Handling L2 Input in Phonological STM: The Effect of Non-L1 Phonetic Segments and Non-L1 Phonotactics on Nonword Repetition

Gábor Kovács

Corvinus University of Budapest

Eötvös Loránd University

Mihály Racsmány

Hungarian Academy of Sciences

University of Szeged

This article reports on an experiment comparing the effects of three discrete types of deviance from native language (L1) phonetics and phonology on verbal short-term memory performance. A nonword repetition task was used to measure the recall of four stimulus types: (a) high-probability L1-sounding nonwords, (b) low-probability L1-sounding nonwords, (c) nonwords containing illegal L1 phoneme sequences, and (d) nonwords containing non-L1 sound segments. Special response assessment criteria were used in order to control for potential production effects such as an accent. Results reveal a major detrimental effect caused by the presence of unfamiliar sound segments in the input. The decrement produced by phonological deviances was only significant in the case of long (six-syllable) stimuli. A model of LTM-STM interaction is proposed in which the supporting effect of phonetic knowledge is restricted to perceptual analysis and the role of phonological/phonotactic knowledge is confined to reconstructive processes.

Keywords phonological short-term memory; sublexical knowledge; wordlikeness; phonotactic probability; second language acquisition; nonword repetition

Differential learning success is one of the key aspects in which first language (L1) and second language (L2) acquisition differ: L2 learners demonstrate

We thank Péter Rebrus and Miklós Törkenczy for their guidance on the phonology of Hungarian. Special thanks to András Golarich for his invaluable assistance in data gathering.

Correspondence concerning this article should be addressed to Gábor Kovács, Budafoki út 26/B, 1111, Budapest, Hungary. Internet: gabor.kovacs@uni-corvinus.hu

far higher variability both in the rate of acquisition and in the level of final attainment. A growing body of evidence suggests that the capacity of phonological short-term memory (PSTM) might be a strong predictor of long-term L2 learning success. In two longitudinal studies, Service (1992) and Service and Kohonen (1995) found a strong relationship between L2-sounding nonword repetition and L2 vocabulary development. Masoura and Gathercole (1999) generalized these findings by showing clear associations between both L1- and L2-sounding nonword repetition and both L1 and L2 vocabulary knowledge. Through the use of a nonword recognition rather than a nonword repetition task, Gathercole, Service, Hitch, Adams, and Martin (1999) established that the correlation between PSTM and vocabulary learning is definitely due to memory performance, instead of articulatory skills, as some critics had previously suggested. A number of studies indicate, however, that this association weakens as the learner's L2 vocabulary expands, so the prominent role of PSTM skills in L2 vocabulary acquisition might be confined to the earlier stages of the language learning process (Chen & Leung, 1989; Cheung, 1996; Masoura & Gathercole, 2005). Ellis (1996) suggested that the role of PSTM in second language acquisition (SLA) is considerably more general than that implied by studies focusing on the acquisition of novel word forms. He argued that sequence information stored in PSTM serves as a basis for the acquisition of collocations and syntax, and the chunking of sequences in PSTM is the key to fluency development. Ellis and Sinclair (1996) demonstrated that the rehearsal of L2 input in the phonological loop[1] promotes not only the acquisition of L2 word forms and phrases but also metalinguistic knowledge of grammatical regularities and grammatical fluency and accuracy.

The concept of PSTM capacity as a general predictor has, however, been found to be an oversimplification, as PSTM appears to operate in a highly language-specific manner (Masoura & Gathercole, 1999; Thorn & Gathercole, 1999). Memory performance is generally higher for L1 or L1-sounding input than for L2 input, the latter developing in line with the language learner's level of proficiency in the L2 (Cheung, 1996). Because PSTM performance for L1 and non-L1 input can greatly differ, PSTM capacity cannot be treated as a unitary measure. The familiarity with the lexical and sublexical properties of a language enhances PSTM capacity for verbal input conforming to the regularities of that language. The relationship between PSTM and long-term representations is, therefore, a reciprocal rather than a unidirectional one. Current evidence suggests that long-term memory (LTM) supports temporary representations at two distinct levels: (a) The higher recall for lists of words as opposed to nonwords (the *lexicality effect*) indicates that lexical items stored in LTM enhance PSTM mechanisms if the input consists of recognizable words (Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, Brown, & Mercer, 1995); (b) if, however, the input comprises novel verbal items, for which no long-term lexical representations are available, LTM might still support PSTM at a sublexical level. Several studies have found that familiarity with the phonotactic tendencies of a language contributes to PSTM performance (the *phonotactic probability effect*; Gathercole, 1995; Gathercole, Frankish, Pickering, & Peaker, 1999; Thorn & Gathercole).

The process of acquiring L2 vocabulary involves the learning of novel phonological sequences. Because—as the phonotactic probability effect suggests—the successful retention of these word forms in PSTM largely depends on the available sublexical LTM support, familiarity with the phonetic and phonological properties of the target language will facilitate the vocabulary learning process. Subsequently, as the range of L2 vocabulary expands, lexical representations in LTM will support the retention of longer L2 utterances and might even boost the acquisition of word order patterns, collocations, and syntax (Ellis, 1996). Therefore, the acquired sublexical patterns appear to play a more basic role in SLA than lexical LTM support, the former being a prerequisite for the latter. This claim is also confirmed by the fact that the strongest correlations between PSTM performance and language learning success were found in studies in which PSTM capacity was measured through nonword repetition where lexical influences are eliminated (Masoura & Gathercole, 1999; Service, 1992; Service & Kohonen, 1995).

Because long-term sublexical L2 knowledge appears to play an important role in SLA, it is essential to gain further insight into its architecture and the nature of its integration with PSTM processes. The models offered by contemporary linguistics reveal a vast and complex knowledge system operating below the lexical (or more precisely: morphemic) level, including articulatory and perceptual phonetics, phonological rules, and phonotactic regularities and tendencies (for an overview, see Carr, 1993 or Kenstowicz, 1994). In cognitive terms, this language-specific knowledge system exists in the form of LTM representations in each speaker of the language. Although approaches vary, three components of this system can be delineated in a relatively theory-independent manner. Disregarding nonsegmental areas (such as stress, metrics, or intonation), the speaker of a language is familiar with (a) the auditory and articulatory properties of the physical sound segments (phones) utilized by the language (i.e., phonetics), (b) phonological rules and phonotactic constraints on possible phoneme combinations (i.e., phonological well-formedness), and (c) the typicality associated with possible well-formed phoneme sequences (i.e., phonotactic probabilities). In other words, a speaker knows the range of speech sounds, what combinations of those sounds are possible, and which of those sequences are typical in the language.

Each of these levels of sublexical knowledge might separately contribute to successful PSTM trace maintenance and recall. To our best knowledge, previous studies focusing on the nature of sublexical knowledge or its influence on PSTM concentrated exclusively on the third component in the above list, namely phonotactic probabilities. The nonword stimuli used in studies such as Baily and Hahn (2001) or Gathercole et al. (1999) are all well-formed phoneme sequences and vary only in terms of probability as measured through wordlikeness questionnaires or frequency counts in word-type databases. The aim of the present study is to compare the detrimental effect on nonword recall produced by low-probability sequences in the input to possible effects created by illegal sequences or the presence of unfamiliar sound segments. The dissociation among these three levels of sublexical knowledge and the comparability across the three distinct effects is achieved through the application of a novel, fully controlled, algorithm-based stimulus-generating method and a scoring scheme devised to control for the effect of potential confounding speech-motor failures.

