LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

The More the Merrier? On the Influence of Indexical Variability on Second Language Vocabulary Learning

Marc Gimeno-Martínez,[a] Rebeca Sánchez,[a] and Cristina Baus[b]

[a]Universitat Pompeu Fabra [b]University of Barcelona

Abstract: We investigated indexical variation as a variable that promotes second language (L2) vocabulary learning across language modalities. In three experiments, we presented Catalan Sign Language signs (Experiments 1a and 1b), pseudowords (Experiment 2), and English words (Experiment 3) to participants in three conditions that varied in the number of people who introduced these stimuli (one, three, or six people). We evaluated learning outcomes in two recall tasks: a picture-to-L2 naming task and a L2-to-L1 translation task. For the sign modality, indexical variation benefitted the immediate recall of signs in the translation task (Experiment 1a) and delayed recall

CRediT author statement – **Marc Gimeno-Martínez**: conceptualization; methodology; formal analysis; writing – original draft preparation; writing – review & editing. **Rebeca Sánchez**: formal analysis; investigation; writing – review & editing (supporting). **Cristina Baus**: conceptualization; funding acquisition; methodology; supervision; writing – original draft preparation; writing – review & editing.

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

This work was supported by the Spanish Ministry of Science and Innovation (RTI2018-096238-A-I00) and the ACELERA research network (RED2018-102615-T). Cristina Baus was supported by the Ramon y Cajal research program (RYC2018-026174-I). We thank Veronica Bernardini for her assistance in Experiment 3.

Correspondence concerning this article should be addressed to Cristina Baus, University of Barcelona, Department of Cognition, Development, and Educational Psychology, 08035, Barcelona, Spain. Email: baus.cristina@gmail.com

The handling editor for this article was Judit Kormos.

DOI: 10.1111/lang.12554

after two weeks in the picture naming task (Experiment 1b). For the oral modality, we observed no effect when participants learned pseudowords (Experiment 2), but variability benefited recall in the translation task when participants learned English words (Experiment 3). We discuss these contrastive results, considering the influence of indexical variation in adult L2 sign and oral vocabulary learning.

Keywords indexical variability; second language; sign learning; second modality learners; recall; translation; replication

Introduction

Learning a second language (L2) is often a difficult task, especially for adults. Despite exceptions, for most learners acquiring the multiple subsystems of a L2 (e.g., vocabulary, phonology, grammar) represents an extraordinary cognitive effort. A complex set of variables influences this process with respect to the learner (e.g., age, aptitude; Ellis, 1986), the context of learning (e.g., classroom, immersion; Cummins, 1999), or the learning process itself (e.g., incidental vs. explicit vocabulary learning; Ellis, 1994). As such, a great deal of scientific and pedagogical work has focused on identifying variables that contribute to the success of adult L2 learning. In our study, we were concerned with L2 vocabulary learning so as to gain a better understanding of the influence of indexical variation. We focused on a sign language to determine whether the reported positive effects of indexical variation in the number of speakers on L2 vocabulary learning might occur in a language that is not acoustically based.

Background Literature

Influence of Indexical Variation on Speech Processing and Memory

Speech contains two primary sources of information: linguistic and indexical. Linguistic information conveys the content of an utterance, including its phonological, lexical, syntactic, and semantic aspects (Pisoni & Levi, 2007). Indexical information is obtained from the voice and conveys information about talker identity, including speaker-specific characteristics such as gender, age, emotional states, or personality traits (e.g., Hagiwara, 1997; Munson et al., 2006; Pisoni & Remez, 2008). For the linguistic content of speech, literature on speech perception has often neglected the role of indexical information. Researchers have accepted the idea that listeners' perceptual system normalizes the speech signal by disregarding any speaker-specific variation (i.e., through what is known as categorical perception of speech). Nevertheless, accumulated evidence has revealed that listeners use speaker-specific information in the process of speech perception. That is, indexical properties of the voice are encoded and retained in memory along with linguistic aspects of speech (e.g., Bradlow et al., 1999; Goldinger, 1996; Goldinger et al., 1991; Johnson, 2006; Martin et al., 1989; Nygaard et al., 1995, 2000).

Several studies have shown that speech processing and memory are affected by variations in indexical information. Of relevance are studies revealing that variation in speaker characteristics (i.e., the number of speakers) has an impact on speech processing and memory. For instance, an inverse relationship between word recognition accuracy and variation in the number of speakers has been reported (e.g., Goldinger et al., 1991; Mullennix et al., 1989; Ryalls & Pisoni, 1997). Recall of words has been found to be more accurate and rapid for lists of words uttered by a single speaker than for lists of words uttered by multiple speakers. A distinctive positive effect of the number of speakers has been reported when sufficient processing time was given to listeners, allowing them to fully encode indexical information from the voice. As an example, Nygaard et al. (1995) found that listeners' memory recall of a list of words improved when the words were produced by multiple speakers in a slow presentation rate but diminished in a fast presentation rate (see also Goldinger et al., 1991; Palmeri et al., 1993). In light of these results, indexical variation in the number of speakers is considered to be a relevant feature of first language (L1) processing and memory, but certain methodological aspects appear to modulate the impact of indexical variation.

Importantly, not all sources of variability are encoded in memory along with the linguistic content (Nygaard et al., 1995; Sommers & Barcroft, 2006). One source of variability is fundamental frequency (F0), which is defined as the lowest rate of repetition of the cycles of air pressure and determines the pitch of a voice. Variations in F0 are lexically contrastive in tonal languages (e.g., Chinese), that is, tonal languages have similar segmental sequences that are only differentiated by changes in F0, and this results in different lexical units with different meanings. The same variation is not relevant in languages such as English or Spanish. Sommers and Barcroft (2007) showed no impact of indexical variation on L1 English word processing when the fundamental frequency of voice (F0) was manipulated as a source of variation. Altogether, studies from the oral modality have revealed that the impact of indexical variation on L1 processing is determined by its relevance in the language.

Influence of Speaker Variability on Adult L2 Learning

Studies on L2 learning have shown a positive influence of speaker variability on memory recall of L2 words. Novel L2 words have been shown to be more accurately learned when they are introduced by multiple speakers than by a single speaker (Barcroft & Sommers, 2005; Sommers & Barcroft, 2011; see Rost & McMurray, 2009, for similar results with novel words and babies). Barcroft and Sommers (2005), for example, examined the effect of speaker variability on the ability of L1 English speakers to learn L2 Spanish words by comparing learning rates in three conditions: no variability (six repetitions of each word in the voice of one speaker), moderate variability (two repetitions of each word in the voice of three different speakers), and high variability (one repetition of each word in the voice of six different speakers). Barcroft and Sommers evaluated learning with two recall tasks: a picture-to-L2 naming task and a L2-to-L1 translation task. Accuracy scores in both tasks showed that L2 vocabulary learning improved systematically as a function of variability. Words in the no variability condition resulted in lower accuracy rates than did words learned in the moderate variability condition, and words in the moderate variability condition obtained low accuracy rates compared to words in the high variability condition (but see Barcroft, 2001, for no effect of speaker variability).

In light of these results and subsequent replications (Barcroft & Sommers, 2014; Sommers & Barcroft, 2007, 2011), several theoretical accounts have described the mechanisms behind the positive influence of indexical variability on L2 vocabulary learning. One of those accounts that has received most attention is the exemplar-based model described by Goldinger (1998). This framework suggests that indexically varied conditions produce more associative "hooks" and more robust representations for lexical entries stored in long-term memory. In the context of L2 learning, indexically varied representations of words to be learned would lead to richer encoding (Barcroft & Sommers, 2005), which subsequently would facilitate retrieval.

As reported for L1, L2 learners only benefit from variability if it targets an acoustically relevant feature in the language. Sommers and Barcroft (2007) showed that L1 speakers of English (a nontonal language) did not benefit from variations in F0. Barcroft and Sommers (2014) expanded these results by comparing the learning outcomes of speakers of Zapotec, a tonal language, and the learning outcomes of speakers of English, a nontonal language. The researchers exposed the participants to 24 Russian auditory words while the participants viewed the corresponding pictures. The researchers experimentally manipulated the F0 by providing six instances of each word, presented in three learning conditions: no variability (six repetitions spoken at one F0), moderate variability (two repetitions of three F0s), and high variability (one repetition of six F0s). Only the participants for whom F0 was a relevant language feature in their L1, that is, the Zapotec speakers, benefitted from F0 variability in L2 learning. These results supported the phonetic-relevance hypothesis according to which L2 learners only attend to acoustic variations if these variations are phonetically relevant in the languages in which the L2 learners are proficient.

