Why learning foreign words is hard: Evidence of shallower encoding for non-native than native sounding words


Published in:
International Journal of Bilingual Education and Bilingualism

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2020 Routledge. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen’s institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Download date: 15. Sep. 2023
Why learning foreign words is hard: Evidence of shallower encoding for non-native than native sounding words

Marta Marecka\textsuperscript{a,b}, Alison McDonald\textsuperscript{a}, Gillian Madden\textsuperscript{a}, Tim Fosker\textsuperscript{a,CA}

\textsuperscript{a}School of Psychology, Queen’s University Belfast, Belfast, Northern Ireland
\textsuperscript{b}Institute of Psychology, Jagiellonian University, Cracow

Corresponding Author:
Dr Tim Fosker
School of Psychology
Queen's University Belfast
David Keir Building
18-30 Malone Road
Belfast, BT9 5BN
t.fosker@qub.ac.uk

Keywords: second language acquisition, vocabulary learning, N300 ERP, foreign accent, phonotactic probability, semantic encoding

Author Contributions: TF and MM conceived the study, designed the experimental procedures, stimuli and programmed the experiment. TF analysed the data, and MM and TF wrote the paper. TF, MM, AMcD and GM tested participants. MM, AMcD and GM were responsible for data entry. All authors read and agreed the manuscript.
Abstract

Research suggests that second language words are learned faster when they are similar in phonological structure or accent to the words of an individual’s first language. Many major theories suggest this happens because of differences in frequency of exposure and context between first and second language words. Here, we examine the independent contribution of accent and phonological structure on the speed of word learning and on the depth of semantic encoding while controlling for frequency of exposure and context. Fifteen participants learned novel words associated with abstract shapes in a paired associates task. The words systematically varied in their accent and phonological structure. Learning speed was measured for each word and the depth of semantic encoding was measured via a novel manipulation of the N300 ERP component in a Picture Recognition Task of the learned items. Both non-native structure and accent slowed word learning and differences in the N300 effect indicated that semantic encoding was shallower for words with a non-native than native phonological structure, despite almost ceiling levels of accuracy. These results are consistent with a model of second language learning that proposes both accent and phonological structure influence how fast and deep new language vocabulary is learnt.
Introduction

Vocabulary learning is crucial for successful second language acquisition. It is a complex process that involves amassing different kinds of knowledge about the spoken and written forms of words, including their meanings and collocations (Laufer & Goldstein, 2004). However, at its most basic, word learning involves two fundamental processes: (1) Encoding of the phonological (spoken) form of the novel word in memory; and (2) creating a semantic association between this phonological form and the item to which it refers (Service & Craik, 1993). Because both these processes necessarily interact with the phonological form of words, the phonological characteristics of the language should be relevant to the speed of vocabulary acquisition (how fast learners acquire the words) as well as the depth of encoding (how well they learn them). This study investigates how phonological structure and accent influence these two aspects of learning.

The Impact of Phonological Structure and Accent on Word Learning

In the current study we define phonological structure in terms of phonotactic probability (Vitevitch & Luce, 1998; 1999), which denotes how common a particular n-gram is in a given language. N-grams are an ordered combination of speech sounds (e.g. /spl/ in English or /pʃ/ in Polish). Different languages vary in their common n-grams: /pʃ/ is a frequent n-gram in Polish, but non-existent in English. Here, we equate native phonological structure with high phonotactic probability (words that contain common n-grams), and non-native phonological structure with low phonotactic probability.

According to the Elementary Perceiver and Memorizer – Vocabulary (EPAM-VOC; e.g. Jones, 2012; Jones, Gobet, & Pine, 2007) and Chunking Lexical And Sublexical Sequences In Children (CLASSIC; e.g. Jones, Gobet, Freudenthal, Watson, & Pine, 2014)
models, words that are phonotactically improbable should be more difficult to learn. These models assume that frequently occurring n-grams have established representations in a learner’s long-term memory (e.g. Jones, 2012; Jones et al., 2007) and that these representations are used to help with the encoding of novel word forms in short-term memory. When learners encounter a new native word, instead of encoding each vowel and consonant in the word separately, they map the known n-grams onto the word and then encode the word form in short-term memory as a sequence of those n-grams (Jones & Witherstone, 2011; Jones, Gobet, & Pine, 2007; Jones, Gobet, Freudenthal, Watson, & Pine, 2014). This process enables the most efficient use of short-term memory resources. This theory predicts that when individuals attempt to learn foreign words without established representations of n-grams in that language, the encoding of foreign word forms in short-term memory and consequently long-term memory should be less efficient. This is a possible explanation of why participants learn phonotactically probable words faster and more accurately than phonotactically improbable ones (Marecka et al., 2018; Storkel, 2001; 2003; 2004; Storkel & Rogers, 2000).