It appears that one of the principal difficulties in acquiring an L2 lies in representing and maintaining non-L1 input in PSTM. However, why should foreign-sounding input produce lower PSTM performance? In terms of the above framework, sublexical LTM support might be disrupted at three distinct levels; therefore, three possible answers are conceivable: (a) because the input contains unfamiliar sound segments, (b) because the input contains sound sequences that are illegal in the learner's L1, and (c) because the sound sequences in the input are untypical in the learner's L1. In the nonword repetition paradigm, if the control and experimental stimuli conform entirely to the sublexical patterns of existing natural languages, these three effects will be inseparable, because any two languages exhibit differences at each of these levels. Because the purpose of the present experiment is to separate and compare the effects associated with the three components, three different experimental stimulus types are used, each of which differ from the control stimulus type (high-probability L1-sounding nonwords) at one particular level.

Our original hypothesis was that in accordance with the hierarchical nature of the proposed framework, the experiment would yield a terraced pattern of results, with recall performance on each stimulus type being lower than the next. First, we expected lower PSTM performance on nonwords containing low-probability sequences than the high-probability control stimuli on the basis of previous research on the effect of wordlikeness and phonotactic probabilities.

Second, we anticipated the lowest performance levels for stimuli containing unfamiliar sound segments. This hypothesis originated partly from our previous study (Kovács & Racsmány, 2006), in which we had found a robust memory decrement due to the presence of foreign speech sounds in the stimuli. The magnitude of the effect was far higher than those reported in wordlikeness studies,[2] even though a liberal scoring scheme was applied for the experimental stimuli in order to control for production factors, such as accent. Apart from our own data, there is also neurolinguistic evidence indicating that the phonetic analysis of verbal input (i.e., categorizing acoustic signals as phonemes) involves an extremely low-level, preattentive process and the acquisition of the new phonemic inventory of an L2 is accompanied by long-term plastic changes in brain mechanisms in the auditory cortex (Winkler et al., 1999). Finally, we speculated that recall performance on nonwords containing illegal phoneme sequences would fall between low-probability and phonetically irregular stimuli. This tentative hypothesis was based on the theoretical argument that absolute constraints on phoneme distribution represent a more fundamental aspect of a language's grammar than phonotactic tendencies; yet, ill-formed sequences will not disrupt the initial phases of speech perception (such as acoustic cue discrimination and phoneme categorization) as long as the stimuli consist of L1 speech sounds.

Method

Participants

Participants were 40 undergraduate students (14 male, 26 female) at Corvinus University of Budapest. Their ages ranged between 18 and 26 years, with a mean of 20.3 years. All participants spoke Hungarian as their L1.[3]

Design

A within-participants design was used. There were two nominal independent variables: stimulus type and item length. The dependent variable was nonword repetition performance.

Apparatus

Stimuli were recorded digitally with a Shure SM87 condenser microphone at 96.0 kHz in the soundproof studio of Corvinus LSP Examination Centre, Budapest. The experimental sessions took place in the same studio. Stimuli were presented through high-quality audio equipment (Bower & Wilkins DM600 S3 speakers) and participants' responses were recorded on a hard disk at 96.0 kHz.

Stimuli

Four types of stimuli were constructed: (a) high-probability L1 nonwords, which entirely conformed to the phonological constraints of Hungarian and utilized common phoneme sequences; (b) low-probability L1 nonwords, which were legal in terms of the phonology of Hungarian but contained low-frequency consonant clusters; (c) nonwords containing illegal consonant clusters that violate obligatory assimilation rules of the Hungarian language; (d) nonwords containing foreign consonants that do not occur in the phoneme inventory of Hungarian. In order to ensure comparability across the four stimulus types, we had to make certain that the items in the four categories differ solely in terms of the relevant variable (i.e., the level at which the stimuli clash with the sublexical properties of Hungarian). One way to achieve this aim would involve the use of items that are matched across the four categories, with each high-probability L1 nonword having a counterpart in each of the other three categories, which are almost identical except for the required contrast realized at particular positions. Although this method might guarantee a high level of comparability across the four lists of items, it has one major drawback: When used in a within-participants design, it will introduce complex sequence effects that are difficult to control. Supposing the stimuli are presented in a random order, the immediate recall of a nonword might be supported by the memory traces potentially remaining from a prior presentation of one of its matched counterparts. Furthermore, the magnitude of this sequence effect might depend on a number of variables, such as the recency of the prior presentation and the properties of the intervening stimuli. For this reason, as in Kovács and Racsmány (2006), we decided to match the stimuli not in terms of actual phonological content but at the level of the nonword generating algorithm. The algorithm is based on the random assignment of elements to C, V, and CC positions[4] in the items from corresponding pools and is described in detail in the following sections. A full list of stimuli is given in Appendix A.

Forty nonwords were used, 10 in each category. Within the categories, five of the nonwords were four syllables in length and five were six syllables in length. We chose this particular range because items with four to six syllables produced the largest effect for non-L1 sound segments in our previous study (Kovács & Racsmány, 2006).[5] The consonant-vowel patterns for four- and six-syllable items were CVCCVCVCCVC and CVCCVCVCCVCVCCVC, respectively. Therefore, the nonwords did not contain hiatuses[6] and single consonants alternated with consonant clusters in each item. Additionally, each nonword had single consonants in both initial and final positions, which means that all consonant clusters were intervocalic. This was important because in Hungarian intervocalic, morpheme-initial and morpheme-final consonant clusters are subject to different phonotactic constraints. If some of the items had contained consonant clusters in initial or final positions, these clusters would have had to be drawn from separate pools and this would have unnecessarily complicated the study. Due to the alternating pattern of Cs and CCs, the introduction of items with an odd number of syllables would have necessitated the use of either initial or final CCs. The fact that the clusters for these positions could not be taken from the same pool as all other (intervocalic) CCs would somewhat reduce comparability across even-syllable and odd-syllable items. For this reason, no odd-syllable nonwords were used.

High-Probability L1 Nonwords (HP)

The structural framework underlying the 10 high probability items is shown in Figure 1. In the framework, each item is a sequence of positions, and for each position, a particular pool of phonological elements (vowels, consonants, and consonant clusters) is specified. In order to enhance the wordlikeness of the items, the model observes the principle of Vowel Harmony, a basic characteristic of Hungarian phonology. In the majority of Hungarian word stems, vowels agree in backness; that is, the vowels in most stems are either all back (*a á o ó u ú* [ɒ aː o oː u uː])[7] or all front (*e é i í ö ő ü ű* [ɛ eː i iː ø øː y yː]). In a minority of words, back and front vowels co-occur, but even these mixed stems cannot contain back vowels and rounded front vowels (*ö ő ü ű* [ø øː y yː]) at the same time (Nádasdy & Siptár, 1994, pp. 94–152; Siptár & Törkenczy, 2000, pp. 63–74).[8] In order for the stimuli to reflect these constraints, three separate vowel pools were created: V_b (back vowels), V_f (front vowels), and V_m (mixed vowels)

C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}				
C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}				
C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}				
C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}				
C_{L1}	V_m	CC_h	V_m	C_{L1}	V_m	CC_h	V_m	C_{L1}				
C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}
C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}	V_b	CC_h	V_b	C_{L1}
C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}
C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}	V_f	CC_h	V_f	C_{L1}
C_{L1}	V_m	CC_h	V_m	C_{L1}	V_m	CC_h	V_m	C_{L1}	V_m	CC_h	V_m	C_{L1}

Figure 1 Schema for high-probability items. C_{L1} = L1 (Hungarian) consonant; CC_h = high-probability two-consonant cluster; V_b = back vowel; V_f = front vowel; V_m = vowel from the mixed-vowel pool.

containing back and unrounded front vowels. Within one item, all vowels were chosen from the same pool. As can been seen in Figure 1, out of the five four-syllable nonwords, two contain back vowels, two contain front vowels, and one contains vowels from the mixed-vowel pool. The same distribution applies to the five six-syllable nonwords. The codes C_{L1} and CC_h stand for L1 consonants and high-probability intervocalic two-consonant clusters, respectively.

