Volume 34 Issue 1 2025

Applied Language Learning

A Practice-Based Journal for Foreign Language Educators

Defense Language Institute
Foreign Language Center
Presidio of Monterey

Applied Language Learning

The mission of Professional Bulletin 65, ***Applied Language Learning*** (US ISSN 1041-679X and ISSN 2164-0912), is to provide a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences.

Applied Language Learning, published by the Defense Language Institute Foreign Language Center and Presidio of Monterey, California, presents professional information. The views expressed herein are those of the authors, not the Department of Defense or its elements. The content does not necessarily reflect the official US Army position and does not change or supersede any information in official US Army publications. ***Applied Language Learning*** reserves the right to edit material.

To access *Applied Language Learning* online:
https://www.dliflc.edu/ojs/ALL/index

To communicate with the Editor:
Send correspondence to the Editor at journals@dliflc.edu

CONTENTS

ARTICLES

BOOK REVIEWS

GENERAL INFORMATION

Applied Language Learning

Research Article

The Importance of Placement Tools in Postsecondary World Language Classes

Pete Swanson
United States Air Force Academy

Jean W. LeLoup
United States Air Force Academy

Appropriate placement of language learners continues to be challenging for postsecondary world language programs. The Seal of Biliteracy is now available in all of the states for high school students. While the Seal is an admirable program, using it in higher education can be problematic because the proficiency levels needed to attain the Seal vary by state, and not all world language programs participate in the Seal of Biliteracy. The purpose of the present study was to compare cadets' proficiency levels in French, Portuguese, and Spanish at the beginning of their second year of study to their proficiency level as reflected by their Seal of Biliteracy achievement/placement. The researchers assessed language learner proficiency using the Adaptive Listening Tests and the Adaptive Reading Tests developed at Brigham Young University. Results indicate that reporting attainment of the Seal of Biliteracy did not necessarily guarantee results at higher levels of proficiency in all three languages. Use of the Seal of Biliteracy alone for placement purposes is thus not recommended as the best assessment of students' language abilities. The findings have implications for world language programs in higher education as they find new ways to place secondary students in their programs.

Keywords: *World Language, Student Placement, Seal of Biliteracy, Quantitative Research, Military Academy*

INTRODUCTION

Having proficiency in world languages (WLs) is crucial for America's military, diplomatic, business, and national security interests (La Corte & Voisine, 2020). The American Council on the Teaching of Foreign Languages (ACTFL) reported several years ago that 90% of businesses surveyed needed employees with skills in languages other than English (ACTFL, 2019). However, "the vast majority of American citizens remain monolingual" (American Academy of Arts and Sciences, 2017, p. vii). Unfortunately, there has been a decrease in student enrollment in WL classes. Recently, Lusin et al. (2021) reported that there was a 29.3% decrease in enrollment between 2009 and 2021; however, before this major downturn, "there had been sustained growth" (p. 3). In part, the student enrollment decline can be attributed to the elimination of WL requirements for high school and postsecondary graduation (Lusin, 2012; Maranon, 2024). According to O'Rourke et al. (2016), only seven states require the study of WLs as a prerequisite for high school graduation. Without high school requirements to study a WL to boost course enrollments, it will be challenging for the U.S. to meet the needs voiced by the military and businesses.

In addition to the shortage of students enrolled in WL classes, there has been a shortage of WL teachers for decades (Swanson, 2008, 2010, 2012, 2022), which led to calls to replace WL learning as a requirement for high school graduation in favor of computer programming (Galvin, 2016). The No Child Left Behind Act (2008) and subsequent reauthorizations of the Elementary and Secondary Education Act (United States, 1965) have both been cited as a factor of the WL teacher shortage (Swanson, 2008, 2012) because these acts require teachers regardless of content area to be deemed *highly qualified*[1]. The requirement was highly problematic because many WL teachers who were once licensed to teach were no longer qualified in the eyes of the federal government at a time of a national WL teacher shortage (Swanson & Moore, 2006). Additionally, No Child Left Behind prioritized instruction and the allocation of resources to the core areas such as science and mathematics (Rosenbusch, 2005; Rosenbusch & Jensen, 2004), which reduced funding to support and enhance WL teaching and learning for US K-12 students.

In order to thwart the shortage of WL teachers, states began to authorize alternative routes to teaching, which provided "woefully inadequate training and support to their [teacher] candidates" (Walsh & Jacobs, 2007, p. 14). Typically, these individuals have not participated in student teaching or been exposed to classroom management techniques, which leads to feeling unprepared and overwhelmed (Redding & Smith, 2016). Research continues to show that for individuals entering the teaching profession through alternate routes, such as emergency certification, teacher turnover is higher than attrition for teachers who enter the progression via full certification (Darling-Hammond et al., 2001; Guthery & Bailes, 2022; Nguyen et al., 2020; Sawchuck, 2016; Spencer, 2019). Rates of attrition for uncertified teachers range between 30% (Reaching Higher NH, 2024) and 60% (Morrison, 2024) within five years compared to 15% for certified teachers (Reaching Higher NH, 2024). Research shows that teacher attrition, regardless

[1] A highly qualified teacher "is one who (1) has a bachelor's degree, (2) is fully certified, and (3) has demonstrated subject-matter competency in each of the academic subjects that she or he teaches" (U.S. Department of Education, 2009, p. 11).

of content area, educational level, and teacher certification status costs school districts between $7 billion and $8.5 billion annually (Breakfast Leadership, 2023; Sutcher et al., 2016).

Not surprisingly, the lowering of professional standards for teachers by offering alternative routes to certification, thereby allowing poorly qualified teachers to enter the profession, can hinder student academic performance. Decades of research continue to show that "fully certified and experienced teachers matter for student achievement" (Cardichon et al., 2020, p. 1). For example, findings from a large-scale study of New York City teachers revealed that student achievement growth in elementary and middle school was most improved by having been taught by a fully certified teacher who graduated from a university-based preservice teacher education program (Boyd et al., 2006). Interestingly, the same study showed that students' achievement was hurt most by having an inexperienced teacher on a temporary license. Other studies have corroborated the aforementioned findings showing that teachers with alternative certification had a negative impact on student performance, particularly at the high school level (Clotfelter et al. 2010; Henry et al., 2014). Teaching is a complex endeavor, and it is clear that "high quality teacher preparation makes a difference in student achievement" (Swanson, 2019, p. 763).

Review of the Literature

Many factors influence how quickly language learners progress along their interlanguage continuum toward proficiency in classes at the tertiary level. Four primary elements are student placement in language classes, teacher performance in the classroom and how it affects student learning, language learner application and dedication to acquiring and using a WL, and prior language learning by students prior to placement in a higher education, here reflected in the attainment or lack thereof of the Seal of Biliteracy. Each of these factors are discussed below.

Placing Postsecondary Students Appropriately in WL Classes

The primary goal of student placement systems is to sort "students into relatively homogeneous language-ability groupings, sometimes within specific skill areas" (Brown, 1989, p. 65). In terms of WL placement, such systems have the goal of assigning language learners in the most appropriate proficiency level. Lord (2022) noted that such a "straightforward goal may lead us to believe that placement is a relatively clear-cut process" (p. 101). Nevertheless, it is rather complex as multiple factors (e.g., K-12 teacher certification and curricula, student background) can play a role in one's placement in college level WL courses. Even though placement constitutes a considerable challenge for many postsecondary WL programs, there is a dearth of research on the topic.

Many placement processes focus on traditionally used indicators of student achievement such as standardized test scores, overall or cumulative grade point average, and even seat time in subject matter courses at the secondary level. The scant research involving subject matter placement at the tertiary level involves primarily the curriculum in math and English, rather than language courses. One common theme in these studies is that subject matter placement at the

tertiary level is complicated at best and often unsatisfactory for several reasons. One consensus is that use of traditional placement methods with its heavy reliance on standardized test results can disadvantage minority groups. This can also underrepresent or understate student capabilities, causing extended time of study for those students that have an additional cost in terms of monetary outlay as well as personal frustration (Bahr, et al., 2019; Denison-Furness et al., 2022; Mechur Karp, 2021).

A few studies have documented some new approaches to placement that show some promising results. These approaches reflect a move toward self-placement of some form or another, incorporating student input based on self-evaluation, academic expectations, and curricular objectives. Said approaches utilize a combination of concrete data in an advisory capacity, extensive communication between students and advisors, and consideration of one's own abilities and goals. Sometimes called merely self-placement, other times directed self-placement, these models empower students to make choices based on factors not previously considered or accepted as important (Denison-Furness et al., 2022; Mechur Karp, 2021). Perhaps the most rewarding outcome of these alternative approaches to placement is the acknowledgement of overall positive results vis-à-vis student grade outcomes in the classes in which they enrolled as a direct result of this approach (Lord, 2022; Mechur Karp, 2021). While there are some caveats with self-directed approaches (e.g., labor-intensity, consumption of time for advisors and students alike), the general results are encouraging and should be considered and further explored.

Given the different methods used to place language learners in appropriate levels of WL instruction, the literature indicates that teacher preparation and teacher effectiveness are directly related to student performance (Cardichon et al., 2020; Darling-Hammond et al., 2001; Darling-Hammond et al., 2005; Darling-Hammond & Wei, 2009; Swanson & Hildebrandt, 2018).

Teacher Performance

The question surrounding teacher effectiveness and how to measure it is a challenge (Hildebrandt & Swanson, 2016). Over the years states relied on students' standardized test scores and portfolio assessments (e.g., edTPA) of student achievement as well as encouraging teacher preparation programs to earn and maintain program accreditation. In WL teaching, the Council for the Accreditation of Educator Preparation (CAEP) partnered with the American Council on the Teaching of Foreign Languages (now known by its acronym, ACTFL) to offer WL teacher preparation programs national accreditation.

One of the six standards WL teacher preparation programs must meet for ACTFL/CAEP accreditation is Standard 1 Language Proficiency: Interpersonal, Interpretive, and Presentational. According to the guidelines, teacher candidates must possess a high level of proficiency in the language they teach. Additionally, teacher candidates must show prowess in other areas via the remaining five standards such as language acquisition theories, instruction and assessment, linguistics and cultures, and integration of the *World-readiness Standards for Learning Languages* (The National Standards Collaborative Board, 2015). Collectively, the six standards "reflect the

profession's expectations for the specific knowledge, skills, and dispositions that WL teachers should have as they enter our classrooms" (ACTFL, 2025a). However, only slightly more than a quarter of the WL teacher preparation programs are nationally certified (Moss & Gambrell, 2023).

From an instructional perspective, teacher candidates must demonstrate proficiency in standards-based, Communicative Language Teaching (CLT) approaches—the signature pedagogy in the field advocated for in the 1980s (Miller, 2018) as a response to the limitations of other methods (e.g., audio-lingual, grammar-translation). In an effort to move away from instructional *methods* per se (e.g., grammar-translation), CLT prioritizes developing students' communication and interaction skills by understanding that communication is the expression, interpretation, and negotiation of meaning, and not simply oral expression (Lee & VanPatten, 2003). Second language research has revealed valuable insights into the nature of language learning (e.g., Ellis, 1997, 2008; Krashen, 1982). Of the many understandings, CLT approaches appear to hold the most promise for acquiring a new language (VanPatten, 2017).

CLT is guided by multiple postulations such as (1) learning a second language can be facilitated through using the target language for communication purposes, (2) such communication should be both meaningful and authentic, (3) a tremendous emphasis should be placed on language use rather than language knowledge, (4) learner risk-taking and autonomy should be encouraged, and (5) fluency and appropriacy in the use of the second language should take precedence over structural correctness (Swanson et al., 2022). Unlike a pedagogical method, CLT includes a number of different techniques and does not follow a structured set of procedures that teachers should follow. While CLT has been promoted for approximately 40 years (Brumfit & Johnson, 1979; Widdowson, 1978), Duquette (1995) reported that CLT approaches were not being used in elementary and secondary classrooms. Several decades later, Troyan et al. (2023) found that veteran WL teachers working as external reviewers of WL teacher candidate edTPA portfolios—a nationally-available subject-specific pre-service teacher portfolio—incorrectly evaluated recent preservice teacher candidates' portfolios on Rubric 8, Subject-Specific Pedagogy. The researchers reported that "the candidates' level of expertise may have exceeded those of the raters" (p. 672), which suggests teachers who were not trained to teach from a communicative stance are teaching *about* the target language instead of teaching students to *acquire* the target language (Krashen, 1981).

To that end, as secondary students matriculate in institutions of higher education and enroll in WL classes either as a requirement or an elective class option, postsecondary instructors find themselves teaching students with a wide range of target language abilities. Subsequently, instructors are charged with developing curricula that challenge language learners and promoting language acquisition that is not boring those who are more advanced and not overwhelming students who lack basic skills.

