Looking for exemplar effects: testing the comprehension and memory representations of r'duced words in Dutch learners of French.

Lisa Morano[1], Louis ten Bosch[1], Mirjam Ernestus[1,2]

[1]Centre for Language Studies, Radboud University, Nijmegen, the Netherlands; [2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

{l.morano; l.tenbosch; m.ernestus} @let.ru.nl

Abstract

In this study, we tested whether second language (henceforth L2) learners can encode in the form of exemplars phonetic variation that does not occur regularly in their native language (henceforth L1). Three groups of Dutch learners of French performed a long-term repetition priming lexical decision task in which words were repeated. The second occurrence (target) of an experimental word either matched or mismatched the pronunciation of its first occurrence (prime). When a target matched its prime, both tokens had a completely devoiced or a completely voiced high vowel in their first syllable. When a target mismatched its prime, the prime had a devoiced high vowel in its first syllable, while the target had a voiced high vowel in its first syllable, and *vice versa*. In condition AA and in condition BB we reused the same token (albeit different tokens per condition) in case of a repetition match. In condition AB, we used different tokens for prime and target. The results show that L2 learners are able to encode phonetic information that does not occur regularly in their L1 in the form of exemplars, showing that exemplars are formed before the L2 phonological filter applies, but only under very limited conditions: when the prime is difficult to process and when the matching and mismatching tokens are easily distinguishable. Contrary to our expectations, we also found that mismatching devoiced primes significantly speeded the recognition of the voiced B targets. We hypothesize that this latter result comes from a higher activation of abstract representations after difficult primes. Our results thus show different processing patterns for identical testing conditions using different tokens (conditions AA and BB). These results question the use of exemplars in everyday speech comprehension, adding to the growing body of evidence that exemplar effects only arise in very restricted unnatural conditions.

Introduction

Many researchers now assume that the mental lexicon is hybrid in nature (Pierrehumbert, 2002; McLennan, Luce, and Charles-Luce, 2003; Goldinger, 2007), containing, for each word, both an abstract representation of the word's pronunciation (*i.e.* a string of abstract symbols such as phonemes), and a cloud of exemplars (*i.e.* occurrences encountered by the listener, each encoding fine acoustic characteristics such as speech rate, the speaker's voice, but also phonetic details). Indeed, purely abstractionist or purely exemplarist models of speech comprehension both fail to account for all the findings in the literature. For example, listeners' ability to adapt to a speaker's specific way of talking such as a lisp (*i.e.* perceptual learning; e.g., Norris, McQueen, and Cutler, 2003), or listeners' ability to generalize a phonological rule to new words (e.g., Cristia, Mielke, Daland, and Peperkamp, 2013), cannot be explained if their mental lexicons only contain exemplars of previously encountered tokens without any degree of abstraction. Evidence for exemplars, on the other hand, comes from priming experiments (e.g., Tulving and Schacter, 1990), in which it has repeatedly been shown that native listeners recognize words faster or more accurately when they occur for the second time in the experiment (as "targets") than when they occur for the first time (as "primes") especially if the two tokens share fine, phonologically irrelevant, acoustic characteristics such as information about the speaker's voice (*i.e.* both the prime and the target are uttered by the same person; e.g., McLennan and Luce, 2005). These specificity (or exemplar) effects suggest that the participants stored the first occurrences of the words with at least some degree of acoustic detail, that is, in the form of exemplars.

Nearly all experiments investigating exemplars have been conducted with native (henceforth L1) listeners. Exemplar research has barely studied second language (henceforth L2) learners. Nevertheless, there is much to gain from research with L2 listeners. First, if exemplars play a substantial role in speech comprehension, as most researchers currently assume, the findings obtained with L1 listeners should generalize to L2 listeners, as it is unlikely that listeners use two different mechanisms for speech comprehension in a L2 and in their L1. Second, research with non-native listeners may provide information about which acoustic detail is exactly stored in exemplars. Are exemplars faithful representations of the acoustic signal or are they affected by the listener's linguistic knowledge? That is, for L2 listeners, are exemplars formed before or after their L1 phonological filter (Troubetzkoy, 1939) applies?

It has been shown that L2 learners' abstract representations diverge from those of natives. Pallier, Colomé, and Sebastián-Gallés (2001) found that even highly proficient Spanish-Catalan listeners treat all minimal pairs specific to Catalan as homophones in a lexical decision task with medium-term auditory implicit repetition priming: for Spanish-Catalan bilinguals, [netə] 'granddaughter', primed equally well [netə] and [nɛtə] 'clean', and *vice versa*. Proficiency appears to play an important role, as was shown in another study. Darcy, Dekydtspotter, Sprouse, Glover, Kaden, McGuire, and Scott (2012) tested intermediate and advanced American English learners of French on two front *vs*. back rounded vowel contrasts in French (/y/-/u/ and /œ/-/ɔ/), which do not occur in English. In a lexical decision task with implicit repetition priming, both the intermediate and advanced learners patterned like the natives on the /œ/-/ɔ/ contrast, albeit with slower reaction times, while the intermediate learners, but not the advanced learners, treated the /y/-/u/ minimal pairs as homophones. This suggests that the intermediate learners did not distinguish /y/ and /u/ in their lexical representations, while the advanced learners did. These studies suggest that L2 phonological variation that is irrelevant in listeners' L1 is not immediately stored in listeners' L2 abstract representations, and that it may, or may not, eventually be stored abstractly at higher proficiency levels.

Exemplars in L2 listeners need not be different from exemplars in L1 listeners since L2 listeners have been shown to remain sensitive to L1 irrelevant contrasts provided the task employed could be performed without requiring lexical processing such as a phoneme categorization task (Sebatian-Galles and Baus, 2005; Diaz, Mitterer, Broersma, and Sebastian-Galles, 2012). That is, L2 listeners are able to perform simple low-level tasks in phonetic mode but as soon as linguistic processing is required, such as for a lexical decision task, then their L1 phonological filter prevents them from processing the stimuli in a native-like fashion (with the notable exception of Darcy et al.'s, 2012, results). If exemplars are formed before the phonological filter applies, L2 exemplars can thus well encode L1 irrelevant variation. If exemplars are formed after the phonological filter applies, L2 exemplars probably encode less L1 irrelevant variation. Our research question was the following: Are L2 intermediate learners able to encode, in the form of exemplars, fine linguistic details about the properties of the prime that are not relevant in their L1, and to subsequently use them for speech comprehension (*i.e.* to comprehend the target)?

As previously mentioned, very little exemplar research has been carried out in L2. We could only find two studies reporting exemplar effects for L2 listeners. Trofimovich (2005) tested American English learners of Spanish in an immediate repetition task. The participants

first listened to a list of 36 prime words uttered by three male and three female speakers (the study phase). The participants then performed a 3-4 minute distractor task, followed by an immediate repetition task (the test phase) in which all the primes were repeated (as targets) either in the same voice as during the study phase, or in a different voice from the opposite gender, along with new words. These tasks were performed twice: once in English and once in Spanish, the task order being counterbalanced over all the participants. In their L2, the participants were faster at repeating the words previously heard in the same voice than words which had not been presented during the study phase, but they were equally fast at repeating words heard for the first time in the experiment as words previously heard in the experiment in a different voice. The participants thus treated L2 words repeated in a different voice just as new items in the test phase.

In their L1, Trofimovich's participants showed priming but no exemplar effects: the participants were faster at repeating English words already heard in the study phase than words which had not been presented in the study phase, but it did not matter whether those words were uttered in the study phase in the same or in a different voice. Although Trofimovich's study did not replicate previous studies which found exemplar effects for native listeners (e.g. Craik and Kirsner, 1974; Palmeri et al. 1993; Luce and Lyons, 1998), it shows that exemplar effects can be found for L2 learners.

Further evidence that L2 listeners can store exemplars was provided by Winters, Lichtman, and Weber (2013). The authors tested three groups of listeners in German: English monolinguals, English learners of German, and German monolinguals in an old/new auditory categorization task. The stimuli were monosyllabic consonant-vowel-consonant (CVC) German words, which varied in frequency of occurrence (low, medium, high), and were uttered by five female voices in one block and five male voices in another block (the order being counterbalanced over participants). Within each block, half of the words were repeated either with the same or a different voice. The authors found that target words presented in the same voice as their primes were correctly classified more often than target words presented in a different voice, irrespective of the listener group.

