Phonological acquisition in multilingual speech

An experimental study of Hungarian learners of English and Spanish[1]

Zoltán G. Kiss

1. Introduction

Wang and Nance (2023) write that "It is estimated that over half of the worldwide population use two or more languages regularly in their daily life. While it is difficult to put an exact number on the total speakers using three or more languages worldwide, a European Commission study reported that 25% of EU teenagers were competent users of three languages. Globally, we know that there are many regions where large numbers of people regularly use multiple languages, as in West Africa, Malaysia, India, for example. Multilingualism is, therefore, relatively common across the world, and is likely experienced by over a billion people. It is also likely that the number of multilinguals is increasing rapidly every year. Despite the large numbers of multilingual language users globally, much of the previous phonological research has focused on monolingual and bilingual speakers, and work on third language (L3) phonological acquisition is limited."

Although third language acquisition has been of growing interest in the past decades, the driving forces behind multilingual phonological acquisition are still in need of research. Studies on third language 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 CLI. Some propose that it is the native language (L1) that has a privileged role and thus serves as a source of transfer (e.g., Hermas, 2015). Others argue that it is the second language (L2), which is acquired later (often in adulthood), and as such is cognitively more similar to L3 that is the main source of CLI (Bardel and Sánchez, 2017). It has been also argued that it is the typologically more similar language that is the primary cause of CLI (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 (Westergard et al., 2017) advocate both positive and negative transfer and claim that it occurs property-by-property rather than wholesale depending on which aspects of the native language or the second (non-native) language are perceived to be more **similar**. The modelling of similarity, however, is not straightforward.

Previous studies on L3 phonological acquisition have identified several factors that might contribute to CLI (see Wang and Nance, 2023 for an overview). They are:

- **L2 status:** the impact of previously learnt languages on L3 phonological acquisition, which has been demonstrated as a factor influencing the source of cross-linguistic effects on L3 perception; L2 transfer has been shown to be greater than L1 transfer at the beginning of L3 phonological acquisition; however, the influence of L2 on L3 phonological acquisition diminishes as L3 experience increases.
- **experience with L3:** this refers to refers to exposure to the L3, such as the length of residence in the L3-speaking environment. Increased L3 experience is suggested to facilitate sound discrimination, especially in the early stage of L3 acquisition.
- **typological/structural similarity:** this refers to the relationship between languages and language families that linguists can formally and objectively define and identify. Previous studies agree that CLI is more likely to occur between languages that are closely related. Specifically, L3 learners are likely to

[1] This text is an abbreviated and rewritten version of an article written together with Zsuzsanna Bárkányi.

establish links between the L3 and prior languages they have acquired. They tend to establish links between languages that have more similarities rather than differences, and the similarities between the L3 and L1/L2 can hinder multilinguals' ability to learn an L3. As well as the diachronic links between language families for example, between Spanish and Portuguese, research has also noted effects on L3 learning stemming from **structural similarities** and differences between unrelated languages.

- **level of proficiency (LoP):** previous studies have also demonstrated that LoP played an influential role in L3 sound production. Previous research suggests that the lower the level of L3 proficiency, the greater the CLI from L2 to L3. This implies that learners at the initial stage of L3 acquisition are likely to transfer L2 phonology directly into their L3 production. However, the influence of the L2 decreases as L3 proficiency increases, and L3 phonological categories are more likely to be influenced by input from the L3. In other words, lower proficiency in either the L2 or L3 makes (input of) the other, more proficient language, become the source of transfer onto L3 phonological categories.

Many studies argue in favour of **property-by-property transfer** in L3 phonology. For example, Benrabah (1991) reports that in the speech of Arabic learners of English consonants are transferred from Arabic, while vowels are form 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.

There is no widely applied L3 **phonological** acquisition model, but current L2 phonology models can potentially be extended to account for L3 speech acquisition (Wrembel et al., 2019). One of the most well-known models is the **Speech Learning Model (SLM)** (Flege, 1995). According to this model, L2 speech learning is shaped by **perceptual** biases caused by the L1 phonetic system, but production and perception evolve side by side during the acquisition process. 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. So, learners must discover the phonetic differences and break the L2-to-L1 perceptual link in order to form a new phonetic category. The delinking process can be speeded up by the growth of the L2 lexicon (i.e., learning more and more words).

Although L2 acquisition models potentially be extended to the context of L3 phonological acquisition, L3 perception and production still differ from those of L2 due to complicated sources of **language interactions**. It is logically possible that L3 acquisition is influenced by L1, or L2, or both. What is without a doubt is that some cross-linguistic transfer **will** happen, as it is natural and expected when someone uses more than one language.

Similarly to research on L2 speech, most previous studies on L3 phonology focus on the acquisition of phonemic categories and contrasts, and how L3 learners categorise speech sounds based on their phonetic properties. Almost all L2/L3 research has focussed on the acquisition of single, contrastive sounds (i.e., phonemes), and research on the acquisition of **allophonic alternations** and **dynamic ("postlexical") processes** has been 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 work in English. The main question we would like to answer is **how Hungarian non-beginner learners cope with the voicing phenomena whose properties are quite different in these three languages,** both in how they **produce** L2/L3 speech and how they **perceive** it.

2. Phonological background

2.1. Voicing in Hungarian, Spanish and English

2.1.1. Regressive Voicing Assimilation in Hungarian

Hungarian is a **voicing** language where the voicing contrast of obstruents is based on negative Voice Onset Time (VOT), i.e., voice lead in voiced stops vs. zero/short-lag VOT in voiceless stops. Hungarian has a symmetrical obstruent system with contrastive voiceless–voiced pairs at each place of articulation. For example, /s/ and /z/ contrast word-initially (1a), word-finally (1b), and within the word (1c).[2] Hungarian has regressive voicing assimilation: **adjacent obstruents must agree in their voicing feature**. This means that voiced obstruents voice preceding voiceless obstruents (2a), and voiceless obstruents devoice preceding voiced obstruents (2b).