Student Performance

Much like the complexity surrounding teacher effectiveness and how to measure it, evaluating student proficiency in the target language can be challenging. In order to place students in WL classes, WL program coordinators in postsecondary education have relied on results from program-specific placement tests, test scores from College Board (e.g., Advanced Placement Spanish), and, most recently, from students reporting that they earned the Seal of Biliteracy (Seal of Biliteracy, 2024).

Historically, Carroll (1967) began to examine postsecondary student proficiency outcomes (*N*= 2,782) of WL majors in French, German, Italian, Russian, and Spanish by using the Modern Language Association test, which was aligned to the government's Interagency Language Roundtable (ILR) scale, which predated the *ACTFL Proficiency Guidelines* (2012) by decades. He reported an average attainment of an ILR rating of 2+, which correlates to an ACTFL Advanced-Mid or Advanced-High proficiency rating. Additionally, he found that several demographic factors contributed to higher levels of proficiency such as heritage language background, study abroad, and elementary school WL study. However, there were no differences between males and females.

Since Carroll's (1967) study of postsecondary WL majors at graduation, program administrators, students and administrators alike have been challenged to establish reasonable expectations due to a dearth of research on WL learner proficiency. For example, Isbell et al. (2018) reported that four semesters of WL study at the university level yielded an oral proficiency outcome of Intermediate-Low to Intermediate-Mid in Chinese, French, Russian and Spanish. Isbell et al. (2018) examined postsecondary student oral proficiency. Results indicated a range of outcomes for WL learners across different institutions, in different levels of courses, and with different backgrounds. The researchers reported that study abroad programs had a positive impact on student proficiency in the areas of listening, reading, and speaking.

With respect to WL learner proficiency examined in the receptive skills (i.e., listening and reading), Tschirner (2016) identified average outcomes of WL students after two, three, four, five and six semesters of study of more than 6,000 language learners. Of the subjects who took both the ACTFL Reading Proficiency Test and the ACTFL Listening Proficiency Test in Spanish (*N*=1,769), second semester learners were found to reach Intermediate-Low in reading and just below Novice-High in listening, while fourth-semester learners reached Intermediate-Mid in reading and almost Intermediate-Low in listening. Swanson et al. (2022) reported similar results for second year language learners of Spanish for listening, and the authors recommended that "there is a strong need to set benchmarks for language proficiency" (p. 2) in order to understand what can be attained after specific sequences of WL study. Clearly, setting proficiency benchmarks can inform individual language learner proficiency and program success as well as determine what program innovations are leading to higher levels of student proficiency. Contrarily, if such targets are not determined and confirmed via valid and reliable assessments, programs may find themselves without a common mission and vision for their language learners. Without such a

vision, the appropriate placement of language learners transitioning from secondary to postsecondary WL programs becomes very challenging.

Reiterating the notion that teacher preparation and effectiveness are directly related to student performance, the researchers believe it is important to investigate if the Seal of Biliteracy can be used as a tool to place students in WL classes at the tertiary level.

Seal of Biliteracy

Reasonable expectations of WL proficiency for language learners to reach after a specific learning sequence of study has challenged the field for decades (Swanson et al. 2022). Since the 1960s, educational stakeholders (e.g., instructors, program directors) have grappled to establish reasonable proficiency benchmarks after various sequences of study (e.g., first year, second year) (Carroll, 1967). However, an initiative started in 2011 in California, and now in all 50 states, holds promise to help postsecondary programs know the level of proficiency students have achieved. Designed to represent an attainment of bilingualism and biliteracy for future employers and universities (ACTFL, 2015), the Seal of Biliteracy is an award adhered to high school diplomas by a state department of education or local district to recognize students who have attained proficiency in English and in one or more other WLs (ACTFL, 2015).

School districts and states determine which WL proficiency assessments (e.g., ACTFL Assessment of Performance toward Proficiency in Languages, Standards-based Measurement of Proficiency, STAMP4S) are acceptable. Proficiency levels for attainment of the Seal of Biliteracy vary widely by state. For example, in Illinois, recipients of the Seal of Biliteracy must demonstrate Intermediate-High proficiency (Illinois Council on the Teaching of Foreign Language, n.d.) based on the ACTFL Proficiency Scale (Language Testing International, n.d.) while South Carolina offers students three tiers of the Seal of Biliteracy based on WL proficiency ranging from Intermediate-Mid (bronze) to Advanced-Low (gold) (South Carolina Department of Education, 2019).

Among the manifold reasons for having a Seal of Biliteracy program is that it can serve to encourage students to study languages, certify attainment of biliteracy skills, recognize the value of language diversity, and prepare students with 21st century skills that will benefit them in the labor market and the global society. With respect to studying WL in higher education, the Seal of Biliteracy can "provide universities with a method to recognize and give credit to applicants for attainment of high-level skills in multiple languages" (Seal of Biliteracy, 2024, p. 1).

Given the importance placed on proficiency by the Seal of Biliteracy, one outcome has been noted as a change in methods of instruction and assessment. With increased emphasis on student learning being assessed by levels of proficiency reached, teaching strategies have adapted accordingly. As a result, student enrollment and retention in language classes have also increased as students are actually learning the language by using it rather than by just learning about it (Davin et al., 2018). Although positive outcomes do appear occasionally, the Seal of Biliteracy does not appear to be a panacea for student placement in language classes at the tertiary level. Because of the wide variety of requirements, implementation, and execution of the

Seal of Biliteracy across all 50 states, it does not have just one meaning or significance. Thus, while laudatory, the Seal of Biliteracy is not necessarily a viable (e.g., reliable and valid) measure of a student's language proficiency (Davin et al., 2022).

Nevertheless, advancing the notion that knowing students' level of WL proficiency upon matriculation into an institution of higher education may be helpful to accelerate student proficiency in the target language(s) by placing language learners at the correct level of WL learning, the following research questions guided the present study:

1. What was the level of proficiency in reading and in listening at the beginning of the academic year for second year language learners of French, Portuguese, and Spanish at a U.S. military institution?
2. For participants in the aforementioned three languages, is there a difference in level of proficiency for those who earned the Seal of Biliteracy and those who did not?

METHODS

Study Context

The United States Air Force Academy is a unique institution of higher education as it is a four-year military academy that prepares the next generation of Air Force officers. Cadets must take two semesters of the same language or validate the credit as part of the Academy's core curriculum. Cadets take first-year WL classes five days per week instead of a typical Monday-Wednesday-Friday or Tuesday-Thursday collegiate offering. WL classes at the second-year level and above meet every other day during the semester. Unlike civilian institutions, the cadets must participate in military and physical training activities daily and are evaluated on performance in three areas (academic, military, physical) each semester. Professors and instructors created Roadmaps to Proficiency and set proficiency benchmarks for cadets to attain at the end of each academic year for each level of instruction of the eight languages taught (Arabic, Chinese, French, German, Japanese, Portuguese, Russian, and Spanish).

Upon completion of the first year of language study for French, Portuguese, and Spanish, cadets are expected to reach between Novice-High and Intermediate-Low. At the end of the second year of study, cadets are expected to reach between Intermediate-Low and Intermediate-Mid. Upon completion of their third year of study, cadets are expected to reach between Intermediate-Mid and Intermediate-High. In order to receive a minor in the language, cadets take a total of five classes above the first-year requirement. At the end of the minor, the proficiency benchmark is set in a band from Intermediate-High to Advanced-Low. To determine their proficiency level, cadets take the Adaptive Listening Test (ALT) and the Adaptive Reading Test (ART). Additionally, they are encouraged to take the Defense Language Proficiency Test (DLPT). If they reach a rating of 2 in either modality of the DLPT (reading or listening), they can take the Oral Proficiency Interview, which gives them a rating for speaking—the gold standard in language acquisition.

With respect to the present study, the benchmark proficiency level for first-year cadets at the end of the academic year in French, Portuguese, and Spanish is a band between Novice-High and Intermediate-Low.

Instruments

The ALT and the ART are valid and reliable proficiency assessments (Clifford & Cox, 2013; Cox & Clifford, 2014) aligned with the ACTFL Proficiency Guidelines (ACTFL, 2012, 2024). These two assessments measure general language proficiency instead of what is learned in a WL class (ACTFL, 2025b). Both tests assess one's level of proficiency between the Novice-Low and Advanced-Low level. Each test item consists of either an authentic reading text or audio passage and one multiple-choice question with one correct answer associated with the text or passage. Each test can take up to 90 minutes and is adaptive. That is, the number of test items to which individual test takers respond will vary, depending on performance. Test items are taken from a pool of test items at specific proficiency levels that assess a broad range of topics including everyday life, current events, education, among others. As students begin to fail at a certain proficiency level, the test concludes (Center for Language Studies, n.d.).

Upon completion of a test, a floor rating (the level at which the test taker has demonstrated sustained performance) and a ceiling rating (the level at which the test taker has demonstrated patterns of breakdown) are computed and assigned as test takers receive separate ratings for reading and listening. Results from the ALT and ART can be helpful for a multitude of purposes such as the placement of higher education students in an appropriate course, measuring proficiency at certain points of the curricula, and informing program evaluation (Center for Language Studies, n.d.).

Procedures

Following Institutional Review Board (IRB) approval for human subjects testing, participants in second year French, Portuguese, and Spanish were administered the ART and the ALT in the respective languages during the first week of instruction in early August 2024 in their respective classrooms. At the end of the first day of class, participants were asked to fill out a participant demographic sheet online using *Google Forms*. The following two days, the participants took both the ALT and ART. Per IRB guidelines in order to avoid bias, instructors in the three languages proctored each other's classes. Personnel from the department's language lab also helped proctor the two tests. Results from the two APPT assessments were merged with the participant demographic information into a single spreadsheet using *Microsoft Access*. The spreadsheet was imported into *SPSS 18* for data analysis.

Participants

One hundred thirteen individuals in second-year French (*n*=44), Portuguese (*n*=16), and Spanish (*n*=53) participated in the present study. Second-year students were chosen since that is when

students start a WL minor. Of the 113 students who participated, only 95 participants took both the ALT and ART due scheduling conflicts; their proficiency data are reported here. Some participants (*n*=14) voluntarily chose not to fill out the demographic survey, which was permitted by IRB guidelines for participation in a study. With respect to the 99 individuals who completed the demographic survey, the mean age of participants was 19.11 years. Thirty-nine percent self-reported as female (*n*=44) where 49% self-reported as males (*n*=55). The majority of the participants identified as being either Caucasian (68%) or Latinx (9%). The remainder self-reported as African American (4%), Asian (3%), or multiracial (4%). Eighty-six participants reported not being a heritage learner of the WL they were studying. When asked about their previous WL classes, 73% reported *learning* about the language (Krashen, 1982). That is, they were taught in English about the grammar and rarely used the target language in class. Twelve percent reported that they were *acquiring* the target language by both teachers and students using the target language in the classroom (Krashen, 1982). The remaining 15% reported a mixture of learning the language and trying to acquire the target language. All of the participants were placed in second-year WL courses due to their performance on the departmental placement exams.

When asked about when they had stopped taking classes in the language they were now studying, 10% reported that they stopped immediately after their freshman year of high school. Eighteen percent reported having stopped immediately after their sophomore year. Twenty-three percent and 37% had stopped immediately after their junior and senior year respectively. Of the 99 participants that filled out the demographic survey, 17% reported having earned the Seal of Biliteracy from their high school. The standards for receiving a Seal of Biliteracy varies from state to state, and none of the participants could report the criteria for earning the seal. No participants reported having dual enrollment (college) credit for Spanish.

FINDINGS

The researchers collected baseline-testing data on cadets studying second-year French, Portuguese, and Spanish at the United States Air Force Academy in the fall of 2024 according to results of departmental placement testing. It is important to keep in mind that the proficiency benchmark before the participants tested was a band between Novice-High and Intermediate-Low.

With respect to the first research question about the level of proficiency in reading and in listening for language learners of French, Portuguese, and Spanish at the start of the second year of language learning, the researchers analyzed the proficiency data using frequencies for each of the ACTFL sublevels (e.g., Novice-High, Intermediate-Low). Table 1 shows that participant ratings in the three languages for listening proficiency (*N*=95) ranged between Novice-Low and Intermediate-High. Closer examination shows that 53% (*n*=50) of the participants in the three languages received proficiency ratings between Novice-High and Intermediate-Low for listening, meeting the benchmark for the end of the first year of study. Twenty-six percent of the participants (*n*=25) were below the benchmark with ratings in the Novice-Low and Novice-Mid

sublevels on the ACTFL scale. However, 21% of the participants (*n*=20), in Spanish, exceeded the benchmark range and were in the Intermediate-Mid and Intermediate-High sublevels.