In sum, indexical variation influences L2 vocabulary learning when this is a relevant property in the language. In our study, we explored indexical variation for the number of signers in L2 adult sign learning to determine how relevant signer variation is in a language that is not acoustically based. The aim of this study was twofold. At the theoretical level, determining the role of signer variability in L2 learning would provide information about indexical aspects of sign processing and how these aspects interact with linguistic content in sign language processing and memory. In this respect, our study would contribute to clarifying whether indexical variation is a general linguistic property that influences vocabulary learning regardless of modality or is restricted to acoustically based languages. Second, at the pedagogical level, these results could inform educational practices that promote L2 sign learning. Exploring L2 sign learning is especially relevant considering the increasing number of people who have chosen to learn a sign language as a L2 in recent years. As an example, in 2016, American Sign Language was the third most frequently taught L2 in the United States (Looney & Lusin, 2018). Given the increasing popularity of learning sign languages as a L2, it is important to know how L2 learning occurs when the L1 and the L2 of the learner are not from the same modality, that is, for second modality L2 learners (Pichler & Koulidobrova, 2016; Schönström, 2021).

Knowledge of which properties are similar between sign and oral languages (i.e., modality-independent) and which properties are determined by the language modality (i.e., modality-dependent) is required for exploring the coupling between linguistic and indexical information in sign languages. At the linguistic level, accumulated evidence has indicated that sign and oral languages are sensitive to the same linguistic phenomena, including lexical frequency (Emmorey et al., 2013; Jescheniak & Levelt, 1994) and categorical perception (Gimeno-Martínez et al., 2020; Kuhl, 2004). This implies that linguistic information is organized and flows across levels of processing (e.g., semantic, lexical, and phonological) similarly in both modalities. Likewise, the same variables described in L2 word learning have been reported to influence L2 sign learning. For example, variables such as learners' L1 vocabulary knowledge have predicted L2 sign learning in hearing adults (Williams et al., 2017). In addition, other variables specific to the signed modality such as visual sonority, handshape markedness (Williams & Newman, 2016), or perceptuomotor abilities of learners, including short-term memory for hand and arm movements (Martinez & Singleton, 2018), have appeared to influence L2 sign learning as well.

In contrast to the description of the linguistic aspects of sign processing and learning, and perhaps because indexical information has been mainly described as referring to the acoustic properties of the voice, its counterpart in the signed modality has remarkably been barely described. Notwithstanding, under the assumption that signers have mental representations of sign forms (Corina et al., 2011) and that the lexicon is similarly organized in signed and oral languages (e.g., Caselli & Cohen-Goldberg, 2014), it is conceivable that signers encode signer-specific perceptual variations (indexical aspects) during sign processing. In a priming study, Corina et al. (2011) tested perceptual viewpoint as a source of variability in sign language processing. Perceptual viewpoint referred to the angle view of the signer, with front, left, or right views. Identical prime and target signs (same sign) were presented either from the same viewpoint (e.g., front–front) or from a different one (e.g., front–side). Repetition priming was larger when signs were presented from the same viewpoint than when they were presented from a different viewpoint (see also Emmorey et al., 2009; Pyers et al., 2015). This suggested that indexical variation in perceptual viewpoint is integrated along with sign representations during sign processing.

The Present Study

Our study focused on signer variability to explore if it is encoded in memory along with linguistic information from the sign and hence positively influences L2 learning. Specifically, our main research aim was to explore whether L2 sign vocabulary learning is enhanced when signs are presented by multiple signers compared to by a single signer. To achieve this, we adapted the Barcroft and Sommers (2005) study to the signed modality. We compared learning outcomes of signs learned in three variability conditions: no variability, moderate variability, and high variability. As in the oral modality (Barcroft & Sommers, 2014), if variation in the number of signers is a relevant indexical property in sign processing, variability effects should be expected in L2 sign vocabulary learning, both in immediate recall (Experiment 1a) and in delayed recall (Experiment 1b).

In addition, to obtain an estimate of the effect of speaker variability on the oral modality in our study population of bilingual Catalan–Spanish speakers, we conducted two further experiments to investigate speaker variation in L2 word learning. This way, we could evaluate whether the influence of indexical variation on L2 vocabulary acquisition is a general linguistic property that is independent of the modality of the language to be acquired. In Experiment 2, a new group of participants learned words from an invented language (pseudowords). In Experiment 3, another group of participants learned L2 English words.

Experiment 1a: L2 Sign Learning

Method

Participants

We recruited 54 Catalan–Spanish speakers (40 females, $M_{\text{age}} = 22.25$ years, range = 18–28) from the Universitat Pompeu Fabra's Center for Brain and Cognition database. All were university students without any hearing difficulty or history of deafness, and they reported no previous knowledge of Catalan Sign Language (LSC) or any other sign language and were not enrolled in sign language courses. The participants completed an informed consent form for image recording and experiment participation before the experiment and were paid for their participation. We excluded three participants because they could not complete the task appropriately and because of technical problems.

Materials

We selected 48 noniconic LSC signs and their related pictures for the experiment. We used sign iconicity ratings ($M = 1.65$, $SD = 0.48$; on a scale where 1 = *low iconic* and 5 = *high iconic*) from 12 hearing nonsigners from Baus and Costa's (2015) study. Signs included different semantic categories and were recorded by seven hearing proficient signers (three males, four females). We asked the signers to record the LSC signs with a neutral face. We retrieved black and white pictures corresponding to the signs from Snodgrass and Vanderwart's (1980) study and from the Multipic database (Duñabeitia et al., 2018).

We divided the 48 stimuli into three sets (see Table 1) corresponding to the three learning variability conditions (16 signs in each set): (a) no variability, with six repetitions of the sign performed by one signer; (b) moderate variability, with two repetitions of the sign performed by three signers; and (c) high variability, with one repetition of the sign performed by six signers. Sign iconicity ratings did not differ across the three stimuli sets ($p = .23$). The same video sign was displayed for all sign repetitions from a given signer to avoid intrasigner variability. We counterbalanced the sets of stimuli and the order in which the conditions appeared throughout the experiment, which resulted in nine experimental lists. In addition, to minimize differences in signer intelligibility across variability conditions, we rotated the six signers

Table 1 Sets of sign stimuli from Catalan Sign Language used in Experiments 1a and 1b

Sign set	Stimuli
Set 1	windmill, firefighter, mailbox, moneybox, sheep, garlic, onion, sock, camel, cherry, ant, kiwi, lettuce, cucumber, pear, lemon
Set 2	spider, tree, folder, vacuum, eggplant, boot, strawberry, melon, pea, hamburger, cookie, deer, lobster, shark, olive, grape
Set 3	tiger, frog, potato, pill, doll, peach, bee, asparagus, light bulb, pineapple, lizard, nun, fox, pepper, brush, watering can

Note. English translations of the signs from Catalan Sign Language are reported.

from the high variability condition across participants in the no variability and moderate variability conditions (Barcroft & Sommers, 2005). Thus, by incorporating a signers' rotation procedure, we prevented the same signer from appearing in the same condition and producing the same set of signs for all the participants. In this way, each of the nine stimulus lists had six variants based on the identity of the signers, which resulted in 54 training lists.

The experimental session included a learning phase and a test phase. To avoid repetition of the same signers in both phases, we used video recordings of six signers (three males and three females) in the learning phase. In the test phase, we presented the participants with signs performed by a different signer.

Procedure

We tested the participants individually and conducted the experiment online. We sent the participants a video including a recording of the experimental session run under the E-Prime (Version 2.0) software (https://pstnet.com/products/e-prime). We asked the participants to record a video of themselves while they were doing the experiment to ensure that they were attentive to the screen and had no external distractions (e.g., looking at the phone, other interruptions) during the learning phase so that we could evaluate the accuracy of their responses offline.

The experimental design was as follows. First, participants were presented with a video recording that corresponded to the learning phase, and then they were required to perform two tasks in the testing phase: a picture-to-L2 naming task and a L2-to-L1 translation task. In the learning phase, the participants were informed that they would see a series of six repetitions of 48 LSC signs along with the pictures associated with their meaning, which yielded 288 trials. The participants' task was to memorize the signs. Each trial began with a fixation asterisk that was presented at the center of the screen for 500 ms followed by a picture for 750 ms on the left part of the screen. The video of the target sign was then displayed (3,000 ms) on the right part of the screen while the picture was still visible and remained 1,250 ms after the video ended. A final blank of 500 ms completed each learning trial.

