



Learning and unlearning voicing assimilation

Zsuzsanna Bárkányi^{1,2*} and Zoltán G. Kiss³

¹School of Languages and Applied Linguistics, The Open University, Milton Keynes, United Kingdom, ²Hungarian Research Centre for Linguistics, Budapest, Hungary, ³School of English and American Studies, ELTE Eötvös Loránd University, Budapest, Hungary

OPEN ACCESS

EDITED BY

Baris Kabak,
Julius Maximilian University of
Würzburg, Germany

REVIEWED BY

Peter Avery,
York University, Canada
Romana Kopeckova,
University of Münster, Germany

*CORRESPONDENCE

Zsuzsanna Bárkányi
✉ zsuzsanna.barkanyi@open.ac.uk

RECEIVED 29 September 2023

ACCEPTED 06 December 2023

PUBLISHED 11 January 2024

CITATION

Bárkányi Z and G. Kiss Z (2024) Learning and
unlearning voicing assimilation.
Front. Lang. Sci. 2:1304666.
doi: 10.3389/flang.2023.1304666

COPYRIGHT

© 2024 Bárkányi and G. Kiss. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

This study investigates how postlexical phonological processes are acquired in multilingual speech, namely, how learners cope with conflicting demands in the production and perception of the voicing patterns in their non-native languages, what impact lexical knowledge has on learner behavior, and to what extent existing speech learning models can account for it. To investigate this, 14 Hungarian native speakers, proficient sequential learners of Spanish and English, took part in two types of experiment. The production experiments examined regressive voicing assimilation between obstruents and when the trigger was a sonorant consonant (presonorant voicing) word-internally and across word-boundary. At word level, we compared various lexical groups: non-cognates, double cognates and triple cognates (inhibitory, facilitative, and cognates with conflicting information). The perception experiments aimed to find out whether learners notice the voicing assimilations mentioned. The results showed that participants failed to learn presonorant voicing and failed to block regressive voicing assimilation despite perceiving the latter as linguistically relevant. Data also revealed that there is no direct link between perception and production, and that cognate status had a limited effect, but in triple cognates the primacy of the native language was dominant. Thus, it is concluded that in laryngeal postlexical processes the native language plays the primary role, neither the other non-native language, nor linguistic proximity seems to be decisive. Our data can be best accounted for by the Scalpel Model extended to phonological acquisition.

KEYWORDS

L3 phonology, cross-linguistic influence, L2 speech acquisition, postlexical processes, regressive voicing assimilation, presonorant voicing, Spanish, Hungarian

1 Introduction

Studies on third language (L3) acquisition aiming to determine the source and direction of cross-linguistic influence (CLI) mostly focus on morphosyntactic features and typically the early stages of acquisition. Several models have been proposed to account for the attested phenomena, but they do not agree on which language has a privileged role as a source of transfer: the native language (e.g., [Herms, 2015](#)), or rather the second language—especially in sequential bilinguals—, which is acquired later, often in adulthood, and as such is cognitively more similar to L3 ([Bardel and Sánchez, 2017](#)); or perhaps the typologically more similar language ([Rothman, 2015](#)). It is widely accepted by now that all previously acquired languages are available for transfer ([Berkes and Flynn, 2012](#)). The Scalpel Model ([Slabakova, 2017](#)) and the Linguistic Proximity Model ([Westergaard et al., 2017](#)) advocate for both positive and negative transfer and claim that it occurs property-by-property rather than wholesale depending on which aspects of the native language (L1) or the second (non-native) language (L2) are perceived to be more similar. Discussion regarding wholesale or piecemeal transfer is still ongoing (see the 2021 special issue to *Linguistic Approaches to Bilingualism*). It is also debatable how complete the full transfer is and what exactly constitutes a property or a block of properties. The question also arises how these models can be extended to multilingual phonologies.

Previous studies on L3 phonological acquisition have identified several factors that might contribute to CLI (see Wang and Nance, 2023 for an overview), such as (perceived) typological similarity (Llama et al., 2010; Cabrelli and Pichan, 2021), and experience with L3 (Cal and Sypiańska, 2020). Many studies argue in favor of property-by-property transfer in L3 phonology. Benrabah (1991) reports that in the speech of Arabic learners of English, consonants are transferred from Arabic, while vowels from French, and this is due to the respective similarity of these subsystems. Archibald (2022) shows that stress patterns follow a mixture of influence from Arabic and French. He claims that the data he gathered can be explained by adopting a contrastive feature hierarchy model which can formally capture linguistic proximity. Wrembel et al. (2020) in a speech perception study with L1 Polish speakers, focusing on the acquisition of rhotics and final (de)voicing, also conclude that acquisition is feature-dependent. Kopečková et al. (2022) in production studies also with L1 Polish learners show that transfer comes from both previously acquired languages based on the perceived structural similarity of the examined features. Wrembel (2021) advocates for a dynamic account of CLI in L3 phonology rather than transfer from L1 or L2 only, wholesale or feature-based, since a multilingual speaker has continuous access to the previously acquired language systems.

Although there is no widely applied L3 phonological acquisition model, current well-established L2 phonology models can potentially be extended to account for L3 speech acquisition (Wrembel et al., 2019). The L2 Perception Model (Escudero, 2005) and its revised version (van Leussen and Escudero, 2015), as well as the Perceptual Assimilation Model (Best, 1994; Best and Tyler, 2007) focus on the acquisition of phonemic contrasts (a less relevant aspect in the current study). Flege's Speech Learning Model (SLM, Flege, 1995) and its revised version (SLM-r; Flege and Bohn, 2021) focus on the acquisition of single sounds, and argue that L2 speech learning is shaped by perceptual biases induced by the L1 phonetic system. The model sees CLI as an equivalence classification where perceptual objects are compared to existing L1 categories. The comparison occurs at the level of position-sensitive allophones. The model predicts that categories that are similar in L1 and L2 are more difficult to acquire as a "new category" because they are equated to an existing L1 category. Learners therefore must discover the phonetic differences and break the L2-to-L1 perceptual link in order to form a new phonetic category. Although sounds are categorized at phonetic level, representations in long-term memory are abstract, consequently, phonetic categories are used to access segment-sized units that are used to activate words (or word candidates) during lexical access. The delinking process can be speeded up by the growth of the L2 lexicon.

While SLM claimed that accurate perception precedes accurate production, in SLM-r this has been revised, and the authors claim that production and perception co-evolve, they are closely linked, and a bidirectional relationship is assumed between the two domains. Some researchers bring evidence for a third scenario, namely, that it is accurate production that precedes accurate perception rather than the other way round (e.g., Baker and Trofimovich, 2006), yet others did not find a direct link between these two domains (Derwing and Munro, 2015). Research into the relationship between perception and production for multilingual speakers is scant. Wrembel et al. (2022) in a study with 12 L1

German and 12 L1 Polish adolescent learners found that accurate perception overall precedes accurate production, but linguistic competence, the learnability of segments (e.g., articulatory difficulty of rhotics), and individual differences also play a role.

Similarly to research on L2 speech, most previous studies on L3 phonology focus on the acquisition of phonemic categories and contrasts and how beginner L3 learners categorize speech sounds based on their phonetic properties. Research on the acquisition of allophonic alternations and dynamic (postlexical) processes that create neutralisations is scarce. The present study hopes to reduce this gap by examining regressive voicing assimilation (RVA) in the speech of Hungarian learners of English and Spanish. RVA works on adjacent obstruent consonants in the speakers' L1 (Hungarian), extends to sonorant triggers in Spanish, and does not operate in English. The question is how Hungarian learners cope with these (partly) conflicting demands in their English and Spanish. Unlike in previous studies, the participants of this study are proficient in both their non-native languages. We also explore the link between speech production and perception in these laryngeal processes, and examine how existing speech learning models can account for our data.

2 Background

2.1 Voicing in Hungarian, Spanish, and English

2.1.1 Regressive voicing assimilation in Hungarian

Hungarian is a true voice language (Beckman et al., 2013) where voicing contrast of obstruents is based on negative Voice Onset Time (VOT) or voice lead in voiced stops vs. zero/short-lag VOT in voiceless stops. The language displays RVA: adjacent obstruents must agree in their voicing feature, that is, voiced obstruents voice preceding voiceless obstruents (1a); voiceless obstruents devoice preceding voiced obstruents (1b); and RVA is right-to-left iterative (1c). Hungarian has a symmetrical obstruent system with contrastive voiceless–voiced pairs at each place of articulation, thus /s/ and /z/ contrast word-initially (2a), word-finally (2b), and within the word (2c); note that /s/ in Hungarian is spelt as "sz" and /z/ is spelt as "z."

- (1)
- a. /tb/ → [db]: *hát-ba* 'back.ILL';
két barát 'two friends'
 - /ʃb/ → [ʒb]: *has-ba* 'stomach.ILL';
hús bevezetése 'introduction of meat'
 - b. /bt/ → [pt]: *láb-tól* 'foot.ABL';
láb tünetei 'symptoms of foot'
 - /zt/ → [st]: *víz-től* 'water.ABL';
víz tárolása 'storing of water'
 - c. /skb/ → [zgb]: *maszk-ban* 'mask.INESS'
- (2)
- a. *szár* 'stem' vs. *zár* 'lock'
 - b. *mész* 'limestone' vs. *méz* 'honey'
 - c. *mászna* 'climb.PL.3.PRES' vs. *máznak* 'gloss.DAT'

Unlike in many surrounding languages (e.g., German, Slovak), word-final obstruents do not devoice in Hungarian (3). Sonorant consonants do not participate in RVA: obstruents maintain the voicing contrast before sonorants both within the word (4a) and across a word boundary (4b).

- (3)
- láb-ak* [b] ‘foot.PL’ ~ *láb* [b] ‘foot’
láp-ok [p] ‘marshland.PL’ ~ *láp* [p] ‘marshland’
méz-ek [z] ‘honey.PL’ ~ *méz* [z] ‘honey’
mesz-ek [s] ‘limestone.PL’ ~ *mész* [s] ‘limestone’
- (4)
- a. *plakát* [pl] ‘poster’, *blöki* [bl] ‘doggy’, *sróf* [ʃr] ‘screw’,
zrí [zr] ‘fuss’
kész-nek [sn] ‘ready.DAT’, *kéz-nek* [zn] ‘hand.DAT’
- b. /tm/ → [tm] (*[dm]): *két mag* ‘two seeds’
 /sl/ → [sl] (*[zl]): *kész leves* ‘ready soup’

According to the traditional generative literature, RVA in Hungarian is categorical, exceptionless, and completely neutralizing (Vago, 1980; Siptár and Törkenczy, 2000), which means that voiceless and devoiced or contextually voiced and underlyingly voiced segments cannot be distinguished on the basis of their phonetic and phonological behavior. More recent acoustic phonetic studies, however, suggest that neutralization might be incomplete with residual traces of the underlying voice feature of the obstruents (e.g., Jansen, 2004; Bárkányi and G. Kiss, 2015).

2.1.2 /s/-voicing in Spanish

Spanish, belongs to the same broader typological group as Hungarian, as it is also a true voice language, where stop phonemes can be either voiced or voiceless, although voiced stops are often realized as voiced approximants (unlike in Hungarian), and fricatives and affricates do not display such a symmetry (they are voiceless, except the palatal fricative).