Table 1

Participant Proficiency Ratings for Listening at the Beginning of Second-Year Study

	French	Portuguese	Spanish
No Measured Ability	0	0	0
Novice-Low	5	0	0
Novice-Mid	11	8	1
Novice-High	17	7	8
Intermediate-Low	4	1	13
Intermediate-Mid	0	0	12
Intermediate-High	0	0	8
Advanced-Low	0	0	0
Advanced-Mid	0	0	0
Advanced-High	0	0	0
Total	37	16	42

With respect to reading proficiency, Table 2 shows that participant ratings in the three languages (*N*=95) ranged between Novice-Low and Advanced-Mid. Five of the participants did not receive a proficiency rating, which could be explained as they did not want to take the test seriously and moved through the readings as quickly as possible so this adaptive test concluded before a rating could be assigned. Nevertheless, 43% (*n*=16), 88% (*n*=14), and 57% (*n*=24) in French, Portuguese, and Spanish, respectively, received ratings between Novice-High and Intermediate-Low for reading, meeting the benchmark for the end of the first year of study. Of those who received a proficiency rating, 22% of the total participants (*n*=21) were below the benchmark with ratings in the Novice-Low and Novice-Mid sublevels on the ACTFL scale. Twelve percent of the total participants (*n*=11) in the three languages exceeded the benchmark range, receiving proficiency ratings from Intermediate-Mid and Advanced-Mid in reading.

Table 2

Participant Proficiency Ratings in Reading at the Beginning of Second-Year Study

	French	Portuguese	Spanish
No Measured Ability	4	0	1
Novice-Low	9	0	2
Novice-Mid	3	1	6
Novice-High	12	11	14
Intermediate-Low	4	3	10
Intermediate-Mid	2	0	2
Intermediate-High	2	1	6
Advanced-Low	1	0	0
Advanced-Mid	0	0	1
Advanced-High	0	0	0
Total	37	16	42

Turning to the second research question regarding if there was a difference in level of proficiency for those who earned the Seal of Biliteracy and those who did not in the three languages, the researchers calculated frequencies for listening and reading. Table 3 shows for listening proficiency that for those participants who reported knowing that they had either received or not received the Seal of Biliteracy in French, the listening proficiency ratings ranged between Novice-Low and Intermediate-Low. While the number of participants who reported not receiving the Seal of Biliteracy was higher than those who reported having earned the Seal of Biliteracy, the proficiency ratings were not higher for those who received it. With respect to Portuguese, more participants reported not having received the Seal of Biliteracy. The two participants who reported receiving the Seal of Biliteracy had a listening proficiency rating of Novice-High while five of the 13 participants without the Seal of Biliteracy received a proficiency rating at the same sublevel and one sublevel higher (Intermediate-Low). Eight participants received a rating one sublevel lower (Novice-Mid) than those who had reported earning the Seal of Biliteracy.

Turing to the participants in Spanish with respect to listening proficiency, the nine participants reporting having earned the Seal of Biliteracy had proficiency ratings in the three Intermediate sublevels. Of those participants who reported not receiving the Seal of Biliteracy (*n*=26), eight had received a proficiency rating below those with the Seal of Biliteracy and having a proficiency rating in the Novice-Mid and Novice-High categories.

Table 3

Comparison of Participant Proficiency Ratings for Listening with Respect to having Earned the Seal of Biliteracy

	French		Portuguese		Spanish	
	Earned Seal	Did not Receive Seal	Earned Seal	Did not Receive Seal	Earned Seal	Did not Receive Seal
No Measured Ability	0	0	0	0	0	0
Novice-Low	1	4	0	0	0	0
Novice-Mid	1	8	0	8	0	1
Novice-High	3	11	2	4	0	7
Intermediate-Low	1	3	0	1	2	7
Intermediate-Mid	0	0	0	0	3	8
Intermediate-High	0	0	0	0	4	3
Advanced-Low	0	0	0	0	0	0
Advanced-Mid	0	0	0	0	0	0
Advanced-High	0	0	0	0	0	0
Total	6	26	2	13	9	26

With respect to reading proficiency to examine differences in levels of reading proficiency for those who earned the Seal of Biliteracy and those who did not in the three languages, Table 4 shows somewhat similar findings when compared to the listening proficiency data. For French, the participants had a wide range of proficiency ratings from Novice-Low to Advanced-Mid. For those who reported not having earned the Seal of Biliteracy, the proficiency ratings were clustered from No Measured Ability to the Intermediate-Mid sublevel, with half (n= 13) at the benchmark for the end of the first year of language learning (Novice-High to Intermediate-Low). While the majority of those who had reported earning the Seal of Biliteracy were in the Novice levels (n=4, 67%), two individuals (33%) had ratings at the Intermediate-High and Advanced-Low levels. In Portuguese, all of the individuals who reported not having the Seal of Biliteracy ranged between the Novice-Mid and Intermediate-Low sublevels. Only one of the participants who reported having earned the Seal of Biliteracy had a higher reading proficiency rating, Intermediate-High. For Spanish, five of the participants (19%) reporting not having the Seal of Biliteracy were below the aforementioned benchmark. The remainder of that group achieved a rating in reading at the benchmark levels or above. Four of the nine participants (44%) who reported earning the Seal of Biliteracy achieved a proficiency rating of Intermediate-High.

Table 4

Comparison of Participant Proficiency Ratings for Reading with Respect to having Earned the Seal of Biliteracy

	French		Portuguese		Spanish	
	Earned Seal	Did not Receive Seal	Earned Seal	Did not Receive Seal	Earned Seal	Did not Receive Seal
No Measured Ability	0	3	0	0	0	1
Novice-Low	2	6	0	0	0	1
Novice-Mid	1	2	0	1	0	4
Novice-High	1	9	1	9	4	8
Intermediate-Low	0	4	0	3	1	8
Intermediate-Mid	0	2	0	0	0	2
Intermediate-High	1	0	1	0	4	1
Advanced-Low	1	0	0	0	0	0
Advanced-Mid	0	0	0	0	0	1
Advanced-High	0	0	0	0	0	0
Total	6	26	2	13	9	26

Viewing findings collectively, fewer participants reported having the Seal of Biliteracy, and the majority of the participants who reported having the Seal of Biliteracy did not have higher proficiency ratings in either modality than those who reported not having the Seal of Biliteracy.

DISCUSSION

The Department of Defense continues to place a premium on language and culture enabled military personnel (United States Air Force, 2024); this study provides important findings with relation to the importance of proper language placement of students and, secondarily, the use of the Seal of Biliteracy as an appropriate tool for such placement in WL classes. As previously stated, the main objective of student placement systems is to group students by similar language ability, aiming to assign learners to the most appropriate proficiency level in WL courses. While this seems straightforward, the process is complex due to various factors such as K-12 teacher certification, curricula, and student backgrounds. Despite being a challenge for postsecondary WL programs, there is limited research on the topic (Hudson & Clark, 2008; Lord, 2022). Traditionally, placement is based on standardized test scores, grade point average, and previous coursework, particularly in math and English. These methods are often inadequate and can disadvantage minority students, leading to longer study times, increased costs, and frustration (Bahr et al., 2019; Denison-Furness et al., 2022; Mechur Karp, 2021).

Recent studies suggest promising alternative approaches to placement, such as self-placement, where students assess their own abilities and goals with guidance from advisors. These methods incorporate student input and offer a more personalized process, potentially leading to better

outcomes in terms of student grades. Though labor-intensive and time-consuming, the positive results from these approaches warrant further exploration (Denison-Furness et al., 2022; Lord, 2022; Mechur Karp, 2021). Regardless of the difficulties of optimal execution of the process, placement of students in appropriate levels of language study is key to developing their proficiency and enabling them to achieve the highest levels of proficiency possible given their particular educational circumstances.

Clearly, setting reasonable expectations for WL proficiency after specific study sequences has been a challenge for decades (Lord, 2022; Swanson et al., 2022). Since the 1960s, educators have tried to establish clear proficiency benchmarks, but consistency has been difficult to achieve. However, the Seal of Biliteracy, an initiative started in 2011 and now adopted across all 50 U.S. states, aims to address this. The Seal of Biliteracy recognizes students who have achieved proficiency in both English and one or more WLs by placing an award on their high school diplomas. Each state or school district determines which WL proficiency assessments are acceptable for awarding the Seal of Biliteracy, and the required proficiency levels vary. As noted earlier, Illinois requires students to demonstrate Intermediate-High proficiency, while South Carolina offers three tiers of proficiency ranging from Intermediate-Mid to Advanced-Low (South Carolina Department of Education, 2019).

The Seal of Biliteracy serves several purposes: it encourages language study, certifies biliteracy skills, highlights the value of language diversity, and prepares students with marketable 21st-century skills. For universities, the Seal of Biliteracy also provides one way to recognize and credit students for their multilingual abilities. However, the wide variation in proficiency levels for attainment of the Seal of Biliteracy across all 50 states makes it problematic to rely on this measure of language proficiency with any sustained confidence (Davin et al., 2022). Such a statement is corroborated by the results of this study, which show disparity in proficiency level results between those who reported earning the Seal of Biliteracy and those without it. Subjects in all three languages were widely dispersed among the proficiency levels, regardless of attainment of the Seal of Biliteracy or not.

Referring to Tables 3 and 4, there were more participants who reported not having received the Seal of Biliteracy than those who reported having earned the Seal of Biliteracy. The researchers would expect participants who reported earning the Seal of Biliteracy to have an Intermediate level of proficiency at the least. However, that was not the case for the three languages. When examining the data from Table 3 for the ALT, few participants who reported having the Seal of Biliteracy achieved a rating of Intermediate-Low or higher in French and Portuguese, However, all of those participants taking Spanish achieved ratings at the Intermediate level. With respect to results on the ART, one third of the participants reporting to have earned the Seal of Biliteracy received a proficiency rating at the Intermediate-Low level or higher whereas one of the two in Portuguese received a rating in the Intermediate levels. Slightly more than half of those in Spanish reporting that they earned the Seal of Biliteracy received a rating in the Intermediate levels. Overall, a higher percentage of those who reported not having earned the Seal of Biliteracy had achieved a higher rating in both listening and reading proficiency.

The researchers concluded that the data did not indicate that by having earned the Seal of Biliteracy participants would have achieved a higher proficiency rating in the two modalities. Such a finding suggests that those participants who did not receive the Seal of Biliteracy yet had achieved a proficiency rating in the two modalities at or above the Intermediate-Low level might be discriminated against in terms of receiving placement credit for prior WL learning. For example, if a WL program were only to place WL learners based on achievement of the Seal of Biliteracy, many language learners may not be placed correctly. The data show that many students of Spanish who reported not having the Seal of Biliteracy achieved at least an Intermediate level of proficiency in the two modalities.

While all 50 states have now recognized the Seal of Biliteracy, some WL programs with dynamic curricula and instruction may not be participating in a Seal of Biliteracy program where their students can possibly receive higher education credit and a more advanced level placement in WL classes. The researchers advance the notion that a Seal of Biliteracy with uniform criteria adopted by the 50 states that promote proficiency in the target language would be beneficial to place language learners in the appropriate levels in higher education. By doing so, language learners can begin to track their path toward higher levels of proficiency as they move from secondary to postsecondary WL programs. At present, there is a need in the United States for more people with second language abilities in the workforce (ACTFL, 2019), and the researchers predict that as the American economy continues to be more global, there will be a need for bilingual individuals in just about every language.

While the topic of how to move WL students up the proficiency ladder was not the central topic of this research, it is important to note that building WL proficiency in the target language hinges on teachers' ability to ascertain and even control execution of CLT approaches in the WL classroom. In other words, the goal is to pinpoint the most effective teaching styles/methods to facilitate language learner movement along their individual interlanguage continuum, to the end of increased language proficiency. WL teachers come from a variety of backgrounds and educational bases, and it would be implausible to expect that they all teach precisely in the same way. Nevertheless, an overall adherence to CLT in the classroom has been shown to be a successful approach to increasing language proficiency among learners (Lightbown & Spada, 2021).

Although this study adds to the literature, it is not without its limitations. First, several of the participants did not receive a proficiency rating for reading. While participants were asked to try their best, it is unknown if they did so. Second, due to privacy concerns regarding participants' academic records, it was not possible to verify if they did or did not actually receive a Seal of Biliteracy. Finally, the participants are enrolled in a military academy and thus may have different motivations and approaches to WL learning than those at a four-year postsecondary institution. In spite of these limitations, the researchers call for further research on placement of WL learners in higher education. It would be informative to design and carry out qualitative studies to determine commonalities and differences between those who earned the Seal of Biliteracy and those who did not. Additionally, future research could focus on productive skills (i.e., speaking,

writing) where the present study focused on receptive skills (i.e., listening, reading), which would be able to be compared to earlier studies (Isbell et al., 2018).