Accent is a more difficult concept to define than phonological structure. A speaker is perceived to have a foreign accent when they produce vowels, consonants and prosodic patterns differently than they are usually produced by native speakers of the language. Example markers of a foreign accent include articulating stop consonants with non-native values of voice onset time (VOT), changing the manner of articulation of some consonants (e.g. the production of English stop consonants as fricatives by Spanish speakers of English; González-Bueno, 1997; Magen, 1998; Riney, Takada, & Ota, 2000), and shifting the quality of vowels and diphthongs towards the parameters typical for the first language of the speaker (e.g. the pronunciation of the English vowel /æ/ as /ɛ/ by Polish speakers of English; Major, 1987; Munro, Derwing, & Flege, 1999). Speech is also perceived as foreign-accented when
speakers employ non-native intonation patterns, rhythm and stress, and when they put additional (epenthetic) vowels into words (de Mareüil & Vieru-Dimulescu, 2006; Jilka, 2000; Major, 1986; Munro & Derwing, 2001; Liu & Lee, 2012).

We propose that accent may play an important role in word learning, because several studies show that foreign-accented speech is more difficult to process and memorize than speech with native (or familiar) accent (Adank & McQueen, 2007; Bent & Bradlow, 2003; Clopper, Tamati & Pierrehumbert, 2016; Floccia, et al., 2006; Grohe & Weber, 2018). Non-native speech is less intelligible than native speech (e.g. Bent & Bradlow, 2003) and words with non-native accent elicit slower responses in auditory lexical decision tasks and animacy decision tasks (Adank & McQueen, 2007; Floccia, et al., 2006). While listeners can adapt relatively quickly to non-native accented speech (Bradlow & Bent, 2008; Witteman, Bardhan, Weber, & McQueen, 2015), for very strong non-native accent, this adaptation process is likely to take longer (Witteman, Weber, & McQueen, 2013). The reduced intelligibility and slower processing of foreign accented words may make these words harder to encode and recall in experimental tasks (Clopper, Tamati & Pierrehumbert, 2016; Grohe & Weber, 2018). This effect may play a role when learners are asked to acquire completely novel words spoken with a foreign accent. Some evidence for this claim comes from a paper by Marecka et al., 2018. In this study, Polish children learned English-accented novel words more slowly than Polish-accented novel words, even though both groups of words had a similar average phonotactic probability.

The impact of Phonological Structure and Accent on Semantic Encoding

As highlighted earlier, the process of learning new words involves two steps – learning the phonological form of the word and learning the association between this form
and the word’s meaning (Service & Craik, 1993). Here, we will call the former *phonological encoding* and the latter *semantic encoding*. According to EPAM-VOC, knowing the phonological structure of the word (*n*-gram sequences in it) should help with *phonological encoding* – faster encoding of the word forms (Jones et al., 2007; 2014). Accent, if it indeed influences word learning, probably does so by impacting the ease of processing word form. If it affects speech intelligibility (Bent & Bradlow, 2003), it should also influence phonological encoding. A question remains if phonological factors also play a role in *semantic encoding*.

From a theoretical standpoint, it is difficult to answer this question, because there are very few models of word acquisition that take into consideration all phases of learning. The main ones that do are: Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll et al., 2010), Bilingual Interactive Activation – developmental (BIA–d; Grainger, Midgley, & Holcomb, 2010), The Revised Hierarchical Model – Repetition, Elaboration, Retrieval (RHM-RER, Rice & Tokowicz, 2019) and The Distributed Lexical/Conceptual Feature Model (DL/CFM, Kroll & De Groot, 1997).