Even though Spanish has RVA, because of the phonotactic restrictions of the language, the segment undergoing assimilation is mostly /s/. Spanish /s/-voicing presents a special case within RVA languages since there is no alveolar voiced fricative phoneme in the language. The Central-Northern Peninsular variety has two voiceless sibilant fricatives, an interdental /θ/ and an apico-alveolar /s/. All the other varieties have only one sibilant fricative /s/, which has a wide range of dialectal and individual realizations from apical to laminal, interdental, etc. (Quilis, 1993). Spanish clearly shows a preference for open syllables—coda obstruents are fragile in the language and there is high variability in their realizations (e.g., Hualde, 2005)—, therefore /s/-voicing in Spanish only occurs in dialects where syllable-final /s/ rarely undergoes aspiration and deletion. In these varieties when /s/ is followed by a voiced consonant—a voiced obstruent (5a) or a sonorant (5b), including glides (5c), within the same word or across a word-boundary—, /s/ becomes partially or fully voiced (Hualde, 2005). Importantly, in Hungarian there is no prenasal voicing (PSV) as in (5b).

- (5)
- a. *esbelto* [zβ] ‘slim’, *es bueno* [zβ] ‘it’s good’
 b. *isla* [zl] ‘island’, *es largo* [zl] ‘it’s long’
 c. *deshielo* [zj] ‘thaw’, *los hielos* [zj] ‘the ices’

TABLE 1 Summary of the three laryngeal systems.

	Hungarian	English	Spanish
Type of laryngeal contrast	True voice	Aspirating	True voice
Laryngeal contrast within the obstruent inventory	Symmetrical	Symmetrical	Symmetrical (but limited to stops only)
RVA	Yes	No	Yes
PSV	No	No	Yes

Similarities highlighted in bold.

Most phonologists who studied /s/-voicing in Spanish found high degrees of individual variation (Schmidt and Willis, 2011), and claim that the process is gradual (e.g., Campos-Astorkiza, 2015), or that gradient data is the result of categorical but optional assimilation (Bárkányi, 2014). Note that although RVA is very common in true voice languages, PSV is much less frequent than preobstruent voicing, prosodic restrictions also seem to apply (Bárkányi and G. Kiss, 2015), and the phenomenon is viewed by some researchers as a result of extended passive voicing as opposed to the spreading of voicing, as in RVA (Jansen, 2004; Strycharczuk, 2012).

2.1.3 The voicing pattern of English

English, just like Hungarian, displays a symmetrical laryngeal obstruent system, but unlike Hungarian and Spanish, English is an aspirating language (Lisker and Abramson, 1964), that is, the contrast of stops is based on aspiration rather than voicing. “Voiced” stops, or as generally referred to in the phonological literature, lenis stops (in initial position) are produced with zero or short-lag VOT, thus phonetically they are typically voiceless and unaspirated, while voiceless, or fortis, stops are produced prevocally with a relatively long-lag VOT (i.e., aspirated). In contrast to true voice languages, in English no systematic laryngeal spreading, i.e., RVA, is attested (Jansen, 2004; Szigetvári, 2020; see (6a)). Similarly to Hungarian, English does not have prenasal voicing either (6b).

- (6)
- a. *matchbox* [tʃb] (*[dʒb]); *anecdote* [kd] (*[gd]);
baseball [sb] (*[zb]); *bonus deal* [sd] (*[zd])
- b. *disloyal* [sl] (*[zl]); *mismatch* [sm] (*[zm]);
business model [sm] (*[zm])

We summarize the relevant features of the three laryngeal systems in Table 1.

2.2 Voicing in multilingual studies

Most studies that deal with the acquisition of laryngeal features in L3 focus on the phonetic realization of voiceless stops, usually by measuring VOT in the speech of multilingual learners. While these studies tested different groups of trilingual speakers (heritage speakers, beginner L3 learners, advanced L3 learners) and employed different methodologies (reading, picture naming;

monolingual sessions, bilingual sessions, etc.) to find out whether learners created separate phonetic categories in their languages, no prevalent conclusions emerged, although realizations similar to L1 were more likely. Wunder (2011) based on the production of voiceless stops /p t k/ in the speech of eight L1 German, L2 English and L3 Spanish speakers found mainly L1 effects on the L3, but more importantly, in half of the cases, tokens displayed values in between the two languages. Llama and Cardoso (2018) and Llama and López-Morelos (2016) also found that L1 plays a more decisive role in L3 pronunciation, but these authors claim that language proficiency and language dominance are also significant factors. Amengual (2021) only examined the acoustic realization of /k/ in bilingual (L1 English–L2 Japanese and L1 Japanese–L2 English) and trilingual (L1 Spanish, L2 English and L3 Japanese) groups. This study also found that VOT values are closer to the L1 values than the target realizations of each language, but also demonstrated that participants produced language-specific VOT patterns which were influenced by language mode and cognate status (see Section 2.3). Very few studies examine perception, or link perception and production. A notable exception is Liu and Lin (2021), who claim that there is no direct link between the perception and production of voicing. In a study with 39 L1 Mandarin Chinese, L2 English and L3 Japanese or Russian learners, where participants had to carry out a reading task and a phoneme identification task targeting voiced and voiceless stops, Liu and Lin (2021) found that voiced and voiceless stops did not behave in the same way. There was a positive correlation between the perception and production of L3 voiceless stops in the initial stages of acquisition, but no correlation was found between the perception and production of L3 voiced stops. This means that, in perception, phonetic similarity led to confusion as predicted by SLM-r, but pre-voicing in stops—which was a novel phonetic feature for these L3 learners—was easily perceived. Participants, on the other hand, had difficulty producing voicing lead. Note that in these cases learners had to map an existing phonemic contrast (voiced vs. voiceless stops) in their L1 to a phonetically different voiced–voiceless contrast.

A different scenario is when learners have to acquire or block a phonetically very different allophone. Cabrelli and Pichan (2021) in a study focusing on the production of intervocalic voiced stops in the speech of early and late bilingual (L1 English–L2 Spanish and L1 Spanish–L2 English) L3 Brazilian Portuguese and Italian learners found an overall trend toward transfer of Spanish-like [+continuant] segments into the typologically similar Romance L3. The authors conclude that transfer was determined by global similarity between L3 and the source language (Spanish) despite this being non-facilitative. Note, however, that almost half of the realizations were either produced with a stop or partial stop closure, which is not fully compatible with the Typological Primacy Model (Rothman, 2015).

Studies on “feature changing” phonological processes where no new phonetic category is created like word-final devoicing (or the lack of it) also seem to indicate that L1 is hard to overcome. Kopečková et al. (2022) in a delayed repetition task with beginner L3 learners (L1 German, L2 English, L3 Polish and L1 Polish, L2 English, L3 German) found that more than half of the realizations showed L1 influence (the realizations were basically identical in all three languages), while the other half was some other sound substitution. The authors, however, did not measure

the amount and proportion of voicing in the final obstruent; rather, they classified obstruents into three categories (voiced, voiceless and partially voiced). Furthermore, the research design did not control for RVA which in several instances could block word-final devoicing. Thus, these results should be interpreted with caution. Wrembel et al. (2020) arrived at a similar conclusion in a perception study with 13 L1 Polish, L2 English, L3 German teenagers. The authors found no significant development in either L2 English or L3 German in the perception accuracy and processing speed of word-final obstruent (de)voicing.

Similarly to word-final devoicing, RVA does not create a new segment either, i.e., learners do not face any articulatory difficulty, but they have to implement or block a postlexical phonological process that applies across the board. As far as we are aware, no studies deal with RVA in L3. Darcy et al. (2007) with French and American English speakers found that L1 English speakers compensated less, i.e., showed lower detection rates for items undergoing RVA (a phonological process absent in their language) than L1 French students. The authors observed that participants compensated more for devoicing, that is, they recognized a voiced phoneme that was realized as voiceless better than a voiceless phoneme realized as voiced, which could be a result of partial word-final devoicing occurring in English (Keating, 1984).

2.3 Cognate status effect and voicing

The facilitation effects of cognates have been extensively studied in psycholinguistics (see Amengual, 2012 for an overview). Research has consistently shown that reaction times are faster for cognates compared to non-cognates, they exhibit quicker and more accurate lexical access, display greater repetition priming effects, and are easier to learn. The detection of the cognate status of words leads to the formation of lexical connections (Ecke and Hall, 2021), which affects the morphosyntactic specification, meaning, as well as speech production of the new lexical item.

Previous research on bilingual speech indicates that the similarity of lexical items—considerable phonological and semantic overlap—might impact on the acoustic realization of segments within them (e.g., Mora and Nadeu, 2009; Amengual, 2016). Studies examining the possible cognate effects in the production of VOT give mixed results. While Flege and Munro (1994) found that English cognates in Spanish were pronounced with longer VOT values than non-cognates, Flege et al. (1998) did not replicate the same results. Amengual (2012), on the other hand, did find a significant effect of cognate status in the speech of bilinguals of different levels of competence. His participants produced a phonetic shift toward the non-target language, they produced /t/ with longer (more English-like) VOT values in the Spanish production of cognates compared to non-cognate words. The author explains this in the framework of the exemplar model of lexical representation (Bybee, 2001; Pierrehumbert, 2001) according to which, due to bilingual lexical connections, cognates facilitate phonetic interference in the bilingual mental lexicon. Amengual (2021) extended the study to trilingual L1 Spanish, L2 English and L3 Japanese learners and observed that although speakers produced different VOT values in the three languages,

when the session was bi- or trilingual, speakers transferred non-target-like phonetic characteristics more than in monolingual sessions. Also, learners in their least dominant language (Japanese) produced lengthened, more English-like, VOT values which the author sees as a transient CLI.

2.4 Hypotheses

The current study aims to address how multilingual speakers handle the conflicting cross-linguistic influences on RVA and PSV in their speech productions. In order to determine the source of CLI, cognates and non-cognates are compared, and to examine the dynamic aspect of these assimilations, data from sandhi contexts (across a word-boundary) are compared to within-the-word realizations.

The following hypotheses are tested in the study:

- Hypothesis 1:** Inhibitory cognates are realized with voicing properties less similar to those in the target language than non-cognates.
- Hypothesis 2:** Facilitative cognates are more likely to be realized with target-like voicing properties than non-cognates.
- Hypothesis 3:** When cognates are contradictory, e.g., L1 is facilitative but L2 is inhibitory, or L1 is inhibitory but L2 is facilitative, it is the L1 pattern that dominates.
- Hypothesis 4:** Sonorants do not trigger voicing assimilation in sandhi in either Spanish (non-target-like) or English (target-like).
- Hypothesis 5:** Obstruents trigger RVA in sandhi contexts in both Spanish (target-like) and English (non-target-like).

3 Materials and methods

The production experiment aimed to investigate the proportion of voicing in the alveolar fricative in regressive voicing assimilation contexts, including pre-sonorant voicing, in the speech of Hungarian learners of English and Spanish.

3.1 Participants

Fourteen young adult subjects (five male, nine female) participated in the experiment. Their ages ranged between 19 and 25 years (average 21.6). They were rewarded a voucher of 5,000 HUF for their participation in the experiment. All the subjects were students majoring in Spanish language and literature at Eötvös Loránd University, Budapest. They were all native speakers of Hungarian who started learning English and Spanish past 11 years of age. Their proficiency in both languages was at least B2 of the Common European Framework of Reference for Languages as they all successfully passed both English and Spanish proficiency exams administered by the University as part of their studies, but none of them spent more than 3 months in an English-speaking or

Spanish-speaking country. This means that they were all proficient sequential trilingual speakers acquiring their L2 and L3 in a non-immersion context. Although they started learning English before Spanish, as both these interlanguages underwent development at the same time and there is no clear dominance difference between English and Spanish for these participants, in this paper we refer to both English and Spanish as L2 or L3. Eleven subjects speak another Romance language, but they consider themselves less proficient in this additional language than in English and Spanish. All of them claim to speak the Peninsular (Northern-Central Peninsular) variety of Spanish; 4 identify with American English, 6 with British English, and 4 claim to speak a mixed variety. None of the participants reported any speaking or hearing disorder.