(1) a. ***sz**ár* 'stem' vs. ***z**ár* 'lock'

b. *mé**sz*** 'whitewash' vs. *mé**z*** 'honey'

c. *má**sz**nak* 'they climb' vs. *má**z**nak* 'to the gloss'

(2) a. /tb/ ⟶ [db]: *há**tb**a* 'into the back'; *ké**t#b**arát* 'two friends'

/ʃb/ ⟶ [ʒb]: *ha**sb**a* 'into the stomach'; *hú**s#b**olt* 'meat shop'

b. /bt/ ⟶ [pt]: *lá**bt**ól* 'from the foot'; *lá**b#t**orna* 'foot exercise'

/zt/ ⟶ [st]: *ví**zt**ől* 'from water'; *ví**z#t**orony* 'water tower'

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, so obstruents maintain the voicing contrast before sonorants both within the word (4a) and across a word-boundary (4b). Thus, a voiceless obstruent does not become voiced before a sonorant: Hungarian does not have presonorant voicing.

(3) *lá**b**ak* [b] 'feet' ~ *lá**b*** [b] 'foot'

*lá**p**ok* [p] 'marshlands' ~ *lá**p*** [p] 'marshland'

*mé**z**ek* [z] 'honeys' ~ *mé**z*** [z] 'honey'

*me**sz**ek* [s] 'limestones' ~ *mé**sz*** [s] 'limestone'

(4) a. ***pl**akát* [pl] 'poster', ***bl**öki* [bl] 'doggy', ***sr**óf* [ʃr] 'screw', ***zr**í* [zr] 'fuss'

*ké**szn**ek* [sn] 'for the ready', *ké**zn**ek* [zn] 'for the hand'

b. /tm/ ⟶ [tm] (*[dm]): *ké**t#m**ag* 'two seeds'

/sl/ ⟶ [sl] (*[zl]): *ké**sz#l**eves* 'ready soup'

According to the traditional generative literature, RVA in Hungarian is categorical, exceptionless, and **completely neutralising** (Vago, 1980; Siptár and Törkenczy, 2000), which means that voiceless and devoiced sounds cannot be distinguished from each other. For example the stem *mé**sz*** 'limestone' with final [s] is realized

[2] Note that /s/ in Hungarian is spelt as 'sz' and /z/ is spelt as 'z'. The symbol '#' stands for a word boundary in the examples. A phonological process is called **sandhi** if it works across the word boundary (e.g., between two words).

with final [z] in the suffixed form *mész**ből*** 'from limestone' (pronounced as [-zb-]). This pronunciation of *mész* is the same as that of the word ***méz*** 'honey' with final [z] (i.e., *mészből* and *mézből* cannot be distinguished from each other, the /s/–/z/ contrast is **neutralized** in these words).

2.1.2. /s/-voicing in Spanish

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

Although Spanish also has RVA, because of the phonotactic restrictions of the language, the phoneme undergoing assimilation is mostly /s/. Spanish /s/-voicing presents a special case within RVA languages since /s/ has no voiced counterpart /z/ in this language (there are no minimal pairs with /s/–/z/). The Central-Northern Peninsular dialect 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 realisations from apical to laminal, interdental, etc. (Quilis, 1993). The voicing of /s/ in Spanish only occurs in dialects where syllable-final /s/ is not deleted. In these dialects, 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 presonorant voicing (PSV)** as in (5b). Thus, for Hungarian learners, PSV could be thought of as an **extension of Hungarian RVA** to include not only the preobstruent position but also the presonorant position.

(5) a. ***esb**elto* [zb] 'slim', ***es#b**ueno* [zb] 'it's good'[3]

b. ***isl**a* [zl] 'island', ***es#la**rgo* [zl] 'it's long'

c. ***deshi**elo* [zj] 'thaw', ***los#hi**elos* [zj] 'the ices'

Most phonologists who studied /s/-voicing in Spanish found high degrees of individual variation, thus some speakers voice /s/ others do not. Even the same speaker may show variation (sometimes s/he voices, sometimes does not). Note that although RVA is very common in voicing languages, PSV is much less frequent than preobstruent voicing, in this sense, PSV is a more "marked" voicing process than RVA.

2.1.3. The voicing pattern of English

English, just like Hungarian, but unlike Spanish, displays a symmetrical laryngeal obstruent system. But unlike Hungarian and Spanish, English is not a voicing language but an **aspirating** language, 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, so phonetically they are typically voiceless and unaspirated, while voiceless, or **fortis**, stops are produced prevocalically with a

---

[3] /b/ in this position is realized as the approximant [β]; this manner change does not affect RVA (i.e., there will be voicing before it, too).

relatively long-lag VOT (i.e., aspirated). In contrast to voicing languages, English does not have RVA. But similarly to Hungarian, English does not have presonorant voicing, either (6b).

(6) a. *ma**tchb**ox* [ʧb] (not *[ʤb]); *ane**cd**ote* [kd] (not *[gd]);
*ba**seb**all* [sb] (not *[zb]); *bonu**s#d**eal* [sd] (not *[zd])

b. *di**sl**oyal* [sl] (*[zl]); *mi**sm**atch* [sm] (*[zm]); *busine**ss#m**odel* [sm] (*[zm])

Therefore, as far as Hungarian L1 speakers of English are concerned, RVA is a rule that they must **block** from applying when articulating English obstruent+obstruent clusters: they must "unlearn" Hungarian and Spanish RVA. The relevant voicing properties of the languages are shown in Table 1. Observe where the languages are similar, and where they are different:

Table 1.

	Hungarian	**English**	**Spanish**
Basis of voicing contrast	voicing	aspiration	voicing
Voicing contrast within the obstruent inventory	symmetrical	symmetrical	symmetrical (but limited to stops only)
RVA	yes	no	yes
PSV	no	no	yes

2.2. Voicing in multilingual studies

Most studies that deal with the acquisition of laryngeal features in L3, focus on the phonetic realisation 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 the results have mostly shown the **impact of L**1. For example, 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, the speech of the participants displayed values in between the two languages. Llama and Cardoso (2018) also found that L1 plays a more decisive role in L3 pronunciation, but language proficiency and language dominance are also significant factors.

2.3. Cognate status effect and voicing

Previous research on bilingual speech indicates that **cognate status,** i.e., the **similarity of lexical items** – considerable phonological and semantic overlap – might impact on the acoustic realisation of segments within them (e.g., Mora and Nadeu, 2009; Amengual, 2016). The impact can be **facilitative** (i.e., positive) when the pronunciation in the target languages is the same as in the other language(s), and so it helps, "facilitates" the intended pronunciation. It can also be **inhibitory** (i.e., negative) when the target pronunciation is different from that in the other language(s), and so it acts against the intended pronunciation.