Here, we focus on the last model, DL-CFM, because arguably it offers the clearest predictions on how phonological knowledge can influence learning. DL-CFM proposes three levels of representations of word: lexical (the level of form), conceptual (the level of meaning), and an intermediate level of language–specific lemmas which connects the two other levels. While there is one lemma per word, Kroll and De Groot posit that the form and meaning are distributed, which means that they do not form whole units, but rather they come about by activating smaller features together. For example, the form of the word “dog” is not seen in this model as one unit, but as a web or cluster of vowels, consonant and speech sequences (/d/, /ɒ/, /g/, /dɒ/, /ɒg/) represented together. The meaning of the word “dog” would also be a web of semantic features (“barks”, “pet”, “four legs”, “fetches wooden sticks”). In the model these features are shared between words and between languages, so the
representations of different words can and often do overlap. For example, a word “bog” would share the sequence /ɒg/ with “dog” so the words will be interconnected on the level of form. The Polish word “pies” (dog) and the English word “dog” in a Polish-English bilingual speaker would share the same semantic features.

The lexicon within this model is a network where one feature can activate neighbouring features. For instance, the word “dog” would activate the similar word “bog”, because they contain the same sequence /ɒg/. Activations of one feature can strengthen the association with a related feature. This happens for instance in the case of cognates – that is words in two languages that share the same meaning and have a very similar form. Seeing a cognate in L2 activates the L1 translation and both representations are strengthened (Dijkstra et al., 2010). Furthermore, according to DL-CFM, in beginner learners the lemma of the acquired L2 words is strongly connected to the lexical features but only weakly to semantic features. This means that the learners first focus on the form and then meaning. With time, the connections with semantic representations become stronger. The higher and more frequent the activation of the representations, the more stable, robust and easily available they become.

On the basis of the DL-CFM assumptions, we predict that learning native words may be easier than learning foreign words both in terms of phonological and semantic encoding. If a person learns a novel native word, they will probably know many native words that share speech sequences with this form. The new word activates the similar words, which in turn might strengthen the activation of the new form. Because the form will be more activated, it will be learned faster (phonological encoding). However, the stronger activation of form will likely trigger also strong activation of meaning. This will cause faster and deeper learning of meaning (semantic encoding). This is the hypothesis we test in this paper.
Event-related Potential (ERP) Measures of Semantic Encoding

Determining the impact of phonological structure and accent on *semantic encoding* independent of *phonological encoding* is difficult to accomplish using behavioural measures. However, specific ERP components are thought to indicate the degree of semantic processing/integration independent of phonological processing (Bakker et al., 2015; Barrett & Rugg, 1990; McPherson & Holcomb, 1999). Two ERP components that are specifically sensitive to the semantic, or more broadly conceptual, aspects of stimuli are the N300 and N400. However, since the N400 is elicited directly by words which necessitate phonological processing, the N300 elicited by images is arguably the most independent measure of semantic processes. The N300 is characterised as a broad negative-going deflection over fronto-central sites and is typically present 200-350 ms after the onset of the visual stimulus of interest. Relatively large N300s are elicited in response to pictures that are incongruent with the context (e.g. a picture of basketball players throwing a watermelon) or incongruent with the preceding stimuli (e.g. when a picture story is followed by an incongruent ending; Barrett & Rugg, 1990; Ganis & Kutas, 2003; McPherson & Holcomb, 1999; Mudrik, Lamy, & Deouell, 2010; West & Holcomb, 2002). The size of the N300 (amplitude) is modulated by the degree of semantic association between stimuli. For example a picture of a cat will elicit a smaller N300 when preceded by the a picture of dog (conceptually related) than a picture of a glass (conceptually unrelated; Barrett & Rugg, 1990; McPherson & Holcomb, 1999). In a similar way, a picture following an associated newly learned word, should elicit a smaller N300 than a picture following an unlearnt word or an unrelated word. The difference in amplitude of the N300 in response to pictures following related (congruent) versus unrelated (incongruent) newly learned words should index how deeply the learned words are
This difference in N300 amplitude between congruent and incongruent stimuli is termed the N300 effect.

The Current Study

Both EPAM-VOC (Jones et al., 2007; 2014) and DL-CFM (Kroll & De Groot, 1997) suggest that phonological encoding is easier for new words that are similar to an individual’s native language in phonological structure and accent than for words with a foreign accent and structure. However, DL-CFM also suggests that native-like words should be also easier to encode semantically than non-native-like words. Essentially, the DL-CFM suggests that the ease of learning both the form and meaning of new words depends on what they sounds like. To the best of our knowledge, no previous studies have tested this claim taking into consideration both familiarity in accent and in structure. Here we examine whether foreign accent and foreign phonological structure will lead to slower learning of non-native words and weaker connections between the form and meaning (shallower semantic encoding). To control for exposure and context, we taught participants a set of new words and their associated concept using a paired associates paradigm. Within the words we controlled separately for accent and for phonological structure. We examined the learning trajectories in the Paired Associates Task to gauge the speed of learning and we investigated the depth of semantic encoding by measuring accuracy, reaction times (RTs) and N300 effect amplitudes in response to a Picture Recognition Task. We hypothesised that:

