EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.17000

Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice

Kevin Papin[a] and Walcir Cardoso[b]

[a]Département de didactique des langues, Université du Québec à Montréal, , papin.kevin@uqam.ca and [b]Department of Education, Concordia University, , walcir.cardoso@concordia.ca

How to cite: Papin, K.; Cardoso, W. (2023). Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice. In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.17000

Abstract

This study investigated the pedagogical use of Google Translate's Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) capabilities for improving L2 French students' pronunciation of orthographic {-ent} endings in French: while {-ent} is silent in verbal constructions (e.g. "(ils) résident" [re.zid] '(they) reside'), the same sequence is pronounced [ã] in other forms such as nouns and adjectives (e.g. "(un) résident" [re.zi.dã] '(a) resident'). Twenty beginner learners of French completed self-regulated homework activities using Google Translate to listen to (to develop sound awareness and aural perception) and produce phrases (to develop oral production) containing the target {-ent} forms. A pretest, immediate posttest, and delayed posttest measured awareness, perception, and production of {-ent} pronunciation. Results of repeated-measures ANOVA indicate no significant improvement in awareness or perception, but a significant increase in production scores from pretest to each of the two posttests. The findings suggest that integrating Google Translate's built-in speech technologies into L2 pedagogy can improve the oral production of French grapheme-to-phoneme rules, such as those observed in {-ent} sequences.

Keywords: *L2 French, Google Translate, speech technologies, Automatic Speech Recognition (ASR), Text-To-Speech synthesis (TTS), L2 pronunciation.*

1. Introduction

Although online translation tools such as Google Translate (GT) are becoming increasingly popular among L2 learners for writing practice (Tsai, 2020), CALL researchers have called for an investigation into the potential of their built-in speech technology features to support L2 pronunciation learning (Derwing et al., 2022). These features include Text-To-Speech synthesizers (TTS) and Automatic Speech Recognition (ASR). TTS can provide machine-generated models of speech that learners can imitate and compare against their own pronunciation (Liakin et al., 2015, 2017). ASR, on the other hand, can detect learners' oral production and convert it to text, providing instant feedback on accuracy via the orthographic output (Cardoso, 2018; Garcia et al., 2020). Integrating these speech technologies into L2 pedagogy through platforms like GT could provide personalized, automated feedback to improve the development of pronunciation and speaking skills efficiently (Papin & Cardoso, 2022; van Lieshout & Cardoso, 2022).

Prior research has often examined TTS and ASR separately. These studies found that their pedagogical use can enhance L2 pronunciation (ASR: McCrocklin, 2016; TTS: Liakin et al., 2015). However, only a handful of

BY-NC-SA 2023, Editorial Universitat Politècnica de València

studies have targeted L2 French (for exceptions, see Liakin et al., 2017; Papin & Cardoso, 2022) and only three studies have examined the benefits of combining these two technologies in a single L2 learning experience (Khademi & Cardoso, 2022; Papin & Cardoso, 2022; van Lieshout & Cardoso, 2022). Moreover, while Papin and Cardoso (2022) found that the use of GT's ASR and TTS capabilities helped L2 French learners improve their production of French liaison (e.g. the re-syllabification of latent consonants when they appear in consonant-vowel sequences across words, as /z/ in *nos amis* [no.za.mi] 'our friends'), no study has yet looked at the combined impact of using both ASR and TTS on rule-based pronunciation phenomena in French. In this study, we focus on the pronunciation of the orthographic {-ent} sequence, which is pronounced [ã], except in verbal constructions, where it remains silent (e.g. while {-ent} is pronounced [ã] in nouns such as (*un*) *résident* [re.zi.dã] '(a) resident', it is not pronounced in verbal constructions such as (*ils*) *résident* [re.zid] '(they) reside').