As an example, let us assume that we want to measure the voicing of English /s/ before a sonorant in the speech of L1 Hungarians, thus, English is the **target language**. In this position, /s/ is pronounced as a voiceless [s] in English, i.e., the intended **target pronunciation is a voiceless [s]**. We can use the word *asleep* to check this. This word does not exist in Hungarian, and so we do not expect a cognate effect on the voicing of /s/. We can also use the word *dyslexia*, which also contains the cluster /sl/, but this word exists in Hungarian as well as in Spanish (with similar spelling and meaning), therefore, this word may show a potential **cognate effect.** We call this cognate a **triple cognate** because the word exists in all **three** languages. In Hungarian, this word is pronounced with a voiceless [s] – just like in the target language English. Therefore, this word in Hungarian has a **facilitative/positive** effect on the pronunciation of the English word to be with [s]. However, the Spanish word *dyslexia* is pronounced with a voiced [z] (due to the rule of PSV), therefore, the Spanish "version" of the word has an **inhibitory/negative** effect on the English pronunciation. We can assume that the Spanish [z]-pronunciation will cause the English pronunciation to be with the "wrong" [z]. As we will see, we will call such cognate words as "SPA–HU+" (= inhibitory/negative in Spanish, facilitative/positive in Hungarian).

3. Hypotheses

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

The following hypotheses regarding the speech of L1 Hungarian speakers will be tested in the study:

Hypothesis 1: **inhibitory** cognates are realised with voicing properties **less similar** to those in the target language than non-cognates.

Hypothesis 2: **facilitative** cognates are more likely to be realised with voicing properties of the target language than non-cognates.

Hypothesis 3: when cognates are **contradictory**, e.g., L1 facilitative but L2 inhibitory, or L1 inhibitory but L2 facilitative, **it is the L1 pattern that dominates**.

Hypothesis 4: **sonorants do not cause presonorant voicing** across a word boundary in either Spanish (where they do) or English (where they do not).

Hypothesis 5: **obstruents cause regressive voicing assimilation** across a word boundary in both Spanish (where they do) and English (where they do not).

4. Materials and methods

The hypotheses were tested by carrying out two experiments: a **speech production** experiment and a **speech perception** experiment. The production experiment investigated how much voicing English and Spanish [s] and [z] contained before a voiced stop, and before a sonorant sound. Voicing was measured both within words and between words (= sandhi context). The perception experiment aimed to investigate if the participants perceived RVA and PSV in English and Spanish.

4.1. Participants

Fourteen young adult subjects (5 male 9 female) participated in the experiment. Their ages ranged between 19–25 years (average 21.6). 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. As there is no clear time sequential or dominance difference between English and Spanish for these participants, we refer to both English and Spanish as L2 or L3. Eleven participants 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.

4.2. Materials

The test words were embedded into 10–13-syllable-long neutral declarative carrier sentences. They always occurred in the first half of the sentence, but were not sentence initial. Some examples of the test sentences that the participants read out are in (7) (the test words are in bold):

(7) *Word-internal position:*
The white **baseball** cap is my favourite.
The **baseline** was painted white.
Es un **disgusto** grande romper con él.
La **dislexia** se diagnostica hoy en día.

(8) *Between words (sandhi) position:*
The **bonus** deal made everybody happy.
This **virus** loves to mutate quickly.
Algunas **casas** baratas aparecieron.
Hay **casas** lindas con jardines.

As we said, the effect of **cognates** was also investigated. To help understand the figures later, the following two tables list all the test words that were used in the production experiment, together with what sound was the target sound, in which context, the type of the cognate, and the coding of the cognate. For example, *dyslexia* was used to measure the voicing of [s] (thus [s] is the target sound) before the sonorant [l] (thus the context is presonorant). It is a triple cognate because it exists in all three languages, and we code it as Spanish −, and Hungarian +, or “SPA–HU+” (it is an inhibitory cognate in Spanish, but a facilitative cognate in Hungarian).

Table 2: Cognate coding of the **English** test words used in the production experiment

				Cognate coding	
Test word	**Target sound**	**Context**	**Cognate status**	**Spanish**	**Hungarian**
asthma, dyslexia, snob	[s]	presonorant	triple	–	+
Iceland, Yasmin	[s]	presonorant	triple	–	–
disloyal	[s]	presonorant	double	–	
asleep, baseline, Christmas, mismatch	[s]	presonorant	noncognate		
Bosnia, Islam	[z]	presonorant	triple	+	–
phantasmal	[z]	presonorant	double	+	
cosmos, plasma	[z]	presonorant	triple	+	+
amusement, dazzling, rosemary, wisely	[z]	presonorant	noncognate		
baseball	[s]	before stop	triple	–	–
disdain, disgust	[s]	before stop	double	–	
crossbar, disbelief, misbehave, misgiving	[s]	before stop			
asbestos, lesbian	[z]	before stop	triple	+	+
Asbora, Osdol	[z]	before stop	double	+	
gooseberry, husband, Thursday, wisdom	[z]	before stop			

Table 3: Cognate coding of the **Spanish** test words used in the production experiment

				Cognate coding	
Test word	**Target sound**	**Context**	**Cognate status**	**English**	**Hungarian**
cosmos, plasma	[z]	presonorant	triple	+	+
asma, dyslexia, esnobismo	[z]	presonorant	triple	–	–
Bosnia, islam	[z]	presonorant	triple	+	–
Islandia, Yasmin	[z]	presonorant	triple	–	+
fantasma, fantasmas	[z]	presonorant	double	+	
desleal, desmontar	[z]	presonorant	double	–	
asno, isla, mismo, trasladar	[z]	presonorant	noncognate		
asbestos, lesbiana	[z]	before stop	triple	+	+
béisbol	[z]	before stop	triple	–	+
desdén, disgusto	[z]	before stop	double	–	
Rasgora, Rosberg	[z]	before stop	double	+	
esde, esbelto, esbozo, rasgo	[z]	before stop	noncognate		
cosmos, plasma	[z]	presonorant	triple	+	+
asma, dyslexia, esnobismo	[z]	presonorant	triple	–	–

4.3. Method of the production experiments

The English and Spanish sessions were kept separate. Half of the participants started with the English session and the other half with the Spanish session. 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. Sessions took place in the soundproof booth of the Hungarian Research Centre for Linguistics. The sentences were read from a monitor screen in randomised order. Each test sentence was read four times. This meant 44 sentences for the English data, and 39 sentences for the Spanish data by 4 repetitions by 14 speakers (thus altogether 83 sentences × 4 times × 14 participants = 4648 recordings were analysed).