1) Words with native accent and structure would be learnt faster than words with non-native accent and structure. Therefore, in the Paired Associates Learning Task we predicted steeper learning trajectories for native than non-native words.
2) Words with non-native accent and structure would have a weaker form-meaning association (shallower semantic encoding) than words with native accent and structure. This would be indexed by a smaller N300 effect in response to words with non-native accent and structure compared with those with native accent and structure in the Picture Recognition task.

Methods

Participants

The participants were 17 native speakers of English (age range: 18 - 27 years), with no previous language or hearing difficulties and no prior musical training. We excluded musically trained participants, because according to recent studies they could have an advantage in processing and learning foreign-sounding words (Dittinger et al., 2016; Dittinger, Chobert, Ziegler, & Besson, 2017). Two participants failed to complete all measures and therefore were not included in any analyses. All participants achieved hearing thresholds above 30 dB HL across octave frequencies from 250-8000 Hz in a brief hearing screen. All the participants scored in the normal or above normal range for both verbal IQ (-0.93 SDs below to 2 SDs above the population mean) and non-verbal IQ (-1.1 SDs below to 1.5 SDs above the population mean), as measured by the Wechsler Abbreviated Scales of Intelligence (WASI-II; Wechsler, 2011). None of the participants identified as proficient early bilinguals according to their responses in the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). All participants listed English as their dominant language with less than <20% of their current exposure to a language other than English. Eight participants identified as completely monolingual,
providing English as their only language on the LEAP-Q. The remaining participants listed some experience of between 1 and 3 additional languages based on the UK modern languages curriculum, including French, German, Spanish, Irish, and one participant identified as an adult learner of British Sign Language. Only two participants identified as fluent in one of the additional languages identified, reporting having attained fluency by age 16 or 17 (self-rated proficiency of spoken comprehension was 9 or 10 out of 10). The remaining participants identified as not fluent in any additional language (self-rated proficiency of spoken comprehension ranging from 1 to 4 out of 10).

Stimuli

Words: Two Native Structure bisyllabic non-words were created for the study. This was done by combining frequent CCV and CV syllables in the English corpus (taken from SUBTLEX-UK: van Heuven, Mandera, Keuleers, & Brysbaert, 2013). Both words had high phonotactic probability (bi-gram frequency) and low neighbourhood density (Coltheart N = 0), as measured via a phonotactic probability calculator (Vitevitch & Luce, 2004), as well as high wordlikeness measures obtained via an online questionnaire (see Gathercole, 1995 for the procedure). There were two Non-native bisyllabic words created by combining infrequent English syllables. Both words had Coltheart N equal to 0 (to match the Native Structure words), but low wordlikeness and low phonotactic probability. One of the words in each set was recorded with a native (British Southern English) accent and the other was recorded with non-native (Polish) accent by an early Polish-English bilingual. Additionally, for each word a distractor was created by replacing one phoneme in a word with a different one (e.g. ‘sinty’ became ‘sinby’). The distractors were similar to the target words in terms of phonotactic probability and neighbourhood density. Each distractor was read with the same accent as the target word. There were three recordings of each word and all three were presented during the
tasks. This was done to minimise the potential of learning the acoustic characteristics of words that typically change on each articulation.

**Objects:** Four abstract silhouette images were constructed to cover the same screen area within a tolerance of 20 pixels.

**Procedure**

Participants performed the hearing screen, WASI-II and Leap-Q as well as two tasks: (1) a behavioural *Word Learning Paired Associates task*, in which they had to learn novel names of four novel objects and (2) a *Picture Recognition Task* testing how well the participant learned the name-object association in task 1. During the Picture Recognition Task EEG was recorded.

**Word Learning Paired Associates Task:** Participants were asked to learn 4 words in a novel language which were paired with novel objects. The task consisted of 11 training blocks, alternated with 11 testing blocks. During a training block, novel images were presented one at a time in a random order. Each trial in the block began with a fixation cross presented for 1 second. Then for 1 second an image was presented simultaneously with the spoken word. This was followed by a blank screen for 1 second. In each training block each word-object pairing was presented once in a random order, giving 4 trials per block.