The nonword-generating algorithm involved two steps: (a) The pools of phonological elements were constructed in such a manner that in each case the number of elements in a pool was equal to the number of corresponding positions in the framework and (b) the elements were assigned to the positions in a random order. The V_b pool included the vowels *a, á, o,* and *u* [ɒ aː o u].[9] Because the 10 items contain 20 V_b positions altogether (see Figure 1), each of the chosen four vowels occurred five times in the pool. A piece of computer software was used to mix the resulting 20 vowels and assign them in a random order to the 20 V_b slots in the framework. The same procedure was followed with all the other pools. The V_f pool contained the vowels *e, é, i, ö, ü* [ɛ eː i ø y] and each of the five vowels occurred four times in the pool.[10] The V_m pool included the vowels *a, á, o, u, e, é, i* [ɒ aː o u ɛ eː i]. In this case, because the number of positions in the framework (10) is not divisible by the number of different elements (7), a completely equal distribution of elements could not be achieved and in such cases, as a compromise, the pool was constructed in such a manner that the difference between the frequency counts of any two elements could not exceed 1. In this particular case, a random subset of three elements (*a, e,* and *i* [ɒ ɛ i]) occurred twice in the pool, whereas the remaining elements were included only once. Frequency data for each pool used in the algorithm are presented in Appendix B.

Generally speaking, the phonology of Hungarian does not impose constraints on the distribution of single consonants in morpheme-initial, intervocalic, or morpheme-final positions (Siptár & Törkenczy, 2000, pp. 98 and 105). Therefore, all Hungarian consonants were included in the C_{L1} pool, with the exception of *ty, gy, ny, h, j,* and *dzs* [tʲ dʲ nʲ h j ʤ].[11] For a list of the remaining18 consonants and their frequency counts, see Appendix B.

Finally, for the CC_h pool we needed a selection of high-probability intervocalic consonant clusters. Based on the frequency statistics provided by Törkenczy (1994, pp. 362–363), a random sample of 25 clusters was taken from the complete set of intervocalic nongeminate CCs that occur in over 15 word stems. It is interesting to note that a universal phonological constraint/tendency, the Syllable Contact Law, according to which the first consonant in these clusters should be more sonorous than the second, appears to be inoperative in

Hungarian (Siptár & Törkenczy, 2000, p. 131). As a result, 40% of the elements in the CC_h pool violate this principle (see Appendix B).

Low-Probability L1 Nonwords (LP)

The method of constructing low-probability nonwords was identical to that applied for high-probability nonwords with the exception that in place of high-frequency intervocalic CCs, low-frequency CCs were used. The consonant clusters in these items were taken from a separate pool (CC_l) that was drawn up on the basis of Törkenczy's frequency data (1994, pp. 362–363). The 25 CCs in this pool occur in extremely few (one or two) Hungarian word stems. The CCs were chosen from the complete set of 48 such clusters after the elimination of combinations that only occurred in (a) words of foreign origin, (b) low-frequency words that might not be familiar to all participants, or (c) words whose monomorphemic status was considered to be doubtful by Törkenczy.[12]

Nonwords With Illegal Consonant Clusters (ICC)

The contrast between high-probability and ill-formed nonwords was again realized in CC positions. When constructing ill-formed items, the algorithm used a pool of illegal consonant clusters (CC_i) that consist of Hungarian phonemes but never occur on the surface because they are eliminated by obligatory assimilation processes. The initial set of illegal clusters included those marked as such in Törkenczy (1994, pp. 362–363).[13] In the selection of the 25 CCs from this broader set, it was an important criterion that the clusters, despite being illegal, should be easily pronounceable and perceivable. Moreover, it was our intention that ICC items target low-level categorical rules of phonological well-formedness but do not at the same time contradict higher level phonotactic tendencies. Therefore, clusters with subminimal sonority distance (Siptár & Törkenczy, 2000, p. 111) and those that can be regarded as phonotactically "adverse" in terms of the antilabial constraint, the antifricative constraint, or the antipalatal constraint (Törkenczy, 1994, pp. 375–386) were eliminated. Elements in the final set fall into four categories: (a) Some of the clusters violate the voicing assimilation rule, which states that obstruent clusters must agree with respect to voicing (Siptár & Törkenczy, pp. 76–82). Despite the large number of potential combinations, we found that only those consisting of a voiceless fricative + a voiced stop are easy to pronounce and perceive; therefore, the following clusters were included in the CC_i pool: *f-d, f-g, s-b, s-d, s-g, sz-b, sz-d, sz-g* [f-d f-g ʃ-b ʃ-d ʃ-g s-b s-d s-g].[14] (b) The CC_i pool also contains clusters that break the obligatory affrication rule whereby all underlying intramorphemic dental stop + sibilant clusters will surface as long affricates

(Siptár, 1994, pp. 210–211). Affrication is suspended in the following CCs: *t-sz, d-z, t-s, d-zs* [t-s d-z t-ʃ d-ʒ]. (c) Two kinds of palatalization processes are distinguished in the phonology of Hungarian: lexical palatalization, which states that /t d n l/ become [tʲ dʲ ɲ j] before /j/, and postlexical palatalization, which is obligatory in those cases in which /t d n/ are followed by /tʲ dʲ nʲ/ and become [tʲ dʲ ɲ] as a result (Siptár, pp. 251–259). The following clusters are ill-formed in terms of these processes and were added to the pool: *t-j, d-j, n-j, l-j, t-ny, d-ny* [t-j d-j n-j l-j t-ɲ d-ɲ]. (d) Finally, in the case of an underlying /l/ + /r/ sequence, the output of the liquid assimilation rule will be [ɾː] (Siptár, p. 202). The consequently ill-formed *l-r* [l-ɾ] cluster is also included in the CC_i pool.

Nonwords With Foreign Consonants (FC)

This stimulus category differs from the previous two in that the items contrast with control items (HP) at the single C positions rather than in the clusters. The CCs were taken from the same high-probability pool that was used in HP nonwords (CC_h). For single consonantal positions, a separate pool of 15 different foreign consonants was used (C_{L2}). These phonetic segments occur in various languages but not in Hungarian. The range of foreign consonants was the same as that used in our previous study (Kovács & Racsmány, 2006) and a full list is given in Appendix C.[15]

Procedure

Participants were told they were going to hear some funny sounding nonsense words and were instructed to repeat them as accurately as possible. In order to control for potential sequence effects due to initial inexperience, participants were first presented with a practice list of eight items (see Appendix D). The items in the practice list were generated by the same method as the nonwords used in the experimental task. The 40 nonwords in the four categories were pooled and presented in a random order through loudspeakers. Participants were encouraged to make an attempt for correct repetition after each nonword. A 6-s pause was inserted after each item, which was sufficient in all (but one) cases for participants to respond including instances of false starts and self-correction. Responses were recorded with high-quality audio equipment (see Apparatus subsection). At the end of each session, the participant was briefly informed of the goal of the experiment. These conversations revealed that participants consciously experienced the varying difficulty of items but were unaware of the items belonging to discrete categories. Sessions lasted for 7–8 min.