L2 listeners are thus able to store details about the speaker's voice in the form of exemplars. This may not come as a surprise since L2 listeners already have ample experience in processing indexical variation in their L1, and it has been shown that the ability to use consistent information about a speaker's voice across items is easily transferable to L2 speech perception (Bradlow and Pisoni, 1999). The question is whether L2 listeners not only store in exemplars indexical information but also phonetic variation that occurs regularly in their L2

but not in their L1. While exemplar effects encoding indexical variation have already been attested by Winters, Lichtman, and Weber (2013) and Trofimovich (2005), to our knowledge, no previous study has found exemplar effects encoding L2 phonetic variation that does not occur regularly in the listener's L1. In this study, we tested whether exemplar effects in L2 listeners can also be found when manipulating regular phonetic variation instead of indexical (or speaker) variation.

One way to study exemplar effects for regularly occurring L1 specific phonetic variation instead of indexical variation is to focus on pronunciation variants of words resulting from reduction. Reduction is the weakening or deletion of phonemes or even whole syllables, occurring in informal connected speech, compared to the words' canonical pronunciations, that is the pronunciations of words in isolation (Ernestus and Warner, 2011). Most previous experiments investigating exemplar effects by manipulating linguistic variation focused on categorical variation, substituting one allophone with another allophone (e.g. [ɛ] with [e] in Pallier et al., 2001; and [t] and [d] with [ɾ] in McLennan, Luce and Charles-Luce, 2003 et al.). It could be argued that in these experiments listeners stored several abstract representations (one for each word pronunciation variant) rather than different exemplars. Using categorical variation thus makes it difficult to attest for the role of exemplars.

Reduction reflecting continuous variation, on the other hand, cannot be stored abstractly. Such reduction may result in an infinite number of realizations, which all activate the same abstract pronunciation variant of the word. Reduction reflecting continuous variation is thus an interesting characteristic to manipulate in order to test for unambiguous exemplar effects. To our knowledge, no previous study has done so.

In our study, we investigated the reduction phenomenon of phrase-medial high vowel devoicing. In casual French, in a noun phrase like *la cité* ([la.si.te] 'the city'), the /i/ can be more or less devoiced (up to completely) as the voicing (*i.e.* vibration of the vocal folds) fails to be re-established in time after the devoiced consonant /s/ (Torreira and Ernestus 2010). Furthermore, phrase-medial high vowel devoicing in French is a gradient phenomenon. In their corpus study, Torreira and Ernestus found that the high vowels were more devoiced or completely absent after certain consonants, the higher the speech rate, and the further away the vowel was from the end of the accentual phrase. Given that the same variables predict presence and amount of voicing, t absence of voicing is the end of a continuum that is reached in extreme devoicing conditions. This phenomenon has never been reported for Dutch, suggesting that it is part of the sound pattern of French but not of Dutch.

Consequently, if Dutch learners of French show exemplar effects in an experiment that manipulates phrase-medial high French vowel devoicing, we can conclude that L2 learners can also store, in the form of exemplars, L2 specific sound patterns.

We wished to use a task that requires deep processing of the stimuli to approach everyday speech processing. In our study, we used a lexical decision task. Although it can be argued that a lexical decision task is a very artificial task to investigate speech comprehension, it ensures a deeper linguistic processing than an old/new categorization task (or continuous recognition memory task) or a shadowing task, which are often used in exemplar studies (e.g. Craik and Kirsner, 1974; Palmeri, Goldinger, and Pisoni, 1993; Goldinger, 1996; Bradlow, Nygaard, and Pisoni, 1999; Trofimovich, 2005; Mattys and Liss, 2008; Winters et al., 2013). The words' forms need to be accessed to elicit responses from the participants: to decide whether a stimuli is a real word or not the participants need to access what the word means, even vaguely.

We tested Dutch intermediate learners of French in a lexical decision task in French in which the experimental words contained a high vowel following a voiceless consonant. The experimental words were all repeated either as a pronunciation match (*i.e.* both the high vowel of the prime and that of the target were devoiced, or both were voiced) or as a pronunciation mismatch (*i.e.* when the high vowel of the prime was devoiced, the one of the target was voiced and vice versa). If participants react faster to a target when it matches than when it mismatches the pronunciation of its prime, we can conclude that L2 participants show exemplar effects, indicating that they are able to store, in the form of exemplars, phonetic information that does not occur regularly in their L1, and to later on reuse those exemplars to comprehend the next token of the word.

We ran the same experiment three times. In condition AB, we used different recordings for prime (a voiced or devoiced token A) and target (a voiced or devoiced token B). As already pointed out by Hanique et al. (2014), using two different tokens (or recordings) for prime and target represents more ecologically valid testing conditions than using identical tokens, given that in daily life, we never hear the exact same token twice: in a conversation, if a person repeats a word, she will produce a new token that will vary slightly from the first one.

We compared this condition with two conditions in which the prime and target were identical in case of a match (like in nearly all the previous studies on exemplar effects): one using only the tokens used in the first condition as primes (condition AA), and one using only the tokens used in the first condition as targets (condition BB).

Method

Participants

We tested 120 Dutch university students who had studied French for four to seven years in high school (intermediate level, or B1-B2 levels of the Common European Framework of Reference for Languages, CEFR, Council of Europe, 2011) and who were paid for their participation. The participants were between 18 and 29 years old (mean: 21.74), 95 were female and 105 were right-handed. None of the participants reported any hearing problems. The participants were randomly assigned to one of the three conditions (AB, AA, BB).

Materials

Our experimental words were selected from the vocabulary of two beginners textbooks used in French classes at Dutch secondary schools (*Franconville* and *Grandes Lignes*). They were bisyllabic words containing a high vowel (/i/, /y/, or /u/) following a voiceless consonant in their first syllable (*cf.* Appendix 1). Out of all possible words, we selected the 24 most frequent words, with a preference for those containing /i/ and /y/ as these vowels are more constricted than /u/, which allows them to be more easily devoiced than /u/ (Meunier, Meynadier, and Espesser, 2008)[1]. The frequency of occurrence of our experimental words in the movie subtitles corpus of Lexique 3.81 (New, Pallier, Ferrand, and Matos, 2001) ranged from 0.71 (per million words) for *cycliste* 'cyclist' to 107.92 for *sujet* 'subject'(mean: 31.40, *cf.* Appendix 1), that is, they were fairly frequent words (most of them ranging between the median at 8 occurrences per million words and the third quartile at 43 occurrences per million), which is normal for beginners' vocabulary words.

We also selected 78 bisyllabic frequent words, without particular restriction, from the above mentioned beginners textbooks to be used as existing-word fillers. Finally, we created 102 bisyllabic pseudo-word fillers by adding a phonotactically legal syllable to the first syllable of all the experimental and existing-word fillers already selected.

All the stimuli, preceded by their definite determiners, were recorded in a sound attenuated booth with a head mounted microphone at 44.100Hz by the first author of this paper, a female French native speaker from Caen. The easiest way to obtain fully devoiced

[1] One of the reviewers attracted our attention to the fact that the participants may not process the devoiced vowel at all despite the remaining durational and formant cues signaling the presence of the vowel (as it has been shown to happen for German natives listening to Japanese accented German; Zimmerer, Rei, and Reetz, 2013). In that case, three items could be confused with other French words (*purée* 'mashed potatoes' could be confused with *pré* 'meadow'; *pilote* 'pilot' with the reduced form of *pelote* 'woolen ball'; and *poulet* 'chicken' with *plaie* 'wound'). However, the occurrence frequencies of the possibly confounded words (*pré* 'meadow'; *pelote* 'woolen ball'; *plaie* 'wound') are all lower than the occurrence frequencies of our stimuli, making it unlikely that our participants knew these words.

high vowels in the first syllables appeared to have the speaker produce all the experimental words without their determiners. In this way, for the devoiced (short for 'containing a devoiced high vowel in the word's first syllable', also in the rest of the text) recordings, the speaker could comfortably whisper the first syllables and then voice the second syllable, while for the voiced (short for 'containing a voiced high vowel in the world's first syllable', also in the rest of the text) recordings, she could just speak out loud the whole words. The first vowel of the devoiced stimuli was always completely devoiced and the first vowel of the voiced stimuli was always fully voiced (*cf.* Figure 1). The speaker also recorded all the experimental words with their determiners. The best devoiced and voiced recordings without determiners were then each paired with their closest voiced recordings with determiner in terms of intonation and duration. The final stimuli were obtained by cross-splicing the voiced determiners with the devoiced and voiced recordings without determiners.