Each training block was followed by a testing block, where participants’ learning was assessed. During the testing block, a fixation cross appeared for 1 second. Then the participants heard the spoken word. Immediately after this presentation, they saw all four novel object images on the screen in a square matrix, numbered 1 to 4. They were required to press the number on the keyboard which corresponded to the matching object, or to press 0 if the spoken word was a distractor. After each response, they received auditory and visual
feedback in the form of a ‘ding’ or ‘buzz’ sound and the word “correct” or “wrong” appearing on the screen for 1 second. During the testing block, participants heard and assessed each of the 4 target words and each of the 4 distractors 3 times (24 trials in total per block). The order of presentation of spoken words within the testing block was randomised, as was the position of the images in the square matrix.

To avoid any incidental associations between the novel object images and the spoken words, the pairings of objects and words were counterbalanced across participants.

**Picture Recognition Task:** In this task, participants were tested on their knowledge of word-image pairings which they learned in the preceding Word Learning Paired Associates Task. Each trial in this task began with a fixation cross lasting 1.5 seconds. Then a spoken word was played. This word was either one of the words learned during the preceding task or a distractor word. Following a 500 ms inter-stimulus interval, an image (also taken from the preceding task) was presented that either matched or mismatched with the heard word. The participants had 2 seconds to decide if the image matched or not. If it did, they pressed one key, and if it did not they pressed another. Response keys were positioned under the left and right index fingers. The side of response was counterbalanced across blocks. The task was divided into 4 blocks of equal trial numbers. Each block consisted of 18 matching trials per condition and 18 mismatching trials per condition (72 match and 72 mismatch trials per each of the four conditions for the whole task). For this task both behavioural data (accuracy and RTs) and continuous EEG data was collected.

**EEG Acquisition, Pre-processing and Analysis**

EEG was acquired from 32 Ag/AgCl electrodes positioned according to the extended 10-20 system at 1 kHz via a Synamps 2 Amplifier (Neuroscan, El Paso). An additional
monopolar pair of electrodes positioned above and below the right eye monitored eye blinks and vertical eye movements. Continuous measurement of scalp activity was recorded referenced to the Cz electrode and filtered online bandpass 0.1 – 200 Hz. Impedances were kept below 10 kΩ during recording and signals were digitally filtered offline lowpass 35 Hz at 48 dB/Oct, zero phase shift. Artefacts from eye blinks were mathematically corrected using the regression approach described by Gratton, Coles and Donchin (1983) before being cut into 900 ms epochs starting 100 ms prior to the stimulus onset time (SOT). Epochs containing activity over ±80 µV were automatically rejected. Epochs were baseline corrected relative to the pre-stimulus baseline and individual condition averages computed for each participant re-referenced to the global average of all the channels. All artefact-free trials irrelevant of participant accuracy were included in the condition averages.

Due to a technical fault, two of the participants included in the analyses only received half the number of trials in the word identification task compared with the other participants (36 trials per condition rather than 72). An average of 59 ± 14 (out of 72) artefact-free trials were available for the match conditions and an average of 60 ± 14 (out of 72) artefact-free trials for the mismatch conditions across participants.

To answer our research question, we examined individual participant ERPs in response to the novel object images (time-locked to the image onset). We computed the N300 effect as the difference in mean amplitude between each mismatch trial condition average and the average of the match trials. Mean amplitudes were computed as the 40 ms window centred on the largest negative peak (local minima) time-locked to Fz for each individual participant. The search interval for analysis of the N300 (220-340 ms post SOT) was decided based on the mean window of difference between the average of the match and the average of the mismatch conditions over the frontal region.
Computation of Word Learning Trajectories and Statistical Analysis

For the Word Learning Paired Associates Task we analysed the learning trajectories. For each of the 11 testing blocks within the task we computed the accuracy rates and then we computed word learning trajectories as the linear slopes in accuracy across learning blocks until the participant reached a plateau in performance. The plateau start was identified by examining the differences between all adjacent blocks sequentially from the latest to the earliest block. Specifically, the plateau start was defined as 11 (the total number of blocks presented) minus the number of consecutive blocks, from the latest to the earliest, where the performance was flat, i.e. there was no difference between the identified block and the previous one. For example, if the last three blocks (9-11) of a participant were identical, then their plateau start was trial 8, if last four blocks were identical, then their plateau start was trial 7 etc. For a plateau to be reached, performance on at least two training blocks had to be identical.