The acoustic analysis was carried out in Praat (Boersma and Weenink, 2022), where the length of the alveolar fricative ([s] or [z]) was measured, and within that length, how long voicing (the vocal fold vibration) was. The length of voicing was compared to the whole length of the fricative. Thus, for example, if the fricative was 33 milliseconds long, and within this interval, voicing was 29 milliseconds long, then this particular articulation of the fricative contained 88% voicing (thus it was almost fully voiced). You will see these percentages in the figures below.

How much voicing in the fricative /s/ makes it to be a "voiced [z]"? Research into the perception of voicing has shown that if the alveolar fricative contains at least 30% voicing (i.e., the vocal folds vibrate during 30% of the articulation of /s/), listeners are more likely to categorise the sound as the "voiced [z]" rather than the "voiceless [s]". This means, that **as far as perception is concerned, a fricative does not need to be 100% voiced to be perceived as "voiced". Around 30% is enough.** Less than 30% voicing means that the fricative is more likely to be perceived as voiceless [s].

4.4. Perception experiment

In order to explore to what extent the production results are mirrored in perception, a perception experiment was also designed. 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 presonorant voicing was employed to mirror the L1 laryngeal patterns of listeners. The same participants as in the production experiments listened to each text three times, so nine texts all together, in random order. After the text finished, instructions appeared on a computer 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).

¤ ¤ ¤

In the following section, we will present the results of the experiments. Take your time to digest the figures. Keep going back to Tables 2 and 3 to make sure you understand which words were in which groups.

5. Results

5.1. Production experiments

5.1.1. Cognate status effect before sonorants

English /s/ before sonorants. In this environment, /s/ is pronounced voiceless in English. Did the participants achieve this articulation target? Figure 1 shows the voicing percentages of presonorant /s/ in four cognate groups. The *Y* axis shows the voicing percentage values from 0% (/s/ had absolutely no voicing) to 100% (/s/ was fully voiced). The *X* axis shows the cognate groups. A reminder: in the group names, 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/. The blue dots indicate the individual percentage values, while the mean of each group is shown by the horizontal blue line.

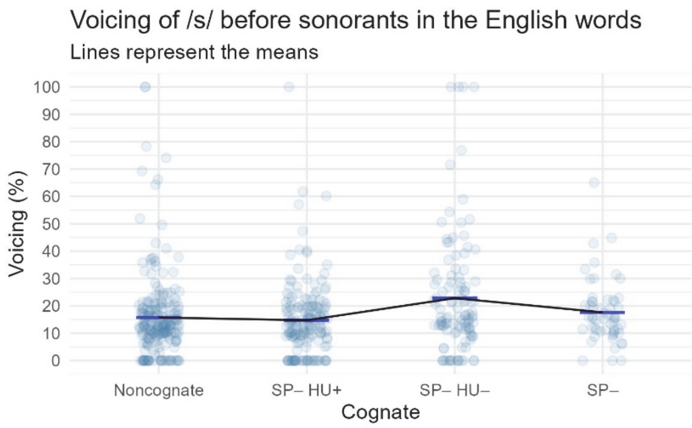
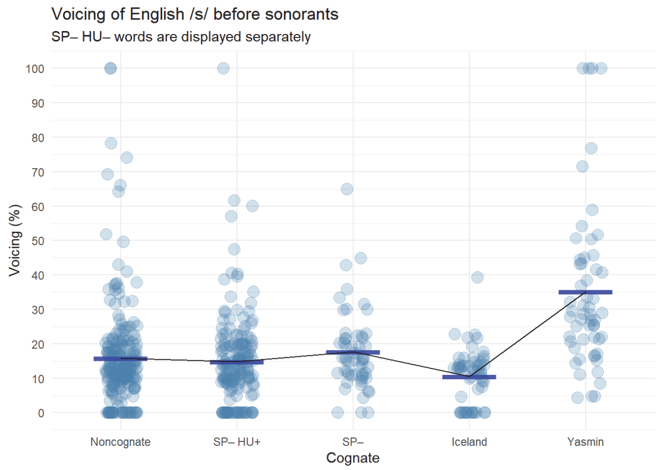


Figure 1

As we can see in Figure 1, most values were rather low, much lower than the 30% voicing threshold (as we said, only values above 30% are likely to be perceived as voiced). We can conclude that **presonorant /s/ had little voicing across the groups**. The SP–HU– group showed the highest average voicing at 22.8%, but it is still a relatively low average. This group is where *both* Hungarian and Spanish have a voiced [z] for the given word, i.e., they can potentially have a "multiplied" negative effect on the English target, which should be voiceless. We do see some cognate effect, but it is only a very small effect: voicing was barely raised in this group.

We can also observe that Spanish cognates cannot raise voicing, neither alone nor in combination with Hungarian. Due to presonorant voicing, all these words are pronounced with voiced [z] in Spanish (thus Spanish is always inhibitory in this case), but this did not influence the voicing of English /s/, it remained relatively voiceless in all three cognate groups.

Overall then, **we cannot observe a strong cognate effect**: none of the three cognate groups were significantly different from the noncognate group. The devil is often in the detail, however, and so it is useful to "open up" the SP–HU– group, which contained the words *Yasmin* and *Iceland*, and check which word was responsible for the slight rise in the voicing. This is shown below:



We can clearly see that the word that was responsible for the slight increase of voicing in the SP–HU– group was *Yasmin*, with a mean voicing of around 35%. The other word, *Iceland* did not raise the voicing of English /s/ much at all despite being pronounced with a /z/ in both Hungarian and Spanish. Thus, *Yasmin* had an inhibitory/negative cognate effect on the voicing of /s/.

Presonorant /z/ in English words. The mean voicing of /z/ in the cognate groups is displayed in Figure 2. We can see that there is considerable voicing in /z/ in all groups (the means are above 30%), except in SP+HU–, i.e., when the word has a cognate in Hungarian pronounced as voiceless [s].