In the testing phase, we employed two recall tasks: first a picture-to-L2 naming task and then a L2-to-L1 translation task. In the picture-to-L2 naming task, the participants were required to perform the LSC sign corresponding to the picture displayed for 10,000 ms on the screen after a fixation asterisk displayed for 1,000 ms. A final blank of 1,000 ms completed the trial. In the L2-to-L1 translation task, after a fixation asterisk of 1,000 ms, the participants were presented with LSC signs displayed for 3,000 ms and were asked to verbally provide their Catalan translation (with a maximum response time of 10,000 ms). A final blank of 1,000 ms completed each trial.

Data Analysis

We binary coded the data (correct/incorrect) in both recall tasks, that is, the picture-to-L2 naming task and the L2-to-L1 translation task, after the experiment by analyzing the participants' video recordings (see Sinkeviciute et al., 2019, for a similar analysis approach). In the picture-to-L2 naming task, we coded each sign production as incorrect if the participants did not recall the sign, if they provided a nontarget sign, or if at least one of their signs' sublexical components (i.e., handshape, location, and movement) deviated greatly from the target (see Ortega et al., 2019, for a similar response coding). For the L2-to-L1 translation task, we excluded from the analysis incorrect responses or trials in which the participants did not respond. We considered trials in which the participants provided a similar word to the expected answer (e.g., "peach" instead of "apricot") as correct responses.

We analyzed the two tasks separately with generalized mixed models (binomial family) using the lme4 package (Bates et al., 2011) for R (R Core Team, 2019). Models converged reliably, including fixed effects for the variability condition (no variability, moderate variability, and high variability) and crossed random effects for participants and items (Baayen et al., 2008). The R code for the final statistical model was: accuracy ~ variability condition + (1 | participants) + (1 | items). We took accuracy in the no variability condition as the intercept to which we compared the moderate and high variability conditions. We considered fixed effects estimates fitted by maximum likelihood (Laplace approximation) to be significant if p was less than .05.

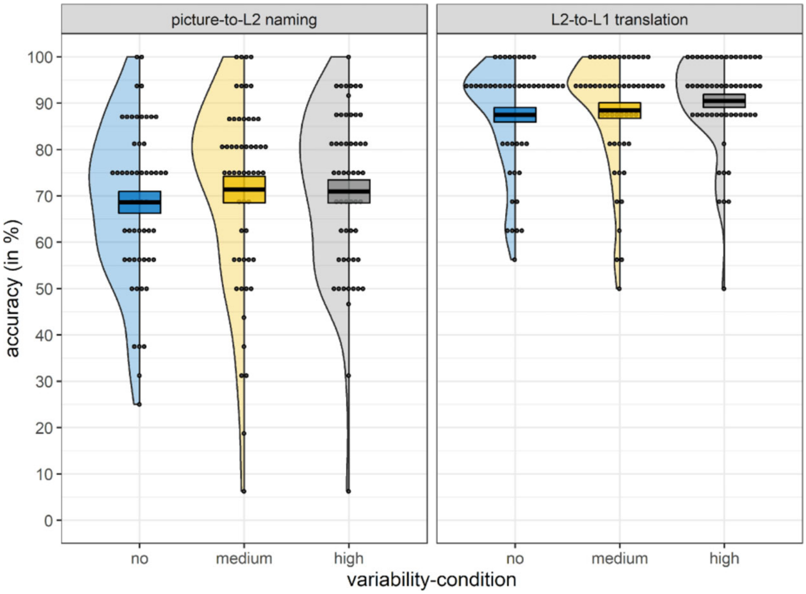

Figure 1 Percentage of correct responses in each variability condition in the two tasks for Experiment 1a. The half violin shape shows the kernel probability density of participants' mean scores. Dots indicate the percentage of correct response for each participant in each variability condition. Box plots indicate mean values and standard error.

Results

Figure 1 and Table 2 show the percentage of correct responses for each variability condition across tasks for Experiment 1a.

Picture-to-L2 Naming Task

The mixed-effects model for the picture-to-L2 naming task (based on 2,575 observations), Akaike information criterion (AIC) = 2,598, $R^2_{marginal} = .001$, $R^2_{conditional} = .42$, revealed no effect of variability: moderate variability, $b = 0.19$, $SE = 0.12$, 95% CI [−0.06, 0.45], $z = 1.56$, $p = 0.12$; high variability, $b = 0.16$, $SE = 0.12$, 95% CI [−0.11, 0.41], $z = 1.31$, $p = .19$. These results indicated that the number of signers had no influence on the participants' sign recall accuracy.

L2-to-L1 Translation Task

The mixed-effects model for the L2-to-L1 translation task (based on 2,592 observations), AIC = 2,598, $R^2_{marginal} = .005$, $R^2_{conditional} = .46$, revealed an effect of variability. The participants more accurately retrieved words learned

Table 2 Descriptive statistics for percentage of correct responses in the four experiments across variability conditions and tasks

	Experiment 1a ($N = 51$)			Experiment 1b ($n = 39^{a}$)			Experiment 2 ($N = 54$)			Experiment 3 ($N = 42$)		
	LSC test			LSC retest			Pseudowords			English words		
Variability condition	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Picture-to-L2 naming task												
No variability	0.69	0.46	[0.81, 0.56]	0.35	0.48	[0.50, 0.20]	0.34	0.48	[0.47, 0.22]	0.65	0.48	[0.79, 0.51]
Moderate variability	0.71	0.45	[0.83, 0.59]	0.42	0.49	[0.58, 0.27]	0.34	0.47	[0.46, 0.21]	0.58	0.49	[0.73, 0.43]
High variability	0.71	0.45	[0.83, 0.59]	0.40	0.49	[0.55, 0.25]	0.38	0.48	[0.51, 0.25]	0.66	0.48	[0.80, 0.51]
Overall accuracy	0.70	0.46	[0.82, 0.58]	0.39	0.49	[0.54, 0.24]	0.35	0.48	[0.48, 0.23]	0.63	0.48	[0.78, 0.48]
L2-to-L1 translation task												
No variability	0.88	0.33	[0.96, 0.79]	0.77	0.42	[0.67, 0.41]	0.54	0.50	[0.67, 0.41]	0.68	0.47	[0.82, 0.54]
Moderate variability	0.88	0.32	[0.97, 0.80]	0.79	0.41	[0.66, 0.39]	0.53	0.50	[0.66, 0.39]	0.77	0.42	[0.90, 0.64]
High variability	0.91	0.29	[0.98, 0.83]	0.77	0.42	[0.65, 0.38]	0.51	0.50	[0.65, 0.38]	0.75	0.43	[0.88, 0.62]
Overall accuracy	0.89	0.32	[0.97, 0.80]	0.78	0.42	[0.66, 0.39]	0.53	0.50	[0.66, 0.39]	0.73	0.44	[0.87, 0.60]

Note. LSC = Catalan Sign Language; L2 = second language; L1 = first language. [a]Subsample of Experiment 1a participants.

in the high variability condition than they did those words learned in the no variability condition, $b = 0.43$, $SE = 0.17$, 95% CI [0.09, 0.80], $z = 2.52$, $p = .01$. We found no effect in the moderate variability condition, $b = 0.14$, $SE = 0.16$, 95% CI [−0.21, 0.46], $z = 0.8$, $p = .40$.

Discussion

Our results revealed an influence of signer variability on the participants in the L2-to-L1 translation task when we compared the no variability condition to the high variability condition. We observed no differences between signs encoded in the no variability condition and in the moderate variability condition. That is, the participants benefitted from variability in the number of signers but only when the number of signers was sufficiently high. In contrast, we did not observe any benefit of signer variability in the picture-to-L2 naming task.

The absence of effects in the picture-to-L2 naming task contrasted with the results of previous studies in the oral modality showing that indexical variation positively influenced L2 vocabulary recall (e.g., Barcroft & Sommers, 2005). Relative to those studies, accuracy in the picture-to-L2 naming task in our study was noticeably high (70% in our study vs. 40% in previous studies), especially considering the number of signs to be learned (48 in our study vs. 24 in previous studies). Therefore, we reasoned that indexical variability might only benefit L2 vocabulary recall at lower levels of accuracy. That is, it is possible that variation only helps when the task is difficult enough. To further explore whether the lack of effects in the picture-to-L2 naming task was due to high accuracy levels, we tested a subset of participants ($n = 40$) again approximately two weeks later ($M = 15$ days, range = 12–18).