Since learning speed can be substantially impacted by differences in initial performance (i.e. how well participants performed in training trial 1), we confirmed no significant differences in initial performance (intercept values) across conditions (p>0.1) before examining the slope values directly. Since the intercept and slope values for some conditions did not conform to a normal distribution, we compared the values for each condition with a factorial non-parametric repeated measured analysis (Friedman test): Accent (native and non-native) and phonological structure (native and non-native).

For the Picture Recognition Task we examined the behavioural data (accuracy and RTs) as well as the ERP data. The accuracy and RT data were submitted to separate factorial repeated measures ANOVAs: Congruence (2 levels: match, mismatch) x accent (2 levels:
native, non-native) x phonological structure (2 levels: native, non-native). For the ERP data, we ran a series of \textit{a priori} one-sample t-tests on the mean amplitude differences between match and mismatch conditions (mismatch - match) of the linearly derived fronto-central region electrode (Fz, FC1, FC2, Cz) to establish whether a N300 effect was present for each condition. Then a factorial repeated measures ANOVA was used to examine for differences in the N300 effect (mismatch – match) mean amplitude between the different word types: Accent (2 levels: native, non-native) x phonological structure (2 levels: native, non-native) x electrode (4 levels: Fz, FC1, FC2, Cz). We only report significant or borderline main effects or interactions. Where sphericity could not be assumed, only effects surviving Greenhouse-Geisser correction of the degrees of freedom are reported.

\textbf{Results}

\textit{Word Learning Paired Associates Task: Behavioural Results}

A significant difference was found for learning speed dependent on the nativeness of the speech tokens learnt ($\chi^2(3) = 9.26, p<0.05$; see Figure 1). \textit{Post-hoc} Wilcoxon signed-rank tests showed that words with native structure were learned faster when they also had a native accent ($Z = -2.27, p<0.05$, two-tailed). Words with native accent were learned faster when they had native structure ($Z = -2.50, p<0.05$, two-tailed). Finally, words with native accent and structure were learned faster than words with non-native accent and structure ($Z = -2.83, p<0.01$, two-tailed).

\textit{Figure 1 here}
**Picture Recognition Task: Behavioural Results**

Mean accuracy scores in the Picture Recognition Task were very high ranging from 94% to 97% across conditions. The factorial repeated measures ANOVA showed no significant main effects or interactions in the accuracy of the participants (all p>0.1). Analysis of RTs revealed a significant main effect of congruence with participants responding significantly faster to matched trials compared with mismatched trials (match = 720 ± 125 ms, mismatch = 786 ± 98 ms; F[1,14]=33.52, p<0.001). There was also a significant main effect of phonological structure with participants responding faster to tokens with a native phonological structure than tokens with an non-native phonological structure (native phonological structure = 740 ± 115 ms, non-native phonological structure = 781 ± 117 ms; F[1,14]=5.47, p<0.05).

**Picture Recognition Task: Event-related Potential Results**

As predicted, we observed a negativity at fronto-central sites peaking at around 250 ms and consistent with an N300 effect (see Figure 2). *A priori* one-sample t-tests on the peak amplitude differences between match and mismatch conditions showed that the N300 was present for each condition (all p<0.05; see Figure 2 bar graph).

*Figure 2 here*

A significant interaction of electrode x structure was found (F[3,42]= 2.92, p < 0.05) and a significant electrode x accent x structure interaction (F[1.67, 22.51] = 5.71, p<0.05, ε = 0.54). Analysis of simple effects showed that a greater N300 effect was found for words with
a native structure compared with those having a non-native structure, at a single electrode (Fz; mean difference $= 0.56 \mu V$, $p<0.05$, Bonferroni corrected). Analysis of the 3-way interaction revealed that differences at the Fz electrode between native and non-native structure were driven by words spoken with a native accent (mean difference $= 1.19 \mu V$, $p<0.01$, Bonferroni corrected).

**Discussion**

We examined the speed and depth of encoding in individuals learning novel native-like and non-native words. Previous studies (Marecka *et al.*, 2018; Storkel, 2001; 2003; 2004; Storkel & Rogers, 2000) and theories such as EPAM-VOC (Jones *et al.*, 2007; 2014) suggest that individuals might learn the form of new native-like words, i.e. ones that have familiar phonological structure and accent, more rapidly than they learn the form of new non-native words. The DL/CFM (Kroll & De Groot, 1997) suggests that individuals also encode more deeply the link between word form and meaning for novel native words than non-native words. We investigated the speed of learning using a Word Learning Paired Associates task and the depth of *semantic encoding* by measuring the N300 effect during a Picture Recognition Task for a set of newly learned words.

