# ASSESSING VARIABILITY IN PRONUNCIATION DEVELOPMENT: CASE STUDIES FROM A COMPLEXITY PERSPECTIVE

**Shannon R. Becker**, Northern Illinois University
**Jessica L. Sturm**, Purdue University

Learning one's target language in a foreign language environment presents particular challenges for both the amount of L2 input and the opportunity to practice. When one considers additionally how little time is dedicated to pronunciation in the typical foreign language classroom (Olson, 2014), it seems inevitable that pronunciation would lag behind other skills in this instructional context. To address this skills gap, the instruction and assessment of pronunciation need to be prioritized in research and practice. In this paper we revisit the results of a longitudinal study following four L1 English learners of L2 French, designed to explore how the limited L2 exposure and pronunciation instruction of a foreign-language environment may impact pronunciation development. The results are reframed here as evidence of the adaptation and change that are characteristic of a developing complex dynamic system, leading to a discussion of alternative ways to present and analyze pronunciation data that consider the inherent variability and non-linearity of linguistic behavior.

**Cite as:** Becker, S. & Sturm, J. (2022). Assessing variability in pronunciation development: case studies from a complexity perspective. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13261

## INTRODUCTION

In the brief time available to instructors in the foreign language classroom, skills must be prioritized and sacrifices made. Confirming the anecdotal experience of many FL instructors, Olson (2014) determined that, in the first two years of post-secondary foreign language instruction in the United States, just a few minutes per week are dedicated to pronunciation instruction. One of the challenges of pronunciation research, then, is to determine how this limited contact time affects pronunciation learning and what tools can be provided to instructors in order to augment pronunciation teaching. In this paper, we will briefly present the results of a longitudinal study designed to address the first of these concerns before moving on to a discussion of the usefulness of a complex dynamic systems theory[1] (CDST) approach in analyzing the resulting data.

### Motivation for the original study

In her 2019 longitudinal study, Sturm assessed the pronunciation development of students without explicit instruction in the skill, i.e. in a typical post-secondary foreign language classroom in the United States. Several threads of research underpinned this study, the first of which is the well-

[1] Though several terms such as complexity, complex systems, complex adaptive systems, and dynamic systems have been used in the literature to refer to a particular set of characteristics as applied to language development, recent studies seem to have settled on Complex Dynamic Systems Theory (CDST), so that is the wording I have chosen here as well. See Fogal, G.G. (2022), Larsen-Freeman, D. (2019), and Stotz & Cardoso (2022) for examples.

documented dearth of pronunciation instruction in the foreign language classroom (Hannahs, 2007; Huensch, 2019; Olson, 2014; Saalfeld, 2011) Given the relatively low amount of time and attention afforded to pronunciation in the FL classroom in combination with the importance of intelligibility to successful communication (Derwing & Munro, 2009; Levis, 2005) as well as the documented effectiveness of explicit pronunciation instruction (Gordon & Darcy, 2016; Lord, 2005; Miller, 2012; Offerman & Olson, 2016; Sturm, 2013), the aim of this study is to determine a baseline for pronunciation development in order to offer suggestions for improvement.

**Complex dynamic systems theory analyses in second language acquisition**

There are several compelling reasons to use complex dynamic systems theory when analyzing second language development. To begin with, CDST foregrounds variability in the data, allowing us to deal in the reality of individual development rather than obscuring it in group averages. While it can be true that certain types of variability in empirical data are indicative of methodological problems, variability can be a sign of dynamic language development and should therefore be carefully analyzed in order to understand the full picture. Focusing on variability and change in this way, it becomes increasingly important to visualize data in ways that probe the nonlinearity of learning processes. In other words, rather than relying solely upon linear representations of language behavior drawn from group averages, it may be necessary to include detailed analyses of individual data in order to flesh out the reality of language development as it happens. This means that qualitative data may have an important role to play in clarifying the nonlinear behavior shown in quantitative analyses by providing additional context regarding *how* language behavior evolves over time.