Scoring

The nonword repetition paradigm has often been criticized for not providing a clean measure of PSTM performance. Early critics noted that apart from memory mechanisms such as storage and retrieval, the task involves speech perception and production processes that might independently influence performance levels (e.g., Snowling, Chiat, & Hulme, 1991). In our case, speech output processes required special attention in the assessment of recalls of ICC and FC items, as these nonwords contain sequences and segments that the participants might be incapable of producing. Mistakes concerning illegal consonant clusters or foreign sound segments might equally be due to their detrimental effect on memory mechanisms or production failure. Because the purpose of the experiment was to compare memory performance on the four item types, in order to ensure comparability across scores we needed to control production factors. This was achieved by applying more liberal scoring schemes for ICC and FC items: For each cluster in the CC_i pool and each consonant in the C_{L2} pool, a range of production variants was established, and substitutions within the predefined range were regarded as instances of "accented" output and were accepted as correct recall (see the subsections *ICC Nonwords* and *FC Nonwords*). For the one-way ANOVA comparing the four item-type conditions, scores were weighted for item length: In all four categories, four points were scored for the correct recall of a four-syllable item and six points were scored for the correct six-syllables items. Therefore, total scores reflect the total number of syllables in all correctly recalled items. For the factorial ANOVA, which explored the interaction between item type and item length, unweighted scores were used: One point was scored for each correct recall.

HP and LP Nonwords

Because these items entirely conformed to the phonology of Hungarian, it was assumed that none of the segments or phoneme sequences induce difficulties in speech production. Therefore, a strict scoring scheme was applied: Only perfect recalls were classified as correct, which contained no phoneme substitutions, additions, or omissions. In some cases in which the mistake was evidently due to production failure, the response was accepted: These included responses with a false start, stuttering, or self-correction (e.g., stimulus: [ʦoɾvaːnɒlmopaːmpɒf]; response: [ʦoɾvaːnɒlno mopaːmpɒf]). In some responses, the word-final stops [p b t d k g] were produced without an audible release obscuring the identity of the segment. In such cases, because we had no reason to doubt it, it was assumed that the segment was stored and retrieved correctly from PSTM. The same principle was followed in cases in which voiced obstruents became partially or fully devoiced at the end of the word.

ICC Nonwords

Although the illegal consonant clusters in these stimuli consisted of familiar L1 sound segments, most participants had difficulty producing these unusual combinations. Because all CC_i clusters were ill-formed due to the suspension of an obligatory assimilation rule, recalls in which the relevant rule was applied were accepted as correct. In the case of clusters that violated voicing assimilation, the homogeneity of voicing could be restored in two ways. For example, with [f-g], in which the first obstruent is voiceless and the second is voiced, two different kinds of adjustments were found in the data: [vg] with two voiced segments and, less commonly, [fk] with two voiceless segments. Both variants were accepted despite the fact that voicing assimilation in Hungarian is strictly regressive (i.e., right-to-left). Appendix E lists all of the accepted substitutions for CC_i clusters.

FC Nonwords

In order to establish the range of acceptable "accented" variants of the foreign segments used in FC stimuli, we relied on the results gained from our previous study (Kovács & Racsmány, 2006). In that experiment, we used a sound segment repetition task in which participants repeated the 15 foreign consonants (which were identical to the ones used in the present study) presented between two [a] vowels (e.g., [aβa aça], etc.). The range of acceptable substitutions was then assembled on the basis of the responses we obtained: All substitutions that were produced by at least 2 out of the 28 participants were licensed. Because participants came from the same population in both studies, we assumed that our original data appropriately reflected the distribution of accented variants for the consonants in question and decided to use it as the criterion in the present study. Strict adherence to the originally defined range, however, produced scores close to floor. Therefore, we decided to liberalize the scoring scheme for FC items even further by additionally licensing all substitutions that occurred at least twice in our present data. FC scores were then recalculated using the more lenient scheme. Appendix F lists all accepted substitutions for each C_{L2} segment.

The application of a slightly altered, more lenient scoring scheme for the ICC and FC stimulus types raises the issue of comparability of scores across the four conditions. The potential validity problems stemming from this feature of the design are, however, partly offset by the one-tailed nature of the hypothesis: Non-L1 features in the input are assumed to affect PSTM functions negatively (to various degrees) or show no detectable effect; it is most unlikely for phonologically illegal sequences or non-L1 sound segments to improve memory performance. In some cases, the errors that are classified as production errors and are therefore disregarded might in fact originate from memory failure, and for this reason, the obtained score might be somewhat higher than a score reflecting a pure measure of PSTM performance. In other words, strictly speaking, scores for the ICC and FC stimuli correspond to a *higher bound* of the "true" score, which cannot be measured directly due to the nature of the task. This results in reducing the probability of making a Type I error (i.e., finding a detrimental effect on memory when it is in fact an artefact of speech production). Admittedly, the risk of a Type II error (i.e., not detecting a minor effect in spite of its presence) is at the same time increased.

Results

Comparison of weighted scores across the four item types (irrespective of length) revealed a terraced pattern. Means and 95% confidence intervals are shown in Figure 2. A one-way repeated measures ANOVA[16] revealed a

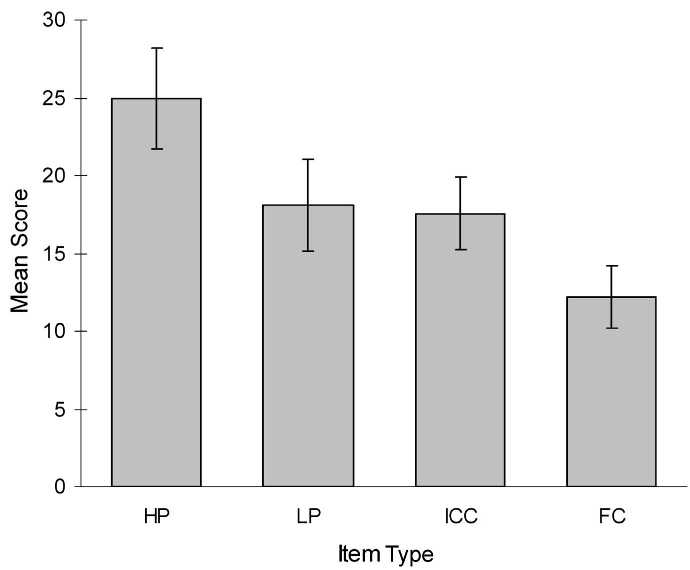

Figure 2 Mean nonword repetition scores and 95% confidence intervals as a function of item type. HP = high-probability L1 nonwords; LP = low-probability L1 nonwords; ICC = nonwords with illegal consonant clusters; FC = nonwords with foreign consonants.

significant effect of item type, $F(3, 117) = 42.23$, $MSE = 26.08$, $p < .001$. For post hoc analysis, paired-sample t-tests were performed for all pairs of item type with the familywise error rate held at $\alpha = .05$. At the adjusted significance level[17] of $\alpha' = .0083$, all pairwise differences were significant ($p < .001$, $.668 \leq r \leq .84$), except for that between LP and ICC stimuli, $t(39) = .48$, $p = .63$, $r = .077$. Therefore, in accordance with our expectations, the highest performance was achieved on high-probability L1 nonword serving as control stimuli. A significant decrement was produced by both improbable (LP) and impossible (ICC) sequences of L1 sound segments. Interestingly, however, no difference in performance can be observed between improbable and impossible items as long as the stimuli consist of familiar L1 speech sounds. Nonwords containing non-L1 segments (FC) appear to produce the largest detrimental effect on PSTM, with the performance on these stimuli being significantly lower than on each of the other three types.