Experiment 1b: Retest L2 Sign Learning

Method

Approximately two weeks after Experiment 1a (range = 12–18 days) had ended, a subset of 40 participants repeated the tasks from the first experiment. In terms of materials and analysis, the design of Experiment 1b was the same as that of Experiment 1a. Importantly, unlike Experiment 1a, in the retest there was no training phase, so the participants completed only the two recall tasks. We excluded the data from one participant from the picture-to-L2 naming task because the video recording was defective and we could not check his responses properly for accuracy.

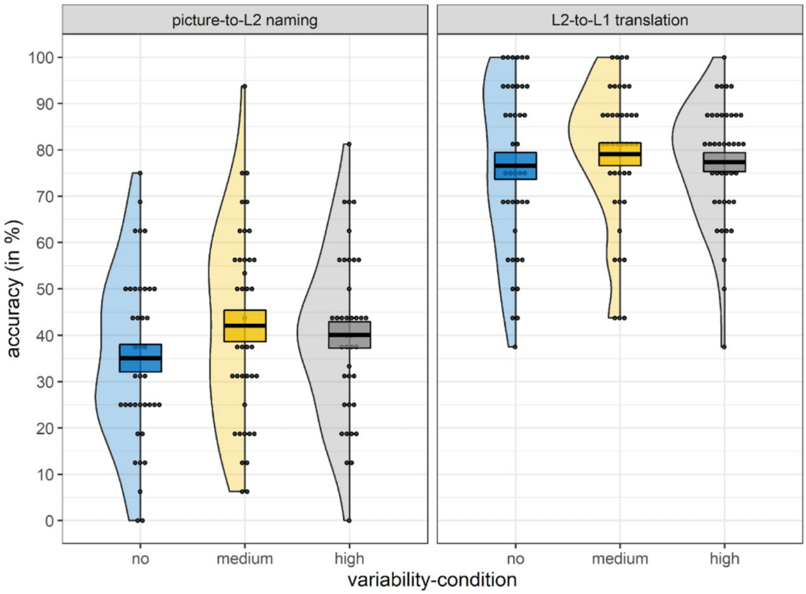

Figure 2 Percentage of correct responses for each variability condition across tasks for Experiment 1b. The half violin shape shows the kernel probability density of participants' mean scores. Dots indicate the percentage of correct responses for each participant in each variability condition. Box plots indicate mean values and standard error.

Results

Picture-to-L2 Naming Task

The mixed-effects model for the picture-to-L2 naming task (2,592 observations), AIC = 2,598, R^2_{marginal} = .005, $R^2_{\text{conditional}}$ = .46, showed an effect of signer variability: moderate variability, $b = 0.35$, $SE = 0.14$, 95% CI [0.09, 0.60], $z = 2.60$, $p = .01$; high variability, $b = 0.27$, $SE = 0.13$, 95% CI [0.01, 0.53], $z = 1.99$, $p = .05$. This indicated that the number of signers influenced sign recall. The participants recalled the signs learned in the no variability condition less accurately than they did those that they had learned in the moderate variability and high variability conditions (see Figure 2 and Table 2).

L2-to-L1 Translation Task

The mixed-effects model for the L2-to-L1 translation task (1,920 observations), AIC = 1,643, R^2_{marginal} = .0007, $R^2_{\text{conditional}}$ = .51, revealed no effect of variability: moderate variability, $b = 0.15$, $SE = 0.16$, 95% CI [−0.18, 0.51],

$z = 0.97$, $p = .33$; high variability, $b = 0.01$, $SE = 0.16$, 95% CI [−0.31, 0.36], $z = 0.07$, $p = .94$. This indicated that the number of signers had no influence on the participants' sign translations (see Figure 2 and Table 2).

Discussion

Our results for both the test (Experiment 1a) and retest (Experiment 1b) indicated that signer variability influenced the participants' L2 sign learning. The participants learned better the signs learned in a context of multiple signers than those that they had learned from a single signer. Once again, our results partially replicated previous results in the oral modality (Barcroft, 2001; Barcroft & Sommers, 2005). In Experiment 1a, we observed the effect of variability only in the L2-to-L1 translation task and only when the number of signers was sufficiently high. In Experiment 1b, we observed the effect of signer variability only in the picture-to-L2 naming task. Both variability conditions (moderate and high) showed better accuracy than did the no variability condition.

Altogether, our results supported the notion that signs contain lexical and indexical information and that both sources of information influence learning. However, our results also showed that signer variability might not be as relevant in L2 sign learning as has been reported in the oral modality. Before we made further conclusions about differences between modalities and the influence of indexical variability on L2 learning, we conducted two experiments testing L2 learning of spoken words with L2 learners from the same population. In Experiment 2, the participants learned words from an invented language, that is, pseudowords, and we manipulated acoustic variability in the number of speakers. In Experiment 3, the participants learned new words in English, their L2, and we manipulated acoustic variability in the number of speakers.

Experiment 2: Learning Words From an Invented Language

Method

Participants

We recruited 54 bilingual Catalan–Spanish speakers (38 females, $M_{age} = 21$ years, range = 18–34) from the Universitat Pompeu Fabra's Center for Brain and Cognition database. None of them had participated in the previous experiment learning LSC signs.

Materials

We used the same set of 48 pictures used in Experiments 1a and 1b for this experiment. In this case, we matched pictures with words from an invented

Table 3 Groups of pseudoword stimuli used for the three different conditions of variability in Experiment 2

Pseudoword set	Stimuli
Set 1	cecefo (*windmill*), minón (*firefighter)*, ina (*mailbox*), pemalero (*moneybox*), anlecalora (*sheep*), rufeso (*garlic*), arpel (*onion*), oraka (*sock*), salana (*camel*), vansusta (*cherry*), tisbilla (*ant*), hosmurcue (*kiwi*), nafleta (*lettuce*), jibi (*cucumber*), leta (*pear*), beceserca (*lemon*)
Set 2	vetruza (*spider*), tisbero (*tree*), suntilla (*folder*), ócemo (*vacuum*), médano (*eggplant*), ricuento (*boot*), aliza (*strawberry*), cacebla (*melon*), percel (*pea*), lepón (*hamburger*), morba (*cookie*), sama (*deer*), cunvo (*lobster*), edo (*shark*), angrebador (*olive*), harniza (*grape*)
Set 3	nívuton (*tiger*), mecosar (*frog*), sorano (*potato*), cerocho (*pill*), faumante (*doll*), acefo (*peach*), cardetus (*bee*), mafralo (*asparagus*), lufón (*light bulb*), jobro (*pineapple*), crena (*lizard*), sible (*nun*), gubra (*fox*), sira (*pepper*), vavecoa (*brush*), miza (*watering can*)

Note. English words referred to the pictures assigned to each pseudoword are reported in parentheses.

language (i.e., pseudowords) instead of with LSC signs (see Table 3). We generated pseudowords based on Spanish subsyllabic elements with the Wuggy pseudoword generator (Keuleers & Brysbaert, 2010). We formed target items by combining the objects' corresponding real names that varied in length from one to four syllables ($M = 2.72$ syllables). Fifteen native Spanish speakers (eight females, seven males) recorded the pseudowords in a soundproof room using the audio recording and editing software Audacity. We asked the speakers to record target pseudowords in a neutral voice type. We constructed stimulus lists following the same criteria as in Experiment 1, that is, the list for the training phase included the rotation of speakers used in the no variability and moderate variability conditions, which resulted in 54 experimental lists.

Procedure

As in our previous experiment, we tested the participants individually and online. The design of the experimental variability conditions was the same as for Experiments 1a and 1b. The sequence and procedure of the tasks (learning phase and test phase) were, with some exceptions, the same as in

Experiment 1a. First, the participants were told that they had to memorize words from a new language. Second, stimuli (pseudowords) were presented in their auditory form. Third, in the test phase, the participants were presented with a combination of nine speakers who were different from those speakers used in the learning phase. The trial structure of the two tasks was the same as in Experiment 1a with one exception. In the L2-to-L1 translation task, the participants listened to a L2 word and translated it into their L1.

Data Analysis

As in Experiments 1a and 1b, we binary coded the data (correct/incorrect) in the picture-to-L2 naming task and the L2-to-L1 translation task after the experiment. In the picture-to-L2 naming task, to maintain the same exclusion criteria adopted for the experiments on sign learning, we considered responses correct only if the participants produced all the phonemes of the pseudoword correctly. Likewise, we considered as correct responses trials in which the participants provided a different but acceptable word for the chosen picture (e.g., participants named a picture of a doll as "doll" or "baby"). We considered other responses, including mispronunciations, intrusions (naming the picture in another language), and no responses, as incorrect responses. For the L2-to-L1 translation task, we considered incorrect responses or trials in which the participants did not respond to be errors. As in Experiments 1a and 1b, we analyzed accuracy with generalized mixed models, including fixed effects for the variability condition (no variability, moderate variability, and high variability) and crossed random effects for participants and items. The R code for the final statistical model was: accuracy ∼ variability condition + (1 | participants) + (1 | items).