To this point, CDST analyses of second language acquisition have fallen into one of three categories: those that dive deeply into the theoretical considerations of applying this paradigm to language learning (Cameron & Larsen-Freeman, 2007; Larsen-Freeman, 1997; Van Geert, 2008), those that tackle the problem of how to adjust our methodological tools to better reflect complexity in the data (van Dijk, Verspoor, & Lowie, 2011), and, less often, those that address individual language skills from a CDST perspective. Of these individual skills, studies have been performed on L2 writing (Baba & Nitta, 2014; Yang & Sun, 2015), morphosyntax (Lenzing, 2015), oral proficiency (Larsen-Freeman, 2006; Opitz, 2012) and syntax (Verspoor, Lowie, & Van Dijk, 2008), among others. To this point, research into these discrete L2 skills seems to be the least common category within L2 CDST literature, with research into the skill of pronunciation being particularly limited.

One available study that addresses phonological attainment using a complex adaptive systems model (Aslan, 2017), is in fact a case study of one individual's attitudes and perceptions regarding his own phonological development. The participant was considered an expert level English learner and, as reported by the author, the study focused on "...his own self-perceived attainment as evidenced by his interactions with other speakers." This was accomplished through semi-structured interviews and reflective essays. Units of meaning were then drawn from the essays and interviews and put into categories that correlate with aspects of complex adaptive systems. Through a detailed analysis of this data, the author concluded that learner perceptions are important in the acquisition process and that the native speaker norm should be rejected.

As a step toward applying the CDST paradigm directly to pronunciation research, the intent of this paper is to demonstrate the usefulness of the paradigm in exploring individual development in production over time. By using CDST to re-analyze data from a traditionally designed longitudinal study, we aim to determine whether pronunciation does indeed follow the trajectory of a complex dynamic system, as evidenced by the variability and nonlinearity of individual performance.

### Research questions

The research question for the original study (Sturm, 2019) was: how does the ability to pronounce various French sounds develop in beginning L2 learners without explicit pronunciation instruction? In the current paper, we were interested in digging deeper into this data to show the messy reality of pronunciation development for individual learners. The research question guiding this paper was thus: how does individual development on two specific phonemes compare with the overall performance of the group?

## METHODS

The Sturm (2019) study followed students over four semesters of typical foreign language instruction, in this case the first four semesters of French language instruction. Only those participants who completed at least three of the four semesters and all of the audio recordings were included in the analysis. This resulted in a small sample size of four students, a number that is not ideal for typical quantitative analyses but that is appropriate for a case study. In order to confirm the reliability of the patterns found in the data of these four main participants (labeled as 4, 32, 80, and 89 according to their randomly assigned numbers in the original dataset), the data of 53 students who completed two semesters were also included in some parts of the analysis.

To ensure consistency of sounds and syllables across participants, a reading task was chosen and was completed at the beginning and end of each semester. The text used for the reading task consisted of a short discussion of the importance of bread to the French diet (144 words) followed by two short sentences containing many French vowel sounds (32 words). In the initial analysis, a count of pronunciation errors was conducted, with the score based on the number of incorrect syllables. Intra-rater reliability was established by repeating the coding several months later on ten randomly chosen files, revealing 91.4% agreement. Subsequent qualitative analysis described in detail the nature of the observed errors.

### Complex dynamic systems theory analysis of Sturm (2019) data

The first step of the CDST analysis was to evaluate nonlinearity in the original data by computing linear trendlines to visually represent the way results are sometimes presented in SLA research. Comparing this to the ups and downs of the individual data serves to disrupt the image of smooth, consistent behavior, and instead emphasizes the progression and regression that occur in individual learning.

The second step was to zoom in from this general picture of development and conduct an additional quantitative analysis of the pronunciation performance of the four participants on two specific phonemes. Focusing more narrowly on certain phonemes at this stage allowed for a more detailed analysis of how those segments were produced and to closely evaluate variation both within the same text and over time. For this analysis, we chose the two monosyllabic words that were most prevalent in the text of the reading task. These were the words *pour* and *pain*, whose target productions are /puʀ/ and /pɛ̃/, respectively. Because the subject of the text was the importance of bread to French people, *pain* was the most frequently used word, appearing nine times. *Pour*, on the other hand, is one of the most common prepositions, and appeared in the text five times. This gave plenty of tokens all within the same phonetic context.