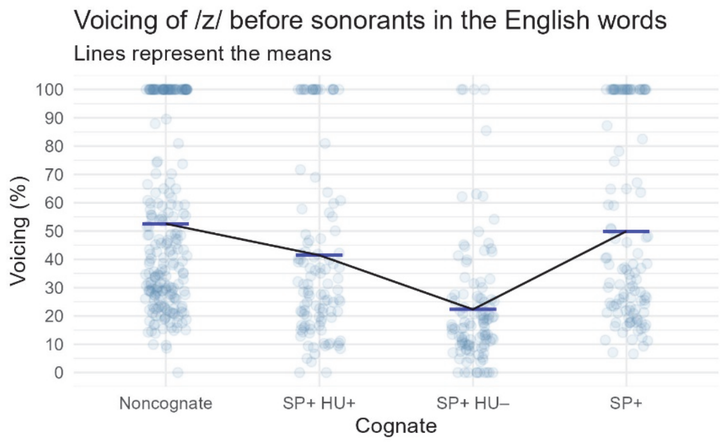


Figure 2

This time the word *Bosnia* contributed to the **cognate effect** in the SP+HU– group the most: the fricative in this word was produced by the participants with a mean of only 13.9%, i.e., pretty voiceless. The difference between the noncognate words and the SP+HU– cognates was significantly different, the Hungarian voiceless pronunciation thus significantly pulled down the voicing of the fricative. Notice that in this group the cognates are contradictory: Spanish has a positive effect (/s/ is voiced in Spanish), while Hungarian a negative effect (the fricative is voiceless), and it turns out that the Hungarian effect “wins”, which supports Hypothesis 3 above.

Spanish words with presonorant /s/. Remember: the target pronunciation is a voiced [z] in Spanish due to PSV. Did the Hungarian participants apply PSV here? The answer is no: the fricative remained relatively voiceless across all groups in the Spanish words (Fig. 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% (this is still little voicing though).

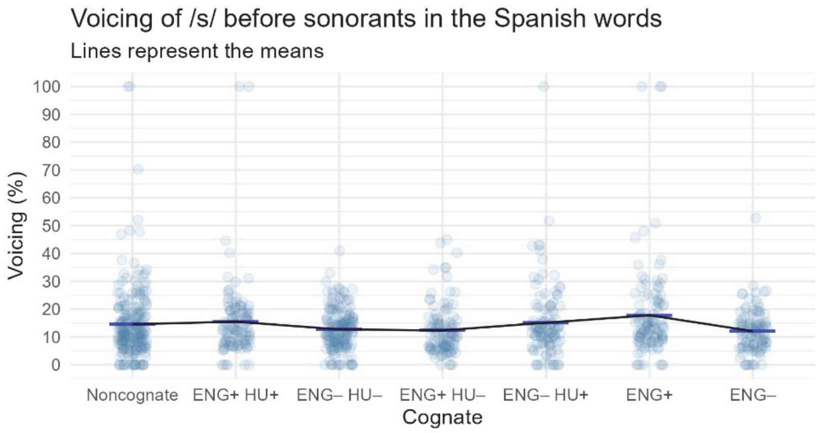


Figure 3

There was no significant difference between the noncognate words and any of the cognate groups, i.e., **we cannot observe any cognate status effect**. There were no significant differences between any of the groups either. This is an interesting result. These advanced learners did not seem to have learnt Spanish presonorant voicing, and they did not even voice /s/ in those words which have a voiced /z/ both in Hungarian and English, such as *cosmos, plasma.* Notice that they voiced the fricative in these words in the English experiment (Figure 2), yet they did not voice it in the Spanish experiment for the same words.

5.1.2. Cognate status effect before voiced stops

Now we turn to the environment where there can potentially be regressive voicing assimilation.

English words containing /s/ plus voiced stops. In this environment, /s/ is pronounced voiceless in English. Again, did the participants achieve this articulation target? As we can see (Figure 4), the "doubly" inhibitory cognate group (SP–HU–) showed the most average voicing (around 60%), but the fricative contained a fair amount of voicing in the other groups, too (close to 50% on average). Note that this group only included one word, *baseball*, thus, this word was pronounced with a voiced [z] on average in all words (cognate or noncognate).

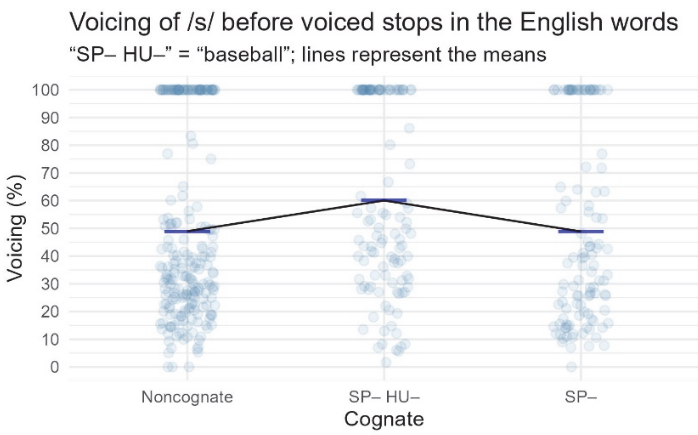


Figure 4.

These results strongly indicate that the participants **did not "unlearn" Hungarian (and Spanish) regressive voicing assimilation** in their English speech, despite being advanced learners.

English words containing /z/ plus voiced stops. In this environment, the fricative is pronounced voiced in English, just as in Hungarian and Spanish. As expected, the participants produced the fricative with a fair amount of voicing on average, in many cases the fricative was 100% voiced:

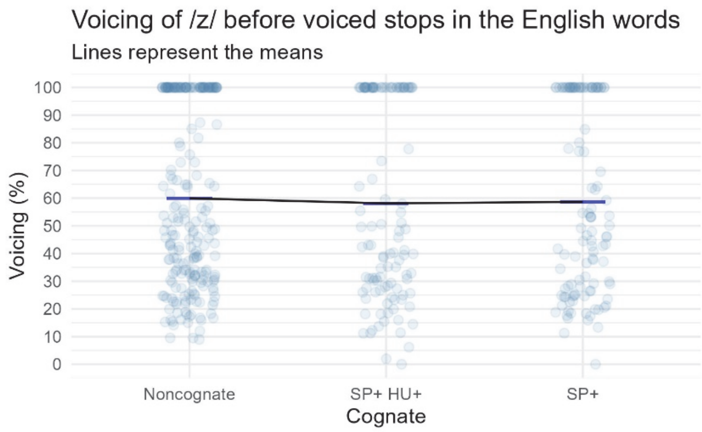


Figure 5.