3.2 Materials

In the Spanish part of the production experiment the target segment was /s/, while in the English part it was /s/ in sandhi context and both /s/ and /z/ within the word, in the following positions (see the [Supplementary material](#) for a complete list of test sentences and the cognate coding of test words, [Supplementary Tables 9, 10](#)):

Sandhi

- Word-finally before a voiced stop /b d g/ across word-boundary: e.g., SP *coches duros* ‘tough cars,’ ENG *bonus deal*.
- Word-finally before a sonorant consonant /m n l/ across word-boundary: e.g., SP *casas modernas* ‘modern houses,’ ENG *business model*.

Word-internally, in triple cognate words

- In facilitative cognates before a voiced stop /b d g/, e.g., SP *lesbiana*, ENG *lesbian*.
- In facilitative cognates before a sonorant consonant /m n l/, e.g., SP *plasma*, ENG *plasma*.
- In inhibitory cognates before a voiced stop /b d g/, e.g., ENG *baseball*.
- In inhibitory cognates before a sonorant consonant /m n l/ e.g., SP *esnobismo*, ENG *Yasmin*.
- In L1 facilitative and L2 inhibitory before a voiced stop /b d g/, e.g., SP *béisbol*.
- In L1 facilitative and L2 inhibitory before a sonorant consonant /l m n/, e.g., SP *Yasmin*; ENG *snob*.
- In L2 facilitative and L1 inhibitory before a sonorant consonant /l m n/, e.g., SP *Bosnia*; ENG *Bosnia*.

Word-internally, in double cognate words English–Spanish

- In facilitative cognates before a voiced stop /b d g/, e.g., SP *Rasgora*, ENG *Asbora*.
- In facilitative cognates before a sonorant consonant /m n l/, e.g., SP *fantasma*, ENG *phantasmal*.
- In inhibitory cognates before a voiced stop /b d g/, e.g., SP *desdén*, ENG *disdain*.

- In inhibitory cognates before a sonorant consonant /m n l/ e.g., SP *desleal*, ENG *disloyal*.

Word-internally, in non-cognate words

- Before a voiced stop /b d g/, e.g., SP *esbelto* ‘slim’, ENG /s/ *crossbar*, /z/ *husband*.
- Before a sonorant consonant /m n l/, e.g., SP *asno* ‘donkey’, ENG /s/ *Christmas*, /z/ *rosemary*.

The selection of cognates presented a number of challenges. It is not easy to determine the degree of similarity (be it orthographic, phonological or semantic) that lexical items have to show in order to induce a cognateness effect. It is also important that the words are relatively frequent in all the languages so that learners are familiar with them. Furthermore, some facilitative or inhibitory combinations were logically impossible. For instance, the /s/+voiced stop sequence in both Hungarian and Spanish trigger RVA thus L1 is always facilitative in L2/L3 Spanish and always inhibitory in L2/L3 English (furthermore in English, these sequences are mostly heteromorphemic). Other clusters should be possible, but we could not find lexical items like facilitative dual cognates for /s/+voiced stop. In these cases, we used an invented proper name in a carrier sentence that suggested that the test word was a loan from the other language. In order to get a full picture of cognate status effect across these three languages, all three dual combinations (HU–ENG, HU–SP, and ENG–SP) should have been tested. However, with Hungarian not being an Indo-European language, there are not enough lexical items with the required segment sequences that are cognates with only one of the other two languages (note that in the sandhi context cognateness was not controlled for, the stimuli were all non-cognates).

Stimuli were embedded into 10–13-syllable-long neutral declarative carrier sentences. They occurred in the first half of the sentence, but were not sentence initial.

3.3 Method of the production experiments

As language mode might have an effect on the acoustic realization of sounds (Amengual, 2021) and language mode was not a variable tested in the present research, English and Spanish sessions were kept separate. Half of the participants started with the English session and the other half with the Spanish session. Before each session they had to read a few sentences in Hungarian to adjust microphone settings and to make sure learners did not have any speech disorder. As sessions had to be recorded on the same day, after the first session participants had a lunch break and came back for the second session which again started with the sentences in Hungarian. As a reviewer pointed out this might have had a risk of L1 priming, a limitation that must be kept in mind when interpreting the data. Sessions took place in the soundproof booth of the Hungarian Research Center for Linguistics.

Sentences and fillers (which formed part of another experiment) were read from a monitor screen in a randomized order, which was generated by SpeechRecorder (Draxler and Jansch, 2004). Each test sentence was read four times. This meant

44 sentences for the English data and 39 sentences for the Spanish one by four repetitions by 14 speakers.

3.4 Measurements

The acoustic analysis was carried out in Praat (version 6.2.23; Boersma and Weenink, 2022). The spectrograms were segmented manually by the authors and a research assistant, and the following measurements were carried out on the basis of the boundaries inserted:

- Duration of the target consonant /s/ or /z/.
- Absolute length of the voiced interval.
- Ratio of the voiced part compared to the total length of the consonant.

In addition to voicing (vocal fold vibration) in the strict sense, a number of other systematically occurring phonetic-acoustic correlates of voicing contrast are attested in the literature. Voiced obstruents are generally shorter than their voiceless counterpart, and variation is also attested in the surrounding sounds. While voiceless consonants are longer, the preceding vowel is typically shorter (Wells, 1982). In the current research, the cognate status of test words had to be taken into account, so we could not control for the quality of the vowel preceding the fricative; therefore, only the proportion of voicing compared to the fricative interval was measured.

The duration of the fricative was determined on the basis of the frication noise. Voicing was measured based on the visual inspection of the spectrograms and oscillograms, and a low-pass filter with a cut-off frequency of 500 Hz was used to securely determine the exact portion of the voicing oscillation.

3.5 Perception experiment

In order to explore to what extent the production results are mirrored in perception, a perception experiment was designed. As most L2 perception studies aim to determine whether learners acquired a certain contrast, they often apply forced-choice tests where participants have to decide whether they hear phoneme *A* or phoneme *B*. However, we did not find it adequate in the present study. Firstly, because listeners might compensate for RVA (Kuzla et al., 2010; Bárkányi and G. Kiss, 2023). Secondly, because they might be biased against a segment that does not form part of the phoneme inventory, that is, they might be reluctant to choose [z] for Spanish; or they might be biased by the orthographic form of test words. As speakers are often unaware of the application of postlexical processes even in their L1, we wanted to leave it open that learners would like to respond “I don’t know,” or that they simply cannot perceive the processes under scrutiny at all or as a linguistically relevant feature. For this reason, we decided to test the perception of RVA and PSV in a more holistic way. A short (approximately one-minute long) story was recorded in both L2/L3 by two phonetically trained bilingual female speakers with native-like proficiency in both languages (Hungarian–English and

Hungarian–Spanish, respectively). Then, the same short story was recorded, but this time in the English text RVA was applied as would be in Hungarian, and in the Spanish text no pre-sonorant voicing was employed to mirror the L1 laryngeal patterns of listeners (another recording was made of the same story as a distractor that formed part of another experiment). Thus, there were three slightly different texts and participants listened to each text three times in a random order, so they had to listen to nine texts all together.

The experiment was carried out in Praat MFC with the same participants as in the production experiments, at the end of the production sessions. The screen was blank while participants listened to a text. Texts were separated by 1.5 s of silence and a 5-second-long bell. After the text finished, instructions appeared on the screen and participants had to rank on a scale from 1 to 5 how native-like the speaker sounded (with 1 not at all native-like and 5 completely native-like).

3.6 Statistical analysis

The statistical analysis (including the generation of the various plots and data tables) was carried out in R (R Core Team, 2022) using various *tidyverse* packages (Wickham et al., 2019). Linear mixed effects models were used to model the production data, using the packages *lme4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), and *broom.mixed* (Bolker and Robinson, 2022). The model function used the default settings (e.g., the Satterthwaite approximation was used to calculate the degrees of freedom for the *t*-distributions). The outcome variable was always the percent of voicing during the fricative constriction. The predictor variables were different depending on the phonological environment of the fricative. The effect of the *cognate status* was analyzed in the *word-internal* data, separately for the English and Spanish words, and separately for the two triggers (sonorants and voiced stops; for details see below). This is because the cognate groups were necessarily different by language and/or phonological environment. The contrast coding of the cognate status factor used the default “dummy” coding (i.e., each cognate group was compared to the non-cognate words). It was possible to analyse the effect of the *trigger environment* (presonorant vs. before voiced stop), the *language* (English vs. Spanish words) as well as their interaction in the *sandhi* environment, i.e., when /s/ was word-final and the trigger was at the beginning of a following word. In the *sandhi* environment, planned orthogonal contrast codings were used for the fixed-effect predictors, so that their main effects (the estimated marginal means) can be calculated and interpreted more easily, in addition to their interaction effect. The random-effect structure of the models contained subject and item (i.e., the words used in the experiments). Which exact model was used in which analysis will be detailed in the results section below. The best-fitting model was selected after carrying out model comparisons employing likelihood ratio tests (using maximum likelihood). A model was retained if the chi-square test was significant. The same procedure was used to test the utility of the random effects, except that in this case restricted maximum likelihood was used in the likelihood ratio tests. Pairwise comparisons using Bonferroni

adjustment to the *p*-values were carried out using the *emmeans* package (Lenth, 2023), while R-squared effect sizes were calculated with the help of the *r2* function of the *performance* package (Lüdtke et al., 2021). We are going to abbreviate conditional R-squared as “R2c”, and marginal R-squared as “R2m.”

The perception experiment fitted cumulative link mixed models to the data using the package *ordinal* (Christensen, 2022). The outcome variable was the rating of nativeness by the participants, which was on an ordinal scale (1, 2, 3, 4, 5). The fixed predictor variable was the type of the recording they listened to (native-like vs. non-native-like). The models included subject as the random effect. The link function used was probit, as it is considered to be more suitable than the logit link function in models that contain random effects (Hahn and Soyer, 2005). The best-fitting model was selected using the same principles and procedure as in the case of the linear mixed effects models used in the analysis of the production data.

4 Results

4.1 Production experiments

4.1.1 Cognate status effect before sonorants

4.1.1.1 English /s/ before sonorants

Figure 1 shows the mean voicing percentage of presonorant /s/ in four cognate groups, the descriptive statistics can be found in Table 2. The negative sign “–” refers to inhibitory cognate status, while “+” refers to facilitative cognate status. Thus, for example, “SP–HU+” refers to English words that have cognates both in Spanish and Hungarian, but in Spanish the /s/ is realized with voicing, so it can potentially voice English /s/. However, as we can see, /s/ had little voicing across the groups, the SP–HU– group showed the highest average voicing at 22.8%.

The best-fitting linear mixed effects model was one with varying by-subject intercepts and slopes, and varying by-item intercepts. According to this model, none of the three cognate groups were significantly different from the non-cognate group. Pairwise comparisons resulted in none of the groups being significantly different from each other, either. Details of the model coefficients and the effect sizes are in Supplementary Table 1. We note that the word that was responsible for the slight increase of voicing in the SP–HU– group was *Yasmin*, with a mean voicing of 35.1% (SD = 23.8%).