We created two tokens for each voicing type, meaning that for each experimental word we obtained four tokens: a voiced token A, a voiced token B, a devoiced token A, and a devoiced token B. Tokens A were on average 805ms long (804ms for the voiced ones, Standard Deviation (henceforth SD) = 106, and 806ms for the devoiced ones, SD = 124) and tokens B were on average 811ms long (796ms for the voiced ones, SD = 134, and 826ms for the devoiced ones, SD = 136). Note that for the B tokens, it is not the case that the devoiced form was always longer than the voiced one (*cf.* Appendix 1 for the durations of all individual tokens). The existing-word fillers and the pseudo-word fillers were not cross-spliced but two tokens were recorded per word-type. The average duration of the existing-word fillers was 719ms (SD = 120) and of the pseudo-word fillers 739ms (SD = 128).

Finally, all the stimuli were scaled to 70 dB of average intensity. All the stimulus recording, editing, and scaling was done in Praat (Boersma and Weenink, 2017).

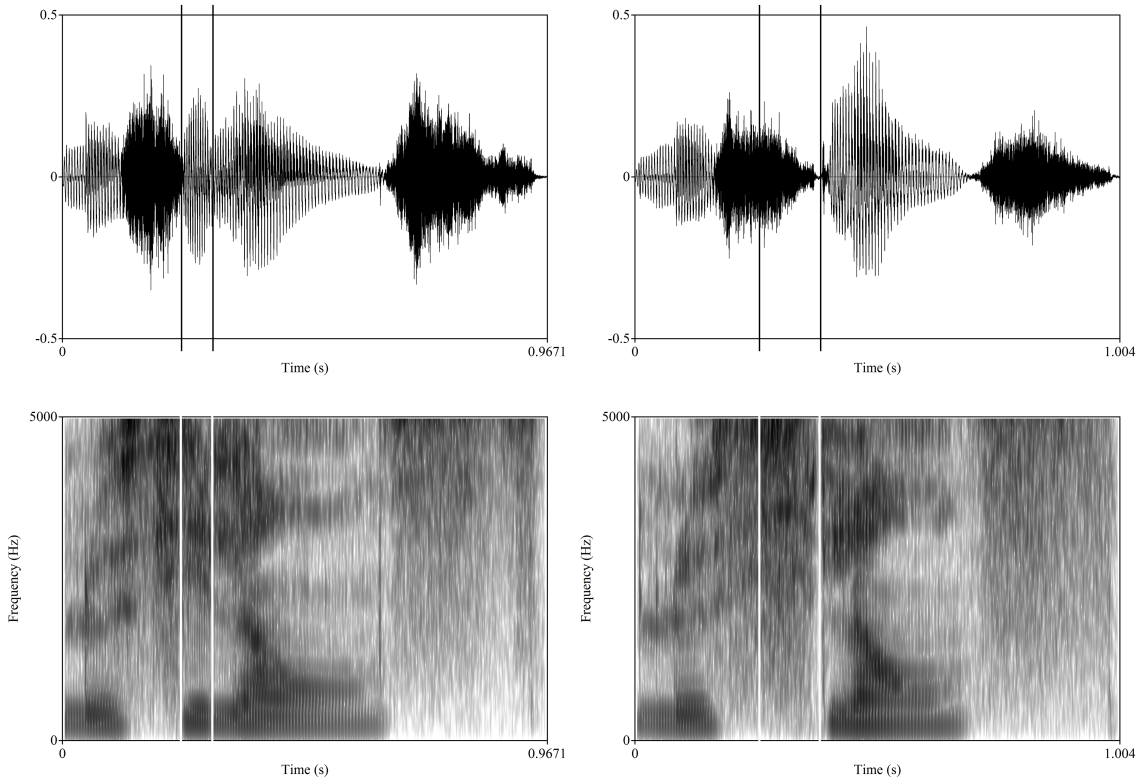

Figure 1. Waveforms (top panels) and spectrograms (bottom panels) of the target word *le silence* 'the silence': voiced token A on the left, and devoiced token A on the right. The high-vowel /i/ boundaries are indicated by the vertical lines.

The lexical decision task consisted of two blocks of 132 trials each. Twelve of the experimental words were presented in the first block and twelve in the second block. Within each block, the experimental words were repeated either as a variant match (*i.e.* both prime and target had either voiced or devoiced vowels) or as a variant mismatch (*i.e.* when the prime was voiced, the target was devoiced, and vice versa). The prime and target were separated by seven to 98 trials (average: 65), replicating the lags used in the first and third experiments of Hanique, Aalders, and Ernestus (2014). Although these lags are not as long as the ones used by Goldinger (1996), who found exemplar effects one week after presentation of the prime, they are long enough to ensure that our results could not stem from the participants holding the primes into their working memories until they could process the target.

The remainder of the trials per block included 36 bisyllabic real-word fillers (of which six were repeated), and 48 bisyllabic pseudo-word fillers (of which 18 were repeated). Finally, six real-word fillers and six pseudo-word fillers were used for practice trials, with two real-word fillers and two pseudo-word fillers being repeated. The practice trials were the

same for all the participants, and they were very similar in frequency of occurrence and phonological structure to the stimuli in the experiment.

We created five pseudo-randomizations of the trials: a block never started with an experimental word; there were never two experimental words in a row; there were never more than eight pseudo-word fillers in a row; and a prime and a target were never separated by more than 100 trials. For each pseudo-randomization, we then created four different stimulus lists that kept the trial order obtained by pseudo-randomization constant and differed only regarding the voicing type of the experimental words. In each of the four stimulus list, the primes and targets of half of the experimental words occurred in the same pronunciation variant (six voiced ones, and six devoiced ones), and those of the other half showed a difference in voicing (six voiced primes followed by devoiced targets, and six vice versa). Consequently, across all four stimulus lists created from one pseudo-randomization, each experimental word was tested for each of the four possible matching and mismatching combinations. Each of the 20 lists created in total was randomly assigned to two participants per condition.

In Condition AB, we used different recordings (or tokens) for the primes and the targets, so that even in case of a match, the prime (token A) and the target (token B) were different recordings. As shown in Table 1, in the condition AB, the primes and targets matched in pronunciation variant but diverged in terms of duration. In condition AA and in condition BB, we only used the tokens A and B, respectively, so that in case of a match, prime and target were the same token and thus did not differ in duration (hence the zeros in Table 1).

Condition	*Match*	*Mismatch*
AB	44 (25)	53 (36)
AA	0	49 (30)
BB	0	50 (35)

Table 1. Average absolute temporal differences (in milliseconds) between primes (voiced and devoiced) and targets (voiced and devoiced) per condition. Standard deviations are given between parentheses.

Procedure

The participants were tested individually in a sound attenuated booth equipped with headphones, a mouse, and a button box with stickers *JA* 'yes' / *NEE* 'no' on the buttons. The participants first signed a consent form and filled in a language background questionnaire, before doing the lexical decision task. The lexical decision task was presented with PsychoPy

(Peirce, 2007). The participants were instructed to indicate as fast as possible with the button-box, and using their dominant hand, whether the word they heard over the headphones was a real word in French or not. The instructions insisted that the participant did not need to know the exact meaning of the word in order to press the 'yes' button but that they had to be certain that the word occurred in French. The next trial initiated 1000ms after the participant's answer or 3500ms after the onset of the preceding stimulus in case the participant did not react. In order to increase motivation and discourage guessing, the participants received feedback in percentage accuracy at the end of each block. The whole experimental session lasted a little less than half an hour.

Results

One participant in condition AB and one participant in condition BB were removed from the dataset since their accuracy on the experimental words in the lexical decision task was below chance level (43.75% and 33.33%, respectively).

We analysed all the data from this study using the software R (R Development Core Team, 2007). All the trials to which the participants did not react were discarded (ten out of the 5664 experimental word trials). Accuracies were analysed by means of a mixed effects model for logistic regression (Jaeger, 2008), for which the dependent variable was the probability of a correct response. Reaction times (RTs; measured from word offset) to correct trials within 2.5 standard deviations from the targets' grand mean (345ms; discarding 52 data points out of 1768; 3% of the data) were analysed by means of mixed effects regression models (Baayen, Davidson, and Bates, 2008). Prior to analysis, all RTs and stimulus durations were log-transformed. Our dependent variable for the linear mixed effect model was thus the log-transformed RT. We used item and participant as crossed random effect factors.