Spanish words containing /s/ plus voiced stops. In this environment, /s/ is pronounced as voiced [z] in Spanish, just like in Hungarian, but unlike in English, where the fricative is voiceless. And so in here, English may show a negative cognate effect.

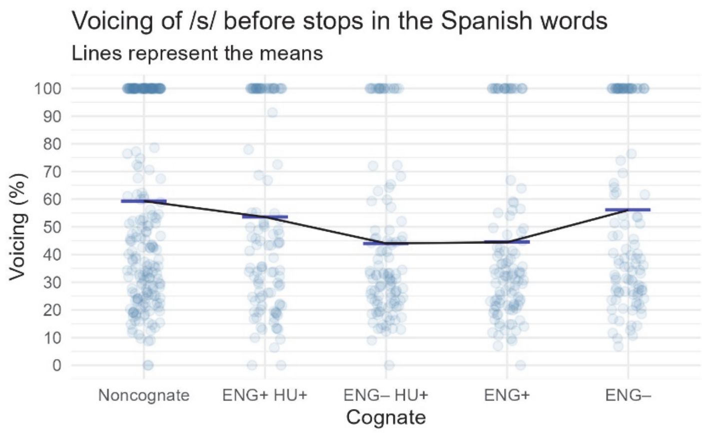


Figure 6.

Just like for English /s/ and /z/, /s/ in the Spanish words was articulated with a considerable amount of voicing (between 44% to 60% on average) (see Figure 6). Voicing in none of the cognate groups was significantly different from that in the noncognate group, and there were no significant differences between the means of any of the groups either. Notice especially how the groups in which English has a potentially negative effect did not bring down voicing significantly. We can conclude that **the Hungarian L1 voicing assimilation pattern dominates** the pronunciation in all word groups.

5.1.3. Speaker variation

The figures above already indicate that the averages hide variation, there are values that are rather far away from the mean. This is also true if we look at the variation the participants showed. There are two types of variation regarding speakers: 1. **inter**speaker variation: this refers to the mean differences **between** speakers; 2. **intra**speaker variation: this refers to the different pronunciations by the **same** speaker. If only one instance is recorded from a speaker, that one value may not be a good representative of the speaker's behavior. This is why the production experiment contained at least four repetitions for each test word from each participant. An example of the type of inter- and intraspeaker variation can be seen in Figure 7, which shows the voicing percentages of /s/ in the English word *baseball* by each speaker (the speakers are represented by numbers).

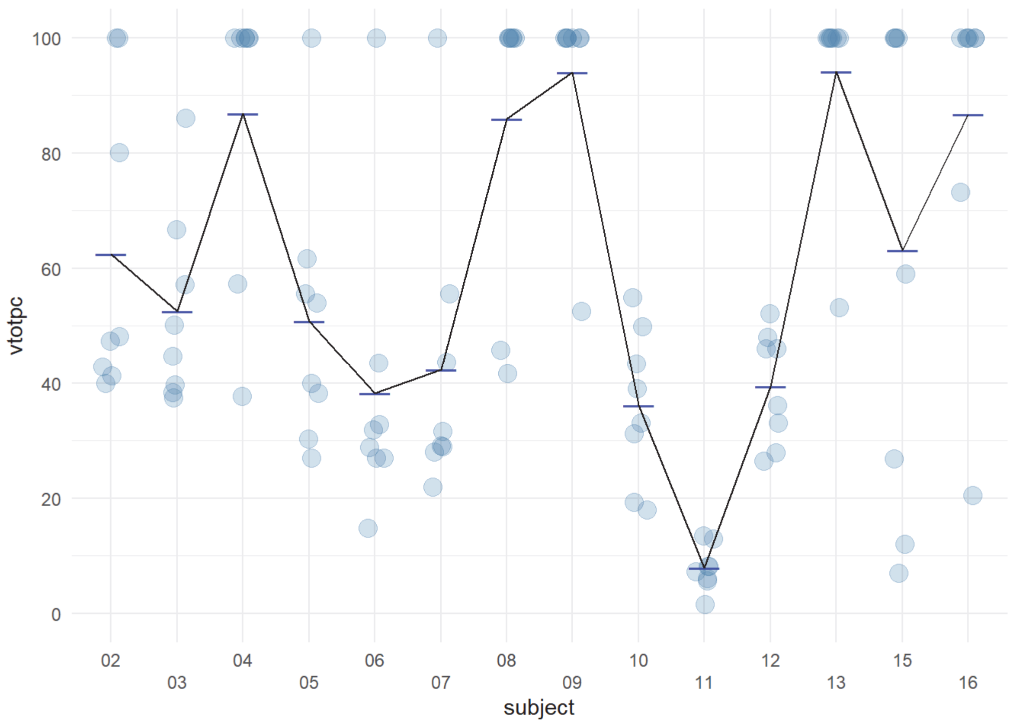


Figure 7

Figure 7 reveals that despite the supposed homogeneity of the group of participants, their averages were different. We can demarcate three types of interspeaker variation: 1. speaker 11 clearly learned the English voiceless pronunciation (all the voicing values are low), 2. while speaker 9 and 11 rather consistently fully voiced the fricative; 3. others had less voicing but still above the 30% threshold. Some in this group even showed relatively high intraspeaker variation. For example speaker 15 sometimes fully voiced the /s/, sometimes pronounced it rather voiceless.

