Influence of research tasks and linguistic factors on phonetic convergence in language alternation

Ernesto R. Gutiérrez Topete[1]

[1]*University of California, Berkeley (USA)*
ernesto.gutierrez@berkeley.edu

Recent studies have reported on the dynamic nature of bilingual speech (i.e., two languages produced in a single discourse event) compared to monolingual speech (i.e., one language produced in a single discourse event) by multilingual speakers. For example, in monolingual speech, multilingual speakers "maintain language-specific phonetic categories"; meanwhile, in bilingual speech, the speakers displayed "phonetic convergence" [1, pg. 1280-1]. Seemingly, the active usage of multiple languages increases phonetic convergence in the production of multilingual speakers. Other studies on the production of voice onset time (VOT) during language alternation (i.e., language switching and code switching) corroborate the convergence effect observed in bilingual speech [e.g., 2, 3, 4, 5]. However, some inconsistencies have been pointed out [6, 7]. Namely, while some studies find bidirectional convergence, others only find effects in one language. And among those studies with a unidirectional effect, some researchers claim the effect occurs only in the speakers' *L2* (as opposed to the L1) while others claim the *dominant language* is more susceptible to convergence. The wide array of research tasks used to study this phenomenon makes it all the more difficult to pinpoint the cause of the aforementioned inconsistent directionality of the convergence effect. For example, some studies relied on word list reading tasks [6] or phrase/passage reading tasks [2, 3], and others used speech spontaneously produced during sociolinguistic interviews [4], inter-subject group conversations [4, 5], or puzzle tasks [5].

In order to shine light on the possible effect of task type on the directionality of convergence effects in bilingualism research, the present study analyzes acoustic productions across four of the most popular research tasks (i.e., word list reading, passage reading, puzzle—spot the difference—and casual interview tasks) from a single group of Spanish-English bilingual speakers to obtain VOT measurements for word-initial voiceless stops /p t k/. A total of 60 Spanish-English bilingual subjects participated in the four tasks, which yielded nearly 100 hours of recorded bilingual speech. Data collection took place in a sound booth at the Berkeley PhonLab. The audio for the word list and passage reading tasks were annotated by hand. The transcriptions for the puzzle and interview tasks were automated with OpenAI's Whisper automatic speech recognition model. All data were then processed with the Montreal Forced Aligner [8], and VOT measurements for voiceless stop-initial words were obtained in both languages using AutoVOT [9].

A mixed-effects linear regression model in R was performed on the VOT values of a majority of the corpus data. The independent variables under examination were task type (word list, passage, puzzle, interview), place of articulation (POA; bilabial, coronal, velar), and across-boundary context in English (/sC/, /nC/, other), to analyze the potential of resyllabification (more precisely, switch sites with /s/-final Spanish words followed by voiceless stop-initial English words, compared to /n/-final Spanish words or words ending in other sounds). Lexical item and subject were included as random intercepts. Study results show an influence of task on English data, with passage reading experiencing the highest level of convergence (i.e., lower VOT values), followed by the interview, puzzle, and word list (see figure 1). Spanish data showed no difference across tasks. POA results revealed that in English /p/ VOT productions are lower than /t k/ VOT, whereas in Spanish /p t/ VOT values are lower than /k/ VOT (see Figure 2). Finally, there is no evidence of resyllabification, given that we do not see shorter VOT values in the /sC/ linguistic context (compared to other contexts), which is linked to shortened VOT productions in American English when found as an onset cluster.

On their own, these results suggest that convergence is linked to attention to speech, where tasks that draw more attention to speech are less prone to convergence, meanwhile those that require less attention to individual sounds present higher levels of convergence. These results suggest that

methodological choice has an influence on the acoustics of (code-switched) speech, potentially resulting in methodological artifacts in studies that do not take this into consideration. Moreover, VOT productions per POA vary in English and Spanish (English: /p/ < /t k/; Spanish /p t/ < /k/), a pattern that can be explained by the distinctions of coronal stops in each language; Spanish has dental stops, and English has alveolar stops. In other words, in Spanish /p t/ are produced closer together than their English counterparts, resulting in more similar VOT productions between these two sounds in Spanish speech; these findings underlie the importance of analyzing articulatory differences when comparing language pairs in language alternation research. Finally, the fact that /sC/ linguistic contexts in switch sites were not correlated with lower VOT values in English stops indicates that speakers do not engage in a resyllabification process across languages during code-switched speech, suggesting that this form of phonotactics (ie, resyllabification) does not transfer from one language to another, even during bilingual speech.

All in all, this study provides (1) a comparative analysis of research methodologies that are commonly used in code-switching studies to uncover the inadvertent task effects in production studies and (2) a better understanding of the language processing mechanisms that are engaged during bilingual speech to better inform our methodological choices.

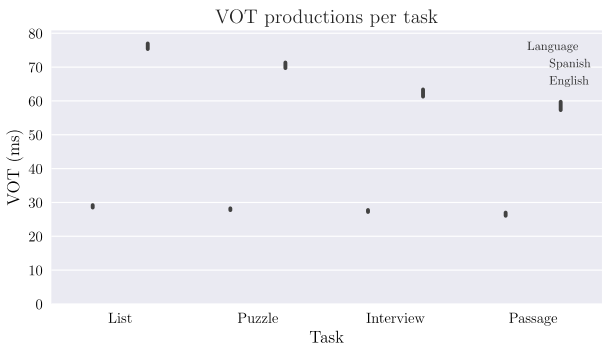

Fig.1 VOT productions per methodological task in English and Spanish

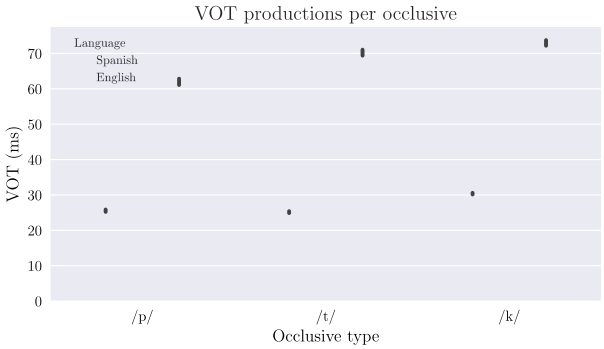

Fig.2 VOT productions per place of articulation in English and Spanish

References

[1] Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149* (2), 1271–1284.

[2] Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K. A. (2005). Perseverative phonetic effects in bilingual code-switching. In R. Gess & E. Rubin (Eds.), *Theoretical and experimental approaches to romance linguistics* (pp. 291–306). John Benjamins Publishing Company.

[3] Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In M. Grantham O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th generative approaches to second language acquisition conference* (pp. 9–16).

[4] Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, *19* (4), 423–443.

[5] Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish–English spontaneous code-switching. *Journal of Phonetics*, *52*, 121–137.

[6] Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in vot production. *Journal of Phonetics*, *41* (6), 407–420.

[7] Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110–137.

[8] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. *Interspeech, 2017*, 498–502.

[9] Keshet, J., Sonderegger, M., & Knowles, T. (2014). AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]. [Version 0.94, retrieved June 2020 from https://github.com/mlml/autovot/].