CONCLUSION

Clearly, WL proficiency is a vital component of success in many sectors of today's society: the military, the business world, the economy in general. Thus, the national WL teacher shortage is of great concern for those involved in language instruction and learning. Any encouragement for learners to continue their language study is welcome. This includes proper placement in language courses to ensure efficient progress along the interlanguage continuum. It also relies on WL instruction stemming from the most current and effective methods and approaches being employed in the classroom. Emphasis on learners reaching their full WL potential by achieving proficiency in their chosen language is a fundamental but essential step. Focusing on the tenets of CLT, employing successful classroom practices, and highlighting the overall importance of mastering a second (or more) language(s) will place learners in an optimal position to thrive in their chosen life path. A well-planned program of K-12 WL articulation is the foundation for these language learners to succeed.

Distribution Statement

Approved for public release: distribution unlimited.

Disclaimer

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the United States Air Force Academy, the Air Force, the Department of Defense, or the U.S. Government. PA#: USAFA-DF-2024-722.

Authors

Pete Swanson, Ph.D., is Professor of Spanish at the United States Air Force Academy and serves as the Division Chief for Spanish and Portuguese in the Department of Languages and Cultures. His research focuses on the identity and practices of highly effective world language teachers. He is the author of six books, scores of peer reviewed articles and book chapters, and is Past President of the American Council on the Teaching of Foreign Languages.

Jean W. LeLoup, Ph.D. (The Ohio State University), is Professor Emerita of Spanish in the Department of Languages and Cultures at the United States Air Force Academy, where she directed the Intermediate Spanish curriculum. Her areas of interest include the immersion classroom, the voseo and the linguistic landscape of Costa Rica, and proficiency benchmarking in the Spanish curriculum.

REFERENCES

ACTFL. (2012). *ACTFL proficiency guidelines 2012*. https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

ACTFL. (2015). *Guidelines for implementing the seal of biliteracy*. https://sealofbiliteracy.org/2015-guidelines

ACTFL. (2019). *Making languages our business: Addressing foreign language demand among U.S. employers*. https://www.actfl.org/uploads/files/general/MakingLanguagesOurBusiness_FullReport.pdf

ACTFL. (2024). *ACTFL proficiency guidelines 2024*. https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL_Proficiency_Guidelines_2024.pdf

ACTFL (2025a). *ACTFL/CAEP program standards for the preparation of WL teachers*. https://www.actfl.org/professional-learning/program-review-services

ACTFL. (2025b). *ACTFL proficiency placement test (APPT)*. https://www.actfl.org/assessments/postsecondary-assessments/actfl-proficiency-placement-test-appt

American Academy of Arts and Sciences. (2017). *America's languages: Investing in language education for the 21st century*. https://www.amacad.org/publication/americas-languages

Bahr, P. R., Fagioli, L. P., Hetts, J., Hayward, C., Willett, T., Lamoree, D., ... & Baker, R. B. (2019). Improving placement accuracy in California's community colleges using multiple measures of high school achievement. *Community College Review*, *47*(2), 178–211.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance, and Policy*, *1*(2), 176–216.

Breakfast Leadership. (2023, April 3). *How burnout and teacher turnover costs the US $8.5 billion annually*. https://www.breakfastleadership.com/blog/how-burnout-and-teacher-turnover-costs-the-us-85-billion-annually

Brown, J.D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly, 23*(1), 65–83. https://doi.org/10.2307/3587508

Brumfit, C., & Johnson, R. (1979). *The communicative approach to language teaching*. Oxford University Press.

Cardichon, J., Darling-Hammond, L., Yang, M., Scott, C., Shields, P.M., & Burns, D. (2020). *Inequitable opportunity to learn: Student access to certified and experienced teachers*. Palo Alto, CA: Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/CRDC_Teacher_Access_REPORT.pdf

Carroll, J.B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, *1*(2), 131–151. https://doi.org/10.1111/j.1944-9720.1967.tb00127.x

Center for Language Studies. (n.d.). *Familiarization manual*. Brigham Young University. Author.

Clifford, R., & Cox, T.L. (2013). Empirical validation of reading proficiency guidelines. *Foreign Language Annals, 46*(1), 45–61. DOI: 10.1111/flan.12033

Cox, T.L., & Clifford, R. (2014). Empirical validation of listening proficiency guidelines. *Foreign Language Annals, 47*(3), 379–403. DOI: 10.1111/flan.12096

Clotfelter C.T., Ladd, H.F., & Vigdor, J.L. (2010) Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655–681.

Darling-Hammond, L., Barnett, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23*(1), 57–77.

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., & Vasquez Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*, 42. https://doi.org/10.14507/epaa.v13n42.2005

Darling-Hammond, L., & Wei, R. C. (2009). Teacher preparation and teacher learning: A changing policy landscape. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *The handbook of education policy research* (pp. 613–636). Routledge.

Davin, K.J., Heineke, A.J., & Egnatz, L. (2018). The Seal of Biliteracy: Successes and challenges to implementation. *Foreign Language Annals*, *51*(2), 275–289.

Davin, K.J., Heineke, A.J., & Hancock, C.R. (2022). The Seal of biliteracy: A 10-year retrospective. *Foreign Language Annals*, *55*(1), 10–34.

Denison-Furness, J., Donohue, S. L., Hamlin, A., & Russell, T. (2022). Welcome/not welcome: From discouragement to empowerment in the writing placement process at central Oregon community college. *Writing Placement in two-year colleges*, 107.

Duquette, G. (Ed.). (1995). *Second language practice. Classroom strategies for developing communicative competence*. Clevedon, England: Multilingual Matters Ltd.

Ellis, R. (1997). *Second language acquisition*. Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition*. Oxford University Press.

Galvin, G. (2016, October 13). Some say computer coding is a foreign language. *Newsweek*. https://www.usnews.com/news/stem-solutions/articles/2016-10-13/spanish-french-python-some-say-computer-coding-is-a-foreign-language#:~:text=When%20it%20comes%20to%20allowing,communicate%20with%20machines%20and%20programs

Guthery, S., & Bailes, L. P. (2022). Patterns of teacher attrition by preparation pathway and initial school type. *Educational Policy*, *36*(2), 223–246.

Henry G.T., Purtell K., Bastian K.C., Fortner C.K., Thompson C.L., Campbell S.L., & Patterson, K.M. (2014). The effects of teacher entry portals on student achievement. J*ournal of Teacher Education, 65*(1), 7–23.

Hildebrandt, S., & Swanson, P. (2016). *Understanding the WL edTPA: Research-based policy and practice*. Information Age Publishing.

Hudson, T., & Clark, M. (2008). *Case studies in foreign language placement: Practices and possibilities*. National Foreign Language Resource Center – Hawaii.

Illinois Council on the Teaching of Foreign Language. (n.d.*). Recognizing language educators & learners*. https://www.ictfl.org/seals-of-biliteracy.html

Isbell, D.R., Winke, P., & Gass, S.M. (2018). Using the ACTFL OPIc to assess proficiency and monitor progress in a tertiary foreign languages program. *Language Testing, 36*(3), 439–465.

Krashen, S.D. (1981). *Second language acquisition and second language learning*. Pergamon Press.

Krashen, S.D. (1982). *Principles and practice in second language acquisition*. Prentice Hall International.

La Corte, M., & Voisine, A. (2020, May 27). Foreign language skills are crucial for America's diplomatic, business, & security interests. *Niskanen Center*. https://www.niskanencenter.org/foreign-language-skills-are-crucial-for-americas-diplomatic-business-security-interests/

Language Testing International. (n.d.). *ACTFL proficiency scale*. https://www.languagetesting.com/actfl-proficiency-scale

Lee, J., & VanPatten, B. (2003). *Making communicative language teaching happen*. McGraw-Hill.

Lightbown, P. M., & Spada, N. (2021). *How languages are learned*. *5th Edition.* Oxford University Press.

Lord, G. (2022). Exploring a compassionate approach to standard placement in language classes. *Second Language Research & Practice, 3*(1), 101–115. https://doi.org/10125/69870

Lusin, N. (2012, March). The mla survey of postsecondary entrance and degree requirements for languages other than English, 2009–10. *The Modern Language Association of America*. https://www.mla.org/content/download/3316/file/requirements_survey_200910.pdf

Lusin, N., Peterson, T., Sulewski, C., & Zafer, R. (2021). Enrollments in languages other than English in US institutions of higher education, fall 2021. *Modern Language Association of America*. https://www.mla.org/content/download/191324/file/Enrollments-in-Languages-Other-Than-English-in-US-Institutions-of-Higher-Education-Fall-2021.pdf

Maranon, R. (2024, May 21). State to require more math to graduate from high school in the future instead of fine arts and foreign language. *Fox23 News*. https://www.fox23.com/news/state-to-require-more-math-to-graduate-from-high-school-in-the-future-instead-of/article_d3ed36d6-17a1-11ef-b380-87b81c3563ad.html

Mechur Karp, M. (2021). Developmental education self-placement approaches: Considerations when designing with equity in mind. *Strong Start to Finish*. https://strongstart.org/wp-content/uploads/2021/09/SSTF-PhaseTwo-502.pdf

Miller, J. (2018). Bridging the gap between pronunciation research and proficiency-based teaching. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*, ISSN 23809566, University of Utah, September, 2017 (pp. 111–119). Iowa State University. https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2018/09/PSLLT-Proceedings-9_9-14-18_2.pdf

Morrison, H. (2024, February 6). Uncertified teachers can't be a permanent solution for school districts. *Medium*. https://heathmorrisonsuperintendent.medium.com/uncertified-teachers-cant-be-a-permanent-solution-for-school-districts-

9f31e414d313#:~:text=Uncertified%20teachers%20leave%20the%20profession,manage%20and%20instability%20for%20students.

Moss, G.D., & Gambrell, J.A. (2023). The WLs professional life survey: K-20 articulation through the 5 cs and the can-dos. *Foreign Language Annals, 56*(1), 8–28.

Nguyen T.D., Pham L.D., Crouch M., Springer M.G. (2020). The correlates of teacher turnover: An updated and expanded meta-analysis of the literature. *Educational Research Review, 31*, 100355. https://doi.org/10.1016/j.edurev.2020.100355

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

O'Rourke, P., Zhou, Q., & Rottman, I. (2016). Prioritization of K-12 WL education in the United States: State requirements for high school graduation. *Foreign Language Annals, 49*(4), 789–800.

Reaching Higher NH. (2024, February 27). *Teacher certification and student outcomes: What the research says.* https://reachinghighernh.org/2024/02/27/teacher-certification-and-student-outcomes-what-the-research-says/

Redding, C., & Smith, T.M. (2016). Easy in, easy out: Are alternatively certified teachers turning over at increased rates? *American Educational Research Journal, 53*(4), 1086-1125. https://doi.org/10.3102/0002831216653206

Rosenbusch, M.H. (2005). The no child left behind act and teaching and learning languages in the U.S. Schools. *The Modern Language Journal, 89*, 250-261.

Rosenbusch, M.H., & Jensen, J. (2004). Status of foreign language programs in NECTFL states. *NECTFL Review, 56*, 26–37.

Sawchuck, (2016, July 19). Shortcut paths to teaching linked to higher turnover. *Edweek*. https://www.edweek.org/teaching-learning/shortcut-paths-to-teaching-linked-to-higher-turnover/2016/07

Seal of Biliteracy. (2024). *State laws regarding the seal of biliteracy*. https://sealofbiliteracy.org/index.php

South Carolina Department of Education. (2019). *South Carolina seal of biliteracy: Overview and guidelines*. https://ed.sc.gov/instruction/standards/world-languages/instructional-resources/south-carolina-seal-of-biliteracy-overview-and-guidelines-revised-2019/

Spencer, L. A. (2019). *Alternative or Traditional Certification: Does Route to Licensure Impact Teacher Effectiveness and Retention?*. Trevecca Nazarene University.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). *A coming crisis in teaching? Teacher supply, demand, and shortages in the U.S.* Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/A_Coming_Crisis_in_Teaching_REPORT.pdf

Swanson, P. (2008). Efficacy and interest profile of foreign language teachers during a time of critical shortage. *NECTFL Review*, *62*, 55–74.

Swanson, P. (2010). The shortage of America's foreign language teachers: A review of the literature. *Journal of Teacher Recruitment and Retention, 1*(1), 1–17.

Swanson, P. (2012). The congruence of vocational interests and the workplace environment: Reducing the language teacher shortage. *Language Teaching Research. 16*(4), 519–537. https://doi.org/10.1177/1362168812455588

Swanson, P. (2019). Teacher education: Past, present, and future. In J.W. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 751–774). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781108333603.032

Swanson, P. (2022). The WL shortage can be solved. *The Language Educator, Spring,* 38–40.

Swanson, P., & Hildebrandt, S.A. (Eds.) (2018). *Researching edTPA Problems and Promises: Perspectives from ESOL, English, and World Language Teacher Education*. Charlotte, NC: Information Age Publishing.

Swanson, P., & Moore, A.D. (2006). Changing student misperceptions about foreign language teaching: A research-based approach to improving recruitment practices. *NECTFL Review, 59,* 6–27.