Our predictors of interest were Voicing (a categorical predictor indicating whether the first high vowel of the stimulus was voiced or devoiced), Condition (AB, AA, and BB) or Token (A or B), and Repetition match (*i.e.* whether the prime and target of the experimental word were of the same pronunciation variant). Since Condition and Token overlap considerably in terms of the variation they explain, for each model reported, we compared two variants of our best model: one using Condition and one using Token in order to select the best of the two predictors. We retained in our final model the predictor which lowered the Akaike Information Coefficient (AIC) of the model by at least two points.

Our control predictors were: log Stimulus duration, Trial number (*i.e.* the position of the trial in the experiment, in order to control for learning or fatigue effects), Distance (lag) between prime and target (in number of intervening trials), log RT to the previous trial (so as to control for local speed effects), and log RT on the prime. The continuous and discrete numerical predictors, that is, all the control predictors, have been centered around the mean.

We first fitted a simple main effects model with all the predictors relevant for the dependent variable. Interactions were then tested between the predictors of interest only. To obtain the most parsimonious yet adequate model, only predictors and interactions which showed significant effects (*i.e. t* or *z* with an absolute value exceeding 1.96) were retained in the final models. Predictors which were significant in an interaction but not as main effects were kept in the models as well. Once the fixed effect structure was finalized, random slopes on item and participant were tested for all fixed effects. A random slope was kept in the final model exclusively when supported by likelihood ratio tests (i.e. $p < 0.05$). Finally, following Baayen (2008), to make sure no significant effect was driven by outliers, the final RT model was refitted: RTs with residual standard errors more than 2.5 standard deviation units were excluded from the dataset of the final statistical model (49 data points were removed out of 1716; 3% of the dataset). No predictor lost significance as a result of this refitting of the model[2]. The p values reported were obtained with the lmerTest package version 2.0-36 (Kuznetsova, Brockhoff, and Christensen, 2017).

Accuracy data

The participants' accuracy was quite high although not at ceiling (83.92% overall, with 85.52% accuracy for the pseudo-word fillers, 86.13% for the real-word fillers, and 75.70% for the experimental words), showing that they took the task seriously. Participants' lower accuracy on the experimental words was probably due to the fact that the experimental words were less frequent than the real-word fillers and thus less familiar to the participants.

First occurrences

We first verified whether the participants were sensitive to the devoicing manipulation. To do so, we looked at the participants' accuracy on the primes only (N = 2824), since the participants' accuracy on the targets might have been influenced by whether

[2] One of the reviewers suggested that we use the Median Absolute Deviation (MAD; Leys, Ley, Klein, Bernard, Licata, 2013) to prune our data instead of first discarding outliers 2.5 Standard Deviations from the targets' mean RT and then discarding again outliers deviating more than 2.5 standard units from the predicted values before re-fitting the model. An analysis of our RT data using the MAD is provided in Appendix 2. Importantly, both analyses find the same predictors significant. Thus, both analyses come to the same conclusions.

the targets matched or mismatched their primes. The results are presented in Table 2. The participants were significantly more accurate on the voiced (75.79%) than on the devoiced (66.42%) tokens A, as indicated by a simple effect of Voicing (*cf.* Table 2), while the difference was not statistically significant for tokens B (75.80% accuracy on the voiced tokens and 73.49% on the devoiced ones), as shown by releveling the variable and rerunning the model ($\beta = 0.16$, S.E. = 0.23, $z = 0.67$, $p>0.1$), and as indicated by the significant interaction between Voicing and Token (*cf.* Table 2). We also found a significant random slope of Voicing on Item, which indicates that the effect of Voicing was significantly larger for some items than others.

In sum, the participants were thus clearly sensitive to the devoicing manipulation for the tokens A, but not for the tokens B. That is, to the participants, the devoiced and voiced tokens A were more distinguishable from one another than the devoiced and voiced tokens B, although this was more the case for some experimental words than for others.

Fixed effects		**B**	**SE**	**t**	**p<**
(intercept)		0.96	0.29	3.36	0.001
Token	B	0.46	0.19	2.41	0.05
Voicing	voiced	0.63	0.19	3.26	0.01
Voicing * token	voiced * B	-0.48	0.21	-2.29	0.05
Random effects			**Variance**	**SD**	
Item	Intercept		1.65	1.28	
	voicing		0.49	0.70	
Participant	Intercept		0.40	0.63	

Table 2: Statistical model fitting the probability of a correct response to the primes. N = 2824. Standard error is indicated by SE. The intercepts represents devoiced A tokens' first occurrences. Predictors and random slopes that did not reach significance at the 5% level were not retained in the model and are not listed in the table.

Second occurrences

Given that the participants were sensitive to the devoicing manipulation (at least for the tokens A), we can now investigate whether the participants were more accurate on matching than on mismatching targets. When only considering the targets whose primes were answered to correctly (N = 2041), there appeared to be no effect of Repetition match on accuracy, neither as a main effect nor in interaction with Condition or Token.

Reaction Time data

The RT data suggest priming across all conditions (*cf.* Figure 2): when the participants correctly classified both the prime and the target of the experimental word as real words, they were on average 106 milliseconds faster on the target (345ms) than on the prime (451ms). Note that all RTs are from word offset.

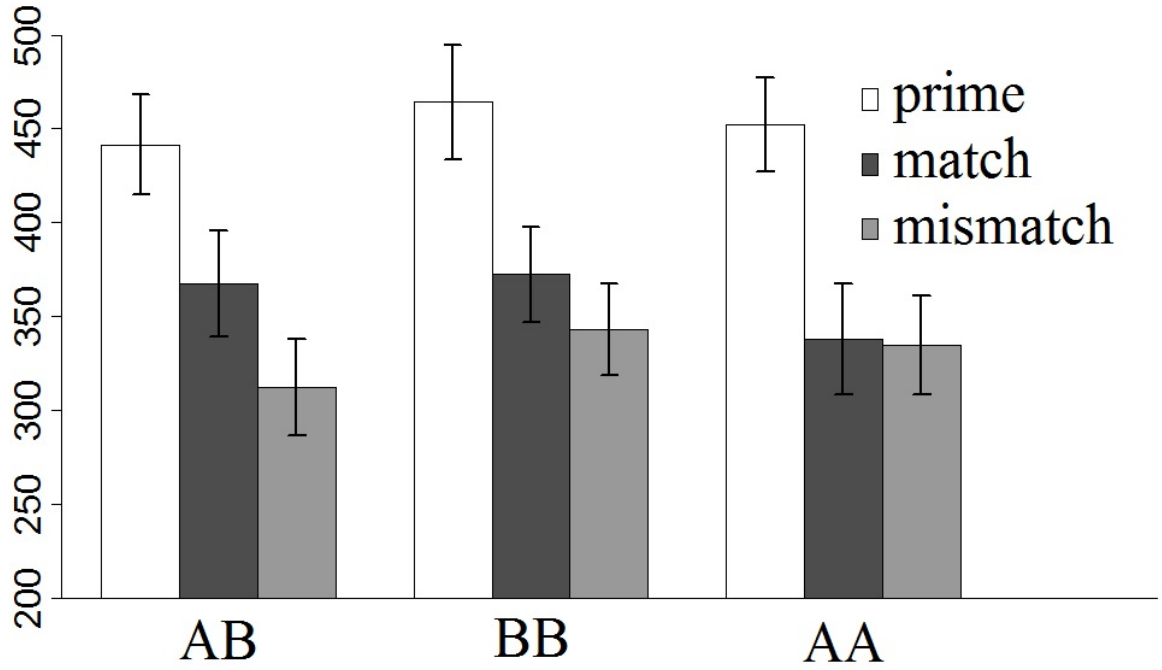

Figure 2: Reaction times (in milliseconds) from word offset for the experimental primes and targets (in match and mismatch cases) when both have been answered to correctly, by condition. Error bars: 95% confidence intervals. N = 3454.