Figure 8 shows speaker variation for the Spanish word *fantasma*, in which the target pronunciation of /s/ should be voiced [z]. As we saw (Figure 3), generally speakers failed to voice presonorant /s/ in the Spanish words but even here two speakers produced a more Spanish-like articulation (speaker 9 and 12) with higher means:

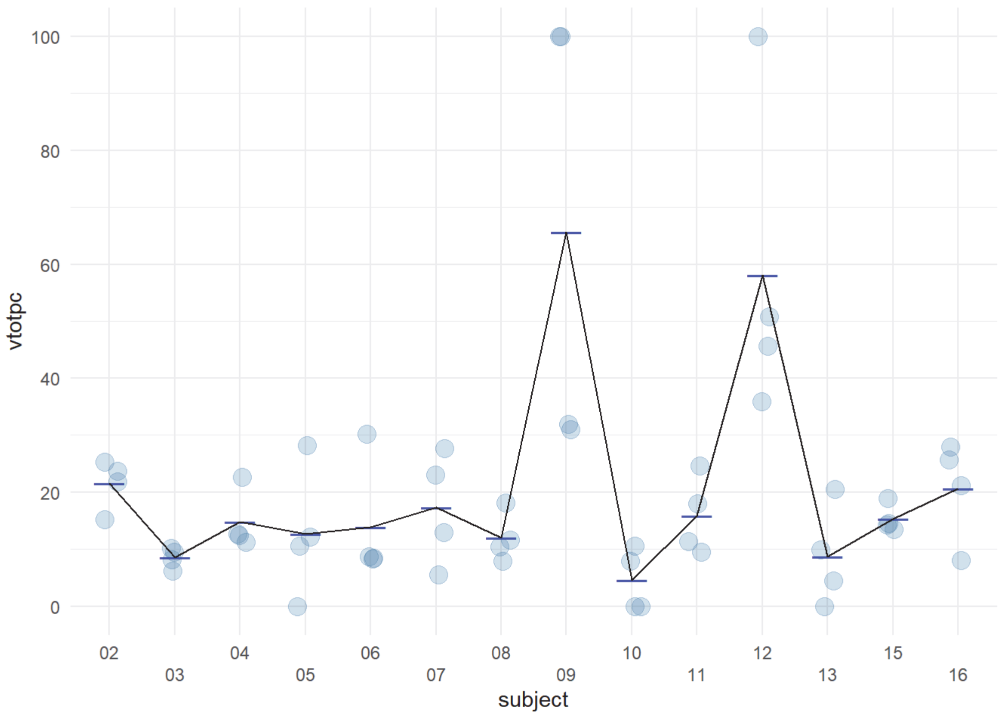


Figure 8

5.1.4. 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 English voicing percentages are shown by blue dots, the Spanish ones with orange dots.

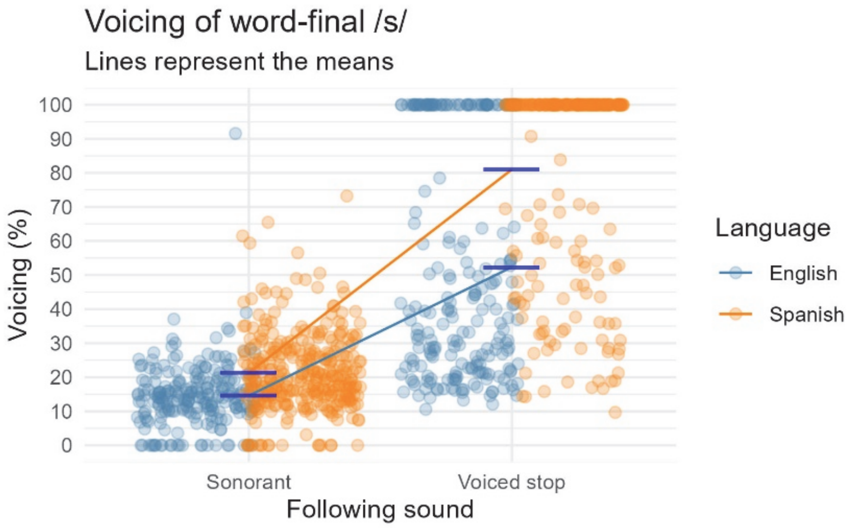


Figure 7.

We can see that word-final /s/ had little voicing before sonorants on average (below the 30% threshold), 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. Remember: in Spanish, word-final /s/ should be voiced [z] before sonorants; while in English, it should be voiceless [s] before voiced stops. This is not what we can observe in Figure 7.

5.2. Perception experiment

We begin with the English results. The left part of Figure 8 displays the ratings of the participants of the two recordings they listen to. "Tom" was the native-like recording, while "TomVoi" was the one where Hungarian-like RVA was applied, i.e., the text was native-like in all aspects, except that /s/ was voiced before voiced stops. A reminder of the ratings: 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, it was rated as 1 (the worst) six times, as two 18 times, while only the native recording was rated with 1 only once. These results indicate that the participants of the experiment **perceived the "misapplication" of the Hungarian voicing assimilation rule,** and rated it worse than the native recording.

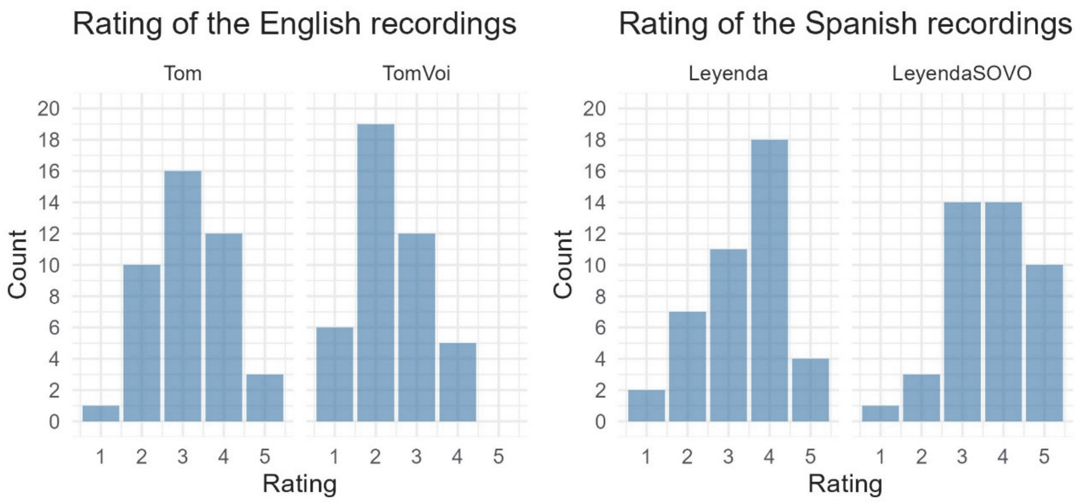


Fig 8.