4.1.1.2 Presonorant /z/ in English words

The mean voicing of /z/ in the cognate groups is displayed in Figure 2, the descriptive statistics are shown in Table 2. We can see that there is considerable voicing in /z/ in all groups, except in SP+HU–, i.e., when the word has a cognate in Hungarian pronounced as voiceless [s].

The word *Bosnia* contributed to the cognate effect in the SP+HU– the most: the fricative in this word was produced by the participants with a mean of only 13.9% (SD = 11.7%), there were no tokens above 54% of voicing at all.

The best-fitting model for this data was the one with varying by-subject intercepts and slopes, and varying by-item intercepts. According to this model, the difference between the non-cognate

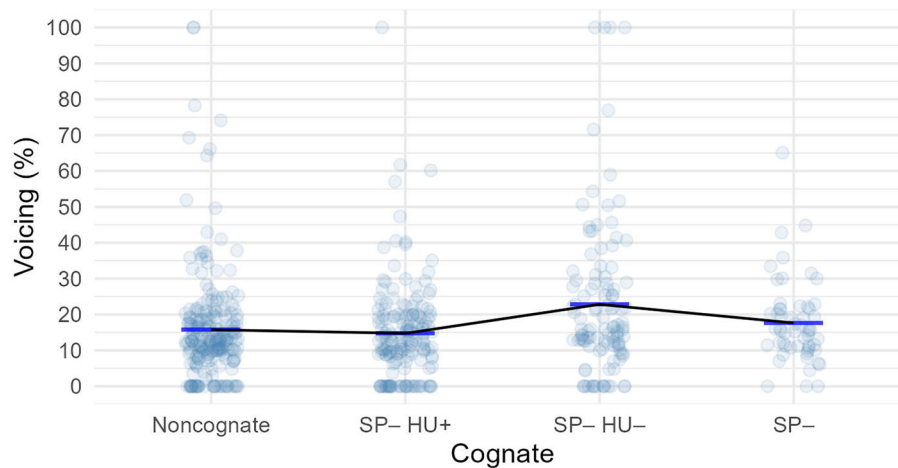


FIGURE 1
Voicing of /s/ before sonorants in the English words. Lines represent the means.

TABLE 2 Descriptive statistics for the voicing of presonorant /s/ and /z/ in English words in the cognate groups.

Sound	Cognate	N	Mean	SD	Median	Min	Max
/s/	Non-cognate	224	15.75	15.12	12.83	0	100
	SP-HU+	168	14.73	13.34	12.71	0	100
	SP-HU-	112	22.81	21.58	15.97	0	100
	SP-	56	17.63	11.53	15.56	0	65.06
/z/	Non-cognate	224	52.57	31.13	42.71	0	100
	SP+HU+	112	41.44	31.33	31.24	0	100
	SP+HU-	112	22.31	21.52	16.40	0	100
	SP+	112	49.95	34.05	35.82	6.62	100

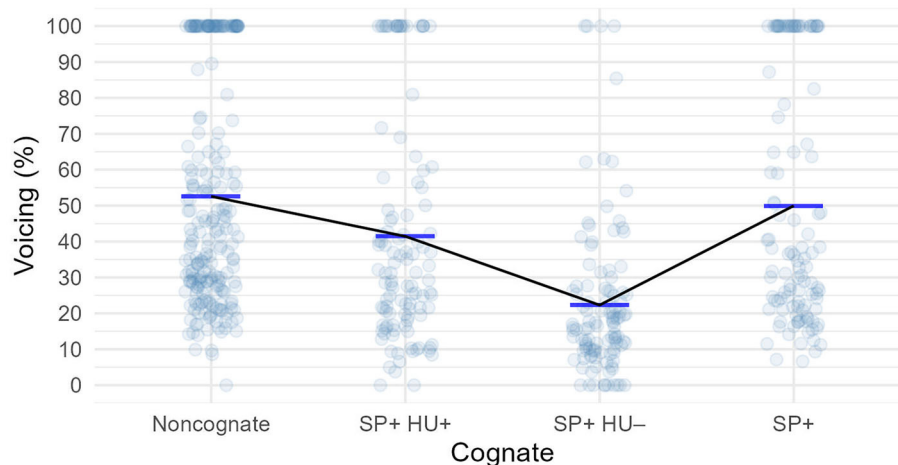
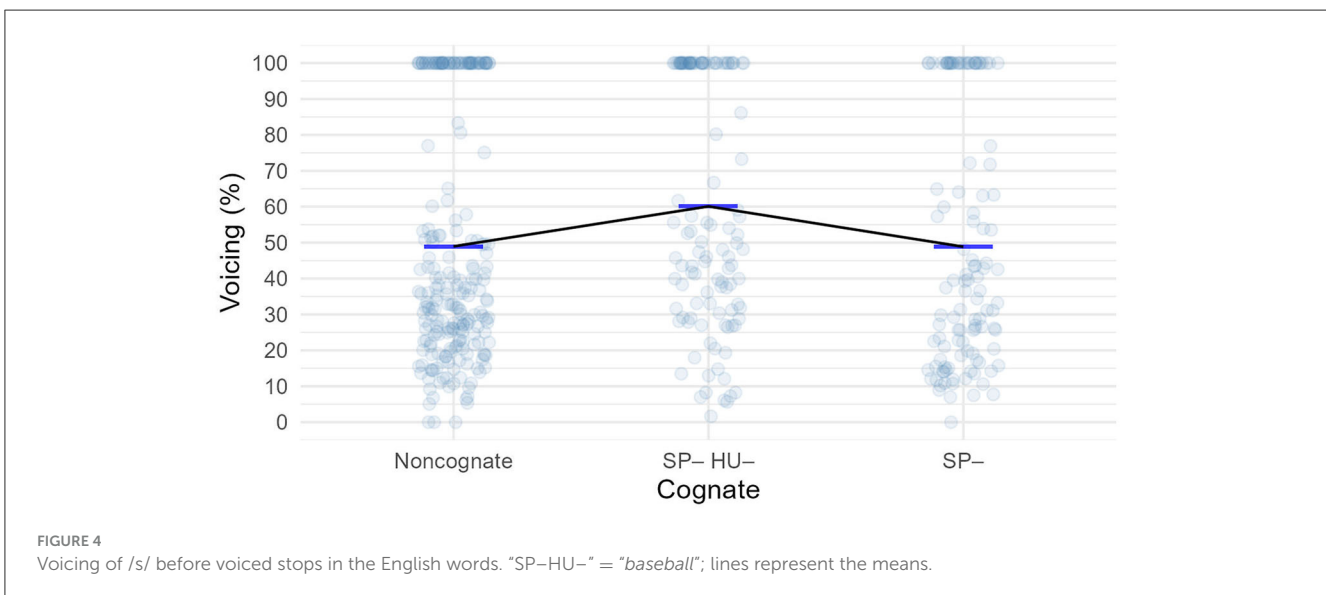
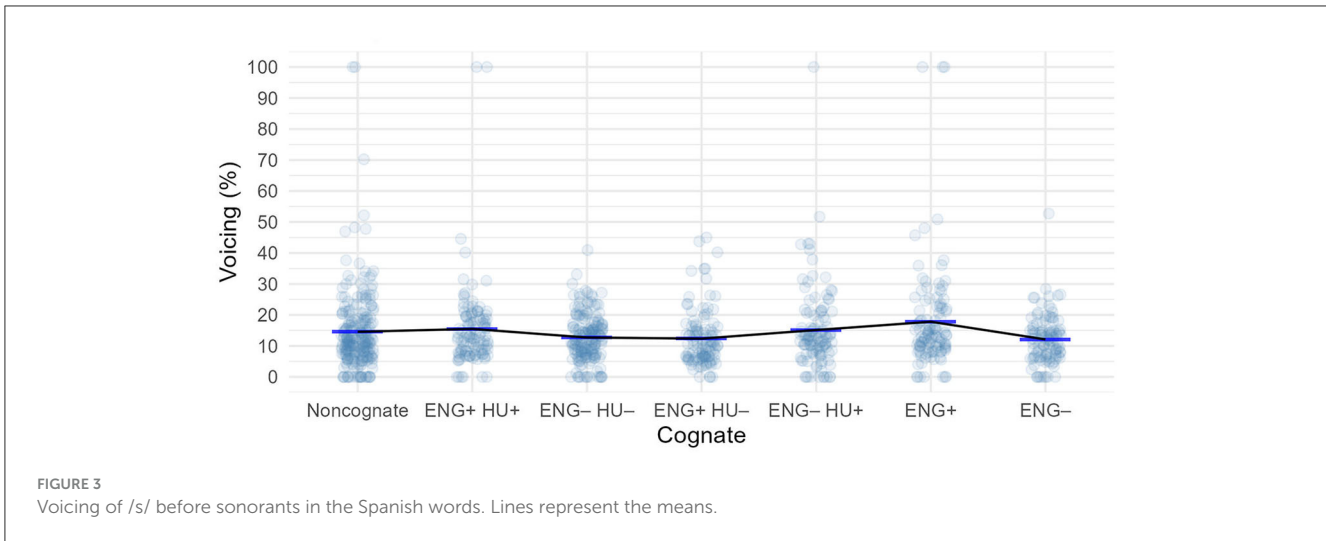


FIGURE 2
Voicing of /z/ before sonorants in the English words. Lines represent the means.

words and the SP+HU- cognates was statistically significant (Supplementary Table 2). Pairwise comparisons did not uncover further significant group differences.

4.1.1.3 Spanish words with presonorant /s/

The fricative remained relatively voiceless across all groups in the Spanish words (Figure 3; mean voicing ranged between 12.1



and 17.8%). Three groups showed a small amount of increase in the mean voicing: ENG+HU+, ENG-HU+, and ENG+, which had the highest mean at 17.8 (still relatively little voicing though).

The best-fitting linear mixed effects model for the Spanish presonorant data was the one with varying by-subject intercepts and varying by-item intercepts. According to this model, there was no statistically significant difference between the non-cognate words and any of the cognate groups, i.e., we cannot observe any cognate status effect (Supplementary Table 3). Pairwise comparisons did not uncover any significant group differences either.