Swanson, P., Peltier, J.P., LeLoup, J.W., Earnest, D., & Malone, M.E. (2022). Proficiency benchmarking in Spanish. *Central States Report,* 1–16.

The National Standards Collaborative Board. (2015). *World-readiness standards for learning languages* (4th ed.). Author.

Troyan, F., Swanson, P., & Russell, V. (2023). Removing barriers to WL teacher certification: A closer look at edTPA ratings. *Hispania, 106*(4)*,* 657–675*.*

Tschirner, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals, 49*(2), 201–223. https://doi.org/10.1111/flan.12198

United States. (1965). Elementary and secondary education act of 1965: H. R. 2362, 89th Cong., 1st sess., Public law 89–10. Reports, bills, debate and act. [Washington]: [U.S. Govt. Print. Off.].

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, *State and Local Implementation of the No Child Left Behind Act*, Volume VIII—Teacher Quality Under NCLB: Final Report, Washington, D.C., 2009. https://www2.ed.gov/rschstat/eval/teaching/nclb-final/report.pdf

United States Air Force. (2024). *Language-enabled airmen bridge cultures, elevate missions.* https://www.aetc.af.mil/News/Article-Display/Article/3818845/language-enabled-airmen-bridge-cultures-elevate-missions/

VanPatten, B. (2017). *While we're on the topic: BVP on language, acquisition, and classroom practice*. ACTFL.

Walsh, K., & Jacobs, S. (2007). *Alternative certification isn't alternative*. Thomas B. Fordham Institute. https://www.nctq.org/research-insights/alternative-certification-isnt-alternative

Widdowson, H. (1978). *Teaching language as communication*. Oxford University Press.

Applied Language Learning

Article

Building and Cleaning Corpora for Linguistic Analysis: A Practical Guide

Ghadi Matouq
Doctoral Candidate, Applied Linguistics, University of Memphis

Hana Alqabba
Doctoral Candidate, Applied Linguistics, University of Memphis

This guide aims to make corpus building and corpus analysis feasible and practical for language instructors and/or researchers who may view building a corpus as difficult or believe that linguistic analysis requires advanced programming skills. Many avoid creating custom corpora due to these perceived barriers, instead relying on existing corpora and basic analysis tools. We present accessible instructions for corpus building, text cleaning, and linguistic analysis based on our coursework and research experience. The guide contains two parts: theoretical foundations covering corpus linguistics definition, research questions in corpus linguistics, and different types of corpora; and practical applications including corpus construction, text preparation, automated annotation, and an introduction to some types of lexical analysis. The guide demonstrates that systematic instruction makes corpus methods accessible to language teachers and novice researchers. We emphasize that hands-on practice is essential for developing corpus research skills and encourage active application of these methods to readers' research questions. We conclude with a discussion on the benefits of corpus analysis for the language classroom.

Keywords: *Corpus Linguistics, Corpus Building, Text Cleaning, Linguistic Analysis, Corpus-Based Research*

INTRODUCTION

Corpus analysis is a valuable skillset for language instructors and researchers, as it provides opportunities for teachers to better understand language features and use (particularly related to vocabulary and grammar) and also design effective materials for classroom instruction or even language assessments. Yet, many language instructors or even researchers view corpus building

as an overly difficult task and text analysis as a highly technical process requiring advanced programming skills. This, however, is not the case. What corpus building and corpus analysis do require is systematic thinking, as well as learning the steps and application options.

This article demonstrates how to build, clean, and analyze a corpus through step-by-step instructions, allowing language instructors and novice researchers to approach these processes with clarity, efficiency, and effectiveness. There are many ways to address the processes and decisions introduced here, but we hope that this practical guide is enough to get interested teachers and researchers started.

This guide is organized into two main parts. The first briefly provides a foundation by defining corpus linguistics, examining research questions in corpus linguistics, and exploring different types of corpora. The second section focuses on practical applications, covering corpus building, importance of text preparation and cleaning, linguistic annotation, and some types of lexical analysis. Corpus building and analysis require hands-on practice to master. Therefore, we encourage readers to follow the instructions in this guide and use the tools to build, clean, and analyze their own corpora, while also exploring other tools and processes that may work for them.

CORPORA AND CORPUS LINGUISTICS

A Definition

As a field of study, corpus linguistics involves studying language in use through a corpus or corpora (plural of corpus). A corpus is a collection of texts stored in digital format for linguistic research. To analyze these texts, corpus linguists use software tools that help them find patterns in the data. For example, they can examine how often a word appears, what words tend to co-occur with other words, or how words or phrases are used, showing every occurrence of a word or a phrase in context. These are just a few examples, as researchers in corpus linguistics are diverse in interests, diverse in aims, and notably creative.

Research Questions in Corpus Linguistics

Corpus linguistics serves as a research methodology that allows researchers to examine different areas of applied linguistics including grammar, lexis, discourse, pragmatics, and second language acquisition. Researchers can address both quantitative and qualitative questions depending on their research goals (Timmis, 2015).

Quantitative corpus analysis focuses on frequency information that can be generated easily from corpus data. Basic frequency questions include: What are the most frequent words in a corpus? How many instances of a given word appear? What percentage of total tokens does a word represent? What are the most frequent collocations or phrases? What are the most frequent grammatical structures? These questions can be applied with specific focus, such as examining

word frequency in different domains (e.g., medical vs. legal texts), demographics (e.g., age or proficiency groups), or modes (e.g., spoken vs. written language).

However, many research questions require qualitative analysis that cannot be answered through automatic corpus analysis alone. For example, determining which meaning of a word like "tip" is most frequent, or identifying when "marvelous" is used sarcastically, requires manual examination of corpus data. Thus, it is important for researchers to read individual instances and interpret their meaning in context. This combination of quantitative patterns and qualitative interpretation makes corpus linguistics an effective methodology for understanding authentic language use across diverse research contexts.

Different Types of Corpora

Texts in corpora can be written, spoken, and multimodal. Written corpora could include books, research articles, newspapers, emails, and blogs, digitized and stored in text files. Some written texts may need initial preparation before the cleaning process begins. For example, scanned and hand-written texts require conversion into machine-readable formats before they can be used reliably. Nevertheless, written corpora are widely used in research as they are often readily available in digital form.

Unlike written corpora, spoken corpora are somewhat less common because they are harder to collect, require transcription, and are notable less 'formalized', as they are comprised of human speech. These include transcripts of conversations, narratives, lectures, debates, and conference talks. The least common type of corpora are those including multimodal components collected through the video recording. These include verbal and non-verbal data, including speech and body language (Foster & Oberlander, 2007). Multimodal corpora represent a development from spoken corpora, offering additional modes of communication including gesture, facial expression, and intonation, synchronized with transcripts. Multimodal corpora expand the kinds of communicative phenomena that can be examined, depending on the research focus (McCarthy & O'Keeffe, 2010).

Corpora can also be categorized in terms of their content, with two main broad categories: general and specialized corpora. General corpora contain a large collection of text from many different sources, genres, and domains, representing the whole language in all its diversity. Therefore, they may include texts coming from newspapers, books, academic writing, conversation, TV shows, fiction, non-fiction, etc. Specialized corpora, on the other hand, are not as large as general corpora, and their size could vary based on the language they represent. They include texts that meet specific criteria, focusing on specific registers and genres.

Several influential corpora have become classic references in corpus research, which will be used here to exemplify the distinction between general and specialized as well as written and spoken corpora through their content and mode. The Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) are two important general corpora in English that are predominantly comprised of written texts, with very few spoken texts (20% of COCA and 10% of

BNC come from spoken language). The TenTen family of corpora, 10+ billion word corpora available in over 50 languages such as Spanish, Japanese, Russian, Arabic, and Greek, represent large-scale web-based general corpora (Jakubíček et al., 2013). The British Academic Written English (BAWE) corpus and the Michigan Corpus of Upper-Level Student Papers (MICUSP) are written corpora representing language use in the academic domain. The British Academic Spoken English (BASE) and the Michigan Corpus of Spoken Academic English (MICASE) are spoken corpora. Both are also examples of specialized corpora representing language use in English academic domains. These existing corpora offer significant savings in time and effort, often include detailed metadata useful for coding and analysis, and can be quite large and therefore enable reliable frequency counts of linguistic features and more diverse word and phrase types than smaller corpora could provide.

However, a significant limitation of utilizing existing corpora is that they may not align perfectly with research questions and target population. The reality is that "no corpus is one size fits all" (Egbert et al., 2020, p. 5), which means that the texts included in any given corpus shape the linguistic population to which the results can be generalized. Therefore, it becomes essential to choose a corpus that matches research needs in two keyways: (1) the types of texts it includes and (2) whether it contains enough data for reliable analysis. Additionally, it is important to identify where mismatches exist and interpret findings within these limitations (Egbert et al., 2020). Further, in language for specific purposes contexts where educators frame language teaching around specific genre competencies, it is likely that publicly available specialized corpora do not exist.

PRACTICAL CONSIDERATIONS

Corpus Building

Researchers may want to compile their own corpus if existing corpora cannot be used to answer their questions. For instance, they may want to focus on a particular author, genre, or variety of language that has not been represented in available corpora. They also may build new corpora to analyze recent language use. If you decide to build your corpus, here is a step-by-step process for corpus compilation. A summarized outline of these steps is provided in Figure 1.

1. **Make a research plan.**
   Start with planning and defining research or teaching goals: begin by asking your research questions, determining the type of language that could address your research questions, and identifying specific linguistic features of interest. For example, a question about differences in the use of reporting verbs (e.g., state, claim, emphasize, discuss) between experts' research articles and students' MA theses. This question could be answered by collecting research articles from reputable journals and MA theses available online.

2. **Design your corpus.**
   Now that you have identified the corpus that answers your research questions, you need to design your corpus.
   a. *Set some criteria to start compiling your corpus.* Corpus compilers use a range of criteria when selecting texts for a corpus to ensure that the resulting dataset is representative, balanced, and suitable for the intended research. Among the commonly used criteria, as described by Sinclair (2004), are mode (spoken/written), type (books/research articles), domain (academic/popular), variety (specific language/languages/language varieties), location (US/UK), and date.
   b. *Determine the corpus size.* While there is no definite answer for how large a corpus should be, Reppen (2022) explains that the size of a corpus could be determined by two main factors: representativeness (ensuring sufficient data to accurately reflect the language being studied) and practicality (considering time constraints). In simple words, you need to balance the ideal with what is feasible. For highly specific research, e.g., the language of a particular author, such as 'the Swale's corpus' compiled by Hyland (2008), a more focused corpus can be fully representative of the language. But for broader studies, e.g., scrutinizing academic language in several disciplines (Hyland, 1999), texts were selected to represent the language under investigation.
   c. *Balance your corpus.* As Sinclair (2004) indicated, balance does not require accuracy in a strict sense. Instead, it demands transparent, reasoned decisions about text proportions, guided by the corpus's goals and the diversity of language. To do so, aim for a balance between different text types and document your inclusion/exclusion decisions. For example, while the text types included in the COCA corpus may not fully capture the range of public discourses in contemporary English, their diversity allows for strong insights.

3. **Collect and prepare your data.**
   There are several cases where researchers should seek and obtain permission before compiling their corpus. For example, approval from an institution's IRB is required for any human subjects research at US institutions. Also, there are cases where researchers need permission to use texts with copyright (Reppen, 2022). In any research, ethical considerations should be respected. You can start your data collection with the required permission, if any.
   a. *Develop protocols for consistent data collection*. Guidelines and detailed procedures are essential to collect reliable data, especially when multiple researchers are involved. For example, you need to decide on what to keep and what to delete from a document in your written corpus (e.g., will you keep references and tables in an academic text or bylines in a newspaper article?). Also, before transcribing texts from your spoken corpus, you need to have a transcription system and decide on how to represent non-verbal data such as laughter, pauses and overlaps, if they will be documented (Reppen & Simpson, 2002). Consider the protocols as a go-to resource for yourself, teams, and the research community. If they need to be changed, then all previously prepared data must be carefully reviewed.
   b. *Create a metadata scheme*. You need to describe your corpus information and keep your data organized and searchable. Think about what information might be useful for

analysis. Information such as context, speakers' identity, gender, age or author, year of publication, and title are some common details, but other information may be of interest to you. Focus on what you need to understand and organize your data clearly, as well as what you may wish to analyze later.

c. *Metadata can be stored inside the text, in the filename, or in a separate file*. To contain information within the text, use headers at the beginning of each file with information about it (Reppen, 2022). You should add headers clearly and consistently. A common style is to use angle brackets < > for each line. Add a clear separator between the header and the main text, like this: <end_header>. Using brackets and separators makes it easy for you and the software to recognize and ignore the header during analysis. Metadata can also be stored separately in a CSV table (e.g., Excel) with each row for one text and each column including metadata information (e.g., title, author, year of publication).