We analysed statistically the RTs to the targets answered to correctly, provided their primes had also been answered to correctly. The results are presented in Table 3. Almost all our control predictors showed significant effects. The participants were faster at answering targets when they also answered quickly on the previous trial; when they had recognized the prime quickly; when the number of intervening trials between prime and target was low; and when the stimuli were short.

More importantly, all of our factors of interest also showed significant effects. The effect of Repetition match differed between the conditions AB *vs.* AA ($\beta = 0.17$, S.E. = 0.05, $z = 3.27$, $p<0.01$) and BB *vs.* AA ($\beta = 0.15$, S.E. = 0.05, $z = 3.00$, $p<0.01$), as shown by releveling the variable and rerunning the model. Given that the conditions AB and BB thus patterned together against the condition AA (*cf.* Figure 2), it is not surprising that Token of

the target (A or B) was a much better predictor than Condition (the model with Token had an AIC ten points lower than the AIC of the model using Condition).

Fixed effects		**β**	**SE**	**t**	**p<**
(intercept)		5.74	0.05	113.00	0.001
Repetition match	Match	-0.12	0.04	-2.82	0.01
Token	B	-0.05	0.05	-1.10	n.s.
Voicing	Voiced	-0.18	0.04	-4.35	0.001
Number of trials between prime and target		0.002	0.0006	2.69	0.01
Stimulus duration (ms logged)		-1.13	0.16	-7.09	0.001
RT to the preceding trial (ms logged)		0.17	0.03	6.23	0.001
RT to the prime (ms logged)		0.30	0.02	15.89	0.001
Repetition match * Voicing	match * voiced	0.12	0.04	2.90	0.01
Repetition match * token	match * B	0.16	0.04	3.68	0.001
Random effects			**Variance**	**SD**	
Item	Intercept		0.02	0.15	
	Voicing		0.02	0.14	
	RT to the preceding trial		0.007	0.08	
Participant	Intercept		0.04	0.19	
	Stimulus duration		0.13	0.37	
	RT to the preceding trial		0.01	0.11	
Residual			0.17	0.42	

Table 3: Statistical model fitting the log-transformed response times (measured from word offset) to the targets provided their corresponding primes had been answered to correctly. N = 1667 after removal of the outliers. Standard error is indicated by SE. The intercept represents the reaction time to a devoiced target A mismatching its prime. Predictors and random slopes that did not reach significance at the 5% level were not retained in the model and are not listed in the table.

We also found a main effect of Voicing (see Table 3), without an interaction of Voicing with Token: participants were slower at processing devoiced targets, independently of whether the targets were token A or token B (*cf.* Figure 3). That is, contrary to the *Accuracy data,* which showed that the participants were only sensitive to devoicing for the tokens A, the RT data show that the participants were sensitive to the devoicing manipulation for both tokens A and tokens B. L2 listeners were thus sensitive to L1 irrelevant information. Interestingly, the significant main effect of Token without an interaction of Voicing and Token indicates that the tokens B were processed significantly faster (*i.e.* were easier to comprehend for the participants) than the tokens A, independently of whether the tokens were voiced or not.

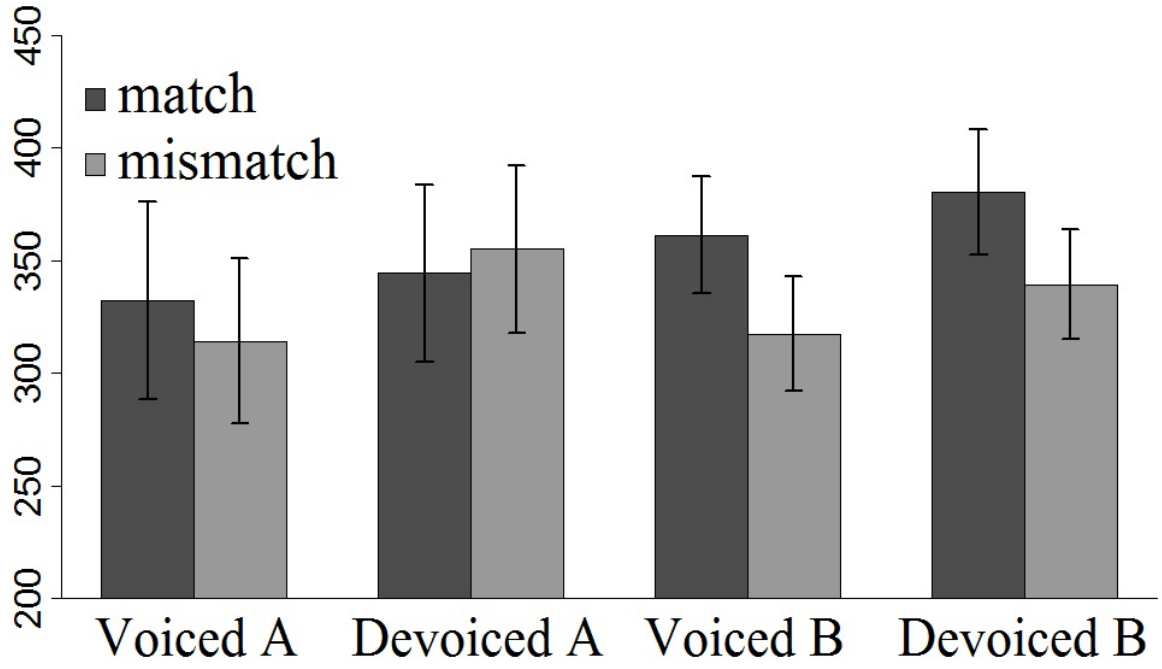

Figure 3: Reaction times (in milliseconds) from word offset for the experimental targets which have been answered to correctly both at prime and target, grouped by voicing and by token. Error bars: 95% confidence intervals per bar. N = 1727.

Repetition match was significant in interaction with Token on the one hand (as previously mentioned) and with Voicing on the other hand. The three way interaction was not significant ($\chi^2(2) = 0.79$, $p>0.1$). The significant simple effect of Repetition match indicates that when the target was the devoiced token A, the participants were faster at answering the target when it matched its prime than when it mismatched it prime.

The significance of Repetition match in the other three cases (*i.e.* when a devoiced B target matched its prime, when a voiced B target matched its prime, and when a voiced A target matched its prime), is difficult to assess from Table 3 given the separate significant simple effects of Voicing and Token on the one hand, and their significant interactions with Repetition match on the other hand. In order to understand the overall effect of Repetition match, we analysed the different contrasts using releveling. By releveling, the model does not change, but the mathematical formulation makes it possible to determine the simple effects in the other three cases. We placed alternatively on the intercept of the model reported in Table 3, the voiced tokens A, the voiced tokens B, and the devoiced tokens B. Only for the voiced tokens B we found a significant effect of Repetition match ($\beta = 0.17$, S.E. = 0.03, $z = 5.02$, $p<0.001$), indicating that the participants were significantly slower when the voiced targets B matched their primes (either voiced primes A or voiced primes B). In other words, the

participants were significantly faster when the voiced targets B mismatched their primes than when the voiced targets B matched their primes. For both the voiced tokens A and the devoiced tokens B, the main effect of Repetition match was not significant.

In sum, the participants were sensitive to the devoicing manipulation as they were less accurate on the devoiced than on the voiced primes A, and they were slower on both the devoiced A and B targets than on the voiced A and B targets. Repetition match showed no effect in the *Accuracy data,* possibly because of lack of statistical power. In the RT data, the A and B tokens patterned differently regarding the effect of repetition match: the devoiced A tokens were answered to faster when they were preceded by a matching prime, while the voiced B tokens were answered to significantly faster when they were preceded by a mismatching prime. In other words, devoiced primes always speeded the participants' RTs on the targets, while voiced primes never led to any significant differences in RTs between a matching and a mismatching target.

General Discussion

This study investigated whether L2 learners show exemplar effects for variation in the acoustic signal that they are not familiar with from their L1. If exemplars are formed after the L1 phonological filter applies, L2 exemplars do not differ from L1 exemplars regarding indexical variation, but only regarding L1 irrelevant linguistic variation.

We tested Dutch intermediate learners of French in a lexical decision task in which words were repeated *(i.e.* using long-term implicit repetition priming) in the same (match) or in a different (mismatch) pronunciation variant. Our experimental words were French words whose first vowel was voiced in one pronunciation variant (henceforth voiced word tokens) and devoiced in the other (henceforth voiceless word tokens). Vowel devoicing is not a characteristic of Dutch and thus linguistically irrelevant for Dutch native listeners.