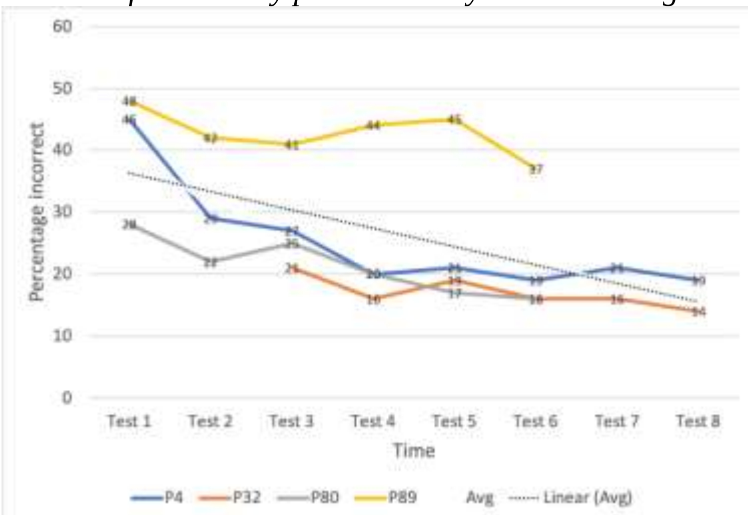
This second analysis focused on the vowels /u/ and /ɛ̃/. Generally speaking, the phoneme /p/ does not represent difficulty for L1 English learners of L2 French aside from the difference in aspiration, which itself does not impede comprehension. Additionally, in the context of the word *pour*, the French /R/ is less salient than it would be in word-initial position. For these reasons, we chose to focus on the vowels. Recordings were re-analyzed for each of the four participants, and the target phonemes were marked as correct or incorrect, following the procedure of the original study.
Since there were many fewer tokens in this analysis compared to the original, which contained 249 syllables, the incorrect sounds in the new analysis were plotted as group percentages. Each participant's overall performance was converted to a percentage as well, in order to accurately compare trajectories. These data were plotted, and additional qualitative assessments were included in order to describe the nature of the errors.

**RESULTS**

In the first part of the analysis, re-evaluating the original data, trendlines were added to the original data of the four participants in order to evaluate the nonlinearity of individual progress. This is demonstrated in Figure 1, which shows the number of incorrectly produced syllables across the eight tests (two per semester for four semesters) of the original study. The dotted trendline falls generally downward, indicating a drop in error counts over time; however, it is apparent in the data of the individual participants that this does not capture the full picture of their development. While they ended up producing fewer incorrect syllables at the time of their last test, their performance did not consistently improve over time. The additional qualitative analysis confirmed this, showing that the errors they produced were likewise inconsistent, i.e. a given token may have been produced correctly on one test and incorrectly on the subsequent one (see Table 1 below).

**Figure 1**

*Number of incorrectly pronounced syllables over eight tests (two per semester)*




This pattern of a general downward trend combined with dynamic individual performance is echoed in the data of the group of 53 students who completed at least two semesters. Additionally, a plateau can be seen from the second to third semester as well as from the third to fourth semester, indicating that while pronunciation may improve slightly when students begin learning French, the improvement tapers off without explicit pronunciation instruction.