4. **Clean the texts.**

Corpus cleaning is an essential step in preparing data for analysis.

a. Work with plain text files and save files in UTF-8 format to ensure compatibility with most corpus analysis tools (Lu, 2014; Reppen, 2022).

b. In a spoken corpus, transcripts should be cleaned by correcting human or machine transcription errors, and identifying and correcting inconsistencies in transcription, such as the use of contractions (e.g., whether to use '*gonna, going to, or both*', '*cannot or can't*' following a previously established criteria), and the presentation of non-verbal elements, ensuring consistency across the dataset (Thompson, 2004).

c. In a written corpus, this involves removing undesirable components from your corpus. This includes headers, footers, page numbers, tables, charts, line breaks, extra white spaces, and copyright notices, unless those things are of interest, and other peculiarities of genres and texts of interest. Also, careful attention should be paid to character changes in letters, punctuation, or symbols. If you convert a PDF to a text file and notice that some letters look strange, it could mean the original PDF uses unusual fonts or encoding, and the converter could not read them correctly; in that case, trying a different converter or using an Optical Character Recognition (OCR) tool might help fix the issue. It is necessary to do manual checking and editing to correct errors that could result from converting the original file or using OCR. Some PDFs that use multiple columns may create similar issues, and attempting to access HTML files or other formats could save considerable time.

d. Note that the analytical focus often influences how the text should be cleaned. Sentence boundaries should be clear and accurate, especially when conducting syntactic analysis and using the sentence as the unit of analysis (e.g., measuring the number of words or clauses per sentence, using tools like TAASSC for syntactic complexity). For example, when cleaning texts with bullet points, the researchers must decide if they will add periods to bullet points that are complete sentences and add semicolons to bullet points that are fragments. Accurate sentence boundary detection is essential for accurate POS tagging (Lu, 2014), which means that sentence boundaries are also important for lexical analyses.

e. Apply consistent name conventions when saving your text files and the original files. Aim for short names with letters and numbers and replace spaces with underscores, as in Lec_1, RA_1 for (lecture 1) and (research article 1). You can include a clear abbreviation for other information, for example, if you have multiple disciplines, you can use CS for Computer Science, BA for Business Administration, ENG for English, and PSY for Psychology.

5. **Conduct a pilot.**
   Before cleaning the full corpus, it is good practice to conduct a pilot using a small, sample, typically around 5-10% of the total data. This subset should reflect the diversity of the full dataset in terms of genre, structure, or source. A pilot can be useful for testing cleaning procedures, such as removing noise, correcting encoding issues, or standardizing formatting. Applying cleaning methods to a sample allows researchers to identify and address potential issues early, including inconsistent spacing, problematic characters, or formatting irregularities. As part of this step, compare the original uncleaned texts to the cleaned text samples, and make sure that the arrangement of the text and paragraphs has not been changed by the conversion process. Based on the findings from the pilot, the cleaning approach can be refined to ensure it is effective. This step improves the overall quality and consistency of the corpus and helps prevent complications during later stages of analysis. See the Appendix for notes on the importance of text preparation and cleaning.

6. **Begin analysis.**
   After ensuring that the corpus is cleaned, the analysis phase can begin. It is important to identify key linguistic features related to the research questions, such as specific words, grammatical patterns, phrase structures, or discourse markers. These features guide the focus of the analysis and help ensure that data interpretation aligns with the study's objectives. Once the relevant features are identified, researchers must select appropriate corpus tools and statistical packages to support their analysis. This may include concordance lines, POS taggers, keywords, or statistical software like SPSS, depending on the complexity of the data and the nature of the analysis. Throughout the process, many researchers seek to balance quantitative findings such as frequency counts or collocation patterns with qualitative insights, including the examination of concordance lines or contextual usage. This approach strengthens the interpretation of the results and allows for a more nuanced understanding of linguistic phenomena in context.

Figure 1
Corpus Compilation Steps

Linguistic Annotation of Corpus Data

After building and cleaning your corpus, begin by adding linguistic information to your texts through *linguistic annotation* if your analysis requires it. This step can be skipped if your research focuses on analyses that can be done without annotation, such as certain frequency analyses, keyword analysis, etc. However, your analysis requires linguistic annotation (such as part-of-speech tagging). You can then conduct various types of linguistic analysis including *lexical analysis* (examining word-level patterns), *syntactic analysis* (investigating grammatical structures), and

more. The type of analysis you choose depends on the research questions you are trying to answer, and there are more types of annotation than those discussed here.

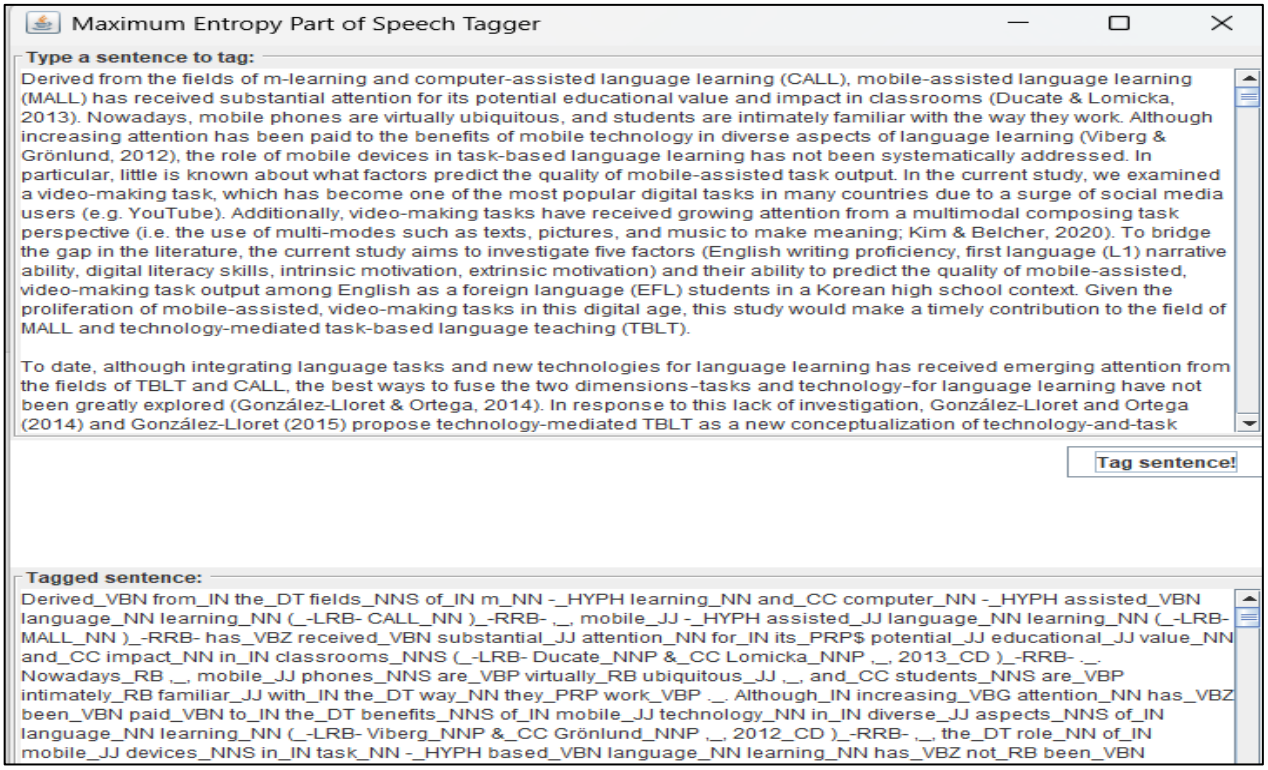
Part-of-Speech Tagging

Part-of-speech (POS) tagging is a widely used type of linguistic annotation that assigns labels or "tags" to each word in your corpus with its grammatical category such as noun, verb, or adjective (Newman & Cox, 2020). The Stanford POS Tagger is highly accurate and provides a user-friendly interface that makes it accessible for beginners in corpus linguistics, but there are other systems. You can use the tagged data as a foundation for investigating grammatical patterns, syntactic complexity, and linguistic variation in your corpus. The Stanford POS Tagger uses the Penn Treebank tagset. The tagger can also be run using Command Line prompts, but this introduces the Graphic User Interface method for novices. This tool works with English, Arabic, Chinese, French, Spanish, and German.

To conduct POS tagging with the Stanford tagger:

1. Go to the Stanford Natural Language Processing Group website and download the Stanford POS tagger (https://nlp.stanford.edu/software/tagger.shtml)
2. Ensure Java 8+ is installed on your computer
3. Extract the downloaded tagger files to a folder on your computer
4. Navigate to the tagger folder and launch the graphical interface application
5. Enter your text in the upper input window (Figure 2)
6. Click the "Tag sentence!" button
7. View the tagged results in the lower window, where each word appears with its grammatical label
8. Save or copy the tagged output in plain text for further analysis

Figure 2
Stanford POS Tagger Graphical User Interface



Lexical Analysis: An Introduction

Texts that have been cleaned can be effectively analyzed using corpus tools to answer research queries. Lexical analysis is an essential area of corpus linguistics that has been associated with the development of corpus linguistics and Sinclair's contribution to the field through the COBUILD (Collins Birmingham University International Language Database) project. Lexical analysis examines the frequency, use, and type of words and phraseological patterns within a text collection. The following subsections provide step-by-step instructions for conducting various types of lexical analysis using specialized corpus tools. Also discussed is corpus analysis with AntConc (Anthony, 2023), a free and feature-rich tool, including an overview of its functionalities and affordances, and phrase-frame analysis using KfNgram. Together, these sections demonstrate how to analyze different linguistic features, from individual word frequencies and collocations to complex discontinuous multi-word patterns.

Corpus Analysis with AntConc

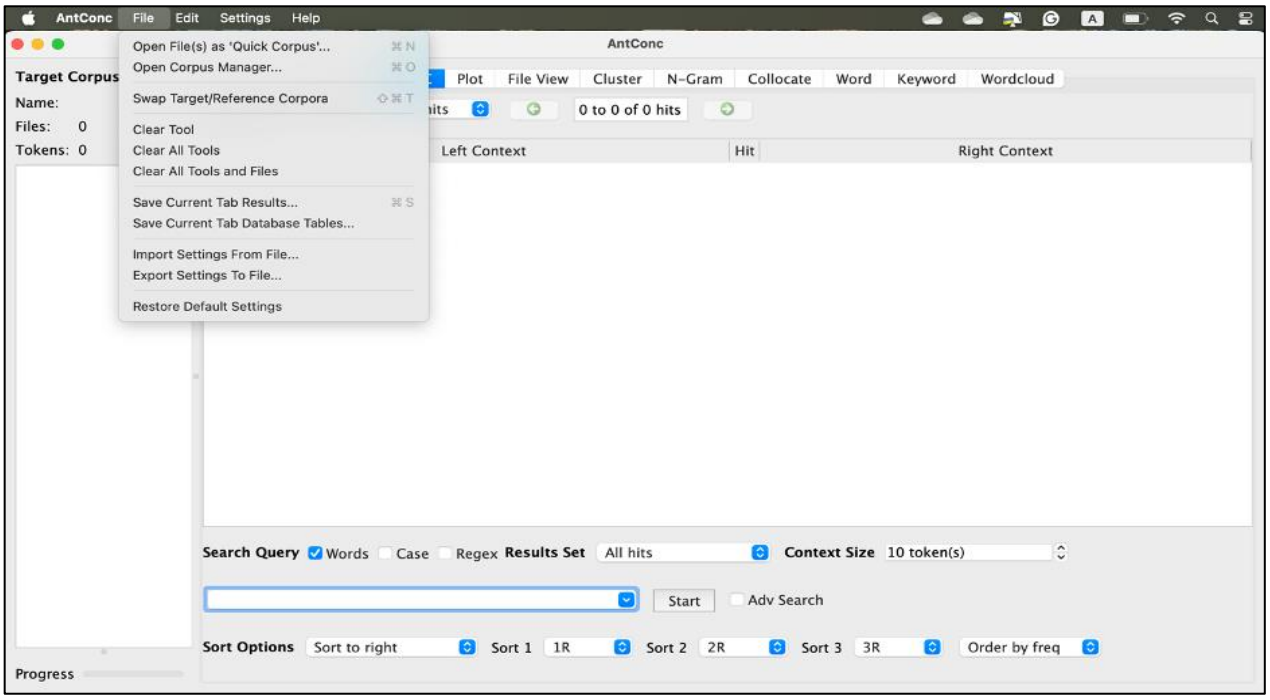
AntConc, developed by Laurence Anthony, is a widely used, freely available tool for corpus analysis, compatible with Windows, macOS, and Linux, that can be downloaded from here https://www.laurenceanthony.net/software/antpconc/.