4.1.2 Cognate status effect before voiced stops

4.1.2.1 English words containing /s/ plus voiced stops

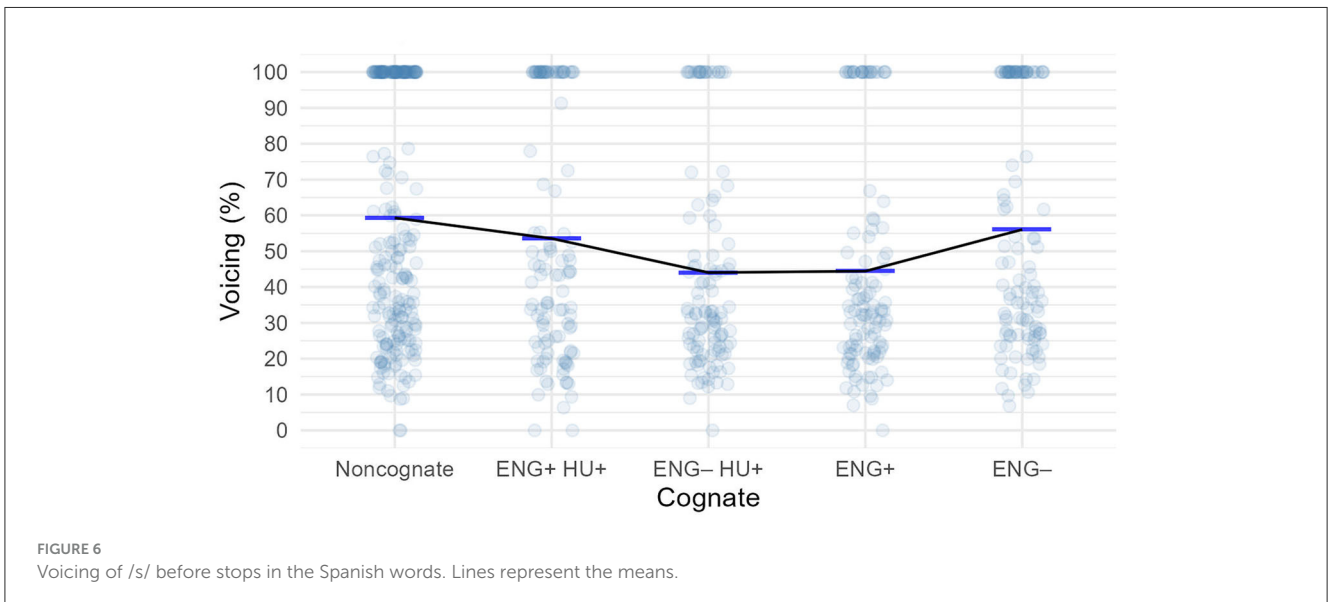
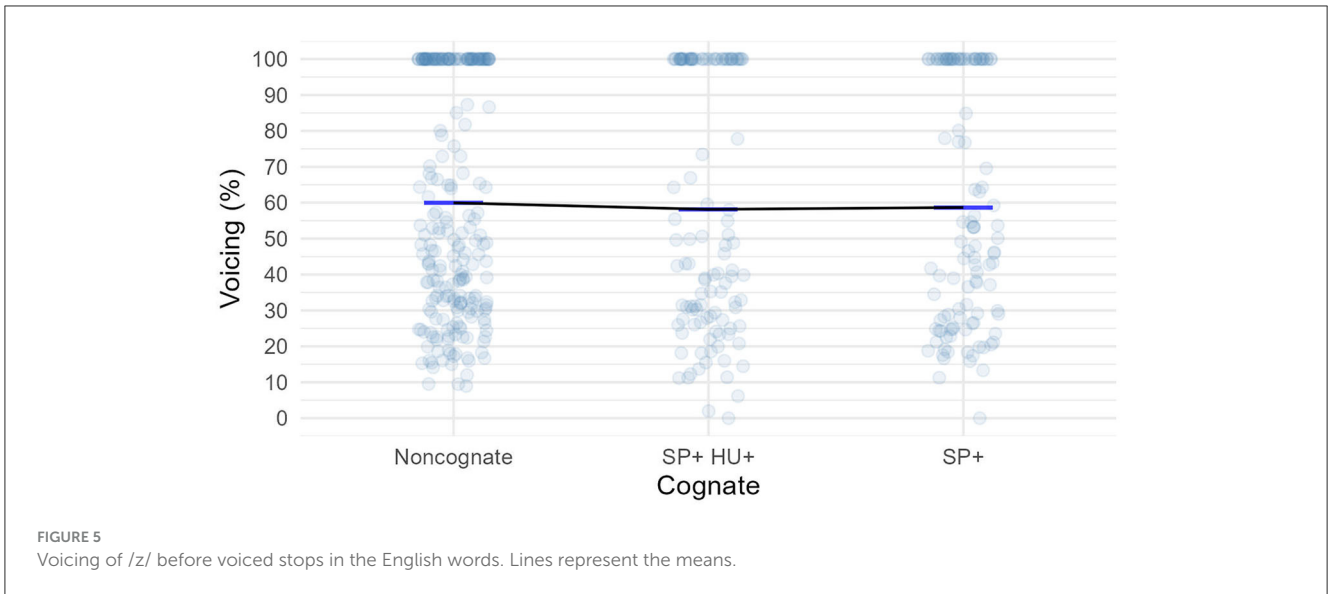
As we can see (Figure 4, Table 3), the “doubly” inhibitory cognate group (SP-HU-) showed the most average voicing (60.1%), but the fricative contained a fair amount of voicing in the other

TABLE 3 Descriptive statistics for the voicing of /s/ before voiced stops in English words in the cognate groups.

Cognate	N	Mean	SD	Median	Min	Max
Non-cognate	224	48.96	33.67	35.77	0.00	100
SP-HU-	112	60.07	33.68	51.19	1.62	100
SP-	112	48.86	34.68	37.02	0.00	100

groups, too (close to 50% on average). We note that this group only included one word, *baseball*.

This data was modeled with a linear mixed effects model that contained by-subject and by-item varying intercepts. According to this model, neither of the cognate groups had a significantly different amount of voicing compared to the non-cognate words (Supplementary Table 4). Neither did the pairwise comparisons show a significant difference between the groups.



4.1.2.2 /z/ before voiced stops in English words

As [Figure 5](#) shows, the fricative contained a fair amount of voicing across all three groups (around 60%) in this case.

The best model fitted to the data was the one with by-subject intercepts and slopes, and by-item varying items. As expected, the model did not uncover any significant differences between the cognate groups and the non-cognate group, or between the cognate groups ([Supplementary Table 5](#)).

4.1.2.3 /s/ before voiced stops in Spanish

Just like for English /s/ and /z/, /s/ in the Spanish words was articulated with a considerable amount of voicing (between 44 and 60% on average; see [Figure 6](#), [Table 4](#)).

The best-fitting model for the Spanish data was the one in which the intercepts were allowed to vary for both subjects and items, but

TABLE 4 Descriptive statistics for the voicing of /s/ before voiced stops in Spanish words in the cognate groups.

Cognate	N	Mean	SD	Median	Min	Max
Non-cognate	224	59.30	34.06	50.12	0.00	100
ENG+HU+	112	53.55	34.17	44.10	0.00	100
ENG-HU+	111	44.04	29.67	32.69	0.00	100
ENG+	112	44.43	30.89	33.12	0.00	100
ENG-	112	56.04	33.60	41.24	6.83	100

not the slopes. According to this model, voicing in none of the cognate groups was significantly different from that in the non-cognate group. Pairwise comparisons did not uncover significant difference between any of the groups ([Supplementary Table 6](#)).

4.1.3 PSV and RVA across a word boundary

Now we turn to the results of the production data that involved the voicing of word-final /s/ followed by a word that began with a sonorant consonant or a voiced stop, i.e., PSV and RVA across a word boundary, in English and Spanish words. The mean voicing percentages can be seen in [Figure 7](#) (the descriptive statistics are tabulated in [Table 5](#)). We can see that /s/ had little voicing before sonorants on average (in the Spanish words mean voicing was somewhat higher but the difference was not significant according to pairwise comparisons), whereas before voiced stops it was far more voiced, especially in the Spanish words in which /s/ contained 81% voicing on average. These results indicate the general absence of PSV and a strong RVA effect in both English and Spanish words.

The best-fitting linear mixed effects model contained trigger (sonorant vs. voiced stop), language (English vs. Spanish), as well as their interaction as fixed-effect predictors, and by-subject varying intercepts and correlated slopes for both trigger and language (but not their interaction), and by-item varying intercepts as random effects. According to this model, the main effect of *trigger* (i.e., the difference between the estimated marginal mean voicing of final /s/ before sonorants and voiced stops) was statistically significant: voicing of /s/ word-finally was significantly greater before the voiced stops compared to before the sonorants. The main effect of *language* (i.e., the difference between the estimated marginal mean voicing in English and Spanish words) was also statistically significant: /s/ had significantly more overall voicing in Spanish than in English. And finally, the *interaction* term was also statistically significant: the difference between PSV and RVA of final /s/ was significantly greater in Spanish words than in English words. [Supplementary Table 7](#) provides a summary of the model.

4.2 Word-internal vs. word-final /s/

We also compared the voicing of word-internal and word-final /s/ in the English and Spanish words in the two environments (before sonorants and before voiced stops, see [Figure 8](#), [Table 6](#)). Since the cognate status was not controlled for in the word-final position, the word-internal group only contained the non-cognate words in this comparison. As we can see, before sonorants, the mean voicing of /s/ is rather similar in the two environments (word-internal: 15.2%, SD = 14.2%; word-final: 18.6%, SD = 11.1%). The mean voicing of /s/ before voiced stops is, however, much higher in both environments (word-internal: 54%, SD = 34.2%; word-final: 66.7%, SD = 34.4%). In addition to this, we can again observe an interaction effect of the language of the words before voiced stops: /s/ had much more voicing word-finally than word-internally in the Spanish words compared to the English words.

We fitted two linear mixed effects models separately for the two trigger sounds (sonorants vs. voiced stops). The best model in both cases was the one that included environment (word-internal vs. word-final), language (English vs. Spanish), as well as their interaction as fixed-effect predictors, and by-subject varying intercepts and correlated slopes for both environment and language, and their interaction. When the trigger sound was a sonorant, the main effects of environment and language were not statistically significant; however, there was a significant

interaction effect: the difference between the mean voicing of /s/ word-internally vs. word-finally was greater in the Spanish words. When the trigger was a voiced consonant, the main effects of environment and language, as well as their interaction, were statistically significant: i.e., overall, there was more voicing in /s/ word-finally, and there was more voicing in /s/ in Spanish overall; in addition, the difference between the mean voicing of /s/ word-internally vs. word-finally was greater in the Spanish words again. [Supplementary Table 8](#) exhibits a summary of the two models.

4.3 Perception experiment

We begin with the English results. The left part of [Figure 9](#) displays the ratings of the participants of the two texts. “Tom” was the native-like recording, while “TomVoi” was the one where Hungarian-like RVA was applied (1 = not at all native-like, 5 = completely native-like). As we can see, the ratings were lower for the non-native-like recording; for example, no participant ranked it with the highest score 5.

The best-fitting cumulative link mixed effects model was the one which included by-subject varying intercepts (but not slopes). According to this model, the non-native text significantly decreased the ratings, i.e., lower ratings were more likely ($b = -1.2$, $SE = 0.27$, $z = -4.46$, $p < 0.001$; $R2m = 0.54$, $R2c = 0.17$). These results indicate that the participants reliably differentiated between the two recordings, and rated the non-native one (with RVA applied) much lower.

The ratings of the Spanish recordings can be seen on the right of [Figure 9](#). “Leyenda” was the native-like recording while “LeyendaSOVO” was the one in which no sonorant voicing was applied in the relevant phonological environments, hence this was the non-native-like recording.

The best-fitting model was the one which included by-subject varying intercepts (but not slopes). According to this model, while the non-native recording increased ratings, this increase was not statistically significant ($b = 0.46$, $SE = 0.24$, $z = 1.9$, $p = 0.056$; $R2m = 0.41$, $R2c = 0.03$). This result indicates that listeners did not reliably differentiate between the native recording (with PSV) and the non-native recording (without PSV).

5 Discussion

This study examined to what extent sequential multilingual speakers produce and perceive postlexical laryngeal processes in their L2/L3. The study also explored if cognates enhance CLI and if so, what properties determine whether it is the L1 or the L2 that has a more significant impact. The implications of the results presented in the previous section are described as follows.