**Figure 2**

*Number of incorrectly pronounced syllables from semester 1 to semester 2.*

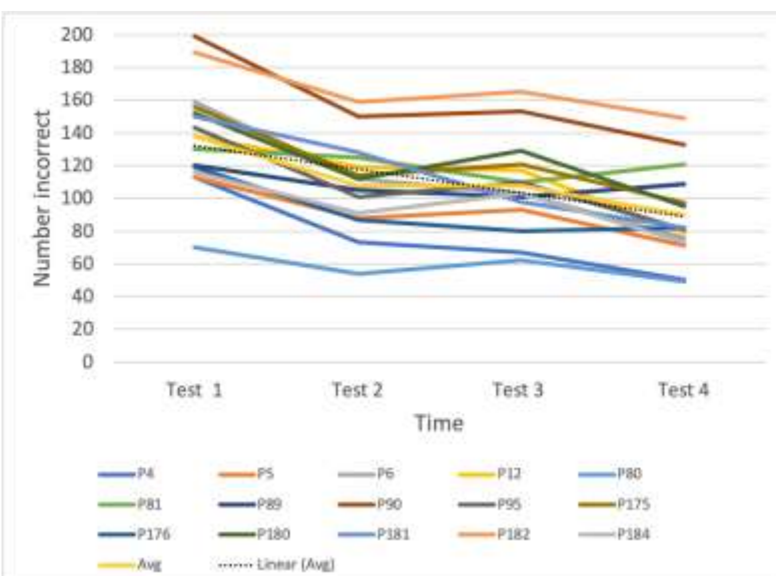


**Figure 3**

*Number of incorrectly pronounced syllables from semester 2 to semester 3.*

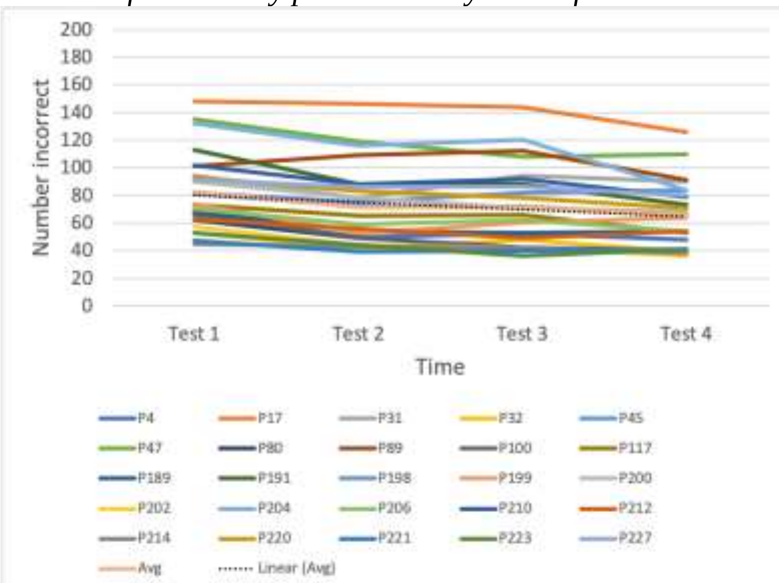
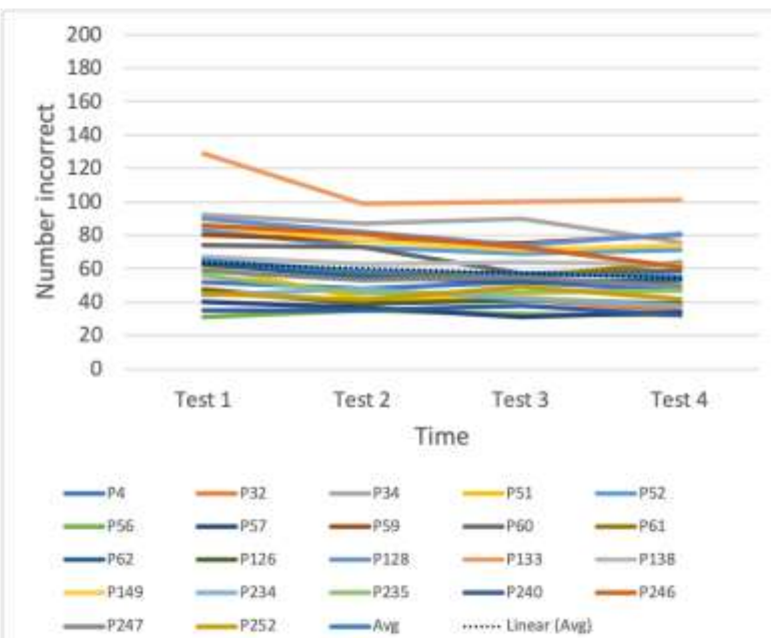