In order to investigate whether the match effect is not only present under the conditions normally tested in exemplar experiments, but also under more ecologically valid conditions, we tested three conditions. In two conditions (AA and BB), the prime and target were identical tokens in the pronunciation match case. These two conditions follow the vast majority of the previous literature on exemplar effects (which reuses the same token). In a third condition (AB), the primes and targets were always different instantiations, so that, even

when an experimental word was repeated as a pronunciation variant match, it was nevertheless a different token, just like in everyday conversations.

Our data suggest an exemplar effect for the devoiced A targets, since the devoiced A tokens were answered to faster when they were preceded by devoiced A primes than by voiced A primes. This match effect shows that L2 listeners are able to encode and store in the form of exemplars phonetic variation that does not occur regularly in their L1 (vowel devoicing). Exemplars thus seem to be formed before the phonological filter applies and to faithfully represent the acoustic signal. The information they encode is probably the same for both native and non-native listeners.

If exemplars are formed before the phonological filter applies, one may wonder whether exemplars are part of the mental lexicon. This question has also been raised by Goldinger (2007), Cutler, Eisner, McQueen, and Norris (2010), Ramus, Peperkamp, Christophe, Jacquemot, Kouider, and Dupoux (2010), and Nijveld, ten Bosch, and Ernestus (2015), among several authors, who hypothesise that exemplars are stored in episodic memory, which is a general type of memory (Tulving, 1985). Episodic traces are detailed memory representations which are context-dependent in the sense that they encode specific events (e.g. listening to a word, watching a movie, hurting one's toe) with their context (e.g. which voice uttered the word, at which row one was seating, how early it was). If exemplars are faithful representations of the acoustic signal, they are likely to be part of episodic memory.

The significant interaction we found between Repetition match and Token indicates that our participants used different processes to comprehend the B and the A tokens. Although conditions AA and BB both used identical tokens for matching primes and targets, they did not pattern in the same way in the participants' RT behaviour on the targets. Rather, the BB condition patterned with the AB condition. In both conditions, there were no exemplar effects. It is thus not the fact that the prime and the target were identical that led to exemplar effects. These results are in contrast with all previous studies on exemplar effects, including Hanique et al.'s (2014), which showed that exemplar effects can also arise when the prime and target are different tokens in the match condition.

Various explanations have been put forward to explain why exemplar effects arise in certain conditions and not in others. One hypothesis is that exemplar effects occur when speech processing is slow, such as when listening to dysarthric speech (Mattys and Liss, 2008), or when real words need to be distinguished from very real word-like pseudowords (McLennan and Luce, 2005). This time-course hypothesis (McLennan and Luce, 2005) can

explain the presence *versus* absence of exemplar effects as the participants were slower on the targets A than on the targets B.

The time-course hypothesis, however, cannot account for mismatch effects. Our data showed one mismatch effect. Participants responded more slowly to voiced B tokens when they were preceded by voiced than devoiced tokens. This raises the question where this effect comes from. This is an important question since it may provide some insight into the conditions leading to exemplar effects, and therefore to the nature of exemplar effects. The difference in results between conditions AA (match effect for devoiced tokens) and BB (mismatch effect for voiced tokens) is the most interesting one, since both conditions used identical tokens for prime and target and it is therefore not obvious what drives the difference in response pattern.

It may be the case that the difference in response pattern is due to subtle acoustic differences between the set of A tokens and the set of B tokens. The voiced and devoiced tokens were probably more different from each other in condition AA than in condition BB. The selection of the tokens for the primes and target for condition AA was made before the selection of the tokens for condition BB and from the same pool of recordings. Consequently, for the cross-splicing of tokens B, the first author had fewer recordings to choose from than for the cross-splicing of tokens A, which probably caused voiced and devoiced tokens A to be better matched than voiced and devoiced tokens B on other acoustic characteristics than devoicing. This was definitely true for stimulus duration (*cf.* Appendix 1): the voiced and devoiced tokens A only differed by 2ms on average, while the voiced and devoiced tokens B differed by 28.5ms on average[3].

To further investigate potential differences between the voiced and devoiced tokens which might have caused our asymmetric results in the AA and BB conditions, we conducted a post-hoc spectral comparison of all voiced and devoiced tokens, using the differences along the Mel Frequency Cepstral Coeficients alignment path, time warped. The results are summarized in Appendix 3. We found that the voiced and devoiced A tokens differed more from each other than the voiced and devoiced B tokens. However, this difference was not significant ($t(45) = -0.27$, $p>0.1$) probably because of lack of statistical power. Consequently, it is possible that the difference between voiced and devoiced vowels stood out less clearly for the B tokens than for the A tokens, especially given the accuracy differences found

[3] This difference in stimulus duration probably stems from a difference in the duration of the high vowel (*cf.* Figure 1). Importantly, 28.5 ms are above the threshold of just noticeable differences for vowel duration (Quené, 2007; Nooteboom and Doodeman, 1980).

between the voiced and the devoiced primes: the participants were about 9%, and significantly, more accurate on the voiced than on the devoiced primes A, but only 2% more accurate on the voiced than on the devoiced primes B, and this latter difference was not statistically significant.

The participants' significantly lower accuracies on the devoiced A primes compared to all other primes, in combination with their significantly lower RTs on both the A and B devoiced targets compared to the voiced targets could explain our pattern of results. On the one hand, the difficulty of processing of both the A and B devoiced tokens could have led the participants' abstract representation to reach a higher level of activation (as activation only increases over time, e.g. Norris and McQueen, 2008) than after the processing of a voiced prime (for which activation stopped to increase as the word was recognized earlier in time). When a voiced target then followed a devoiced prime, the ease of processing of the voiced forms combined with the high activation of the abstract representation, led to a quicker answer on a mismatching than on a matching target. On the other hand, the fact that the devoiced A primes were particularly difficult to comprehend could have led to stronger individual memory traces (or exemplars) being encoded for the devoiced A than for the devoiced B primes. In turn, these highly activated exemplars would then be easy to retrieve and to match to the particularly distinguishable devoiced A targets. When both the prime and target were devoiced tokens, the participants could thus more easily use the exemplar formed with the prime in condition AA than in condition BB. This would explain why there was only a match (exemplar) effect in condition AA with devoiced targets.

This explanation of our asymmetric results would be in line with other studies which propose that listeners may display exemplar effects only under testing conditions that encourage participants to rely on their recent (or episodic) memory. Luce and Lyons (1998) found exemplar effects in an old/new categorisation task, which explicitly requires the participants to make use of their recent memory, but not in a lexical decision task. Hanique et al. (2014) only found exemplar effects in a lexical decision task when it was crystal clear to the participants that tokens were repeated (when the percentage of repeated tokens was high and the number of intervening trials between the prime and the target remained low). Moreover, they only found exemplar effects when manipulating only linguistic and not both linguistic and indexical variation within one experiment. Thus, if the stimuli included too much variation, like the tokens B in our experiment did, no exemplar stood out from the other episodic traces, and consequently no exemplar could be reused in the matching conditions.

Other types of variation have been shown to influence the presence of exemplar effects. For example, confusability between vowels categories has been shown to hinder the benefits of High-Variability training on vowels' identification (Wade, Jongman, and Sereno, 2007), while High-Variability training benefits are traditionally explained with more exemplars creating a more robust category as a cloud than individual exemplars. It thus seems that to produce effects, exemplars need to be clearly recognized or labelled by the listener as belonging to two separate clouds or categories.

So far, we explained our results within models assuming hybrid lexicons. Some other recent models of speech perception answer the problem of the lack of invariance of the speech signal by focusing on how listeners integrate incoming information from the input with their own predictions over the same speech signal, depending on the situation. For example, in their 'ideal adapter' framework, Kleinschmidt and Jaeger (2015) propose that listeners constantly learn from details in the speech signal to immediately adapt their expectations about the incoming input. Whereas this framework accounts well for adaptations to differences among individual speakers stemming from regular and suprasegmental variation within the speech input, it is less clear which predictions it would make with regard to adaptation to irregular phonetic variation. In our study, participants probably noticed that words were repeated, however, they could certainly not predict whether the target would match or mismatch its prime. In the absence of certainty, we may expect listeners not to adapt, and thus to rely on their abstract representations, representing the full forms. Consequently, voiced B targets should benefit from a matching voiced prime (meeting the listeners' long-term expectations of the listeners). However, this is not what we found. In the AB and BB conditions, a mismatching prime speeded the recognition of its voiced target. Our design, however, is not best suited to test the predictions of the 'ideal adapter' model. More studies manipulating irregular phonetic variations with more predictable stimuli are needed to test predictive models of speech perception.