The ratings of the Spanish recordings can be seen on the right of Figure 8. "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 results indicate that listeners did not reliably differentiate between the native recording (with PSV) and the non-native recording (without PSV): the non-native recording was not rated significantly differently from the native recording, hence **they did not reliably perceive the presence vs. absence of presonorant voicing.**

6. Discussion of the results

This study examined to what extent multilingual speakers produce and perceive voicing 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. Some of the implications of the results presented in the previous section are described as follows.

6.1. 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 voicing and regressive voicing assimilation. In the presonorant voicing context, English /s/ – in a non-target-like manner – displays slightly increased voicing in cognates where L1 is inhibitory (= Hungarian has a /z/ before the sonorant), this supports Hypothesis 1. 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 voiced [z]. The other inhibitory cognate, *Iceland* was probably not perceived by the participants as similar enough to produce an impact since vowel quality is too different (a diphthong in English while /i/ in both Hungarian and Spanish).

On the other hand, in the realisation of presonorant English /z/, a cognate status effect was observed. Cognates with an inhibitory L1 influence were realised with significantly less voicing than the other lexical groups. Thus, **English production data overall 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**.

On the contrary, **the Spanish production data do not show any cognate effects**. There is a steady absence of presonorant voicing, thus **(partly) refuting Hypothesis 1**. The results also reveal that facilitative cognates do not differ from non-cognate realisations, thus **Hypothesis 2** (i.e., that facilitative cognates are acoustically more target-like than non-cognates) **for presonorant voicing must be rejected.**

The question arises why the participants behaved differently in this respect in their two non-native languages. The answer may lie in phonemic encoding during the acquisition of these lexical items. While both English and Spanish voiced realisations are acoustically more similar and should be identified with or mapped to Hungarian [z], and English and Spanish voiceless realisations 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 the participants were 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 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 realisation 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. So, the results indicate **that any potential cognate effects tend to be overridden by RVA.**

6.2. Voicing assimilation across the board

The experiments were specifically designed to explore whether voicing assimilation as a dynamic phonological process has been learned or 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 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 across-the-word-boundary 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 participants seem to have acquired presonorant voicing rather well. 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 leaners. There are several factors that could have contributed to this. An appealing explanation is the considerable **similarity of the Hungarian and Spanish voicing systems**, namely, that both are voicing languages, both display RVA between adjacent obstruents, thus they are treated by the participants as having identical voicing systems. In addition, the fact that sonorant voicing in Spanish is variable by native speakers as well might not serve as sufficient and salient input for learners to be "discovered".

Unlike PSV, the non-target-like application of RVA in English was perceived by the participants of this study; however, their productions did not mirror it: /s/ before a voiced stop was predominantly voiced (around 50%). This means that participants failed to block RVA in their English speech, which **supports Hypothesis 5**.

6.3. The link between perception and production

The results of the perception experiments also support the hypothesis that **PSV in the Spanish speech of the participants remained unnoticed** (Figure 8). 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 favour of any of the scenarios described in Section 1 (Introduction), 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 can consider it as an indication that the acquisition of dynamic phonological processes is different from the acquisition of contrastive segments (phonemes).

The presence of RVA in the English speech 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 (remember: participants applied RVA in English where they "shouldn't" have). This could be compatible with the predictions made by SLM (Speech Learning Model), namely, that only accurate perception can be transferred to production. The participants could be at the stage of accurate perception which has not yet been transferred to production.

Overall, one important finding of this study is thus that in dynamic phonological processes such as voicing assimilation, **L1 plays the primary role**, and L2 does not appear to exert a strong enough influence.

References

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', Frontiers in psychology, 7: 617.

Archibald, J. (2022) 'Phonological parsing via an integrated I-language: The emergence of property-by-property transfer effects in L3 phonology'. https://doi.org/10.1075/lab.21017.arc.

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 T. Angelovska and A. Hahn (eds) L3 Syntactic Transfer: Models, new developments and implications. Amsterdam: John Benjamins Publishing Company. 85–102.

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. John Wiley. 137–167.

Boersma, P. and Weenink, D. (2022) 'Praat: doing phonetics by computer'. [Computer program]

Darcy, I., Tremblay, A. and Simonet, M. (2017) 'Editorial: Phonology in the Bilingual and Bidialectal Lexicon', Frontiers in Psychology, 8. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.

Flege, J.E. (1995) 'Second-language speech learning: Theory, findings, and problems', in W. Strange (ed.) Speech perception and linguistic experience: Issue in cross-language research. Timonium, MD: York Press. 157–232.

Hermas, A. (2015) 'The categorization of the relative complementizer phrase in third-language English: A feature re-assembly account', International Journal of Bilingualism, 19: 587–607. https://doi.org/10.1177/1367006914527019.

Hualde, J.I. (2005) The Sounds of Spanish. New York: Cambridge University Press.

Llama, R. and Cardoso, W. (2018) 'Revisiting (Non-)Native Influence in VOT Production: Insights from Advanced L3 Spanish', Languages, 3: 30. https://doi.org/10.3390/languages3030030.

Mora, J.C. and Nadeu, M. (2009) 'Experience effects on the categorization of a native vowel contrast in highly proficient Catalan-Spanish bilinguals.', The Journal of the Acoustical Society of America, 125: 2775. https://doi.org/10.1121/1.4784755.

Quilis, A. (1993) Tratado de fonología y fonética españolas. Madrid: Gredos.

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: Language and Cognition, 18(2): 179–190. https://doi.org/10.1017/S136672891300059X.

Siptár, P. and Törkenczy, M. (2000) The Phonology of Hungarian. Oxford: Oxford University Press.

Slabakova, R. (2017) 'The scalpel model of third language acquisition', International Journal of Bilingualism, 21(6): 651–665. https://doi.org/10.1177/1367006916655413.

Vago, R.M. (1980) The Sound Pattern of Hungarian. Washington D.C.: Georgetown 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', International Journal of Bilingualism, 21(6): 666–682. https://doi.org/10.1177/1367006916648859.

Wunder, E.-M. (2011) 'Crosslinguistic Influence in Multilingual Language Acquisition: Phonology in Third or Additional Language Acquisition', in G.D. Angelis and J.-M. Dewaele (eds) New Trends in Crosslinguistic Influence and Multilingualism Research. (Multilingual Matters). 105–128.