The results of the Word Learning Paired Associates task confirmed that native words are learned much faster than non-native words. The analysis of accuracy slopes indicated that both non-native accent and non-native phonological structure contributed to slowing down of the learning process. The facilitatory effect of native phonological structure on word acquisition is consistent with previous studies (Marecka *et al.*, 2018; Storkel, 2001; 2003; 2004; Storkel & Rogers, 2000). It is also consistent with the theoretical frameworks identified in the introduction. Both EPAM-VOC (Jones *et al.*, 2007; 2014) and DL/CFM (Kroll & De
Groot, 1997) predict that the presence of unfamiliar speech sequences in target words may have led to slower learning because learners could not encode the word as a sequence of familiar segments, as they can for native words (EPAM-VOC, Jones, 2011; Jones et al., 2007; Jones & Witherstone, 2011) or do not get a boost in activation from the similar words (DF/CFM). Non-native accent, on the other hand, may have slowed down the access to phonological structure, because phonological units spoken with a foreign accent sound less prototypical, and thus their recognition may be more difficult. Support for this hypothesis comes from evidence that non-prototypical speech sounds more strongly activate brain regions believed to be responsible for processing the acoustic phonetic detail of speech than prototypical speech sounds do (Myers, 2007). This could mean that foreign-accented words require additional auditory processing because they contain less prototypical speech sounds than native speech. This is consistent with studies indicating that foreign accented words are more difficult to comprehend (Adank & McQueen, 2007; Bent & Bradlow, 2003; Floccia, et al., 2006) and recall (Clopper, Tamati & Pierrehumbert, 2016; Grohe & Weber, 2018).

In the Picture Recognition Task we found that words with non-native phonological structure elicited slower RTs and a smaller N300 effect than words with native-like structure. This suggests that non-native words were encoded less deeply than the native ones, even though participants were very accurate in the Picture Recognition Task. This effect could not be attributed to differences in learning context or frequency of exposure, but only differences in phonological structure – specifically by how familiar the learners were with the n-grams within the word form. Previous studies show that novel word forms become more entrenched in long-term memory when the learner already knows many similar words (Storkel et al., 2006; Vitevitch & Storkel, 2013). The novel word becomes a part of a phonological network, connected to similar word forms. The partial overlap with other words in the network strengthens the activation of the novel word form, speeds its learning, and strengthens its
representation. Based on DL/CFM (Kroll & De Groot, 1997), we predicted this would also lead to faster and deeper semantic encoding. The activation from the form should spread to the meaning of the word. The whole representation of the word, i.e. form, meaning and lemma, should become more activated and, as a result, stronger, deeper and more easily accessible.

The finding that foreign-accented words are learned more slowly than the native accented ones explains why individuals experience particular difficulties when they attempt to acquire a second language not phonologically similar to their native language. It also has implications for foreign language teaching strategies. Somewhat surprisingly, this finding suggests that second language learners may benefit from interactive multi-cultural (multi-accent) classrooms where individuals get the opportunity to hear new words pronounced in the accent of their first language by non-native speakers. Hearing those words in their native accent could in fact help learners acquire the words faster.

Finally, our results also indicate that paired-associate learning tasks, used in many software-based language learning tools may lead to only shallow semantic representations. Foreign language learners may perform well on standardised vocabulary tests (similar to our Picture Recognition task), even when they have not fully internalised the words. Therefore, it is of crucial importance that foreign language teachers use a variety of tasks and strategies that will encourage deep semantic learning across a range of accents and discourage using paired-associate strategies when learning non-native vocabulary.

**Limitations of the Study and Further Research**

There are three limitations of the current study that should be addressed in future research. First of all, the current study tested a relatively small sample of participants, since we wanted to eliminate as far as possible confounding variables such as extensive access to
additional languages, high exposure to additional languages and foreign accents, as well as early acquisition of a second language (prior to the teenage years). We mitigated the impact of the small number of participants by requiring relatively large sampling at the level of the individual. The word learning Paired Associates Task consisted of 264 learning test trials per participant, and the Picture Recognition Task measure consisted of 576 total trials. Nevertheless, future studies should examine the effects of accent and phonological structure on learning words in larger groups of participants.