Finally, our results strongly support Hanique et al. (2014)'s claim that exemplars probably play a very limited role in everyday speech comprehension given that in our study, not only exemplar effects arose in very limited conditions, but we also found significant mismatch effects (*i.e.* the use of abstract representations), even in the very conditions which were expected to trigger exemplar effects. It is currently assumed that exemplars are used for speech comprehension. However, given Hanique et al.'s result, our results, and the many null results reported in the exemplar literature (e.g. Luce and Lyons, 1998; McLennan et al., 2003; Mattys and Liss, 2008; Hanique et al., 2014, Nijveld et al. 2015), it is quite clear that

exemplar effects are not so robust. Researching the exact conditions which can consistently trigger exemplar effects is essential to find which role exemplars actually play in everyday speech perception.

Conclusion

Exemplar effects can also be found for L2 learners, even when the prime and target encode phonetic information that does not occur regularly in the learners' L1. This shows that exemplars can encode information that the phonological filter usually discards, and exemplars must therefore be formed before the phonological filter applies. Exemplars are thus probably not part of the mental lexicon. Interestingly, we also found that participants displayed different response patterns when presented with different tokens of the same words in exactly the same testing conditions. This finding particularly questions the robustness of exemplar effects. Hanique et al. (2014) already warned that exemplars are probably not used in everyday speech comprehension given the limited conditions under which exemplar effects arise. Our study supports this conclusion and extends it to L2 listeners for which the conditions under which exemplar effects arise appear even more limited.

Acknowledgements

This work was supported by the European Research Council under Grant ERC-2011-StG and the Netherlands Organization for Scientific Research under VICI grant 277-70-010, both awarded to the third author.

References

Baayen, Harald: *Analyzing linguistic data: A practical introduction to statistics using R*, Cambridge University Press, Cambridge, 2008.

Baayen, Harald / Davidson Doug / Bates Doug: "Mixed-effects with crossed random effects for subject and items". Journal of memory and language 59(4), 2008, pp.390-412, retrieved 25.10.2107, from doi:10.1016/j.jml.2007.12.005.

Boersma, Paul / Weenink, David: "Praat: doing phonetics by computer [Computer program]. Version 6.0.31". 2017, retrieved 21.08.2017, from http://www.praat.org/.

Bradlow, Ann / Nygaard Lynne / Pisoni, David: "Effects of talker, rate, and amplitude variation on recognition memory for spoken words". Perception & Psychophysics 61(2), 1999, pp.206-219, retrieved 25.10.2017, from doi:10.3758/BF03206883.

Bradlow, Ann/ Pisoni David: "Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors". Journal of the Acoustical Society of America 106, 1999, pp.2074-2085, retrieved 25.10.2017, doi:10.1121/1.427952.

Council of Europe: "Common European framework of reference for languages: Learning, teaching, assessment", 2011, retrieved 25.10.21017, from https://www.coe.int/en/web/common-european-framework-reference-languages/.

Craik, Fergus / Kirsner, Kim: "The effect of speaker's voice on word recognition". Quarterly Journal of Experimental Psychology 26(2), 1974, pp.274-284.

Cristia Alejandrina / Mielke Jeff / Daland Robert / Peperkamp Sharon: "Similarity in the generalization of implicitly learnt sound patterns". Laboratory Phonology 4(2), 2013, pp. 259-285, retrieved 25.10.2017, from doi:10.1515/lp-2013-0010.

Cutler Anne / Eisner Frank / McQueen James / Norris Dennis: "How abstract phonemic categories are necessary for coping with speaker-related variation". In: Fougeron Cécile / Kühnert Barbara / d'Imperio Mariapaola / Nathalie Vallée (eds.), *Papers in Laboratory Phonology 10*. Mouton de Gruyter: Berlin, 2010, pp.91-111.

Darcy Isabelle / Dekydtspotter Laurent / Sprouse Rex / Glover Justin / Kaden Christiane / McGuire Michael/ Scott John (2012) "Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition". Second Language Research 28(1), 2012, pp.5-40, retrieved 25.10.2017, from doi:10.1177/0267658311423455.

Díaz, Begoña / Mitterer Holger / Broersma Mirjam / Sebastián-Gallés Núria: "Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access". Learning and Individual differences 22, 2012, pp.680-689, retrieved 25.10.2017, from doi: 10.1016/j.lindif.2012.05.005.

Ernestus Mirjam / Warner Natasha: "An introduction to reduced pronunciation variants" [Editorial]. Journal of Phonetics 39(SI), 2011, pp. 253-260, retrieved 25.10.2017, from doi: 10.1016/S0095-4470(11)00055-6.

Franconville. Frans voor havo/vwo, Livre de textes 3, Tweede druk, Thieme Meulenhoff: Utrecht/Zutphen, 2004.

Goldinger Stephen: "Words and voices: Episodic traces in spoken word identification and recognition memory". Journal of Experimental Psychology: Learning, Memory and Cognition 22(5), 1996, pp.1166-1183.

Goldinger Stephen: “A complementary-systems approach to abstract and episodic speech perception”. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 2007, pp.49-54.

Grandes Lignes. 1 Livre de textes A. Nordhoff: Groningen/Houten, 2008.

Hanique Iris / Aalders Ellen / Ernestus Mirjam: “How robust are exemplar effects in word comprehension?”. The mental lexicon 8(3), 2014, pp. 269-294, retrieved 30.01.2018, from doi:10.1075/ml.8.3.01han.

Jaeger Florian: “Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models”. Journal of Memory and Language 59(4), 2008, pp.434-446, retrieved 25.10.2017, from doi:10.1016/j.jml.2007.11.007.

Kleinschmidt Dave / Florian Jaeger: “Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel”. Psychological review, 122(2), 2015, pp.148-203, retrieved 23.02.2018, from doi: http://dx.doi.org/10.1037/a0038695

Kuznetsova Alexandra / Brockhoff Per / Christensen Rune: “lmerTest Package: Tests in Linear Mixed Effects Models”. *Journal of Statistical Software*, **82**(13), 2017, pp. 1–26, retrieved 23.02.2018, from doi: 10.18637/jss.v082.i13.

Leys Christophe / Ley Christophe / Klein Olivier / Bernard Philippe / Licata Laurent: “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median”. Journal of Experimental Social Psychology 49(4), 2013, pp. 764-766, retrieved 23.02.2018, from doi: https://doi.org/10.1016/j.jesp.2013.03.013.

Luce Paul / Lyons Emily: “Specificity of memory representations for spoken words”. Memory and Cognition 26(4), 1998, pp.708-715.

McLennan Conor / Luce Paul: “Examining the time course of indexical specificity effects in spoken word recognition”. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 2005, pp.306-321, retrieved 25.10.2017, from doi:10.1037/0278-7393.31.2.306.

McLennan Conor / Luce Paul / Charles-Luce Jan: “Representation of lexical form”. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(4), 2003, pp.539-553, retrieved 25.10.2017, from doi: 10.1037/0278-7393.29.4.539.

Mattys Sven / Liss Julie: “On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality”. Perception and Psychophysics, 70(7), 2008, pp.1235-1242, retrieved 25.10.2017, from doi:10.3758/PP.70.7.1235.

Meunier Christine / Meynadier Yohann / Espesser Robert: “Voyelles brèves en parole conversationnelle”. Actes, Journées d'Etude sur la Parole (JEP) 27, Avignon, France, 2008, pp.97-100, retrieved 25.10.2017, from http://hal.archives-ouvertes.fr/hal-00292408

New Boris / Pallier Christophe / Ferrand Ludovic / Matos Rafael: “Une base de données lexicales du français contemporain sur internet: LEXIQUE, http://www.lexique.org”. L'Année Psychologique 101, 2001, pp.447-462.

Nijveld Annika / ten Bosch Louis / Ernestus Mirjam: “Exemplar effects arise in a lexical decision task, but only under adverse listening conditions”. In: Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow: University of Glasgow, 2015.