**Figure 4**

*Number of incorrectly pronounced syllables from semester 3 to semester 4.*




The above comparisons of individual trajectories versus group averages serve to highlight the importance of individual variation over time as an indicator of behavior change. Even these visualizations, however, tend to obscure individual pronunciation development, since they group the accuracy of all sounds together. In order to further investigate this variability, we proceeded to the second part of the analysis wherein we conducted a more detailed analysis of two specific phonemes.

This second step of the CDST analysis involved re-analyzing the data of the four main participants in order to compare their trajectories of development on two specific phonemes, the vowels contained in the words *pour* and *pain*, to that of their overall development. The error rates for participants #4, 32, 80, and 89 on each of these phonemes are presented in Figures 5 through 8.

**Figure 5**

*Target phonemes, participant 4.*

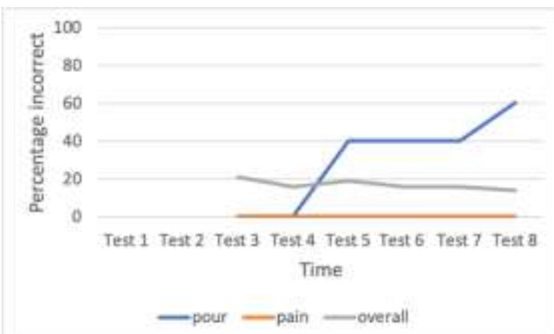



**Figure 6**

*Target phonemes, participant 32.*

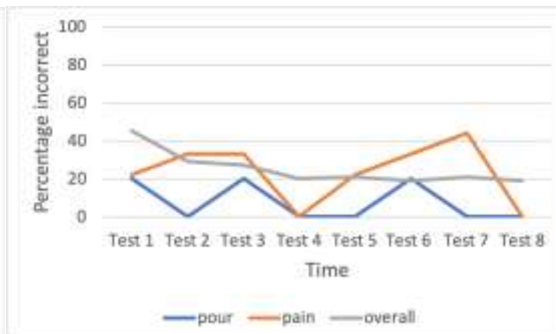



**Figure 7**

*Target phonemes, participant 80.*

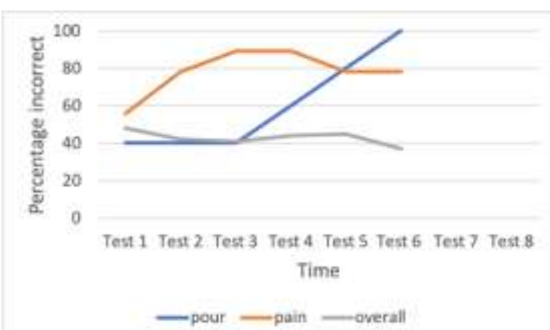
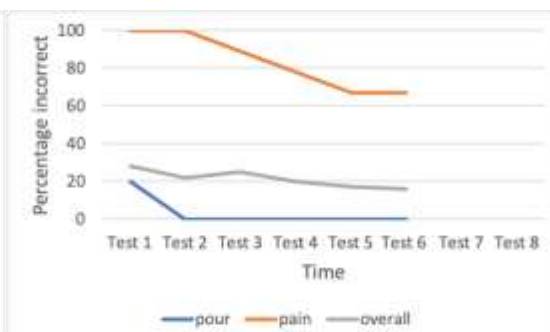



**Figure 8**

*Target phonemes, participant 89.*




The blue lines in Figures 5-8 represent the overall percentage of incorrect syllables over the course of the original study, while the orange and gray lines represent the error rate on the vowels in *pour* and *pain*, respectively. The difference between global performance and performance on the target phonemes is striking. Specifically, it is apparent at a glance that the modest decrease in errors over time represented by their overall performance was not echoed in the production of the two sounds analyzed here. Participant 4, for example, displays the progression and regression typical of development. Participant 32, on the other hand, saw their error rate on *pour* increase over time while their correct production of *pain* remained stable.