In order to explore the interaction between item type and item length, a two-way repeated measures factorial ANOVA was carried out on the unweighted scores (see Scoring subsection). The analysis revealed significant main effects of both type, $F(3, 117) = 37.42$, $MSE = .55$, $p < .001$, and length, $F(1, 39) = 403.69$, $MSE = .98$, $p < .001$, and a significant Type × Length interaction, $F(3, 117) = 19.25$, $MSE = .69$, $p < .001$. It can be seen in Figure 3 that the interaction is predominantly due to the unusual behavior of ICC stimuli (dotted line).[18] Performance on short (four-syllable) ill-formed nonwords (ICC4) was in fact higher than performance on short control (HP4) items. On the other hand, long (six-syllable) ill-formed nonwords (ICC6) were recalled at a lower level than any other item category. This pattern of results might initially seem incompatible with our predictions. Post hoc comparisons, however, revealed homogeneous subsets of item types within each item length category. We conducted six paired-sample t-tests between the pairs of means in both item length groups. Because 12 comparisons were made altogether, in order to keep the familywise error rate at $\alpha = .05$ the adjusted critical value for each comparison was $\alpha' = .004$. With item length held constant at four syllables, there were no significant differences among HP, LP, and ICC stimuli, HP4–LP4: $t(39) = 0.93$, $p = .36$, $r = .147$; HP4–ICC4: $t(39) = -2.29$, $p = .028$, $r = .344$; LP4–ICC4: $t(39) = -2.84$, $p = .007$, $r = .414$. Each of these three measures, however, differed significantly from FC4 ($p < .001$, $.573 \leq r \leq .748$). Therefore, in the case of short items, only the presence of non-L1 phonetic segments produced a significant simple effect. With six-syllable items, the difference between performance on ICC6 and FC6 nonwords, both being close to floor, was nonsignificant, $t(39) = -0.53$, $p = .6$, $r = .085$. All other differences among

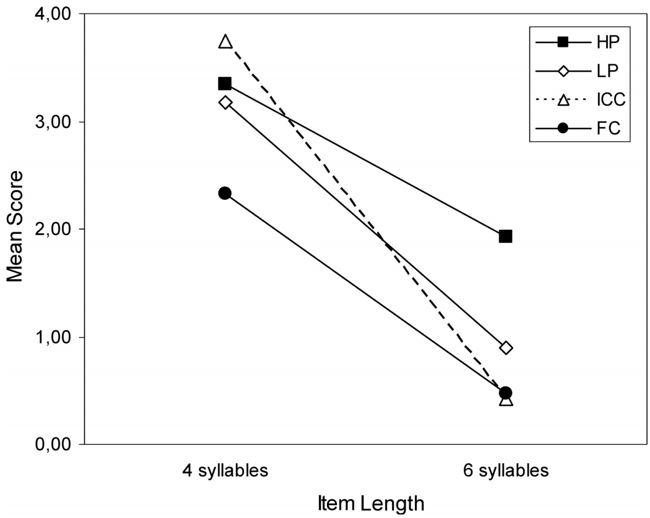

Figure 3 Mean nonword repetition scores as a function of item type and item length. HP = high-probability L1 nonwords; LP = low-probability L1 nonwords; ICC = nonwords with illegal consonant clusters; FC = nonwords with foreign consonants.

six-syllable item types were, however, significant, with $p \leq .002$ and $.468 \leq r \leq .826$ in each case.

To sum up, the analyses of both main and simple effects reveal the following hierarchy between the four stimulus types:

(a) Main effects	HP	>	LP	=	ICC	>	FC
(b) Simple effects (four syllables)	HP	=	LP	=	ICC	>	FC
(c) Simple effects (six syllables)	HP	>	LP	>	ICC	=	FC

where X > Y signifies that performance on Type X was significantly higher than performance on Type Y, while X = Y indicates no significant difference between the two measures.

Discussion

The terraced pattern of results we had originally anticipated on the basis of neurolinguistic studies, linguistic theory, and our own research can be represented as HP > LP > ICC > FC. The main effects we found in the data only partially fit this pattern, as no significant recall advantage was established for LP over ICC stimuli. The analysis of simple effects for short and long stimuli, however, reveals that the reason for this partial fit is that different effects manifest themselves at different levels of PSTM load. For four-syllable stimuli, it was only the presence of non-L1 phonetic segments that produced a significant impairment: No difference could be detected among the recall of high-probability, low-probability, and phonologically ill-formed nonwords. A different pattern was found in the case of six-syllable stimuli: The data show significantly superior recall for high-probability over low-probability nonwords as well as for low-probability over phonologically ill-formed nonwords. Because recall levels for ill-formed stimuli and stimuli containing foreign sounds were both close to floor at six syllables, no significant difference could be established between these two categories.

It is important to note that in no case is the presumed hierarchy of recall advantages reversed. This provides general support for our initial hypothesis that, in accordance with the findings of contemporary linguistic theory, phonetics, categorical phonology, and probabilistic phonotactics constitute successively higher levels within the long-term sublexical knowledge system. Generally speaking, more basic components of this system appear to provide more fundamental support for PSTM representations. The relative contribution made by these layers of knowledge, however, also depends on the amount of verbal material to be maintained in PSTM.

The finding that non-L1 phonetic segments in the input produce the largest decrement regardless of memory load is fully compatible with neurolinguistic evidence on the role of phonetic knowledge in speech perception processes. The incoming auditory signal initially undergoes phonological perception processes whereby the sensory representations are segmented and categorized. The resulting phonemic representations enter the phonological store. The presence of non-L1 sound segments disrupts this perceptual analysis, as the input cannot be unambiguously transformed into a sequence of familiar phonemes (Winkler et al., 1999). Therefore, the representations entering the phonological store will already be inaccurate, unstable, or both. The acoustic features of the input initially stored in auditory sensory memory are subject to extremely fast decay (Frankish, 1996); therefore, once an inaccurately categorized phoneme is recorded in the phonological store, it will either retain its quality as an inaccurate or fuzzy phonemic trace or decay even further. This explains the fact that the highest detrimental effects were produced by foreign sound segments in the stimuli regardless of the length of the presented item.

The dependency of the effect of "hard" (deterministic) phonological and "soft" (probabilistic) phonotactic knowledge on PSTM load suggests that the influence of these knowledge systems is qualitatively different from that of phonetic knowledge and is possibly associated with a different phase in the cascade of subprocesses involved in nonword repetition (see Gathercole, 2006, p. 533). One way to account for the load-dependent effect of these higher forms of sublexical knowledge is to posit that they influence a later stage in the process: The interaction between non-L1 phonology/phonetics and item length might be due to the effect operating at the reconstruction (redintegration) stage at retrieval rather than the initial stage of phonological analysis. A possible model proposed here is based on the integration of the phonological loop construct of Baddeley's Working Memory model (Baddeley, 1986, 2001) and the reconstructive accounts of the wordlikeness effect (Gathercole et al., 1999).