5.1 Cognate status effect and voicing assimilation

The results of the production task revealed a somewhat complex picture of cognate status effect in relation to presonorant

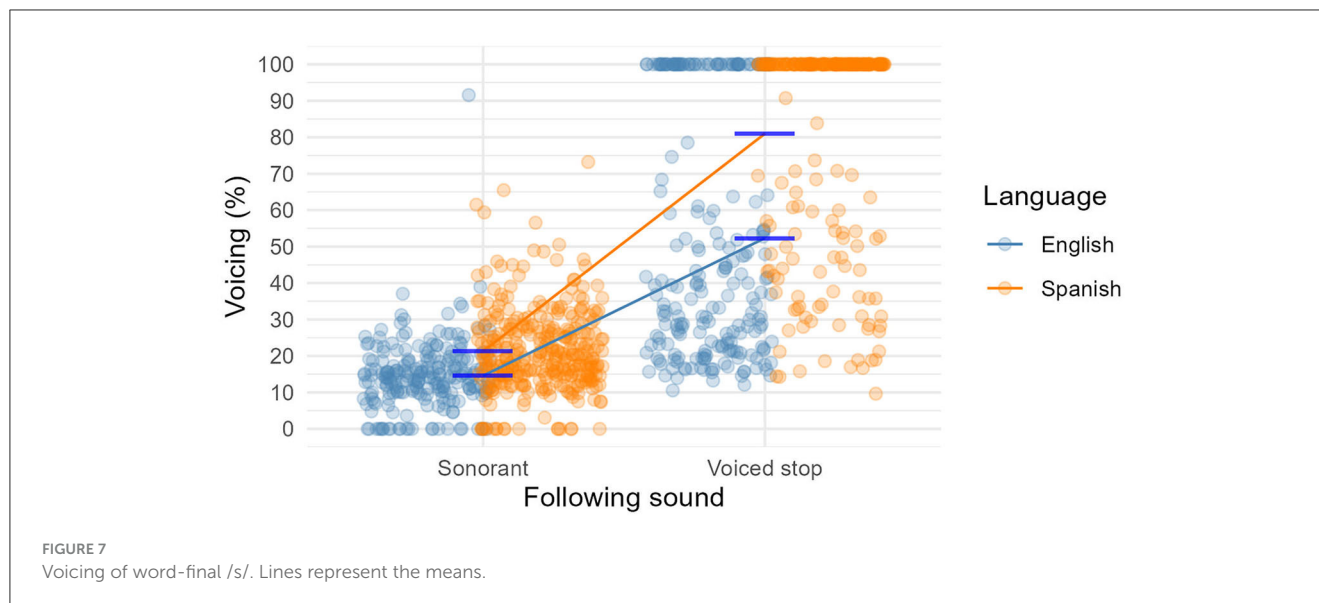


TABLE 5 Descriptive statistics for the voicing of word-final /s/ before stops and voiced stops in English and Spanish words.

Trigger	Language	N	Mean	SD	Median	Min	Max
Sonorant	English	220	14.58	9.41	14.17	0.00	91.57
Sonorant	Spanish	336	21.31	11.31	19.49	0.00	73.22
Voiced stop	English	219	52.32	33.41	39.62	10.55	100.00
Voiced stop	Spanish	220	80.99	28.97	100.00	9.65	100.00

voicing and regressive voicing assimilation. In the pre-sonorant voicing context, English /s/—in a non-target-like manner—displays increased voicing in cognates where L1 is inhibitory (Hungarian has a /z/ before the sonorant), thus supporting Hypothesis 1, although statistical analysis did not yield a significant difference here between the different lexical groups. This result calls for further caution as the increased voicing might be due to methodological reasons. The word that was responsible for it was *Yasmin*, which in some varieties of English is pronounced with a voiced fricative. The other inhibitory cognate, *Iceland* was probably not perceived as similar enough to produce an impact since vowel quality was too different (a diphthong in English while /i/ in both Hungarian and Spanish). On the other hand, in the realization of pre-sonorant English /z/, where pre-sonorant voicing applies vacuously, a cognate status effect was observed. Cognates with an inhibitory L1 influence were realized with significantly less voicing than the other lexical groups. Thus, English production data seem to support Hypothesis 1. It also means that the facilitative effect of L2 Spanish in these triple cognates could not counterbalance the inhibitory effect of L1 Hungarian, thus supporting Hypothesis 3. The reverse scenario (L1 facilitation and L2 inhibition) could not be tested in the latter phonological context since Spanish is always facilitative here because of PSV.

On the contrary, the Spanish production data do not show any cognate effects. There is a steady absence of pre-sonorant voicing, thus (partly) refuting Hypothesis 1. The results also reveal that facilitative cognates do not differ from non-cognate realizations,

thus Hypothesis 2 (i.e., that facilitative cognates are acoustically more target-like than non-cognates) for pre-sonorant voicing must be rejected. The reason for this could be that non-cognates already show a target-like realization which reached a ceiling with no possibility for further improvement. This, however, is not borne out for PSV in Spanish. Spanish pre-sonorant /s/ in non-cognates is as voiceless (i.e., non-target-like) as English pre-sonorant /s/ in non-cognates (which is target-like).

The question arises why our participants behaved differently in this respect in their two non-native languages. We hypothesize that the answer lies in phonemic encoding during the acquisition of these lexical items. While we think that both English and Spanish voiced realizations are acoustically more similar and should be identified with or mapped to Hungarian [z], and English and Spanish voiceless realizations are more similar to and should be mapped to Hungarian [s], the acquisition of a phoneme inventory is closely linked to the acquisition of a lexicon that includes minimal pairs (Darcy et al., 2017). As our participants are proficient speakers, they are likely to have acquired a stable phoneme inventory for both L2/L3 and have formed only one alveolar fricative category for Spanish, which is voiceless because Spanish does not display a /s/-/z/ contrast. In their Spanish speech, they implement only this voiceless segment across the board.

Turning to regressive voicing assimilation between adjacent obstruents, in neither of the two languages did participants treat cognates significantly differently from non-cognates, although some tendencies could be observed. In English, /s/ was on average

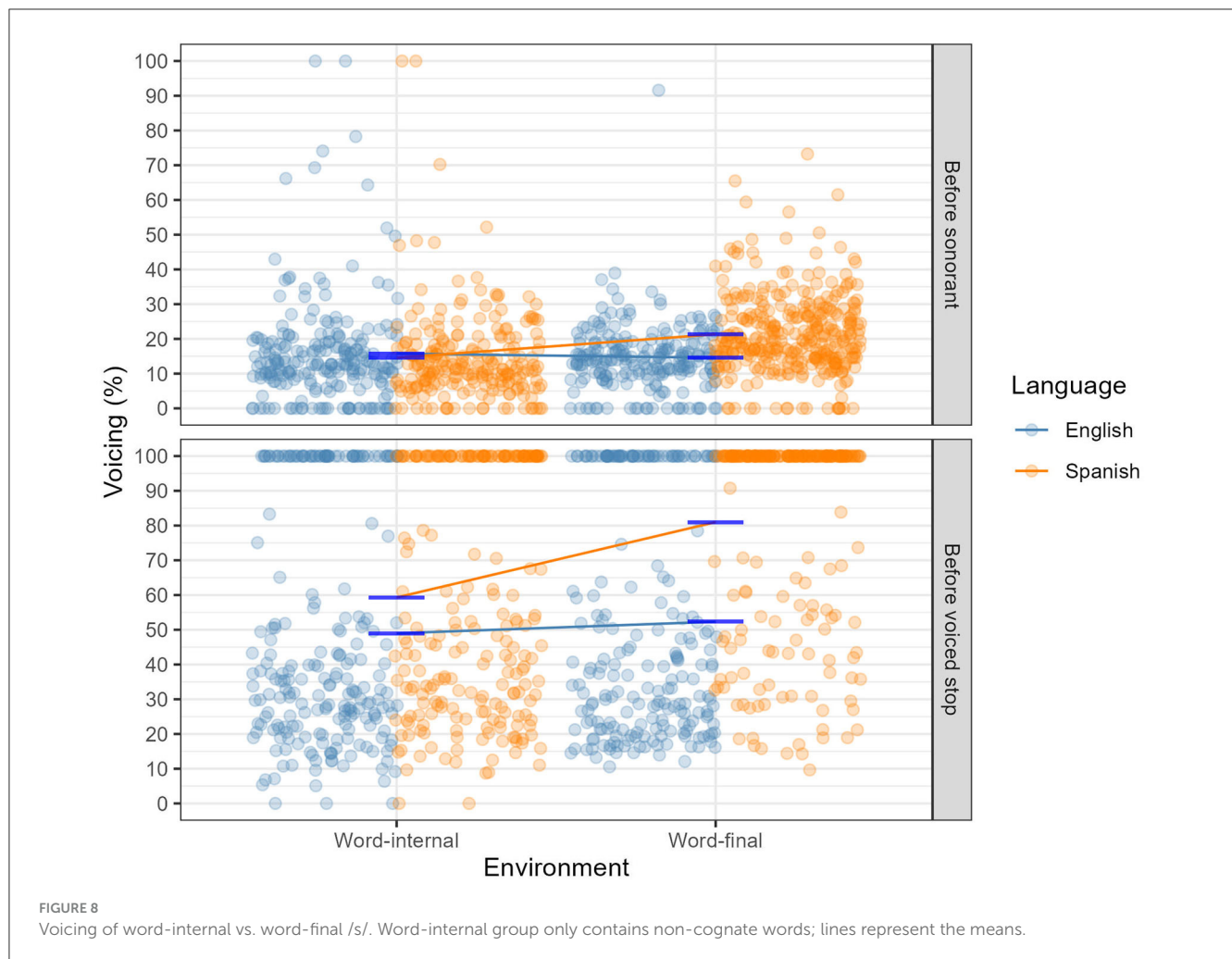
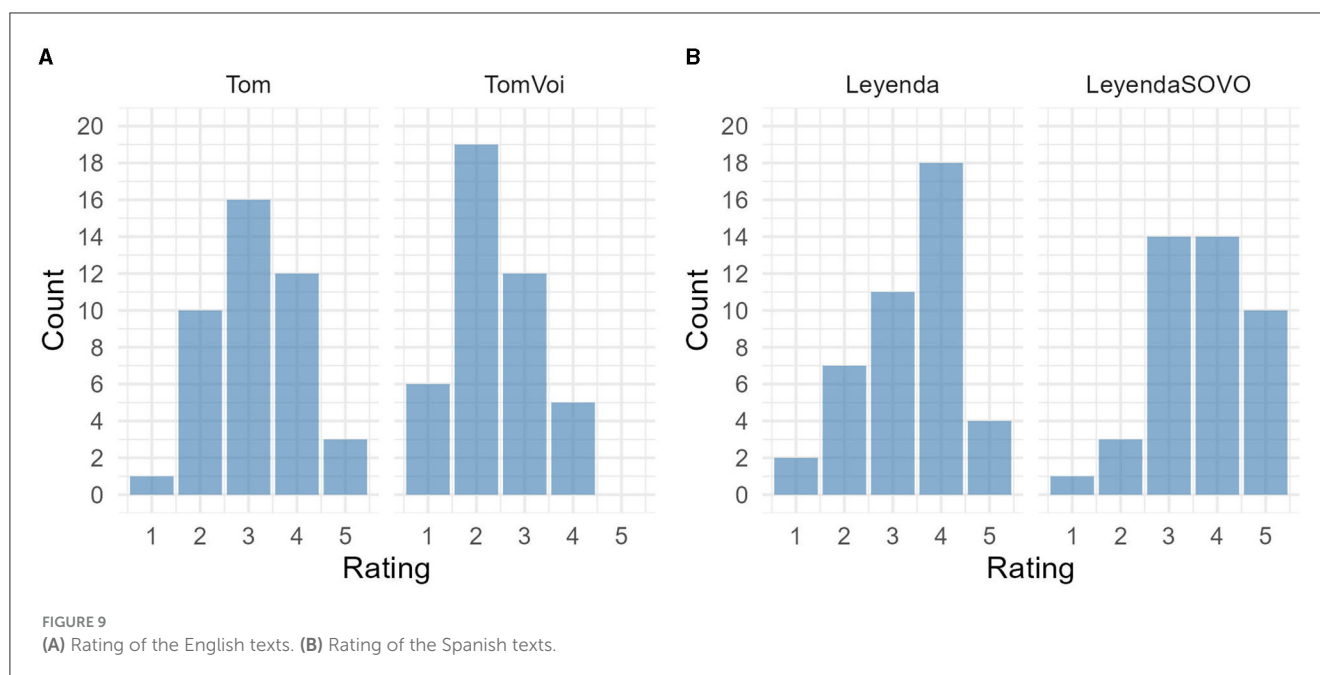


TABLE 6 Descriptive statistics for the voicing of word-internal vs. word-final /s/ before stops and voiced stops in English and Spanish words.