The final step in the CDST analysis was a qualitative description of each participant's errors on the target phonemes, as summarized in Table 1. These observations echo those from the original analysis; namely, that production of the target phonemes was highly inconsistent. For instance, where *pour* was produced five times in the text, it was not produced correctly all five times. Even if production improved in subsequent tests, it was not necessarily correct on the same tokens as in the previous test. Additionally, there was not consistency in the same participants across tokens. For example, while participant 32 displayed stable production of the nasal vowel /ɛ̃/, their production of the oral vowel /u/ actually deteriorated over time and was not stable across tokens. Likewise, while participant 89 saw an increase in errors for the *pain* tokens, their production of *pour* improved drastically from the beginning to the end of the third semester.

**Table 1**

*Qualitative analysis of vowel production in pour and pain by participant.*

| | **P4** | **P32** | **P80** | **P89** |
|---|---|---|---|---|
| pour /puʁ/ | Mostly correct; where errors occurred, they were different tokens each time | Incorrect production increased over time; tokens varied | Highly stable production of /u/ phoneme | Very rapid speech resulting in replacement of /u/ with /ə/ |
| pain /pɛ̃/ | Production is unstable; different tokens produced incorrectly over time | Production of this nasal vowel is highly stable | Production of /ɛ̃/ improved slightly over time; correct production varied by token | Occasionally correct; variable by token with one exception |

## DISCUSSION

Setting up the contrast between group averages and individual results, as visualized above in the figures with added trendlines, is helpful in the initial data analysis in order to clarify that development is much messier and less linear than group averages would have us believe. In the case of the longitudinal data presented in this paper, this strategy helps expose the "one step forward two steps back" nature of language learning. When it is further broken down into results from semester to semester, development is clarified to a greater degree. In Figures 2-4, the individual lines show more of a plateau by the end of the four semesters than would be indicated by the overall averages for each participant.

Breaking the results down to the level of individual phonemes provides an even more stark visual representation of the nonlinearity of pronunciation development, allowing us to see in detail how the production of certain segments varies in accuracy both over time and within a given text. In essence, the analysis has continued to zoom in further at each step in order to decipher what is happening at the most basic levels of the learner's developing language system.

Though we have spent a good deal of time in this paper referring to the importance of variability and of recognizing the nonlinearity of language development, analyses such as those presented here have the potential to elucidate more advanced CDST concepts as well. For example, as complex dynamic systems move through all their possible states, they sometimes end up in what are referred to as attractors or attractor basins, which are stable states that require an influx of energy for change to occur. This could be the case when a learner is consistently producing errors in a certain segment or in certain phonological contexts. The energy needed to escape this attractor state could be provided by explicit pronunciation instruction.

In addition to giving us innovative ways to frame and talk about language development, applying CDST to second language acquisition also provides us with new and different ways of analyzing data both quantitatively and qualitatively. The visualizations outlined in this paper are a good first step in analyzing data and prioritizing context. There are also, however, several descriptive and statistical methods, as well as modeling and simulations, that can sometimes be applied to data in a longitudinal study such as this. Examples of such analyses include moving min-max graphs and resampling techniques such as the Monte Carlo method (van Dijk, Verspoor, & Lowie, 2011). These methods weren't applicable to the current study because of a lack of *density* in the data; another term for this is microgenetic analysis. Both of these terms mean that testing is done at short intervals of time in order to capture development more accurately as it happens. Future studies will be designed with testing at regular intervals in order to take advantage of the statistics afforded by a deeper CDST analysis.

What we have observed in this paper is that there are many ways in which traditional data analyses fail to show the reality of development. A CDST theory perspective encourages us to embrace that reality and look at our data in ways that explain actual development. Far from muddling the picture, these analyses give us the ability to visualize data in ways that value variability and nonlinearity for the information they provide. Rather than indicating a lack of progress, noise in the data, or faulty design, this highly variable information can be indicative of positive change.