The orthographic system of a language can have significant effects on how L2 learners pronounce words in that language (Bassetti et al., 2020; Bürki et al., 2019). Specifically, the regularity and consistency between spelling and pronunciation impacts the ease or difficulty with which learners acquire target pronunciations. The phonology of languages with more regular (transparent) orthographic systems, like Spanish, tend to be easier for learners to acquire than languages with more complex spelling-sound correspondences (opaque), like English. This is because regular orthographies provide clearer cues to guide pronunciation. Irregularities in spelling, on the other hand, can lead learners to mispronounce words (e.g. incorrectly producing [re.zi.dã] instead of the intended [re.zid] in *ils résident*), negatively affecting phonological development. Similar inhibitory effect of orthography on learners' productions has been demonstrated in several studies. For instance, Bassetti et al. (2020) found that Italian learners of English lengthened double consonants (digraphs) of words such as *fi<u>n</u>ish* and *Fi<u>nn</u>ish*, treating each set as minimal pairs in oral production (e.g. producing *fi<u>n</u>ish* with a singleton [n]: ['fɪ.nɪʃ] and *Fi<u>nn</u>ish* with a geminate [n:]: ['fɪ.nːɪʃ]). Given these potential effects, it is important for researchers to closely examine how orthography influences L2 phonological development.

Using a mixed-method approach to data collection, this study addressed the following research question:

- Can the pedagogical use of GT's speech technologies (TTS and ASR) lead to improved pronunciation of the French {-ent} orthographic sequence by French L2 learners?

2. Method

Participants were 20 adult learners enrolled in a beginner L2 French course in an English-speaking Canadian university. To reflect aspects of the participants' pedagogical realities (Erlam & Tolosam, 2022; i.e. the completion of homework assignments), students were asked to complete guided but self-regulated homework activities using GT's TTS and ASR capabilities successively, as part of their course activities. The participants were not presented with any metalinguistic knowledge during the experiment.

The activities targeted the pronunciation assigned to orthographic {-ent} in French, a rule-based orthographic phenomenon that poses difficulty to beginning learners because this morphemic form is homographic in French: while {-ent} is silent in verbal constructions (e.g. (*ils*) *résident* [re.zid]), the same sequence is pronounced [ã] in derived forms such as nouns and adjectives (e.g. (*un*) *résident* [re.zi.dã] '(a) resident').

During the three-week treatment, participants were asked to copy-and-paste a set of phrases containing the target {-ent} forms in GT and press the speaker/playback button (TTS) to listen to the synthesized voice. Participants were then asked to click on the microphone button (ASR) to orally produce the target forms. Finally, they were instructed to verify whether the ASR orthographic output matched what they intended to say (e.g. was intended *(ils) étudient* [e.ty.di] transcribed correctly*?*). The relevant interface of GT and its speech capabilities are illustrated in Figure 1, which also portrays the type of feedback it provides. Here, it displays an incorrect pronunciation for *ils étudient* 'they study', transcribed as produced by the user ([e.ty.djɑ̃] 'student'), not as intended ([e.ty.di]).