Results

Picture-to-L2 Naming Task

The mixed-effects model for the picture-to-L2 naming task (2,592 observations), AIC = 2,580, $R^2_{\text{marginal}} = .002$, $R^2_{\text{conditional}} = .52$, revealed no effect of variability: moderate variability, $b = -0.04$, $SE = 0.13$, 95% CI [−0.01, 0.49], $z = -0.30$, $p = .76$; high variability, $b = 0.23$, $SE = 0.13$, 95% CI [−0.28, 0.19], $z = 1.87$, $p = .06$, with only a trend for participants' recalling pseudowords learned in the high variability condition more accurately than those pseudowords that they learned in the no variability condition (see Figure 3 and Table 2).

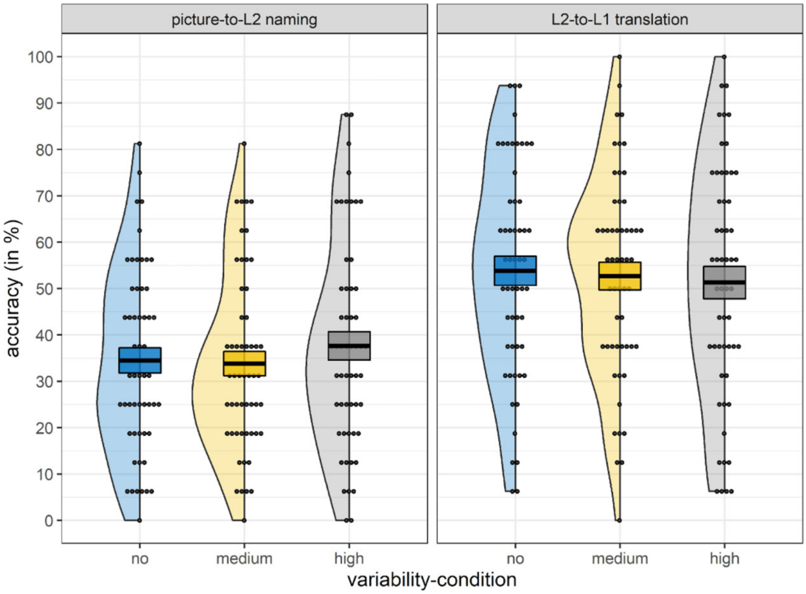

Figure 3 Percentage of correct responses for each variability condition across tasks for Experiment 2. The half violin shape shows the kernel probability density of participants' mean scores. Dots indicate the percentage of correct responses for each participant in each variability condition. Box plots indicate mean values and standard error.

L2-to-L1 Translation Task

The mixed-effects model for the L2-to-L1 translation task (2,592 observations), AIC = 2,928, $R^2_{marginal}$ = .0007, $R^2_{conditional}$ = 0.43, also revealed no effect of variability: moderate variability, $b = -0.07$, $SE = 0.12$, 95% CI [−0.31, 0.16], $z = -0.63$, $p = .53$; high variability, $b = -0.16$, $SE = 0.12$, 95% CI [−0.39, 0.07], $z = -1.40$, $p = .16$. This indicated that the number of speakers had no influence when the participants translated pseudowords into their L1.

Discussion

The results from Experiment 2 showed no differences between variability conditions. That is, the participants' learning rate was not modulated by variation in the number of speakers. Even though we evaluated speaker variability in the oral modality, our results were at odds with previous results in the literature on spoken language that has revealed a systematic increase in L2 recall accuracy with an increased number of speakers (e.g., Barcroft & Sommers, 2005; Barcroft & Sommers, 2014). Differences in learning outcomes between our studies could not be attributed to methodological differences, considering the rotation procedure across conditions (see Barcroft, 2001, for no effect of variability when a rotation procedure was not applied). In our study, following Barcroft and Sommers (2005), we rotated different speakers across conditions.

However, we must acknowledge other relevant differences between the studies. First, the number of words to be learned was twice as high in our experiment as it was in previous studies. Most of the previous experiments tested 24 items (Barcroft & Sommers, 2005), and here we used 48 items. This might have reduced the learning rate, which could have obscured the effect of variability. However, in Experiment 1b, we observed that variability influenced learning outcomes in the picture naming task only when accuracy rates were reduced in delayed recall. Second, in our experiment, the stimuli were pseudowords that we had generated from Spanish phonemes/syllables, Spanish being one of the participants' two native languages, as e tested bilingual Catalan–Spanish speakers. In that sense, Experiment 2 may not have matched the conditions of learning vocabulary in an unknown language but may have required the participants to acquire new Spanish words for existing concepts. To our knowledge, only Runge et al. (2017) tested speaker variability in L1 recall and obtained no evidence for it. However, Runge et al. interpreted their findings as the result of task difficulty—because words were paired with written definitions—rather than an effect of testing L1 words.

Given these experimental differences in number of items and in word status (i.e., pseudowords generated from a language in which the participants were proficient instead of being unknown L2 words), we conducted a new experiment in which we brought our design as close as possible to that of Barcroft and Sommers (2005). To achieve this, we reduced the number of items for the participants to learn and used words in the L2 of our participants, that is, English.

Experiment 3: L2 (English) Words

Method

Participants

We recruited 42 participants (31 females, 11 males; M_{age} 22.30 years, range = 18–40) from the Universitat Pompeu Fabra's Center for Brain and Cognition database. None of them had participated in the two previous experiments. All of them were bilingual Catalan–Spanish speakers who had learned English as their L2 and had a B1 level of English according to the Common

Table 4 Groups of English word stimuli used for the three conditions of variability in Experiment 3

Word set	Stimuli
Set 1	rake, pickle, whip, thimble, crib, sideburns, acorn, gown
Set 2	chalk, owl, elbow, faucet, crutch, muffler, skunk, dreadlock
Set 3	funnel, sling, apricot, peacock, stapler, plunger, clover, crane

European Framework of Reference for Languages, corresponding to intermediate proficiency, the minimum level required to undertake undergraduate studies in Spain.

Materials

We divided a set of 24 English words into three groups (see Table 4) and selected related pictures for them for the experiment. The words were concrete nouns from different semantic categories (animals, fruits, vegetables, tools, and vehicles) that we selectively chose to avoid the use of Catalan/Spanish–English cognates. To avoid words that the participants already knew, we chose words that the CELEX database (Baayen et al., 1995) classifies as low-frequency ($M = 3.45$, $SD = 4.26$). The words ranged in number of syllables from one to three ($M = 1.71$), but we controlled this across word sets. We confirmed the appropriateness of the set of selected words by presenting these words and their corresponding pictures to a different group of Catalan–Spanish bilingual participants. Eight speakers recorded the words in a soundproof room using the Audacity software. We used six speakers for the learning phase (three females, three males), whereas we used the remaining two speakers (one female, one male) for the testing phase to ensure the use of novel voices that were the same for all participants. The speakers were all native speakers of American English. The experimental design followed the same rotation procedure as in Experiments 1a, 1b, and 2, that is, we counterbalanced each word set across variability conditions. This yielded six subvariations to rotate speakers' identity in the no variability and moderate variability conditions (Barcroft & Sommers, 2005).

Procedure

We tested the participants individually in a soundproof cabin in front of the computer. We randomly assigned the list used for each participant. We executed the stimuli presentation in the learning phase through the E-Prime

(Version 2.0) software (https://pstnet.com/products/e-prime). We used the DMDX display system (Forster & Forster, 2003) in the testing phase, and we subsequently checked responses with the CheckVocal software (Protopapas, 2007). Prior to beginning the experiment, we presented the list of pictures to the participants and asked them to state the word in English if they knew it. For each participant, we noted preknown words and excluded them from the analysis (1.8% of the data on average).

The procedure of the experimental session was as follows. In the learning phase, the participants saw an asterisk on the screen for 500 ms, and then they were presented a picture for 4,250 ms. This picture was accompanied by an audio recording of the word that the picture represented 750 ms after the onset of the picture presentation. Finally, the participants saw a blank screen for 500 ms. After the learning phase, we first administered the picture-to-L2 naming task to avoid the participants' hearing the L2 words before the pictures had been named. In the picture-to-L2 naming task, we presented the participants with pictures, and they had to provide the corresponding English names. An experimental trial comprised first an asterisk that was present 500 ms in the screen, then a blank of 300 ms, followed by the picture presentation that remained on the screen for a maximum of 10,000 ms. In the L2-to-L1 translation task, the participants heard the English word that they had to translate into Spanish. A fixation asterisk was presented for 500 ms, then a blank of 300 ms, followed by the auditory presentation of the word in English. When the word finished, a blank screen appeared for a maximum of 10,000 ms.