Second of all, we tested only a population of adult learners. It would beneficial to examine participants with wider language experience, including bilingual adults and children or children acquiring L2 in a classroom. To the best of our knowledge, the last group has only been examined in one study with Polish 9-year old children learning English as a second language at school (Marecka et al, 2018). In a paired associates tasks, participants, learned pseudowords with native-like structure and accent faster than pseudowords with English-like structure and accent, which in turn were learned faster than pseudowords with completely foreign structure and accent. However, that study did not directly examine the impact of the phonological characteristics of the words on their semantic encoding.

The third limitation of the present study is that it examined only speakers of English and exposed them to only one foreign accent (Polish). This is a limitation, because different languages are not comparable in the complexity of their phonological structure and pronunciation and such difference may impact the relationship between the importance of phonological characteristics for learning a language. For example, native speakers of a language that has a very phonologically complex structure, may not have much of an issue with the complexity of phonological structure when they learn a foreign language. This is what was found by Marecka et al., 2020, where native speakers of Polish were required to learn pseudowords with native-like and non-native phonological structure, pronounced with
Polish and Russian accent. While in the current study both accent and phonological structure influenced the speed of learning words, in their study with Polish participants Marecka et al. found an effect of accent only and not structure. This may be because the native language of the participants in the current study is English, which has a much simpler phonological structure than Polish, especially when it comes to consonant clusters. English allows 46 double consonant clusters and 11 triple consonant clusters in the initial position of the word (Trnka, 1966), while in Polish around 160 double clusters, about 100 triple clusters and around 20 quadruple clusters are possible word initially (Dobrogowska 1992). Polish learners may have fewer problems than English learners with acquiring foreign phonological structures because of their previous experience with high phonological complexity. For this reason, we need replications of these types of studies with participants speaking different native languages and presented with a range of non-native accents and phonological structures.

Conclusion

Despite some limitations, our study is unique in exploring how phonological characteristics can influence word learning. First of all, we have shown that both the accent and phonological structure of words influences how fast they are learned. Specifically, words with native-like phonological structure and accent are easier to learn than words with non-native phonological structure and/or accent. Second of all, we found that participants encoded the words more deeply when those words had a familiar phonological structure. Our study therefore points to two possible factors that can help or hinder learning foreign vocabulary: Phonological structure and accent. The fact that words pronounced with familiar accent were learned faster in our experiment suggests that learners might benefit from learning a foreign
language in environments where they can hear new words pronounced in a range of accents, including their native accent. Therefore, smartphone and computer applications for learning vocabulary in a second language, should consider presenting words in a range of accents.
References


24


of Experimental Psychology, 67(6), 1176–1190.

http://doi.org/10.1080/17470218.2013.850521

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in

http://dx.doi.org/10.1111/1467-9280.00064

activation in spoken word recognition. Journal of Memory and Language, 40, 374–408.

probability for words and nonwords in English. Behavior Research Methods,

Vitevitch, M. S., & Storkel, H. L. (2013). Examining the Acquisition of Phonological Word
Forms with Computational Experiments. Language and Speech, 56(4), 493–527.

http://doi.org/10.1177/0023830912460513

San Antonio, TX: NCS Pearson.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level
semantic integration of complex pictures. Brain Research. Cognitive Brain Research,

familiarity with an accent codetermine speed of perceptual adaptation. Attention,
Perception, & Psychophysics, 75(3), 537–556. http://doi.org/10.3758/s13414-012-0404-
**Figure Legends**

**Figure 1.** Word learning trajectories for each factorial condition. The solid lines depict learning slopes. The circles represent individual participant accuracy with the size of the circles showing the percentage of participants performing at the corresponding level of accuracy.

**Figure 2.** Region averaged ERPs in response to the novel object images over the fronto-central region. The thick black line depicts the region averaged ERP across mismatch trials. The thick grey line depicts the region averaged ERP across match trials. The grey ribbons depict the standard error of the ERP across the four mismatch conditions (surrounding the thick black line) and four match conditions (surrounding the thick grey line). Topographic plots show the average voltage difference (mismatch – match) across the scalp during the specified time window 220-340 ms post SOT. The white dots on topographic plots indicate the location of the channels used in the region average. Bar graphs show the mean amplitude difference between match and mismatch trials (mismatch - match) for each condition (non-native accent, non-native phonological structure – nNAnNS; non-native accent, native phonological structure – nNANS; native accent and non-native phonological structure – NAnNS; native accent and native phonological structure – NANS). The y-axis displays negative amplitude difference. Error bars depict the standard error of the mean.
Figure 1
Figure 2