According to the Working Memory model, the material fed into the phonological store will decay in approximately 2 s. Phonological information can, however, be maintained in this passive store for longer periods through the support of an active component, the articulatory rehearsal process, which is capable of refreshing fading memory traces by reading them off and feeding them back to the store. The phonological store and the rehearsal process jointly form a module called the phonological loop. Because rehearsal operates in real time, it constantly competes with trace decay in the phonological store, another real-time process. If the verbal material is relatively short, rehearsal will tend to win out and the traces can be kept intact. If, however, the phonological store is overloaded, rehearsal will only be partially successful, leaving incomplete traces in the store by the time of retrieval. In our case, this explains the fact that no significant differences could be detected among high-probability, low-probability, and ill-formed stimuli for four-syllable items: Due to the virtually perfect trace maintenance in the phonological loop, no reconstructive processes were needed during retrieval. If the contribution of higher level (nonphonetic) knowledge is restricted to reconstruction rather than phonological analysis or storage, these layers of the sublexical knowledge system are inoperative in the case of short stimuli. By contrast, the resources of the phonological loop appear to be insufficient for the perfect retention of longer (six-syllable) nonwords. The representations of these stimuli tend to be partially decayed at the time of retrieval, and reconstructing processes are involved to "guess" the original identity of incomplete traces (Gathercole et al., 1999). In order for reconstruction to be successful, the original input must match the sublexical rules and tendencies stored in LTM. In our data, the decrement produced by six-syllable low-probability and ill-formed nonwords can be explained by the joint effects of trace decay caused by PSTM overload and the failure of subsequent reconstruction. Figure 4 provides a schematic overview of these interactions between the subprocesses involved in nonword repetition and the three levels of long-term sublexical knowledge.

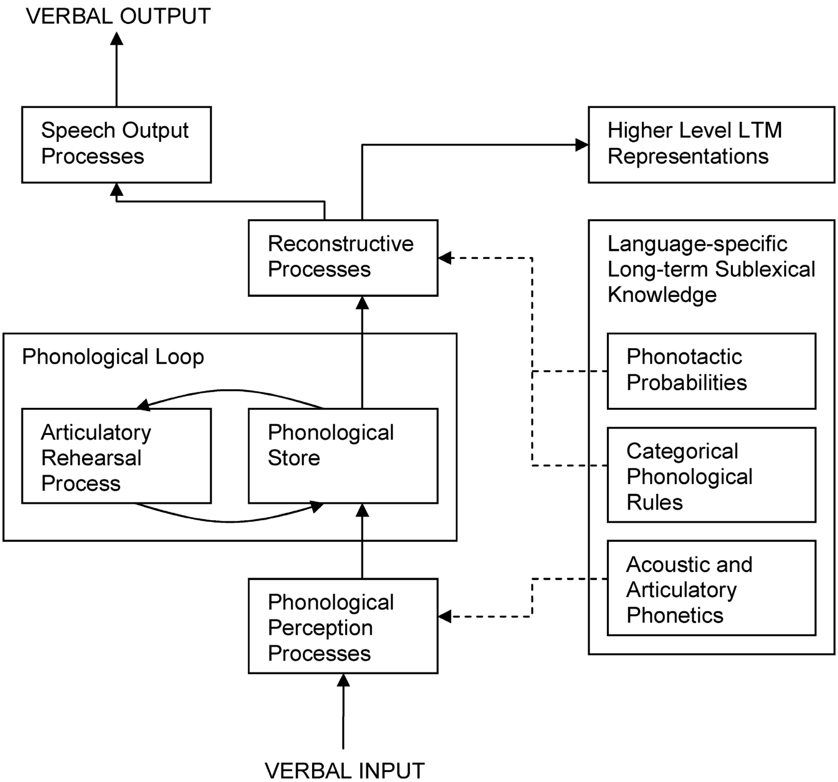

Figure 4 Model of long-term sublexical knowledge supporting PSTM processes.

The late account proposed above is in line with Gathercole et al.'s (1999) original findings. Based on the analysis of the proportions of incorrect, partial, and correct recalls given by children in an immediate serial recall task, the authors concluded that the phonotactic probability effect is due to the reconstruction of incomplete traces rather than the integration of phonotactic knowledge and perceptual analysis. It is important to note, however, that this view is in contrast with some more recent trends in conceptualizing the subprocesses of nonword repetition. Interestingly, Thorn, Gathercole, and Frankish (2005) failed to replicate Gathercole et al.'s results with adult participants in spite of using the same paradigm. As a result, the authors suggested that the influence of phonotactic knowledge should—at least in part—involve a prereconstruction effect (i.e., increasing the level of activation of high-probability items during storage; Thorn et al., p. 151). They noted at the same time that their results do not rule out the possibility of phonotactic knowledge operating during the reconstruction phase as well (Thorn et al., p. 152). In contrast, Gathercole's (2006) recent model of nonword repetition associates the effect of phonotactic probability exclusively with the phonological analysis stage and does not consider redintegration to be operative on nonwords, its role being restricted to the reconstruction of known lexical items (pp. 532–533).

The data yielding these conclusions were obtained from immediate serial recall tasks, which have been shown to tap somewhat different mechanisms than the nonword repetition task used in the present study (see Gathercole, 2006, pp. 534–535); therefore, the results might not be directly comparable to ours. In general, despite the trend in the literature toward an early account, the localization of the phonotactic probability effect still appears to be an open issue: an area in which more research will be necessary to disambiguate conflicting findings. Furthermore, apart from the present study, we know of no research that has targeted the effect of categorical phonological knowledge on PSTM; thus, further replications and extensions of our findings are required before firm conclusions can be drawn.

In summary, the results of this study suggest that different levels of long-term sublexical knowledge interact with PSTM in different ways. At least two distinct levels can be delineated: a phonetic level producing a load-independent effect and a phonological/phonotactic level operating in a load-dependent manner. It appears that the chief difficulty involved in the short-term storage, retention, and retrieval of L2 input lies in its phonetic deviances from the L1, but phonological/phonotactic discrepancies between the L1 and the L2 also play a role, particularly at higher levels of PSTM load.

Revised version accepted 10 August 2007

Notes

1 The verbal component of the Working Memory model (Baddeley, 1986; Baddeley & Hitch, 1974)

2 Mean scores: $M = 53.11$, $SD = 13.2$ for L1-sounding stimuli; $M = 18.11$, $SD = 12.4$ for stimuli containing foreign speech sounds.

3 Data were originally gathered from 41 participants. One of the participants' L1, however, was Ukrainian. Because the purpose of the experiment was to measure the effect of deviations in the input from L1 sublexical patterns, this participant's data were discarded from further analysis.

4 C = consonant; V = vowel; CC = two-consonant cluster.

5 At eight syllables, the effect was nonsignificant due to a floor effect produced by foreign-sounding stimuli and, as we will see later, nonwords with an odd number of syllables are incompatible with the design of the present study.