2023, Editorial Universitat Politècnica de València BY-NC-SA

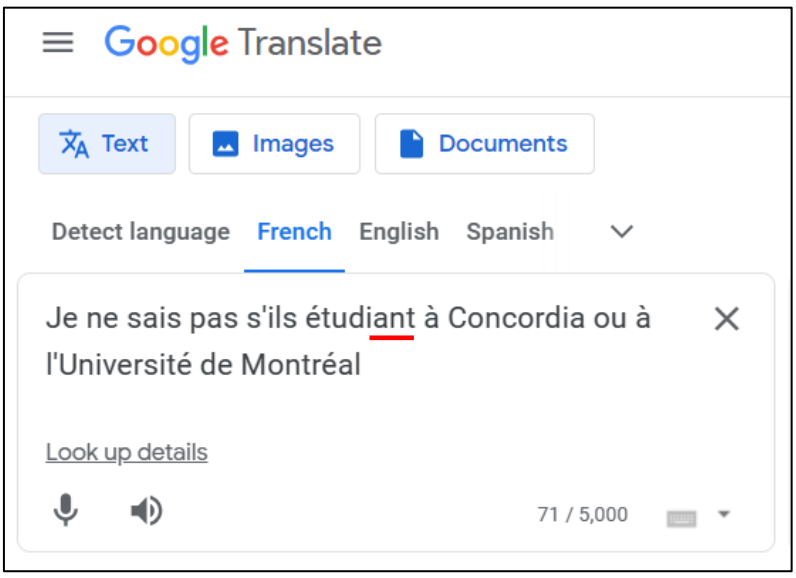

Figure 1. Activity sample in GT: Focus on French {-ent}

To measure the impact of the intervention on the participants' pronunciation development of {-ent}, a pretest and two posttests (immediate and delayed; the latter two weeks after the intervention) were administered. Following Celce-Murcia et al.'s (2010) framework for instructing and evaluating L2 pronunciation, the assessment of phonological development in this study incorporated three distinct levels of analysis (slightly simplified here due to space limitations):

(1) sound awareness (*Awareness*; to assess the participants' level of awareness to the grapheme-to-phoneme rules that characterize homography in {-ent}; e.g. *Is the -ent ending pronounced the same in these two examples: J'ai du talent. vs. Il y a du vent? Can you explain?*);
(2) aural discrimination (*Perception*; to ascertain the participants' ability to differentiate between the two {-ent}forms; participants listened to a short sentence and then decided on whether they heard the target -ent: e.g. *Is the -ent pronounced in the following sentence: La tomate est un aliment bon pour le corps?*); and
(3) oral production (*Production*; to gauge the participants' ability produce the homographic {-ent}; these consisted of reading aloud a set of sentences [e.g. *Les réponses du groupe varient beaucoup*] and a controlled question-and-answer oral task).

3. Results

One-way repeated-measure ANOVAs were performed separately for sound awareness, aural perception, and oral production. The statistical analysis found no significant changes over time for awareness, $F(2, 38) = 2.16$, $p = .287$, partial $\eta 2 = .06$ or perception, $F(2, 38) = .31$, $p = .734$, partial $\eta 2 = .02$. However, the intervention led to statistically significant changes in production over time, $F(2, 38) = 6.52$, $p = .004$, partial $\eta 2 = .26$. Because the study involved multiple comparisons (i.e. three test scores: awareness, perception, production, across three time points: pretest, posttest, delayed posttest), post hoc analysis with a Bonferroni adjustment was conducted, showing a significant increase between the mean scores from pretest to posttest, 1.82 (95% CI, .03 to 3.64), $p = .045$, $d = .50$, and from pretest to delayed posttest, 2.08, (95% CI, .22 to 3.95), $p = .025$, $d = .60$, but no significant increase from posttest to delayed posttest, .27 (95% CI, -.96 to 1.50), $p = 1.00$, $d = .07$. Table 1 summarizes the scores for the three phonological tests at pretest, posttest, and delayed posttest, while Figure 2 provides an illustration of these results (converted to percentages for comparison purposes), emphasizing the significant results in production (where * = $p = .045$ and ** = $p = 025$).

BY-NC-SA 2023, Editorial Universitat Politècnica de València

Table 1. Descriptive Statistics: three phonological tests over time (means and standard deviations)

Phonological Levels	Pretest		Posttest		Delayed Posttest	
	M	SD	M	SD	M	SD
Awareness (/3)	1.75	0.72	2.15	0.67	2.00	0.80
Perception (/12)	9.30	1.84	9.55	1.67	9.25	1.67
Production (/25)	15.63	3.14	17.45	4.16	17.71	3.82