Before beginning any analysis, you need to do the following (see Figure 3):

1. Download and install AntConc from the official website
2. Prepare your corpus files in plain text (.txt) format, although it does accept other file forms
3. Organize your files in a folder
4. Load your corpus: open AntConc → click File → open corpus manager → corpus source → raw file(s) → name your corpus → add files from your corpus folder → click on create

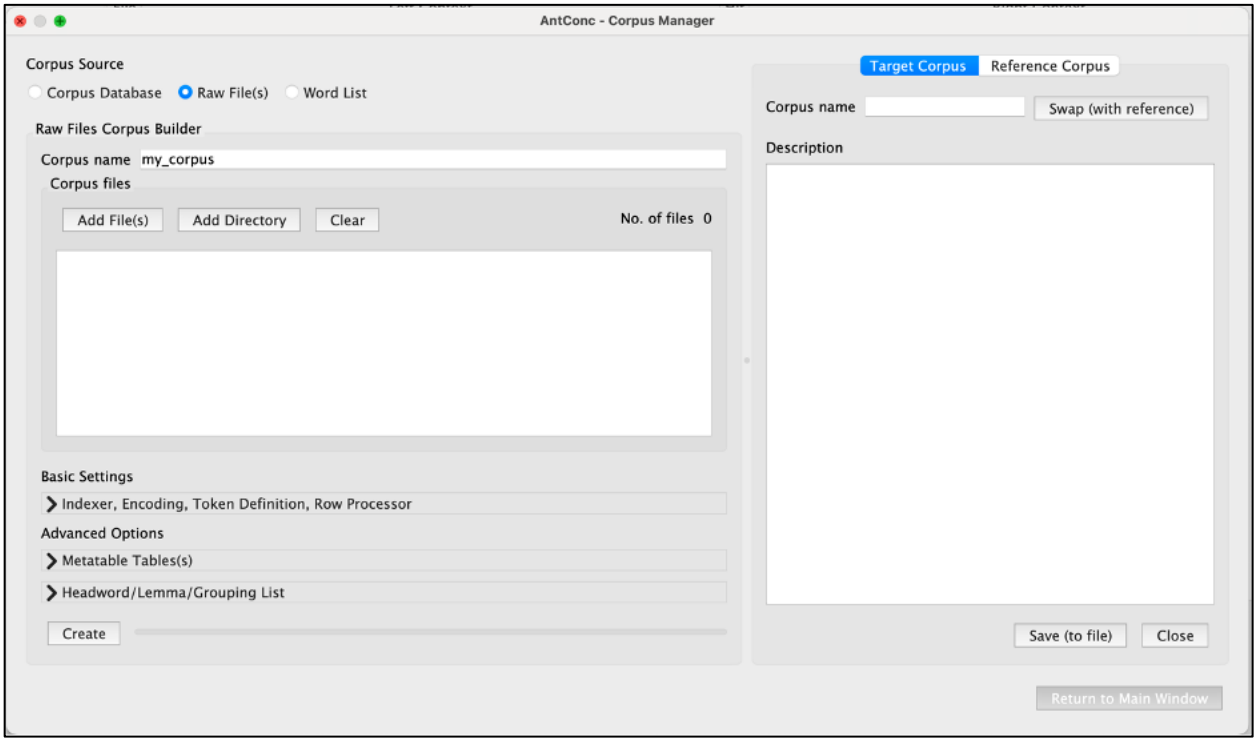
Figure 3

AntConc: Corpus Loading



Under File, you have two options to load a corpus: Open File(s) as a Quick Corpus or Open Corpus Manager. Use Quick Corpus File(s) to open text files for immediate analysis, but use Corpus Manager if you want to do repeated analyses as you can access your corpus later from the current database (any corpus you create using Corpus Manager will be listed there); or if you need to do comparative analysis using target and reference corpora as will be explained in the keyword analysis (see Figure 4).

Figure 4
AntConc: Corpus Manager



Before discussing analytical approaches, it is essential first to explain the functions of the interface tools in AntConc. More information about these tools and the menu options can be found at https://www.laurenceanthony.net/software/antconc/releases/AntConc424/help.pdf

KWIC (Key Word in Context): This tool shows search words in their immediate context, resulting in concordance lines with your search word in the center. You can use KWIC to examine how words are used in context, study patterns of usage and meaning, and conduct qualitative analysis of language use.

Plot: This tool visualizes the distribution of search words across texts. It shows search results as barcodes, where each bar represents a text. The text length is normalized to fit the bar width, and a vertical line on the bar shows the hits. You can use plot to see if keywords appear in specific sections, compare distribution patterns across different texts, and identify topics that appear throughout vs. in specific sections.

File View: This tool displays full text files with search words highlighted. You can use File View to understand broader discourse patterns, examine document structure, and compare and verify concordance findings in a fuller context.

Cluster: This tool identifies word combinations that appear around a specific search word, resulting in lists of word clusters around your search word. You can use Cluster to study the behavior of particular terms and do targeted lexical analysis.
N-Gram: This tool displays contiguous word sequences in the entire corpus (lexical bundles). It results in lists of chunks with frequency data. You can use the N-Gram to identify lexical bundles, examine their frequency, and examine their distribution.

Collocate: This tool finds words that frequently co-occur with search words, showing statistical associations between words appearing within a defined span. You can use this tool to conduct collocation analysis, do semantic prosody analysis (positive/negative associations), identify semantic fields and domains, and critically analyze word choices in particular contexts.

Word: This tool generates frequency lists of all words in a corpus, providing a list of words ranked by frequency, range, and percentage. You can use this tool to identify high-frequency and low-frequency vocabulary, compare vocabulary across different corpora, and study lexical density and complexity.

Keyword: This tool can be used to compare a target corpus against a reference corpus and show statistically significant words that distinguish your corpus. You can use this tool to identify distinctive vocabulary and compare corpora of different genres or periods.

Wordcloud: This tool provides a visual representation of word frequencies. You can use this tool for creating word clouds of words in your corpus by selecting the source of the data (e.g., Word, Keyword), or by inserting a text in the scratchpad.

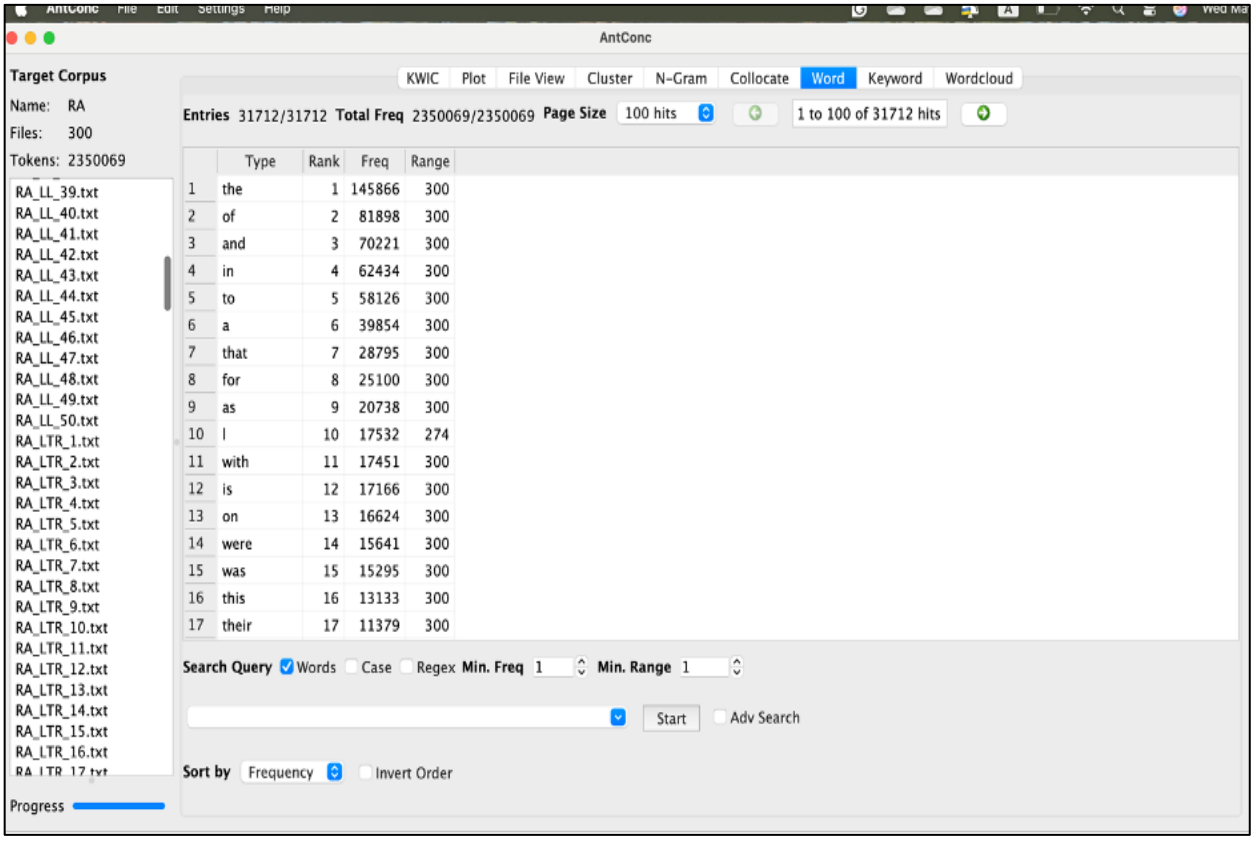
Frequency Analysis

Frequency analysis reveals which words appear most frequently in your corpus, providing insights into words more common or even specific to the corpus.

To start a frequency analysis:

1. Load your corpus into AntConc
2. Click the Word tab (Figure 5)
3. Set the options for the search, minimum range, and frequency
4. Click Start to generate the frequency list
5. Selecting “Frequency” as the sort option shows the results from most to least frequent. Choosing “Range” arranges results based on how widely the words occur across texts.

Figure 5
AntConc: Word



In this image, you can see the corpus information: name, files, and tokens. Word, the third tap from the last one, is selected. You can specify your search by choosing the options next to the search query. You will probably notice that function words, including prepositions and conjunctions, always ranked as the topmost frequent in the list. Researchers are most likely interested in content words rather than function words. Several studies used frequency analysis to identify domain-specific words and compare word patterns across different text types.

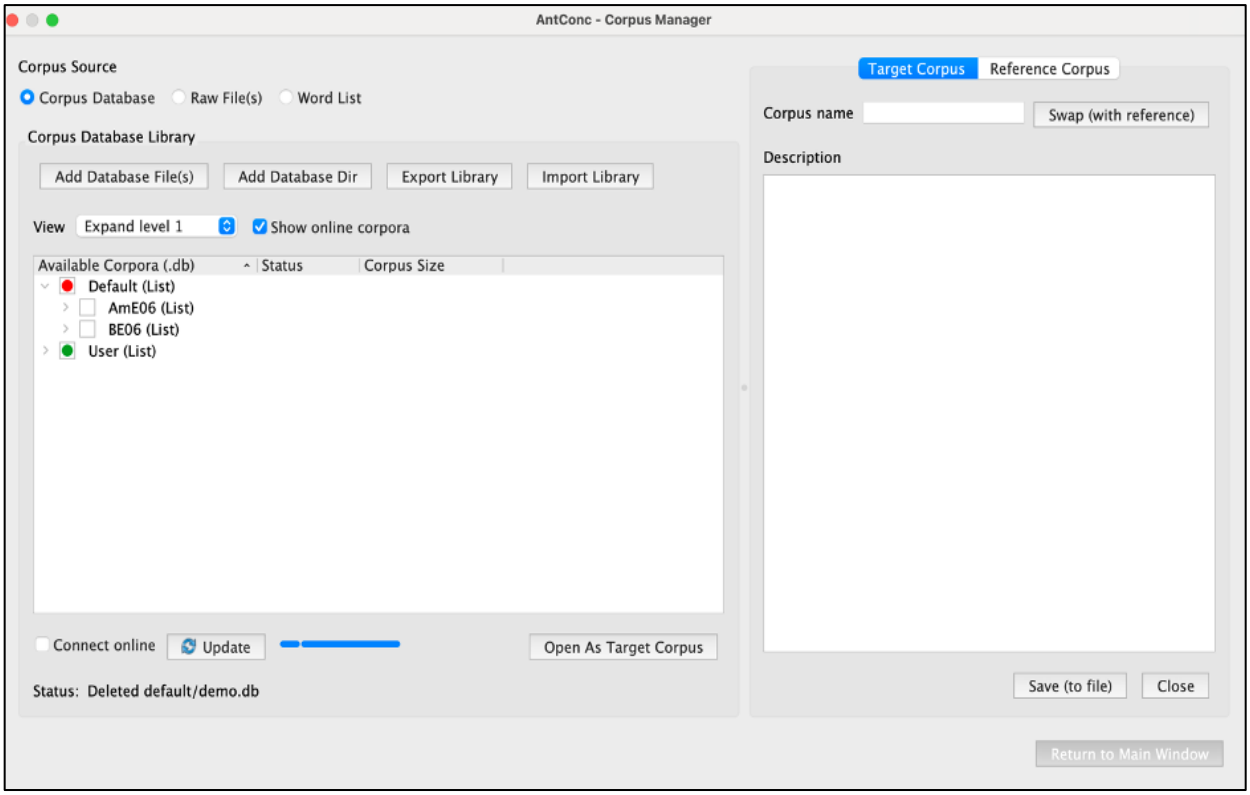
Keyword Analysis

Keyword analysis shows words occurring more frequently in a corpus than in another. Hence, it involves using two corpora: the target corpus and the reference corpus. You could compare a collection of texts on a specific topic/discipline (target corpus) with a general corpus (reference corpus) to identify keywords that are relevant to that topic/ discipline.