Data Analysis

As in Experiments 1a, 1b, and 2, we binary coded the data (correct/incorrect) in both the picture-to-L2 naming task and the L2-to-L1 translation task after the experiment. In the picture-to-L2 naming task, we considered responses correct only if the participants produced all the phonemes of the English word correctly or had only one incorrect phoneme in a single syllable. We considered other responses, including mispronunciations, intrusions, synonymous, and no responses, as incorrect responses. For the L2-to-L1 translation task, we excluded from the analysis incorrect responses or trials in which the participants did not respond. As in the previous experiments, we analyzed accuracy with generalized mixed models, including fixed effects for variability conditions (no variability, moderate variability, and high variability) and crossed random effects for participants and items. The R code for the final statistical model was: accuracy $\sim$ variability condition + (1 | participants) + (1 | items).

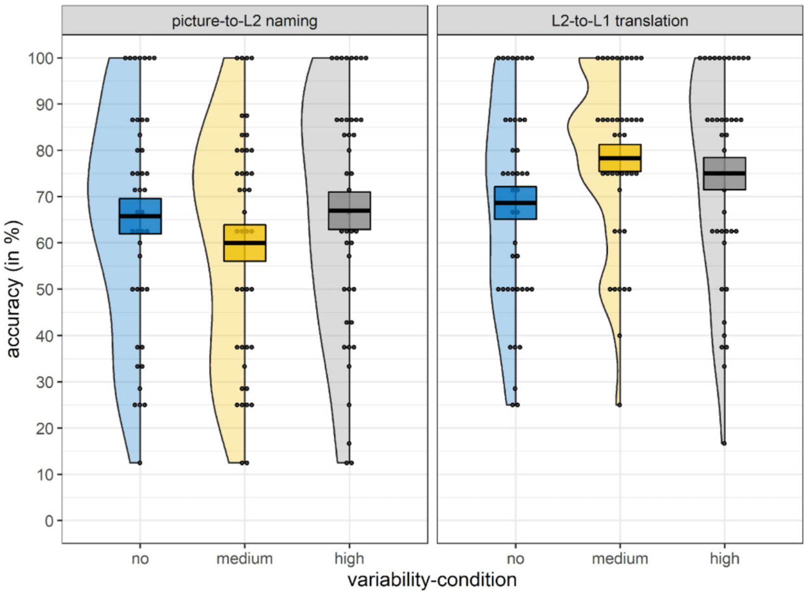

Figure 4 Percentage of correct responses for each variability condition across tasks for Experiment 3. The half violin shape shows the kernel probability density of participants' mean scores. Dots indicate the percentage of correct responses for each participant in each variability condition. Box plots indicate mean values and standard error.

Results

Figure 4 and Table 2 show the percentage of correct responses for each variability condition across tasks.

Picture-to-L2 Naming Task

The mixed-effects model for the picture-to-L2 naming task (895 observations), AIC = 1,068, $R^2_{\text{marginal}} = .006$, $R^2_{\text{conditional}} = .32$, revealed no effect of variability: moderate variability, $b = -0.32$, $SE = 0.19$, 95% CI [−0.76, 0.09], $z = -1.67$, $p = .10$; high variability, $b = 0.09$, $SE = 0.19$, 95% CI [−0.32, 0.52], $z = 0.47$, $p = .64$. This showed that the number of speakers had no influence on the recall of new English (L2) words.[1]

L2-to-L1 Translation Task

The mixed-effects model for the L2-to-L1 translation task (895 observations), AIC = 955, $R^2_{\text{marginal}} = .012$, $R^2_{\text{conditional}} = .30$, revealed an effect of variability: moderate variability, $b = 0.57$, $SE = 0.20$, 95% CI [0.18, 0.98],

$z = 2.80$, $p = .005$; high variability, $b = 0.41$, $SE = 0.20$, 95% CI [0.01, 0.83], $z = 2.03$, $p = .04$. The participants more accurately translated English L2 words that they had learned in the moderate and high variability conditions than they did those English L2 words that they had learned in the no variability condition.

Discussion

The results of this experiment revealed an effect of speaker variability limited to the L2-to-L1 translation task. In this task, the participants more accurately translated words from their English L2 to their Spanish L1 for words encoded in moderate and high speaker variability contexts than for words encoded without such variability. In the picture-to-L2 naming task, we did not observe the positive effect of variability that we had found in the translation task. The absence of indexical variation effects in the picture-to-L2 naming task differed once again from the results of previous studies in L2 spoken vocabulary learning (e.g., Barcroft & Sommers, 2005). Even though we intended the design of Experiment 3 to be a close replication of the Barcroft and Sommers (2005) experiment, one difference between the studies was notable. The participants in our study had preexisting knowledge of English as a L2 when they performed the task. That is, unlike the participants in the Barcroft and Sommers (2005) study, our participants were learning new words in a familiar language. However, it is unlikely that preexisting knowledge of the L2 was responsible for the differences between the studies considering the results of Experiments 1a and 1b. The participants had no preexisting knowledge of LSC, and we found an effect of indexical variability in the translation task but not in the picture naming task. In any case, what seemed clear was that the influence of variability in L2 learning was not determined by the modality in which L2 vocabulary was acquired (i.e., signs vs. spoken words). Thus, an explanation for the pattern of results obtained should include both modalities.

General Discussion

The reason for our study was to investigate the influence of signer variability on L2 sign learning. This was based on observation in the oral modality that indexical variability in the number of speakers boosts L2 vocabulary learning in adults (e.g., Barcroft & Sommers, 2005). These results were important in revealing that linguistic and nonlinguistic (indexical) information are codified in parallel in speech processing and that nonlinguistic variation enhances encoding of lexical representations in a new language (Goldinger, 1998). The extrapolation of this phenomenon to the signed modality was our main objective. In an attempt to determine whether nonlinguistic information is encoded from signs (as in speech) and influences memory and L2 learning, we evaluated the impact of signer variability on L2 sign learning in adults. Overall, our results indicated that indexical information is encoded along with linguistic information from signs (phonological parameters, meaning) and influences the learning of signs. The participants more accurately recalled L2 signs from memory when they had encoded the L2 signs from multiple signers than when they had encoded them from one signer. The effect of variability remained for days, as revealed in Experiment 1b, which involved posttesting, when the participants had not trained with the materials but performed only the two recall tasks.

The results on indexical variation in L2 sign learning only partially replicated previous reports in the oral modality (Barcroft & Sommers, 2005; Sinkeviciute et al., 2019; Sommers & Barcroft, 2007, 2011). First, we did not consistently obtain the effect of variability in the two recall tasks that we employed. In Experiment 1a, we obtained a positive effect of variability in the L2-to-L1 translation task but not in the picture-to-L2 naming task. Conversely, in Experiment 1b, the retest, we observed the effect of variability in the picture-to-L2 naming task but not in the L2-to-L1 translation task. Researchers have often interpreted null effects of speaker variability on L2 learning as a ceiling effect (i.e., task too easy; Sinkeviciute et al., 2019; Uchihara et al., 2021) or a floor effect (i.e., task too difficult; Runge, 2018) in overall performance. As we have described, numerically, accuracy in the picture-to-L2 naming task of Experiment 1a was far better than has previously been reported in L2 oral languages (e.g., Barcroft & Sommers, 2005, 2014). Thus, we interpreted our results in Experiment 1a as a ceiling effect that could have affected the variability effect in the picture-to-L2 naming task. Indeed, in Experiment 1b, accuracy in the picture-to-L2 naming was notably reduced (0.70 vs. 0.39, see Table 2), and signer variability facilitated retrieval. Although relatively high accuracy is a variable that could mask the benefits of indexical variation, it noticeably does not entirely fit with the results of the L2-to-L1 translation task. Between Experiments 1a and 1b, accuracy was only slightly reduced (0.89 vs. 0.79), but variability only influenced recall in the L2-to-L1 translation task in Experiment 1a. Second, we did not observe a systematic increase in accuracy with variability. In the L2-to-L1 translation task, we observed a significant difference only when the variation was sufficiently high (six signers vs. one signer). The lack of variability effects in the moderate variability condition has been interpreted as insufficient variation for the L2 form–meaning connections to be established

(Rott, 1999; van Zeeland & Schmitt, 2013). Once again, this explanation does not fully apply to our data since, in the picture-to-L2 naming task in Experiment 1b, both variation conditions benefitted production, and if anything, the effect was larger in the moderate variability condition (7% gain) than in the high variability condition (5% gain).