Nooteboom Sieb / Doodeman G: “Production and perception of vowel length in spoken sentences”. The Journal of the Acoustical Society of America 67(1), 1980, pp.276-287, from doi: 10.1121/1.383737

Norris Dennis / McQueen James: “ShortlistB: A Bayesian model of continuous speech recognition”. Psychological Review 115(2), 2008, pp. 357-395, retrieved 01.03.2018, from doi: 10.1037/0033-295X.115.2.357.

Norris Dennis / McQueen James / Cutler Anne: “Perceptual learning in speech”. Cognitive Psychology 47, 2003, pp.204-238.

Pallier Christophe / Colomé angels / Sebastián-Gallés Núria: “The influence of native-language phonology on lexical access: exemplar-based vs. abstract lexical entries”, Psychological Science 12(6), 2001, pp445-449.

Palmeri Thomas / Goldinger Stephen / Pisoni David : “Episodic encoding of voice attributes and recognition memory for spoken words”. Journal of Experimental Psychology: Learning, Memory, and Cognition 19, 1993, pp.309-328.

Peirce Jon: “PsychoPy - Psychophysics software in Python”. Journal of Neuroscience Methods 162(1-2), 2007, pp.8-13, retrieved 25.10.2017, from doi:10.1016/j.jneumeth.2006.11.017.

Pierrehumbert Janet: “Word-specific phonetics”. In: Gussenhoven Carlos / Warner Natasha (eds), Laboratory Phonology VII, Mouton de Gruyter: Berlin, 2002, pp.101-139.

Quené Hugo: “On the just noticeable difference for tempo in speech”. Journal of Phonetics 35(3), 2007, pp.353-362, from doi:10.1016/j.wocn.2006.09.001.

R Development Core Team: “R: A Language and Environment for Statistical Computing”, http://www.R-project.org. R Foundation for Statistical Computing, Vienna, 2007.

Ramus Franck / Peperkamp Sharon / Christophe Anne / Jacquemot Charlotte / Kouider Sid / Dupoux Emmanuel: “A psycholinguistic perspective on the acquisition of phonology”. In: Fougeron Cécile / Kühnert Barbara / d’Imperio Mariapaola / Vallée Nathalie (eds.), Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation, Mouton de Gruyter: Berlin, 2010, pp.311-340.

Sebastián-Gallés Núria / Baus Cristina: “On the relationship between perception and production in L2 categories, in Twenty-First Century Psycholinguistics: Four Cornerstones”. In: Cutler Anne (ed), Erlbaum: New York, NY, 2005, pp.279–292.

Torreira Francisco / Ernestus Mirjam: “Phrase-medial vowel devoicing in spontaneous French”. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010), Makuhari, Japan, 2010, pp.2006-2009.

Trofimovich Pavel: “Spoken-word processing in native and second languages: An

investigation of auditory word priming”. Applied Psycholinguistics 26, 2005, pp.479-504, retrieved 25.10.2017, from doi:10.1017.S0142716405050265.

Troubetzkoy Nikolaï: “Principles of Phonology”. Translation Baltaxe Christiane, University of California Press: Berkeley and Los Angeles, 1939/1969.

Tulving Endel / Schacter Daniel: “Priming and human memory systems”, Science 247(4940), 1990, pp.301-306.

Tulving Endel: “How many memory systems are there?”, American Psychologist 40(4), 1975, pp.385-398.

Wade Travis / Jongman Allard / Sereno Joan: “Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds”. Phonetica 64, 2007, pp. 122-144. From doi: 10.1159/000107913.

Winters Stephen / Lichtman Karen / Weber Silke: “The Role of Linguistic Knowledge in the Encoding of Words and Voices in Memory”. In: Selected proceedings: Second Language Research Forum (SLRF), 2013.

Zimmerer Frank / Yasuda Rei / Henning Reetz: “Architekt or Archtekt? Perception of devoiced vowels produced by Japanese speakers of German”. 14th Proceedings of Interspeech, Lyon, 2013, pp. 417-420.

Appendix 1. Experimental word-types (with their translations) classified by their high vowel, with their token durations (in ms) and frequencies of occurrence (per million words) as reported for movie subtitles in the database Lexique3. Standard Deviations from the mean are reported between parentheses.

		A		B			Frequency
High-vowel	**Word-types**	voiced	devoiced	voiced	devoiced		Freqfilm2
/i/	le chinois *the Chinese language*	660	730	601	694		21.88
	la cité *the city*	753	650	728	682		14.55
	le citron *the lemon*	681	655	613	599		8.10
	le cycliste *the cyclist*	943	923	919	884		57.46
	le kilo *the kilo*	921	871	907	955		24.77
	le pilote *the pilot*	944	897	979	895		70.70
	la piscine *the swimming pool*	933	1028	910	972		85.08
	le silence *the silence*	898	966	925	1019		18.76
	le ticket *the ticket*	903	865	963	959		0.71
/y/	la cuisine *the kitchen*	691	655	632	665		19.91
	la culture *the culture*	853	884	925	921		25.73
	la fumée *the smoke*	660	710	668	693		5.19
	le futur *the future*	883	958	846	903		29.10
	la purée *the mashed potatoes*	933	1028	910	972		22.19
	le succès *the success*	821	763	811	862		14.85
	le sujet *the subject*	700	787	762	741		32.33
	le surnom *the nickname*	740	752	654	794		22.05
	la tulipe *the tulip*	765	763	762	871		5.74
/u/	la couleur *the colour*	967	1003	1012	1071		105.53
	le couloir *the corridor*	766	683	704	740		39.58
	le courage *the courage*	725	713	694	709		107.92
	la poubelle *the garbage (can)*	690	677	641	632		6.20
	le poulet *the chicken*	644	626	616	664		13.62
	la poupée *the doll*	868	804	903	872		1.53
Average (SD)		**806 (110)**	**808 (130)**	**795 (136)**	**824 (137)**	**808 (127)**	**31.40 (31)**

Appendix 2. Statistical model fitting the log-transformed response times (measured from word offset) to the targets whose corresponding primes have been answered to correctly. N = 1647 after removal of the outliers that are 2.5 absolute deviations lower or higher than the median. Standard error is indicated by SE. The intercept represents the reaction time to a devoiced target A mismatching its prime.

Fixed effects		**β**	**SE**	**t**	**p<**
(intercept)		5.76	0.05	117.40	0.001
Repetition match	Match	-0.09	0.04	-2.05	0.05
Token	B	-0.02	0.05	-0.41	n.s.
Voicing	Voiced	-0.16	0.04	-3.96	0.001
Number of trials between prime and target		0.001	0.0006	2.23	0.05
Stimulus duration (ms logged)		-1.00	0.14	-6.93	0.001
RT to the preceding trial (ms logged)		0.14	0.02	7.17	0.001
RT to the prime (ms logged)		0.26	0.02	11.05	0.001
Repetition match * Voicing	match * voiced	0.10	0.04	2.31	0.05
Repetition match * token	match * B	0.12	0.04	2.68	0.01
Random effects			**Variance**	**SD**	
Item	Intercept		0.02	0.14	
	Voicing		0.02	0.14	
Participant	Intercept		0.03	0.16	
	RT to the prime		0.02	0.13	
	RT to the preceding trial		0.01	0.10	
Residual			0.17	0.42	

Appendix 3. Average spectral differences along the Mel Frequency Cepstral Coeficient alignment path between primes (voiced and devoiced) and targets (voiced and devoiced) time warped per condition. Standard Deviations are reported between parentheses.

Condition	***Prime***	***Target***	***Condition***	***Spectral differences***
AB	Voiced A	Voiced B	Match	755 (109)
	Devoiced A	Devoiced B	Match	813 (120)
	Voiced A	Devoiced B	Mismatch	1091 (207)
	Devoiced A	Voiced B	Mismatch	1092 (256)
AA	Voiced A	Voiced A	Match	0
	Devoiced A	Devoiced A	Match	0
	Voiced A	Devoiced A	Mismatch	1044 (197)
	Devoiced A	Voiced A	Mismatch	
BB	Voiced B	Voiced B	Match	0
	Devoiced B	Devoiced B	Match	0
	Voiced B	Devoiced B	Mismatch	1028 (218)
	Devoiced B	Voiced B	Mismatch	