6 vowel + vowel combinations
7 Hungarian orthography (in italics) + IPA notation (in brackets).
8 Exceptional cases are restricted to a handful of loan words such as *nüansz* [nyɒns] "nuance" and *sofőr* [ʃofø:r] "driver."
9 We decided to discard long *ó* and *ú* [o: u:] because they only differ from their short counterparts in length and this lack of quality difference might have introduced unwanted complexities in the scoring process. Because, as we will see later, all phoneme substitutions in participants' responses were classified as errors, we needed to minimize ambiguity in the identification of phonemes in the audio data. If both the short and long variant of the same vowel quality had been used, it would have been difficult to judge the phonemic identity of half-long realizations [oˑ uˑ] in the responses.
10 Long *í, ő,* and *ű* [i: ø: y:] were eliminated from the model in order to simplify response assessment (cf. Note 9).
11 We excluded the palatals *ty, gy,* and *ny* [tʲ dʲ nʲ] because during the testing of a preliminary version of the algorithm we found that the multiple occurrence of palatals in a single stem greatly reduced the wordlikeness of an item, according to our intuition as native speakers (e.g., ?*tyalnyegy* [tʲɒlnʲɛdʲ]). The phoneme *h* [h] was eliminated because of its hardly audible voiced intervocalic allophone [ɦ] (as in *tehén* "cow" [tɛɦe:n] or [tɛe:n]) and its lexically conditioned behavior in word-final positions (e.g., the /h/ is dropped in *düh* "anger" [dy] but is realized as a velar fricative in *doh* "mustiness" [dox]). The phoneme *dzs* [ʤ] was discarded due to its occurrence being restricted to very few words of foreign origin and, finally, *j* [j] was omitted from the pool, as it is difficult to produce and perceive as a separate segment in the neighborhood of the front vowels *i* and *é* [i e:].
12 Another way to conceptualize phonotactic probability is in terms of lexical neighborhood density, in which case phonotactic knowledge is viewed as a phenomenon directly stemming from the lexicon rather than an independent, sublexical knowledge system (Baily & Hahn, 2001). The operationalization of phonotactic probability in the present study is equally compatible with both accounts.
13 A similar table is presented by Siptár and Törkenczy (2000, p. 129), in which illegal CCs are marked by dashes.
14 The hyphen signifies the fact that an obligatory assimilation rule is suspended. The cluster *f-b* was excluded because apart from violating voicing assimilation, it also breaks the antilabial constraint mentioned earlier.
15 In our previous study (Kovács & Racsmány, 2006) we found no relationship between familiarity with English and German as a foreign language and the repetition accuracy of nonwords containing non-L1 consonants present in these languages ([ð ŋ ɹ θ] and [ç], respectively). Therefore, the present design categorizes speech sounds in terms of the L1/non-L1 dichotomy, ignoring the potential effects of different levels of familiarity with different L2s. Although this approach is

admittedly an oversimplification of the probable state of affairs, it enabled us to focus more narrowly on the key distinction between features conforming to the L1 sublexical knowledge system and those violating it.

16 Following Howell's recommendations (2002, pp. 108–109), conservative two-tailed significance tests were used in spite of the hypothesis being essentially one-tailed.

17 $\alpha' = \alpha/n$, where n is the number of comparisons; in our case, $n = 6$.

18 With ICC stimuli excluded from the analysis, the interaction between length and the remaining three item types is still significant, $F(2, 78) = 4.47$, $MSE = .81$, $p = .015$.

References

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56*, 851–864.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baily, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*, 568–591.

Carr, P. (1993). *Phonology*. London: Macmillan.

Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a non-native language. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(2), 316–325.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*(5), 867–873.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91–126.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49A*, 234–250.

Frankish, C. R. (1996). Auditory short-term memory and the perception of speech. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 179–208). Hove, UK: Psychology Press.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23*, 83–94.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65–77.

Howell, C. H. (2002). *Statistical methods for psychology*. Pacific Grove, CA: Duxbury.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527–536.

Kenstowicz, M. (1994). *Phonology in generative grammar.* Oxford: Basil Blackwell.

Kovács, G., & Racsmány, M. (2006). Munkamemória és idegennyelv-elsajátítás: Az idegen beszédhangok hatása a verbális rövid távú emlékezetre [Working memory and second language acquisition: The effect of foreign speech sounds on phonological short-term memory]. *Magyar Pszichológiai Szemle, 61*, 399–431.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*(5/6), 383–388.

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3/4), 422–429.

Nádasdy, Á., & Siptár, P. (1994). A magánhangzók [The vowels]. In F. Kiefer (Ed.) *Strukturális magyar nyelvtan 2: Fonológia* (pp. 42–182). Budapest, Hungary: Akadémiai Kiadó.

Service, E. (1992). Phonology, working memory and foreign language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21–50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign-language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155–172.

Siptár, P. (1994). A mássalhangzók [The consonants]. In F. Kiefer (Ed.) *Strukturális magyar nyelvtan 2: Fonológia* (pp. 183–272). Budapest, Hungary: Akadémiai Kiadó.

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford: Oxford Universtiy Press.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12*, 369–373.

Thorn, A. S. C., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology, 52A*(2), 303–324.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50*, 133–158.

Törkenczy, M. (1994). A szótag [The syllable]. In F. Kiefer (Ed.) *Strukturális magyar nyelvtan 2: Fonológia* (pp. 273–392). Budapest, Hungary: Akadémiai Kiadó.
Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*, 638–642.

Appendix A

List of Stimuli

Type[a]	No. of syllables	Vowel harmony	Hungarian graphemes[b]	IPA notation[c]
HP	4	Back	cuncsubolfuv	ʦuntʃubolfuv
	4	Back	suzdázsoltof	ʃuzdaːʒoltof
	4	Front	rökvöbüpresz	ɾøkvøbypɾɛs
	4	Front	szöszmümindél	søsmymindeːl
	4	Mixed	kigzégurbár	kigzeːguɾbaːɾ
	6	Back	corvánalmopámpaf	ʦoɾvaːnɒlmopaːmpɒf
	6	Back	csakszázángatadrul	tʃɒksaːzaːngɒtɒdrul
	6	Front	sedmöcsetvékeglid	ʃɛdmøtʃɛtveːkɛglid
	6	Front	zirképésztümisküv	ziɾkeːpeːstymiʃkyv
	6	Mixed	dirzsegolcenatlat	diɾʒɛgolʦɛnɒtlɒt
LP	4	Back	dutypálálzsáf	dutʲpaːlaːlʒaːf
	4	Back	csozsdaropjás	tʃoʒdɒɾopjaːʃ
	4	Front	sifnüszifjöz	ʃifnysifjøz
	4	Front	düjfömezslin	dyjfømɛʒlin
	4	Mixed	nihnevuknyág	nihnɛvukɲaːg
	6	Back	támszamozsgyurojpuz	taːmsɒmoʒdʲuɾojpuz
	6	Back	cajcapuzvaluhlosz	ʦɒjʦɒpuzvɒluhlos
	6	Front	vetyhezsicmekétyvöcs	vɛtʲhɛʒiʦmɛkeːtʲvøtʃ
	6	Front	pültyébélszöbüsztyék	pyltʲeːbeːlsøbystʲeːk
	6	Mixed	cépfotesnifatylag	ʦeːpfotɛʃnifɒtʲlɒg
ICC[d]	4	Back	dun-jákuf-dot	dun-jaːkuf-dot
	4	Back	szos-gavot-jab	soʃ-gɒvot-jɒb
	4	Front	züsz-deföl-jet	zys-dɛføl-jɛt
	4	Front	ced-nyörid-zib	ʦɛd-ɲøɾid-zib
	4	Mixed	zson-jekid-zsél	ʒon-jɛkid-ʒeːl
	6	Back	csasz-bunud-jagul-ram	tʃɒs-bunud-jɒgul-ɾɒm
	6	Back	zát-sopád-zsocáf-dár	zaːt-ʃopaːd-ʒoʦaːf-daːɾ