In sum, our data generally support the idea that multiple signers benefit L2 vocabulary learning. However, they also suggest that other undescribed variables related to the learner, such as movement and/or visuospatial short-term memory (Martinez & Singleton, 2018) or phonological short-term memory (Martinez & Singleton, 2019), to the items to be learned (e.g., L1 or L2 items, number of items), and to the tasks employed, such as the use of novel words associated with pictures or definitions (Runge, 2018), are interwoven with signed variation in the process of L2 memory encoding, which influences the learning outcomes. Importantly, as our data show, the modality of the language to be learned (sign/oral) does not seem to interact with variation in L2 learning.

The results from Experiments 2 and 3 also suggest some limitations in the effects of speaker variation on L2 learning. When we tested pseudowords in Experiment 2, we observed no variability effects in any of the recall tasks. At first, we hypothesized that the absence of variability could be due to the fact that we were not evaluating L2 vocabulary learning but rather learning new words from a language in which the participants were very proficient. Pseudowords were constructed following the phonology and morphology of Spanish, a language in which the participants had a native or nativelike proficiency. Thus, even though the participants had been told that their task was to learn words in an invented language, it is possible that the participants treated the pseudowords as new L1 Spanish words. However, considering previous evidence from L1 studies, one would expect a negative influence of variability, similar to negative effects reported in L1 processing (Choi et al., 2018; Magnuson et al., 2021; Martin et al., 1989). Thus, a more plausible explanation for the null effect of variability when our participants were learning pseudowords relates to the low accuracy reached in the experiment (0.35 in the picture-to-L2 naming task and 0.51 in the L2-to-L1 translation task). For instance, no benefits for speaker variability were found when participants learned novel L1 words or L2 words via written definitions (Runge et al., 2017) or embedded in written/auditory sentences (Runge, 2018). Runge et al. (2017) suggested that accessing a word's meaning through a multiword description or definition entailed increased difficulty and increased demands on working memory, and this cancelled out variability effects. In line with this observation, the low accuracy rates that we observed in Experiment 2 might have indicated greater working memory demands that learning a large number of new words entails, thus limiting resources to encode indexical variation that benefits later memory and learning.

Experiment 3 was the closest replication of Barcroft and Sommers (2005). However, the results did not fully replicate the benefit of multiple speakers on L2 vocabulary learning. Here, we observed a variability effect in the L2-to-L1 translation task but not in the picture-to-L2 naming task. These results replicated those obtained in Experiment 1a, which revealed a benefit of multiple signers/speakers in the L2-to-L1 translation task but not in the picture-to-L2 naming task. Only when we tested L2 vocabulary a second time (signs in Experiment 1b), did multiple signers benefit in the picture-to-L2 naming task. To account for the results in both modalities, we built upon Jiang's (2000) psycholinguistic model of adult L2 vocabulary learning. According to this model, L2 lexical learning undergoes at least two stages. In an initial stage, new L2 words are mapped to their L1 translations and not directly to meaning. Therefore, each time a L2 word is encountered, its L1 translation is activated, and meaning is only accessed through L1 activation. As suggested, during this initial stage, L2 learners experience more difficulties in retrieving L2 word/sign forms than they do in retrieving meanings (e.g., Ortega & Morgan, 2015; VanPatten, 1990). As experience in L2 increases, L2 words rely less on L1 translations to access meaning, and direct mappings between L2 forms and meaning are created (see also de Groot, 1992; Kroll & Stewart, 1994).

In our experiments, we exposed the participants to a set of new L2 signs/words that they had never seen before. After a few minutes of exposure to new vocabulary, it is likely that sign forms/acoustic representations were still fragile, in the sense that they entailed fuzzy lexical representations (Gor et al., 2021), leaving the participants with limited resources for establishing direct L2 form–meaning mappings. This imprecise encoding of L2 forms would explain the differences found in the influence of indexical variability between the picture naming task, a task requiring production in L2 of the learners, and the translation task, a task requiring production in L1. As shown by Kroll and Stewart (1994), the picture-to-L2 naming task and the L2-to-L1 translation task differ in the degree to which form-to-meaning mappings are emphasized in the task. While the picture-to-L2 naming task is conceptually mediated, the L2-to-L1 translation task relies on lexical links between the two languages. In this context, if the L2-to-L1 translation task relies on lexical links between the two languages, then it could be a more sensitive task for detecting the effects of variability at the initial stages when a word or a sign is learned.

Within this framework, accounting for the results obtained at retest in the picture-to-L2 naming task (Experiment 1b) would necessarily require assuming that between test and retest, the L2 form–meaning mappings were sufficiently strengthened to reveal effects of variability in the picture-to-L2 naming task. Perhaps, because of memory consolidation, novel L2 sign meanings might have been sufficiently integrated in the semantic system. This would result in more sensitivity to signer variation in the task tapping into semantics. Several studies have provided evidence that offline consolidation and sleep facilitates novel word integration (Davis et al., 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Tamminen & Gaskell, 2013). For instance, Tamminen and Gaskell (2013) reported that, despite the recall rate's declining over time, priming effects as an index of integration into the lexicon increased over time of consolidation (see also Clay et al., 2007). In this realm, our results in Experiment 1 might indicate that from test to retest, L2 signs benefitted from time for being integrated into the semantic system despite a decline in the overall recall performance. Although this is an interesting possibility, it requires further work to elucidate the effect of offline consolidation on L2 learning and its interaction with indexical variation.

Limitations and Future Directions

The results of our study across experiments and tasks are not fully consistent with previous studies on the influence of indexical variation on L2 vocabulary learning, which makes it difficult to develop a theoretical framework that encompasses present and past results. We did not design our experiments to be full replications of previous experiments but to cover broad aspects of L2 vocabulary learning so as to draw common lines between studies. In doing so, some methodological differences were warranted (languages of learners, number of items) that could have influenced the pattern of results observed within our experiments and between our research and previous studies. Likewise, other learner-related variables might have impacted achievement in L2 vocabulary learning (Martinez & Singleton, 2019). As an example, Perrachione et al. (2011) reported that individual differences in pitch perception influenced whether participants benefitted from high variability training of phonological contrasts. Thus, individual differences in sign/word perception might have modulated the extent to which our participants benefitted from high variability training.

Further experiments considering methodological and individual differences are needed to provide a better understanding of the strength of the indexical variability effect or the aspects that may influence it. For methodological differences, experiments that replicate the same design where only the variable of interest (e.g., number of items) is changed would be useful for establishing direct comparisons. To explore individual differences, accuracy scores related to indexical variability could be correlated with cognitive measures that influence sign language learning. Variables such as movement short-term memory and visuospatial short-term memory (Martinez & Singleton, 2018) and fluid intelligence and sign phonological short-term memory (Martinez & Singleton, 2019) have been reported as contributing to sign learning. Thus, it is possible that effects of indexical variation interact at individual level with these variables.

Conclusion

Our results provide evidence that signs and words are composed of lexical and indexical information. Both sources of information interact during processing and memorization. In a series of experiments, we showed that indexical variation in the number of signers is a relevant cue that influences L2 sign learning in adults. In terms of sign language teaching practices, our study addressed a question appropriate for effective L2 vocabulary instruction: Is it beneficial to use different signers when new signs are presented? Overall, our data suggest that learners benefit from seeing multiple signers when they learn signs. With a closer look at the results, we did not observe a robust benefit of signer variability across experiments and tasks. Thus, we remain cautious about drawing strong conclusions and giving pedagogical suggestions at this point. Finally, our results indicate some limitations of the positive effect of variability on learning pseudowords and L2 words and suggest that variability effects might interact with L2 lexical development.

Final revised version accepted 18 October 2022

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/upgma. All proprietary materials have been precisely identified in the manuscript.

Note

1 We also analyzed learning outcomes following the scoring procedure (0, 0.5, 1 points) of Barcroft and Sommers (2005), giving partial credit to productions that were missing or used one incorrect phoneme within a single syllable. The results revealed the same pattern of no effect of variability that we had observed with binary scoring (0, 1): moderate variability, $b = 0.01$, $SE = 0.03$, 95% CI [−0.05, 0.07], $t = 0.56$, $p = .57$; high variability, $b = 0.04$, $SE = 0.03$, 95% CI [−0.02, 0.11], $t = 1.60$, $p = 0.11$.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [Database]. University of Pennsylvania, Linguistic Data Consortium.