Trigger	Environment	Language	N	Mean	SD	Median	Min	Max
Sonorant	Word-internal	English	224	15.75	15.12	12.83	0.00	100.00
Sonorant	Word-internal	Spanish	224	14.57	13.15	11.81	0.00	100.00
Sonorant	Word-final	English	220	14.58	9.41	14.17	0.00	91.57
Sonorant	Word-final	Spanish	336	21.31	11.31	19.49	0.00	73.22
Voiced stop	Word-internal	English	224	48.96	33.67	35.77	0.00	100.00
Voiced stop	Word-internal	Spanish	224	59.30	34.06	50.12	0.00	100.00
Voiced stop	Word-final	English	219	52.32	33.41	39.62	10.55	100.00
Voiced stop	Word-final	Spanish	220	80.99	28.97	100.00	9.65	100.00

11% more voiced before voiced stops in triple cognates than in non-cognates or English–Spanish cognates, again pointing in the direction of L1 having a larger impact on the phonetic realization of cognates than L2 (thus supporting Hypothesis 3) but only in the case of inhibitory cognates, thus, supporting Hypothesis 1; Hypothesis 2 must be rejected for RVA, too. In the Spanish data dispersion was slightly greater (15%), and no clear trend could be observed. Note that L1 is always facilitative in this context, just like in English words with /z/+voiced obstruent sequences. It is important to bear in mind that /s/ was produced with a fair amount

of voicing in all these contexts which is non-target-like for English /s/ and target-like for English /z/ and Spanish. Thus, our results indicate that any potential lexical effects tend to be overridden by RVA. The voicing proportion measured in the present data is in line with research on RVA in Hungarian. In Bárkányi and G. Kiss (2019), the proportion of voicing in /s/ before voiced stops was on average 65.4% and before voiceless stops it was 15.1%. It has also been demonstrated that around 30% of voicing during the fricative is enough to induce voiced categorization, that is, an alveolar fricative with 30% of voicing proportion is more likely to



be categorized as /z/ by speakers of Hungarian (Bárkányi and G. Kiss, 2023).

5.2 Voicing assimilation across the board

The experiments were specifically designed to explore whether voicing assimilation as a dynamic process has been learned and unlearned. In order to test this, it is crucial to examine RVA and PSV across a word-boundary with the target segment being at the end of one word and the trigger in the next word. The patterns we observed are similar to those within the word. PSV does not seem to be applied in either of the two non-native languages. /s/ contains little coarticulatory voicing, which is expected and target-like in English, but had PSV been acquired, more voicing would be expected in Spanish. This experimental data supports Hypothesis 4 (i.e., sonorants do not trigger voicing assimilation in sandhi). It is interesting to note that although Spanish word-final /s/ was fairly voiceless (only 21.3% on average), it was voiced significantly more than word-internal /s/ or English /s/ in the same sandhi position. The reason for this might be that despite the fact that participants store the Spanish lexical items with a voiceless fricative and generally produce it as such at word level, two learners seem to have acquired PSV to some extent. They consistently produced tokens with 40–60% voicing in sandhi contexts. Overall, however, we can claim that there is little evidence for PSV of either within the word or across a word-boundary in the Spanish interlanguage of these multilingual advanced learners. There are several factors that could have contributed to this. An appealing explanation is the considerable similarity of the Hungarian and Spanish laryngeal systems, namely, that both are true voice languages, both display RVA between adjacent obstruents, thus they are treated by our participants as having identical laryngeal systems. In addition, the fact that sonorant voicing in Spanish is variable might not serve as sufficient and salient input for learners to be “discovered.”

Unlike PSV, the non-target-like application of RVA in English is perceived by the participants of this study; however, their productions do not mirror it: /s/ before a voiced stop was predominantly voiced (around 50%). This means that participants failed to block RVA in their English interlanguage, which supports Hypothesis 5. It is interesting to note that Spanish word-final /s/ resulted in significantly more voiced realizations than English /s/ in the same position (the word-internal fricative was also slightly more voiced in Spanish than in English). This might suggest that participants do aim to block RVA in English, but they are not very successful. This point is in need of further research as we cannot explain why the sandhi context in Spanish triggered so much more voicing than the same context within the word.

5.3 The link between perception and production

The results of the perception experiments also support the hypothesis that PSV in the Spanish interlanguage of the participants remained unnoticed (Figure 9). This does not necessarily mean that learners cannot hear voicing itself, but even if they do, they perceive it as random noise rather than a language specific phonological process. These data do not provide evidence in favor of any of the scenarios described in the Introduction (Section 1), we can only state that overall, PSV was not acquired in any domain, which is somewhat surprising since the participants of the present study are advanced learners. We consider it as an indication that the acquisition of dynamic phonological processes is different from the acquisition of contrastive segments, static inventories. We found intra- and interspeaker variation in the data (as mentioned in 5.2), we leave the exploration of individual learning patterns for future research.

The presence of RVA in the English interlanguage of the participants shows that learners do hear the non-target-like

application of the process, but fail to block its implementation in their productions. This could be compatible with the predictions made by SLM—assuming this model can be extended to phonological processes, too—namely, that only accurate perception can be transferred to production. Our participants could be at the stage of accurate perception which has not yet been transferred to production. As learners in this case did not have to acquire any non-native segments, the possibility of a motor-articulatory difficulty blocking correct production can be discarded, as would be the case, for instance, when acquiring a trilled rhotic by speakers whose L1 does not have one. However, as our participants are highly proficient speakers of English, it is quite unlikely that they will suppress RVA at a later stage. Thus, these results are also compatible with the claim that there is no direct correlation between perception and production, as reported by Liu and Lin (2021). Wrembel et al. (2020) found that Polish speakers did not show significant development in the acquisition of voicing and devoicing in English and German in the perception domain. This again, indicates that the acquisition of dynamic processes might have a different pattern and a ceiling effect might be reached earlier than in the case of learning contrastive segments.

5.4 Theoretical implications and future research

The findings of the present study can probably be best accounted for by Slabakova's (2017) Scalpel Model (SM). This model has been proposed for L3 morphosyntactic acquisition, just like the Linguistic Proximity Model, and also sees linguistic proximity as a decisive factor in determining transfer, but explicitly claims that additional (cognitive and experiential) factors can have a significant impact on CLI. Such factors include structural linguistic complexity, construction frequency, misleading input, negative evidence and prevalent language activation. We can discard negative evidence for now as learners are rarely corrected for the “erroneous” application of postlexical processes. As far as language activation is concerned, L1 is the language predominantly used by these learners, but there is no clear usage difference between their L2/L3. When extending SM for L3 phonology, we can add articulatory and perceptual complexity next to structural linguistic complexity. As mentioned in 5.3, articulatory complexity is not likely to play a role here, while perceptual salience (or the lack of it) might be important. Its impact on production needs further research: whether allophonic alternations that create salient novel segments that are easy to articulate are more likely to be implemented in production. The question of insufficient or “misleading” input due to the variable nature of PSV in Spanish has been dealt with in 5.2, and is a plausible reason for the lack of acquisition, which might be supported by sequence frequency, although frequency alone is not likely to have an impact on the acquisition of PSV and RVA. Based on CORPES (XXI), the average of the normalized frequency (per million words) of /s/+sonorant (/m, n, l/) and /s/+voiced stops (/b, d, g/) is as follows. While the sequence of /s/+sonorant is slightly more frequent than that of /s/+voiced stop within the word (1,346.5 as opposed to 1,041.4), across the word boundary, there are twice as many /s/+voiced stop occurrences than /s/+sonorant ones (9,906.7 vs. 4,897.8).

The role of occurrence frequency in postlexical processes awaits further research.

Participants failed to learn PSV, a typologically uncommon process. However, they also failed to unlearn RVA, a typologically common process, a default process for true voice languages, but absent from aspirating languages. This might indicate that they simply transferred their L1 laryngeal system into their subsequent languages. Already Eckman (1977) pointed out that German learners could not suppress laryngeal neutralization in English (but English learners had less difficulty in learning the laryngeal properties of German), which the author explains with typological markedness. It would be worth testing whether learners whose native language displays both PSV and RVA are able to block the former more readily than the latter. The primacy of L1 was corroborated by the inhibitory transfer in triple cognates too. Therefore, one important finding of this study is that in postlexical phonological processes, L1 plays the primary role, and L2 does not appear to exert a strong enough influence. It is to be clarified by further research in what ways the acquisition of phonemic contrasts and static phonological features differ from the acquisition of dynamic phonological processes, and whether there is a parallelism between these and the acquisition of lexical items and grammatical knowledge.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the studies involving humans because, it is not mandatory at Eötvös Loránd University. All participants received an information sheet about the purpose, data collection and anonymization in relation to the present experiments. They all gave their written informed consent. The studies were conducted in accordance with the local legislation and institutional requirements.

Author contributions

ZsB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing—original draft, Writing—review & editing. ZGK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present research has been supported by the National Research, Development and Innovation Grant: K142498.