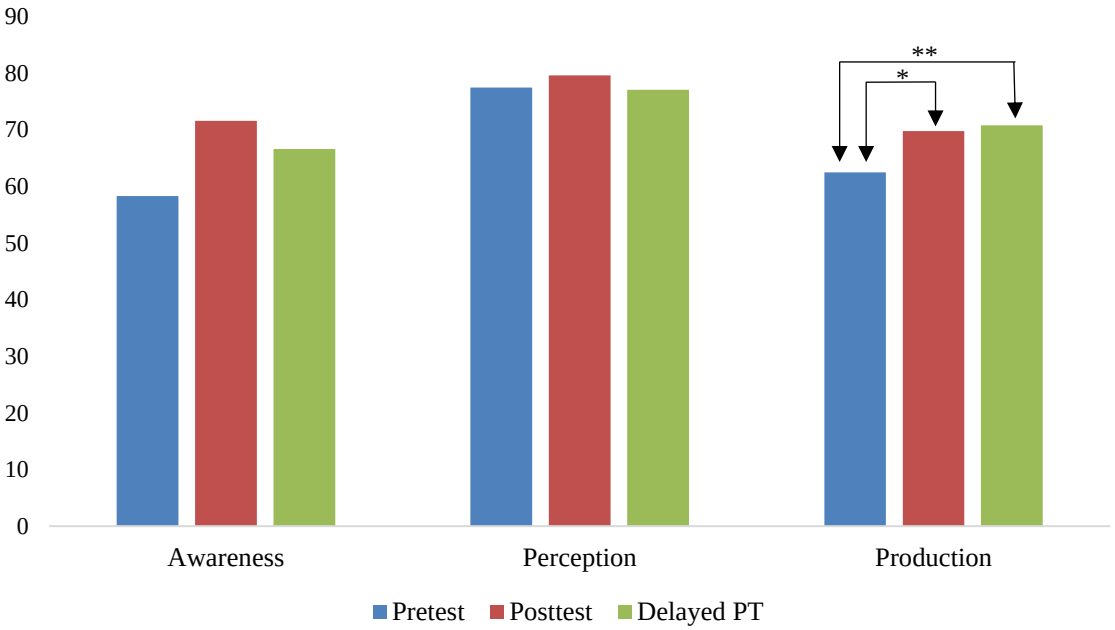

Figure 2. Scores across three tests over time (%).

4. Discussion and Conclusions

This study examined whether the self-regulated use of Google Translate's built-in TTS and ASR could improve L2 French learners' pronunciation of {-ent} endings, an orthographic sequence that is opaque to grapheme-to-phonemes rules: it can be phonetically realized either as [ã] (e.g. in nouns), or remain silent (in verbal forms). As indicated above, the findings only partially confirmed our initial hypothesis: while no gains were detected in the participants development of sound awareness and aural perception, learners' production scores significantly increased from pretest to the two posttests.

These developmental patterns mirror those found for French liaison in Papin and Cardoso (2022), using the same learning platform where the authors found an identical developmental path in which only production benefitted from the proposed treatment, not awareness or perception. In their discussion, the authors attribute this pattern to a ceiling effect in both awareness and perception: since their participants had advanced knowledge of *aural* French liaison before the experiment, indicated on prettest results, there was little room for improvement on those measures compared to production. In the current study, a ceiling effect may also explain the lack of gains

 BY-NC-SA

in perception, as participants scored nearly 80% accuracy on the pretest for this measure, not allowing much opportunity for scores to increase further. However, this explanation does not apply for awareness, since pretest awareness scores were comparable to those obtained for production. Specifically, the score for awareness on the pretest was approximately 58.3% (1.75/3), indicating comparatively low knowledge of the {-ent} pronunciation rules. Other factors that may have contributed to the lack of gains in awareness could stem from methodological issues or insufficient practice. For example, the study's measurement of awareness may have been too narrow, failing to capture more subtle gains in explicit knowledge. To assess awareness, one of the questions asked participants to explain the target pronunciation rule. The reliance on metalinguistic explanation as a measure of awareness assumes that this type of explicit knowledge will manifest in verbalizable rules. However, learners may develop awareness that they cannot readily articulate. Additionally, the three-week training period may not have provided the participants enough pronunciation practice of the target {-ent} form so that they could significantly improve in this level of phonological knowledge.