Barcroft, J. (2001). Acoustic variation and lexical acquisition. *Language Learning*, *51*(4), 563–590. https://doi.org/10.1111/0023-8333.00168

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, *27*(3), 387–414. https://doi.org/10.1017/S0272263105050175

Barcroft, J., & Sommers, M. S. (2014). Effects of variability in fundamental frequency on L2 vocabulary learning: A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition*, *36*(3), 423–449. https://doi.org/10.1017/S0272263113000582

Bates, D., Maechler, M., & Bolker, B. (2011). *Linear mixed-effects models using S4 classes*. (R package version 0.999375-42) [Computer software]. http://lme4.r-forge.r-project.org

Baus, C., & Costa, A. (2015). On the temporal dynamics of sign production: An ERP study in Catalan Sign Language (LSC). *Brain Research*, *1609*, 40–53. https://doi.org/10.1016/j.brainres.2015.03.013

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, *61*(2), 206–219. https://doi.org/10.3758/BF03206883

Caselli, N. K., & Cohen-Goldberg, A. M. (2014). Lexical access in sign language: A computational model. *Frontiers in Psychology*, *5*(May), Article 429. https://doi.org/10.3389/fpsyg.2014.00428

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, and Psychophysics*, *80*(3), 784–797. https://doi.org/10.3758/s13414-017-1395-5

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(5), 970–976. https://doi.org/10.1037/0278-7393.33.5.970

Corina, D., Grosvald, M., & Lachaud, C. (2011). Perceptual invariance or orientation specificity in American Sign Language? Evidence from repetition priming for signs and gestures. *Language and Cognitive Processes*, *26*(8), 1102–1135. https://doi.org/10.1080/01690965.2010.517955

Cummins, J. (1999). Research news and comment: Alternative paradigms in bilingual education research: Does theory have a place? *Educational Researcher*, *28*(7), 26–41. https://doi.org/10.3102/0013189x028007026

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803–820. https://doi.org/10.1162/jocn.2009.21059

de Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. *Advances in Psychology*, *94*, 389–412. https://doi.org/10.1016/S0166-4115(08)62805-8

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Ellis, R. (1986). *Understanding second language acquisition*. Oxford University Press. https://doi.org/10.4324/9780203777282

Ellis, R. (1994). Factors in the incidental acquisition of second language vocabulary from oral input: A review essay. *Applied Language Learning*, *5*(1), 1–32. https://doi.org/10.1075/sibil.17.06ell

Emmorey, K., Bosworth, R., & Kraljic, T. (2009). Visual feedback and self-monitoring of sign language. *Journal of Memory and Language*, *61*(3), 398–411. https://doi.org/10.1016/j.jml.2009.06.001

Emmorey, K., Petrich, J. A. F., & Gollan, T. H. (2013). Bimodal bilingualism and the frequency-lag hypothesis. *Journal of Deaf Studies and Deaf Education*, *18*(1), 1–11. https://doi.org/10.1093/deafed/ens034

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gimeno-Martínez, M., Costa, A., & Baus, C. (2020). Influence of gesture and linguistic experience on sign perception. *Journal of Deaf Studies and Deaf Education*, *25*(1), 115–125. https://doi.org/10.1093/deafed/enz031

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(5), 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152–162. https://doi.org/10.1037/0278-7393.17.1.152

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Frontiers in Psychology*, *12*(November), Article732030. https://doi.org/10.3389/fpsyg.2021.732030

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America*, *102*(1), 655–658. https://doi.org/10.1121/1.419712

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. https://doi.org/10.5070/p78kf3z3s9

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/JMLA.1994.1008

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Looney, D., & Lusin, N. (2018). *Enrollments in languages other than English in United States institutions of higher education, summer 2016 and fall 2016: Preliminary report*. Modern Language Association. https://files.eric.ed.gov/fulltext/ED590075.pdf

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2021). Talker familiarity and the accommodation of talker variability. *Attention, Perception, and Psychophysics*, *82*(4), 1842–1860. https://doi.org/10.3758/s13414-020-02203-y

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 676–684. https://doi.org/10.1037/0278-7393.15.4.676

Martinez, D., & Singleton, J. L. (2018). Predicting sign learning in hearing adults: The role of perceptual-motor (and phonological?) processes. *Applied Psycholinguistics*, *39*(5), 1–27. https://doi.org/10.1017/S0142716418000048

Martinez, D., & Singleton, J. L. (2019). Individual differences in lexical learning across two language modalities: Sign learning, word learning, and their relationship in hearing non-signing adults. *Acta Psychologica*, *198*, Article 102892. https://doi.org/10.1016/j.actpsy.2019.102892

Mullennix, J. W., Pisoni, D. B., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, *85*(1), 365–378. https://doi.org/10.1121/1.397688

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, *34*(2), 202–240. https://doi.org/10.1016/j.wocn.2005.05.003

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1228–1244. https://doi.org/10.1037/0278-7393.26.5.1228

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*(7), 989–1001. https://doi.org/10.3758/BF03205458

Ortega, G., & Morgan, G. (2015). Input processing at first exposure to a sign language. *Second Language Research*, *31*(4), 443–463. https://doi.org/10.1177/0267658315576822

Ortega, G., Özyürek, A., & Peeters, D. (2019). Iconic gestures serve as manual cognates in hearing second language learners of a sign language: An ERP study. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(3), 403–415. https://doi.org/10.1037/xlm0000729

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. https://doi.org/10.1037/0278-7393.19.2.309

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and

training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366
Pichler, D. C., & Koulidobrova, H. (2016). Acquisition of sign language as a second language (L2). In M. Marschark & P. E. Spencer (Eds.), *The Oxford handbook of deaf studies in language* (pp. 218–230). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190241414.013.14
Pisoni, D. B., & Levi, S. V. (2007). Representations and representational specificity in speech perception and spoken word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 3–18). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0001
Pisoni, D. B., & Remez, R. E. (2008). Introduction. In D. B. Pisoni & R. E., Remez (Eds.), *The handbook of speech perception* (pp. 1–3). John Wiley & Sons. https://doi.org/10.1002/9780470757024
Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–862. https://doi.org/10.3758/BF03192979
Pyers, J. E., Perniss, P., & Emmorey, K. (2015). Viewpoint in the visual-spatial modality: The coordination of spatial perspective. *Spatial Cognition and Computation*, *15*(3), 143–169. https://doi.org/10.1080/13875868.2014.1003933
R Core Team. (2019). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org
Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x
Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 589–619. https://doi.org/10.1017/s0272263199004039
Runge, N., Sommers, M. S., & Barcroft, J. (2017). Semantic processing interferes with the benefits of talker variability in vocabulary learning [Poster]. Presented at Annual Meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.
Runge, N. (2018). *The effect of talker and contextual variability on memory for words in sentences*. Arts & Sciences Electronic Theses and Dissertations, 1726. https://doi.org/10.7936/db4h-dz76
Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*(3), 441–452. https://doi.org/10.1037/0012-1649.33.3.441
Schönström, K. (2021). Sign languages and second language acquisition research: An introduction. *Journal of the European Second Language Association*, *5*(1), 30–43. http://doi.org/10.22599/jesla.73

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*(4), 795–820. https://doi.org/10.1017/S0272263119000263

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Sommers, M. S., & Barcroft, J. (2006). Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *Journal of the Acoustical Society of America*, *119*, 2406–2416. https://doi.org/10.1121/1.2171836

Sommers, M. S., & Barcroft, J. (2007). An integrated account of the effects of acoustic variability in first language and second language: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics*, *28*(2), 231–249. https://doi.org/10.1017/s0142716407070129

Sommers, M. S., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, *32*(2), 417–434. https://doi.org/10.1017/S0142716410000469

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology*, *66*(5), 1001–1025. https://doi.org/10.1080/17470218.2012.724694

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2021). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, *44*(2), 357–380. https://doi.org/10.1017/S0272263121000218

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, *12*(3), 287–301. https://doi.org/10.1017/S0272263100009177

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Williams, J., Darcy, I., & Newman, S. (2017). The beneficial role of L1 spoken language skills on initial L2 sign language learning: Cognitive and linguistic predictors of M2L2 acquisition. *Studies in Second Language Acquisition*, *39*(4), 833–850. https://doi.org/10.1017/S0272263116000322

Williams, J. T., & Newman, S. D. (2016). Impacts of visual sonority and handshape markedness on second language learning of American Sign Language. *Journal of Deaf Studies and Deaf Education*, *21*(2), 171–186. https://doi.org/10.1093/deafed/env055

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary