing rising to a level comparable to voicing items as soon as gradient stimuli are presented (see Darcy & Kügler, *in preparation*).

From a learnability perspective, this second view might be more plausible than the first, as listeners do still need to cope for gradient and subphonemic variations in the form of words. It makes therefore some sense to expect that a high sensitivity to input patterns (in terms of categorical voicing and gradient devoicing) would result in a different representation for both kinds of alternations (Saffran, Newport & Aslin 1996). The evidence provided by these experiments is still preliminary, so I will refrain from exaggerated speculation.

An unresolved question is the level of encoding at which native listeners have represented this phonological knowledge, and what format such a representation might have. Several possibilities are compatible with the pattern of compensation seen in these experiments. Listeners could make use of an abstract pre-lexical knowledge to change the way they have interpreted segments before they recognize words. An assimilated /b/ perceived and categorized as a [p] for example, should be reinterpreted as /b/ in appropriate contexts in order to have access to the lexical representation. A different level at which such knowledge of alternations could come into play would be directly that of the lexical representations. It could be the case that listeners have encoded several variants of each lexical entry (for example one form [ʀop], another [ʀob], etc.), and activate them accordingly, depending on the upcoming context. The experiments presented here cannot distinguish between these possibilities. Similarly, the data obtained with both groups of learners are rather preliminary, because it is not clear from these types of experiments what their exact representation for alternations in a second language look like. The results, though, strongly suggest that these might be different. The goal of this paper was not to provide a complete characterization of the learners’ representations. Rather, the type of evidence I presented uncovers the need for further investigation, given that learners might indeed have a different, less detailed representation of phonological knowledge than native listeners.

When learning to compensate for assimilation in a second language, learners are faced with two tasks. First, they need to learn the phonological alternation, in order to reinterpret assimilation sounds in the appropriate contexts only, as do French native speakers. Second, they need to develop native-like representations and processing strategies. The pattern observed here suggests two conclusions: on the one hand, advanced learners are globally and statistically not different from French listeners in compensating for voicing assimilation. They seem to have acquired the knowledge of alternations appropriate for their L2, as the difference between less and more advanced learners shows. On the other hand, the implementation (see Weber & Cutler, 2006) of this acquired knowledge during speech perception seems to be different from what native speakers do, even in more advanced learners.

The asymmetry between voicing and devoicing observed in French listeners could arise in learners’ compensation pattern if they are sensitive to the asymmetry observed in French production data, leading them to represent both alternations accordingly. However, no asymmetry was visible in the learners' data. This suggests a different strategy in learning, or a difference in representation of alternations. They either did not learn the difference between voicing and devoicing, or they do not process categorical assimilation occurrences in the same way as French listeners do, due to different, less precise representations. They might have represented the assimilation rule in a more abstract way than the French, which would lead to a more efficient or overly efficient compensation for both processes equally. Such a trend is visible in advanced learners' compensation for voicing in both the viable and non-viable conditions, where the difference in detection rate for voicing as compared to devoicing is significant for the non-viable condition. They might have developed a wrong generalization about the voicing alternation in French. Indeed, this is what individual results suggest, even if individual data are difficult to interpret: Four participants behave asymmetrically in a French-like way on the viable condition, but they maintain this compensation pattern for words that are assimilated in a context that does not justify the assimilation. It is tempting to conclude that they might not have yet discovered the exact context for voicing assimilation in French. The results do not allow for conclusive answers to these questions, but give rise to many different questions. Further research will be needed to understand such important issues of how second language learners represent the knowledge they acquire, and whether they may be able to develop comparable representations and master processing routines which are as efficient as those used by native listeners for sounds (Sebastian-Gallés et al., 2006), for phonotactics (Weber & Cutler, 2006) and for phonological processes.

References

Cho, Y.-M. Yu (1999). *Parameters of consonantal assimilation*. München: Lincom Europa.

Cohn, Abigail C. (1995). Privativity, underspecification, and consequences for an adequate theory of phonetic implementation. *Working Papers of the Cornell Phonetics Laboratory* 10, 39 – 65.

Content, Alain, Mousty, Pierre & Radeau, Monique (1990) BRULEX Une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, 90, 551-566.

Crowther, C. S., & Mann, Virginia (1992) Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America* 92 (2), 711-722.

Darcy, Isabelle (2006) *Assimilation phonologique et reconnaissance des mots* [phonological assimilation and word recognition]. Bern, Berlin: Peter Lang.

Darcy, Isabelle & Kügler, Frank (in preparation) Compensation for Voicing and Devoicing Assimilation in French: evidence for perceptual asymmetry.

Darcy, Isabelle, & Kügler, Frank (2007) Phonological context effects for voicing and devoicing in French. In Trouvain, Jürgen and William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1257-1260). Saarbrücken.

Darcy, Isabelle, Peperkamp, Sharon, & Dupoux, Emmanuel (2007) Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. In Cole, Jennifer and José I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Darcy, Isabelle, Ramus, Franck, Christophe, Anne, Kinzler, Katherine & Dupoux, Emmanuel (to appear) Phonological knowledge in compensation for native and non-native assimilation. In: F. Kügler, C. Féry & R. van de Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology*. Mouton De Gruyter.

Féry, Caroline (2003). Gradedness in the segmental correlates of focus and prosodic phrasing in French. In T. Meisenburg, & M. Selig (Eds.), *Nouveaux départs en phonologie*. Tübingen: Narr.

Gaskell, Gareth, & Marslen-Wilson, William D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology, Human Perception and Performance,* 22, 144-158.

Goldsmith, John A. (1990). *Autosegmental and metrical phonology*. Cambridge, MA : Blackwell.

Gow, David W. 2002. Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology, Human Perception and Performance,* 28, 163-179.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science*, 15, 314–318.

Lombardi, Linda (1991). *Laryngeal Features and Laryngeal Neutralization*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Mack, Molly (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America,* 71, 173-178.

Marian, Viorica & Spivey, Michael J. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24, 173–193.

Mester, Armin & Itô, Junko (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language,* 65, 258 – 293.

Mitterer, Holger & Blomert, Leo (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics* 65, 956-969.

Pallier, Christophe, Colomé, Angels & Sebastián-Gallés, Nuria (2001). The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries. *Psychological Science*, 12, 445–449.

Saffran, Jenny R., Newport, Elissa L., & Aslin, Richard N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language,* 35, 606-621.

Sebastián-Gallés, Nuria (2005). Cross-language speech perception. In David B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Oxford, UK: Blackwell.

Sebastián-Gallés, Nuria, Echeverría, Sagrario, & Bosch, Laura (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52, 240–255.

Sebastian-Gallés, Núria, Rodríguez-Fornells, Antoni, de Diego-Balaguer, Ruth, & Díaz, Begoña (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience,* 18, 1277–1291.

Snoeren, Natalie, Hallé, Pierre, & Seguí, Juan (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics,* 34, 241-268.

Spivey, Michael J., & Marian, Viorica (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281–284.

Strange, Winifred (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Weber, Andrea & Cutler, Anne (2006) First language phonotactics in second language listening. *Journal of the Acoustical Society of America* 119, 597-607.

Weber, Andrea & Cutler, Anne (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25.

Wetzels, W. Leo, & Mascaró, Joan (2001). The typology of voicing and devoicing. *Language* 77, 207 – 244.