Overall, our findings corroborate those of van Lieshout and Cardoso (2022), whose participants were able to acquire L2 Dutch vocabulary and pronunciation on a short-term basis, also using the two technologies adopted in this study. As such, the results connect to evidence that TTS can provide helpful modeled speech for learning (Liakin et al., 2017), while ASR can offer speaking practice opportunities and instant feedback on learner output (Cardoso, 2018; Liakin et al., 2015). This study uniquely combined these tools for self-regulated practice, operationalized as the completion of homework activities to learn about French pronunciation. As such, it contributes to prior research on utilizing speech technologies for L2 pronunciation instruction.

The study has certain limitations that need to be addressed in future research. The small sample size of 20 learners reduces generalizability. Also, the short-term nature of the pedagogical intervention provides only initial evidence of the adopted tool's capabilities. Additionally, the lack of a control group limits the ability to isolate the pedagogical effects of the target speech technologies. Future research should utilize larger sample sizes, implement a longer-term intervention, and adopt a longitudinal design comparing technology-supported training to control groups undergoing alternative pronunciation instruction approaches.

Acknowledgements

We are grateful to Carol Johnson for her assistance with the statistical analysis. This project received funding from the *Social Sciences and Humanities Research Council* (Papin: 430-2022-00846; Cardoso: 430-2022-00512).

References

Bassetti, B. Mairano, P., Masterson, J., & Cerni, T.. (2020). Effects of orthographic forms on second language speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning, 170*(4), 1218-1256. https://doi.org/10.1111/lang.12423

Bürki, A., Welby, P., Clément, M. & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond "friend or foe?". *The Journal of the Acoustical Society of America. 145*(4), EL265–EL271. https://doi.org/10.1121/1.5094923

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds.), *Short papers from EUROCALL* (pp. 165–170). Research-Publishing.net. https://doi.org/10.14705/rpnet.2018.26.806

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). Teaching pronunciation: A course book and reference guide. Cambridge University.

Derwing, T., Munro, M., & Thomson, R. (Eds.). (2022). *The Routledge Handbook of Second Language Acquisition and Speaking*. Routledge. https://doi.org/10.4324/9781003022497

BY-NC-SA 2023, Editorial Universitat Politècnica de València

Erlam, R. & Tolosam, C. (2022). *Pedagogical realities of implementing task-based language teaching*. Benjamins. https://doi.org/10.1075/tblt.14

Garcia, C., Nickolai, D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction: An empirical study. *Calico Journal, 37*(3), 213-232. https://doi.org/10.1558/cj.40379

Khademi, H., & Cardoso, W. (2022). Learning L2 pronunciation with Google Translate. *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022* (pp. 228-233). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1463Papin

Papin, K., & Cardoso, W. (2022). Pronunciation practice in Google Translate: focus on French liaison. *Intelligent CALL, granular systems, and learner data: short papers from EUROCALL 2022* (pp. 322-327). Research-publishing.net. https://doi.org/10.14705/rpnet.2022.61.1478

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, 32(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis: Focus on French liaison. *Computer Assisted Language Learning*, *30*(3-4), 325–342. https://doi.org/10.1080/09588221.2017.1312463

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Tsai, S.-C. (2020). Chinese students' perceptions of using Google Translate as a translingual CALL tool in EFL writing. *Computer Assisted Language Learning*, *35*(5–6), 1250–1272. https://doi.org/10.1080/09588221.2020.1799412

van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning & Technology, 26*(1), 1–19. http://hdl.handle.net/10125/73460

2023, Editorial Universitat Politècnica de València BY-NC-SA