Type[a]	No. of syllables	Vowel harmony	Hungarian graphemes[b]	IPA notation[c]
	6	Front	szis-döfél-rölüs-béd	siʃ-døfeːl-ɾølyʃ-beːd
	6	Front	cset-szégüsz-gümét-nyip	tʃɛt-seːgys-gymeːt-ɲip
	6	Mixed	vet-szasis-basáf-gun	vɛt-sɒʃiʃ-bɒʃaːf-gun
FC	4	Back	ɣukvuʈalmaʎ	ɣukvuʈɒlmɒʎ
	4	Back	ɖorkuɹaltaχ	ɖoɾkuɹɒltɒχ
	4	Front	ðilfüβügzöç	ðilfyβygzøç
	4	Front	ðizdeɬészmüŋ	ðizdɛɬeːsmyŋ
	4	Mixed	çagluɸésztaɖ	çɒgluɸeːstɒɖ
	6	Back	χoncsoɸulcáθongáʁ	χontʃoɸultsaːθongaːʁ
	6	Back	ʒárváθondaʈumpáɹ	ʒaːɾvaːθondɒʈumpaːɹ
	6	Front	βédröʈirbeθötvüɬ	βeːdɾøʈiɾbɛθøtvyɬ
	6	Front	ʒerzséʁekszéɣisköç	ʒɛɾʒeːʁɛkseːɣiʃkøç
	6	Mixed	ʎotliʁádmeʒipreŋ	ʎotliʁaːdmɛʒipɾɛŋ

[a]HP = high probability L1 nonwords; LP = low-probability L1 nonwords; ICC = nonwords containing illegal consonant clusters; FC = nonwords containing foreign consonant sounds.
[b]Hungarian graphemes are used for all L1 phonemes, but in FC items, non-L1 segments are given in IPA.
[c]The notation is, in some cases, phonemic rather than phonetic. In particular, the /ng/ sequence surfaces as [ŋg] due to place assimilation. The phonemic notation /ng/ is, however, preserved in order to avoid confusion between the Hungarian phoneme /n/ and the foreign segment [ŋ] used in FC stimuli. (Although [ŋ] is an allophone of /n/ in Hungarian, it is regarded as a foreign segment because its distribution is restricted to positions before /k g/ and it cannot occur morpheme-initially, in intervocalic positions, or morpheme-finally as in FC items.) Additionally, the phoneme /h/ might surface as [x] in the clusters /hl hn/.
[d]The hyphen in ICC nonwords signifies the suspension of an obligatory assimilation rule.

Appendix B

Frequency of Phonological Elements in Stimuli

Pool of elements[a]	Stimulus types				Hungarian graphemes	IPA notation	Freq.
V_b	All				a á o u	ɒ aː o u	5
V_f	All				e é i ö ü	ɛ eː i ø y	4
V_m	All				á o u é	aː o u eː	1
					a e i	ɒ ɛ i	2
C_{L1}	HP	LP	ICC		p b t d k g f v sz z s c cs m n l r	p b t d k g f v s z ʃ ʦ ʧ m n l ɾ	2
					zs	ʒ	1
CC_h	HP			FC	mp lt rk zd ng lc rb szt ncs lf rv sk nd lm rzs pr tl dm ksz tv dr gl szm gz kv	mp lt ɾk zd ng lʦ ɾb st nʧ lf ɾv ʃk nd lm ɾʒ pɾ tl dm ks tv dɾ gl sm gz kv	1
CC_l		LP			tyv tyl tyh cm fn fj msz lty lzs jp hn hl pf pj typ kny szty sn zv zsd zsgy zsl lsz jc jf	tʲv tʲl tʲh ʦm fn fj ms ltʲ lʒ jp hn hl pf pj tʲp kɲ stʲ ʃn zv ʒd ʒdʲ ʒl ls jʦ jf	1
CC_i			ICC		f-g s-d s-g sz-b sz-d sz-g d-z t-s t-j d-j l-j t-ny d-ny	f-g ʃ-d ʃ-g s-b s-d s-g d-z t-ʃ t-j d-j l-j t-ɲ d-ɲ	1
					f-d s-b t-sz d-zs n-j l-r	f-d ʃ-b t-s d-ʒ n-j l-ɾ	2
C_{L2}				FC		β ð ʎ ɬ ŋ ɸ ɹ χ ɖ ɣ	2
						ç θ ɮ ʁ ʈ	3

[a] V_b = back vowels; V_f = front vowels; V_m = mixed vowels; C_{L1} = L1 consonants; CC_h = high-probability L1 consonant clusters; CC_l = low-probability L1 consonant clusters; CC_i = ill-formed consonant clusters; C_{L2} = foreign consonant sounds.

Appendix C

Foreign Consonants Used in FC Items

β	Voiced bilabial fricative; Spanish intervocalic *b/v*
ç	Voiceless palatal fricative; German *ich*
ð	Voiced dental fricative; English *mother*
ʎ	Voiced palatal lateral; Italian *gl* or Spanish *ll*
ʁ	Voiced uvular fricative; French *r*
ɬ	Voiceless alveolar lateral fricative; Welsh *ll*
ŋ	Voiced velar nasal; English morpheme-final *ng*
ɸ	Voiceless bilabial fricative; Japanese *f*
ɹ	Voiced postalveolar approximant; English *r*
θ	Voiceless dental fricative; English *thin*
χ	Voiceless uvular fricative; Welsh *ch*
ɮ	Voiced alveolar lateral fricative
ʈ	Voiceless retroflex plosive
ɖ	Voiced retroflex plosive
ɣ	Voiced velar fricative; Spanish intervocalic *g*

Appendix D

Practice List

Type	No. of syllables	Vowel harmony	Hungarian graphemes[a]	IPA notation
HP	4	Back	lulmásarbacs	lulmaːʃɒɾbɒtʃ
HP	6	Front	pöngireszmédédmüg	pøngiɾɛsmeːdeːdmyg
LP	4	Mixed	zafnánimszof	zɒfnaːnimsof
LP	6	Back	bucmopálszavasnot	butsmopaːlsɒvɒʃnot
ICC	4	Front	mösz-gekün-jed	møs-gɛkyn-jɛd
ICC	6	Mixed	széf-delid-zsaced-juzs	seːf-delid-ʒɒtsɛd-juʒ
FC	4	Back	θuncsuɸorzsáʈ	θuntʃuɸoɾʒaːʈ
FC	6	Front	ɖitvüʎérkeðösztiʁ	ɖitvyʎeːɾkɛðøstiʁ

[a]Except for non-L1 segments, which are given in IPA.

Appendix E

Accepted Substitutions in ICC Stimuli (IPA)

CC_i cluster	Accepted substitutions
f-d	vd ft
f-g	vg fk
ʃ-b	ʒb ʃp
ʃ-d	ʒd ʃt
ʃ-g	ʒg ʃk
s-b	zb sp
s-d	zd st
s-g	zg sk
t-s	ʦs ʦː ʦ
d-z	ʣz ʣː ʣ
t-ʃ	ʧʃ ʧː ʧ
d-ʒ	ʤʒ ʤː ʤ
t-j	tʲj tʲː tʲ
d-j	dʲj dʲː dʲ
n-j	ɲj ɲː ɲ
l-j	jː j
t-ɲ	tʲɲ
d-ɲ	dʲɲ
l-ɾ	ɾɾ ɾ

Appendix F

Accepted Substitutions in FC Stimuli (IPA)

C_{L2} segment	From Kovács & Racsmány (2006)	From the present data
β	w b ʊ v	
ç	ʃ xj xç	x
ð	d v	z
ʎ	j l lj	
ʁ	ɾ ɾɾ	
ɬ	s θ ʃ	x h f l hl kl
ŋ	n ɲ ŋg	
ɸ	f	
ɹ	ɾɾ	ɾ
θ	s	f
χ	ħ xɾ h x χɾ	
ɮ	z ð ʒ	l
ʈ	t	
ɖ	d t	
ɣ	g ʁ	v