## ABOUT THE AUTHORS

**Shannon R. Becker** (Ph.D., 2015, Purdue University) is Associate Professor of French Linguistics at Northern Illinois University. Her research interests include L2 French pronunciation and listening comprehension development, complex/dynamic systems approaches to second language acquisition, and L2 pedagogy.
Email sbecker@niu.edu

Northern Illinois University
Watson Hall 115
1425 W. Lincoln Highway
DeKalb, IL 60115
219-765-0944

**Jessica L. Sturm** (Ph.D., 2008, University of Illinois at Urbana-Champaign) is Associate Professor of French and Applied Linguistics at Purdue University where she serves as Chair of French and supervises the French Language Program. Her research interests include classroom second language acquisition, acquisition of L2 French pronunciation, and technology in the L2 classroom.
Email jsturm@purdue.edu

Purdue University
640 Oval Drive
West Lafayette, IN 47907
Stanley Coulter 186
765-496-9692

**REFERENCES**

Aslan, E. (2017). Doing away with the 'native speaker': A complex adaptive systems approach to L2 phonological attainment. *The Language Learning Journal, 45*(4), 447-465.

Baba, K. & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning, 64*(1), 1-35.

Cameron, D. & Larsen-Freeman, D. (2007). Complex systems and applied linguistics. *International Journal of Applied Linguistics, 17*(2), 226-240.

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.

Fogal, G. G. (2022). Second language writing from a complex dynamic systems perspective. *Language Teaching, 55*, 193-210.

Gordon, J. & Darcy, I. (2016). A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation, 2*, 56-92.

Hannahs, S. J. (2007). French phonology and L2 acquisition. In D. Ayoun (Ed.), *French Applied Linguistics* (pp. 50-74). Amsterdam and Philadephia: John Benjamins.

Huensch, A. (2019). The pronunciation teaching practices of university-level graduate teaching assistants of French and Spanish introductory language courses. *Foreign Language Annals, 52*, 13-31.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590-619.

Larsen-Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal*, *103,* 61-79.

Lenzing, A. (2015). Exploring regularities and dynamic systems in L2 development. *Language Learning*, *65*(1), 89-122.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39,* 369-377.

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania, 88,* 257-267.

Miller, J. S. (2012). Teaching French pronunciation with phonetics in a college-level beginner course. *NECTFL Review*, *69*, 47-68.

Offerman, H. M., & Olson, D. J. (2016). Visual feedback and second-language segmental production: The generalizability of pronunciation gains. *System, 59*, 45-60.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*, 47-68.

Opitz, C. (2012). A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *International Journal of Bilingualism, 17*(6), 701-715.

Saalfeld, A. K. (2011). Acquisition of L2 phonology in advanced learners: Does instruction make a difference? In J. Levis and K. LeVelle (Eds.) *Proceedings of the 2nd pronunciation in language learning and teaching conference* (pp. 144-152). Ames, IA: Iowa State University.

Stotz, Q., & Cardoso, W. (2022). Applying complex dynamic systems theory to identify dynamic properties of plurilingual repertoires. *TESL Canada Journal, 38*(2), 140-170.

Sturm, J. L. (2019). Current approaches to pronunciation instruction: A longitudinal case study in French. *Foreign Language Annals*, *52*(1), 1-13.

Sturm, J. L. (2013). Explicit phonetics instruction in L2 French: A global analysis of improvement. *System*, *41*, 654-662.

van Dijk, M., Verspoor, M., & Lowie, W. (2011). Variability and DST. In M. Verspoor, K. de Bot, and W. Lowie (Eds.) *A Dynamic Approach to Second Language Development* (pp. 55-84). Amsterdam and Philadelphia: John Benjamins.

Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, *92*(2), 179-199.

Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, *92*(2), 214-231.

Yang, W. & Sun, Y. (2015). Dynamic development of complexity, accuracy and fluency in multilingual writers' L1, L2 and L3 writing. *Theory and Practice in Language Studies*, *5*(2), 298-308.