Acknowledgments

We would like to thank Firas Shbeeb for his help with the experiments and acoustic analysis, Anna Hamp for her help in the preparation of materials, Péter Siptár for his comments on earlier versions of this paper, and the editor and the two reviewers for their very valuable comments that made this paper much better.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Amengual, M. (2012). Interlingual influence in bilingual speech: cognate status effect in a continuum of bilingualism. *Bilingualism* 15, 517–530. doi: 10.1017/S1366728911000460
- Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Front. Psychol.* 7, 617. doi: 10.3389/fpsyg.2016.00617
- Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *J. Acoust. Soc. Am.* 149, 1271–1284. doi: 10.1121/10.0003559
- Archibald, J. (2022). Phonological parsing via an integrated I-language: the emergence of property-by-property transfer effects in L3 phonology. *Linguist. Approach. Bilingual.* 13, 743–766. doi: 10.1075/lab.21017.arc
- Baker, W., and Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: the role of individual differences. *Int. Rev. Appl. Linguist.* 44, 231–250. doi: 10.1515/IRAL.2006.010
- Bardel, C., and Sánchez, L. (2017). “The L2 status factor hypothesis revisited: the role of metalinguistic knowledge, working memory, attention and noticing in third language learning,” in *L3 Syntactic Transfer: Models, New Developments and Implications*, eds T. Angelovska and A. Hahn (Amsterdam: John Benjamins Publishing Company), 85–102.
- Bárkányi, Zs. (2014). “Reflexiones sobre la asimilación de sonoridad de la l. Introducción,” in *GPS 60: Köszöntő kötet Giampaolo Salvi 60. születésnapjára. Studi di linguistica neolatina per i 60 anni di Giampaolo Salvi* GPS 60, ed Z. Fábán (Budapest: ELTE), 25–36.
- Bárkányi, Zs., and G. Kiss, Z. (2015). “Why do sonorants not voice in Hungarian? And why do they voice in Slovak?” in *Approaches to Hungarian: Volume 14: Papers from the 2013 Piliscsaba Conference*, eds K. É. Kiss, B. Surányi, and É. Dékány (Amsterdam: John Benjamins Publishing Company), 65–94.
- Bárkányi, Zs., and G. Kiss, Z. (2019). A fonetikai korrelátumok szerepe a zöngékontraszt fenntartásában. Beszédprodukción és észlelés eredmények [The role of phonetic correlates in voicing contrast. Results from speech production and perception]. *Általános Nyelvészeti Tanulmányok* 31, 57–102.
- Bárkányi, Zs., and G. Kiss, Z. (2023). Production and perception of voicing contrast in assimilation contexts in Hungarian. *J. Uralic Linguist.* 2, 5–49. doi: 10.1075/jul.00013.bar
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Beckman, J., Jessen, M., and Ringen, C. (2013). Empirical evidence for laryngeal features: aspirating vs. true voice languages. *J. Linguist.* 49, 259–284. doi: 10.1017/S0022226712000424
- Benrabah, M. (1991). “Learning English segments with two languages,” in *Actes du XIIème Congrès International des Sciences Phonétiques* (Aix-en-Provence: Université de Provence, Service des Publications), 334–337.
- Berkes, É., and Flynn, S. (2012). “Multilingualism: new perspectives on syntactic development,” in *The Handbook of Bilingualism and Multilingualism* (Hoboken, NJ: John Wiley & Sons, Ltd), 137–167.
- Best, C. T. (1994). “The emergence of native-language phonological influences in infants: a perceptual assimilation model,” in *The Development of Speech Perception: Theoretical and Methodological Issues in Cross-Language Speech Research*, ed H. C. Nusbaum (Timonium, MD: York Press), 171–204.
- Best, C. T., and Tyler, M. D. (2007). “Commonalities and complementarities: nonnative and second-language speech perception,” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds O. S. Bohn and M. J. Munro (Amsterdam: John Benjamins Publishing Company), 13–34.
- Boersma, and Weenink, D. (2022). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.2.23*. Retrieved from <https://www.praat.org> (May 2, 2023).
- Bolker, B., and Robinson, D. (2022). *Broom.mixed: Tidying Methods for Mixed Models*. Retrieved from <https://CRAN.R-project.org/package=broom.mixed> (May 2, 2023).
- Bybee, J. (2001) *Phonology and Language Use*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics).
- Cabrelli, J., and Pichan, C. (2021). Initial phonological transfer in L3 Brazilian Portuguese and Italian. *Linguist. Approach. Bilingual.* 11, 131–167. doi: 10.1075/lab.18048.cab
- Cal, Z., and Sypiańska, J. (2020). The interaction of L2 and L3 levels of proficiency in third language acquisition. *Poznan Stud. Contemp. Linguist.* 56, 577–603. doi: 10.1515/pscl-2020-0019
- Campos-Astorkiza, R. (2015). “Segmental and prosodic conditionings on gradient voicing assimilation in Spanish,” in *Hispanic Linguistics at the Crossroads: Theoretical Linguistics, Language Acquisition and Language Contact. Proceedings of the Hispanic Linguistics Symposium 2013*, eds R. Klassen, J. M. Liceras, and E. Valenzuela (Amsterdam: John Benjamins Publishing Company), 127–144.
- Christensen, R. H. B. (2022). *Ordinal: Regression Models for Ordinal Data. R Package Version 2022.11-16*. Retrieved from <https://CRAN.R-project.org/package=ordinal> (May 2, 2023).
- CORPES XXI. *Real Academia Española: Corpus del Español del Siglo XXI*. Available online at: <https://www.rae.es/banco-de-datos/corpes-xxi> (accessed November 2, 2023).
- Darcy, I., Peperkamp, S., and Dupoux, E. (2007). Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. *Pap. Lab. Phonol.* 9, 9.
- Darcy, I., Tremblay, A., and Simonet, M. (2017). Editorial: phonology in the bilingual and bidialectal lexicon. *Front. Psychol.* 8, 2017. doi: 10.3389/fpsyg.2017.00507
- Derwing, T., and Munro, M. J. (2015) *Pronunciation Fundamentals*. Amsterdam: John Benjamins Publishing Company.
- Draxler, C., and Jänsch, K. (2004). *SpeechRecorder: a Universal Platform Independent Multi-Channel Audio Recording Software*. Retrieved from <https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/> (May 2, 2023).
- Ecke, P., and Hall, C. J. (2021). The Parasitic Model: lexical acquisition and its impact on morphosyntactic transfer. *Linguist. Approach. Bilingual.* 11, 45–49. doi: 10.1075/lab.20088.eck
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Lang. Learn.* 27, 315–330. doi: 10.1111/j.1467-1770.1977.tb00124.x
- Escudero (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/flang.2023.1304666/full#supplementary-material>

- Flege, J., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)", in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002
- Flege, J. E. (1995). "Second-language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 157–232.
- Flege, J. E., Frida, E. M., Walley, A. C., and Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Stud. Sec. Lang. Acquisit.* 20, 155–187. doi: 10.1017/S0272263198002034
- Flege, J. E., and Munro, M. J. (1994). The word unit in second language speech production and perception. *Stud. Sec. Lang. Acquisit.* 16, 381–411. doi: 10.1017/S0272263100013437
- Hahn, E. D., and Soyer, R. (2005). Probit and logit models: differences in the multivariate realm. *J. Royal Stat. Soc.* 1–12.
- Hermas, A. (2015). The categorization of the relative complementizer phrase in third-language English: a feature re-assembly account. *Int. J. Bilingual.* 19, 587–607. doi: 10.1177/1367006914527019
- Hualde, J. I. (2005). *The Sounds of Spanish*. New York, NY: Cambridge University Press.
- Jansen, W. (2004) *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian, and Dutch*. (Doctoral dissertation), University of Groningen, Groningen, Netherlands.
- Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286–319.
- Kopečková, R., Gut, U., Wrembel, M., and Blas, A. (2022). Phonological cross-linguistic influence at the initial stages of L3 acquisition. *Sec. Lang. Res.* 2022, 1–25. doi: 10.1177/02676583221123994
- Kuzla, C., Ernestus, M., and Mitterer, H. (2010). "Compensation for assimilatory devoicing and prosodic structure in German fricative perception," in *Laboratory Phonology 10*, eds C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée, (Berlin, New York: De Gruyter Mouton), 731–758. doi: 10.1515/9783110224917.5.731
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lenth, R. V. (2023). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. Retrieved from <https://CRAN.R-project.org/package=emmeans> (May 2, 2023).
- Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830
- Liu, J., and Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *SAGE Open* 11, 10. doi: 10.1177/2158244020985510
- Llama, R., and Cardoso, W. (2018). Revisiting (non-)native influence in VOT production: insights from advanced L3 Spanish. *Languages* 3, 30. doi: 10.3390/languages3030030
- Llama, R., Cardoso, W., and Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *Int. J. Multilingual.* 7, 39–57. doi: 10.1080/14790710902972255
- Llama, R., and López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *Int. J. Multilingual.* 13, 444–58. doi: 10.1080/14790718.2016.1217602
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., and Makowski, D. (2021). An R package for assessment, comparison and testing of statistical models. *J. Open Sour. Softw.* 6, 3139. doi: 10.21105/joss.03139
- Mora, J. C., and Nadeu, M. (2009). Experience effects on the categorization of a native vowel contrast in highly proficient Catalan-Spanish bilinguals. *J. Acoust. Soc. Am.* 125, 2775. doi: 10.1121/1.4784755
- Pierrehumbert, J. B. (2001). "Exemplar dynamics: word frequency, lenition and contrast," in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Amsterdam: John Benjamins Publishing Company), 45, 137–157.
- Quilis, A. (1993) *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing. Retrieved from: <https://www.R-project.org/> (May 2, 2023).
- Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: timing of acquisition and proficiency considered. *Bilingualism* 18, 179–190. doi: 10.1017/S136672891300059X
- Schmidt, L., and Willis, E. (2011). "Systematic investigation of voicing assimilation of Spanish /s/ in Mexico City," in *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, ed S. M. Alvord (Somerville, MA: Cascadilla Proceedings Project), 1–20.
- Siptár, P., and Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford: Oxford University Press.
- Slabakova, R. (2017). The scalpel model of third language acquisition. *Int. J. Bilingual.* 21, 651–665. doi: 10.1177/1367006916655413
- Strycharczuk, P. (2012). *Phonetics-Phonology Interactions in Pre-sonorant Voicing*. (Doctoral dissertation), University of Manchester, Manchester, United Kingdom.
- Szigetvári, P. (2020). Emancipating lenes: a reanalysis of English obstruent clusters. *Acta Linguist. Acad.* 67, 39–53. doi: 10.1556/2062.2020.00004
- Vago, R. M. (1980) *The Sound Pattern of Hungarian*. Washington, DC: Georgetown University Press.
- van Leussen, J. W., and Escudero (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6, 1000. doi: 10.3389/fpsyg.2015.01000
- Wang, D., and Nance, C. (2023). Third language phonological acquisition: understanding sound structure in a multilingual world. *Lang. Linguist. Compass* 17, e12497. doi: 10.1111/lnc3.12497
- Wells, J. C. (1982) *Accents of English*. Cambridge: Cambridge University Press.
- Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: the Linguistic Proximity Model. *Int. J. Bilingual.* 21, 666–682. doi: 10.1177/1367006916648859
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* 4, 1686. doi: 10.21105/joss.01686
- Wrembel, M. (2021). Transfer vs. dynamic cross-linguistic interactions: exploring alternative avenues in L3 research. *Linguist. Approach. Bilingual.* 11, 109–115. doi: 10.1075/lab.20092.wre
- Wrembel, M., Gut, U., Kopečková, R., and Balas, A. (2020). Cross-linguistic interactions in third language acquisition: evidence from multi-feature analysis of speech perception. *Languages* 5, 52. doi: 10.3390/languages5040052
- Wrembel, M., Gut, U., Kopečková, R., and Balas, A. (2022). The relationship between the perception and production of L2 and L3 rhotics in young multilinguals; an exploratory cross-linguistic study. *Int. J. Multilingual.* 2022, 1–20. doi: 10.1080/14790718.2022.2036158
- Wrembel, M., Marecka, M., and Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *Int. J. Multilingual.* 16, 513–533. doi: 10.1080/14790718.2019.1583233
- Wunder, E. M. (2011). "Crosslinguistic influence in multilingual language acquisition: phonology in third or additional language acquisition," in *New Trends in Crosslinguistic Influence and Multilingualism Research*, eds G. D. Angelis and J. M. Dewaele (Bristol: Multilingual Matters), 105–128. doi: 10.21832/9781847694430