Keyword analysis reveals words that appear more frequently in the target corpus compared to the reference corpus based on the relative corpus size and some statistical calculations. AntConc's Corpus Manager allows you to manage and select target and reference corpora for analysis. It also offers online wordlist corpora, under the default list, that you can download and

use as a reference corpus. These include the BE06 corpus and the AME06 corpus word frequency lists (see Figure 6).

Figure 6
AntConc: Keyword



To start keyword analysis:

1. Open Corpus Manager and load your target corpus and reference corpus; use the "Target Corpus" and "Reference Corpus" tabs at the top right of the interface to switch between them. The Target Corpus is the corpus you want to analyze or compare; the Reference Corpus is the corpus that serves as a benchmark for comparison.
2. Go back to the main interface, and click on the "Keyword" tab.
3. Go to "sort by" and select "Likelihood". This option ranks Keywords by the statistical significance of the difference. Based on the corpus size and the results, you may want to focus on the top 20-50 keywords, or you may have a long list of significant results.
4. Change your sort options to reveal different layers of your data. For example, search "students" then flip between Likelihood and Range (Tar) to see if this keyword is both distinctive and consistently distributed across your corpus.

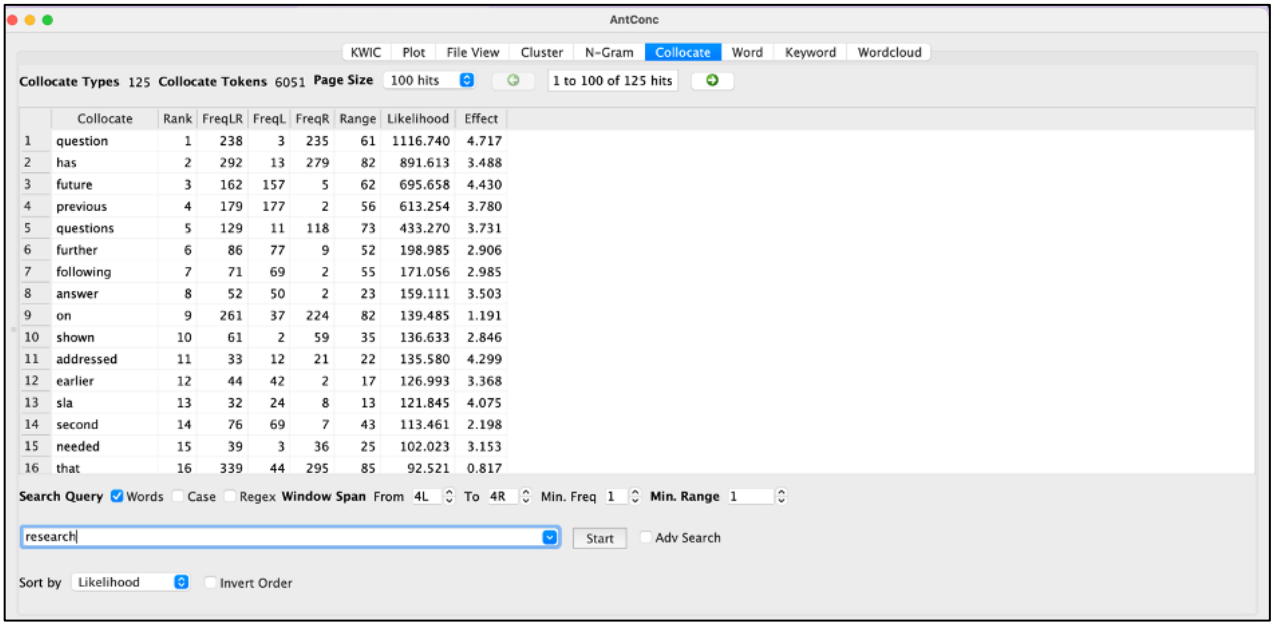
Collocation Analysis

Collocation refers to the tendency of certain words to co-occur more frequently than would be expected by chance. You can look at words that occur directly next to your search word (adjacent collocations) or those occurring within a span (window collocations). As explained by Sinclair et al. (2004), the analysis of collocation involves: node (the target word being studied), collocate (words that appear near the node within a certain distance), span (the window of words examined around the node, e.g., four words before and after), and strength (using statistical calculations to determine whether word pairs co-occur together significantly more frequently than would be expected by chance alone, based on their frequencies and the size of the corpus). Researchers have identified collocations using a collocation-via-concordance approach, where the tool is used for generating the list and the analysis is completed manually without testing for statistical significance, or collocation-via-significance (McEnery & Hardie, 2011).

Here is how to start this analysis:

1. Load your corpus into AntConc
2. Click the Collocate tab
3. Type your node word in the search box. The choice of the word can be driven by the research question, results of frequency or keyword analysis, or informed by literature, technical terminology in a domain, or any other purpose.
4. Set statistical parameters: set window size (e.g., 4L, 4R), minimum frequency (e.g., five occurrences), and sort results by Log-likelihood (i.e., built-in statistical measure used to determine which words are significant collocates)
5. Click ‘start’ to run the analysis.
6. Set the same statistical parameters except for the sorting option: click effect to list the words based on the strength of association (words with a high effect size are strong collocations)
7. Click ‘start’ to run the analysis. When searching for a word’s collocates, a long list of co-occurring words may appear. However, not all of them are meaningful or relevant, so it’s important to filter the results using statistical measures and by reviewing concordance lines in context (Figure 7).
8. Consider those collocates with high Log-likelihood and high effect size
9. Check the concordance lines of the collocates of the given node, and look for patterns

Figure 7
AntConc: Collocate



	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	question	1	238	3	235	61	1116.740	4.717
2	has	2	292	13	279	82	891.613	3.488
3	future	3	162	157	5	62	695.658	4.430
4	previous	4	179	177	2	56	613.254	3.780
5	questions	5	129	11	118	73	433.270	3.731
6	further	6	86	77	9	52	198.985	2.906
7	following	7	71	69	2	55	171.056	2.985
8	answer	8	52	50	2	23	159.111	3.503
9	on	9	261	37	224	82	139.485	1.191
10	shown	10	61	2	59	35	136.633	2.846
11	addressed	11	33	12	21	22	135.580	4.299
12	earlier	12	44	42	2	17	126.993	3.368
13	sla	13	32	24	8	13	121.845	4.075
14	second	14	76	69	7	43	113.461	2.198
15	needed	15	39	3	36	25	102.023	3.153
16	that	16	339	44	295	85	92.521	0.817

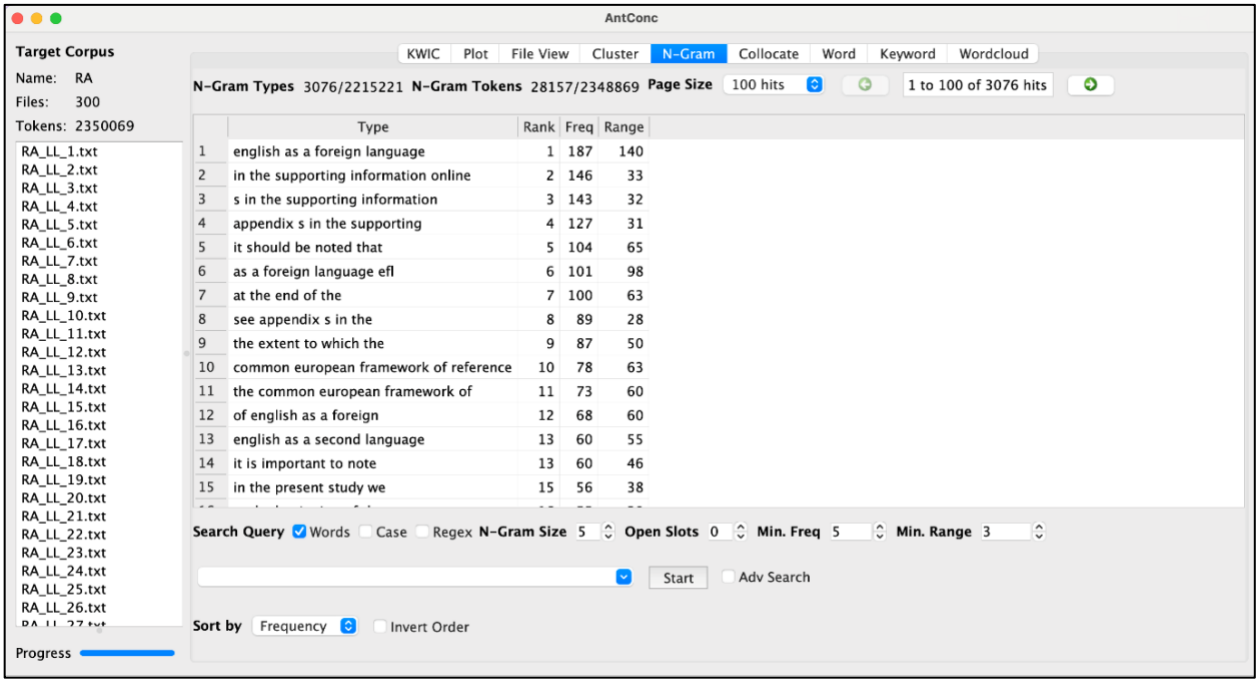
Lexical Bundles

Lexical bundles are continuous sequences of three or more words that occur with high frequency (Biber et al., 1999). You can use the N-Gram tool to extract lexical bundles in your corpus.

Here is how to start this analysis:

1. Load your corpus into AntConc
2. Click the N-Gram tab
3. Set your parameters: decide on the N-gram size (usually 3-6 words, 4 and 5 grams are most common), minimum frequency (e.g., if you are searching for bundles with more than 5 occurrences, set your threshold at 6), and range (distribution of bundles or the number of texts in which the identified bundles occur)
4. Generate the list: a frequency-ranked list of all n-grams meeting your criteria
5. Sort by frequency to see the most common bundles
6. The result will often show a long list of bundles (Figure 8). Copy and paste the results into an Excel file to help organize and filter the data. From the total list of the retrieved lexical bundles, you could exclude those that are incomplete or cross sentence/clausal boundaries. Also, identify and manage overlapping sequences to avoid inflating frequency counts (Chen & Baker, 2010).
7. Analyze the bundles based on their structure, function, or both, and use the concordance function to examine them in context

Figure 8
AntConc: N-Gram



POS-tagged Corpus Analysis

After preparing your POS-tagged corpus files, you can load them into AntConc for part-of-speech analysis. This section shows how to set up AntConc to work with POS tags annotated via the Stanford Parser.

Setting up POS-tagged corpus in AntConc

1. Open AntConc and navigate to File → Open Corpus Manager
2. Create a new corpus by clicking "Corpus Source" → "Raw File(s)"
3. Name your corpus (e.g., RA_POS_Tagged)
4. Add your POS-tagged files from your corpus folder using "Add File(s)" or "Add Directory"
5. In Basic Settings, click the arrow next to "Indexer, Encoding, Token Definition, Row Processor"
6. Change the indexer from "simple_word_indexer" to "simple_word_pos_headword_indexer"
7. Click "Create" to establish the corpus with POS processing enabled and return to main window
8. Go to Settings → Global Settings → Tags → configure Display Type options:
    - **Type:** Shows only the word (e.g., student)
    - **Type+POS:** Shows word with POS tag (e.g., student_NN)
    - **Type+Headword:** Shows word with its base form (e.g., running_run)
    - **Type+POS+Headword:** Shows word, POS tag, and base form (e.g., running_VBG_run)
    - **POS:** Shows only the POS tag (e.g., NN)
    - **Headword:** Shows only the base form (e.g., run)

- **Headword+POS:** Shows base form with POS tag (e.g., run_VBG)

For most POS analysis tasks, select "Type+POS"

9. Click "Set for all tools" to apply the configuration across AntConc
10. Click "Apply" to save the settings

Using POS Tags for Analysis

With POS-tagged corpora loaded, you can perform targeted analyses using specific search patterns:

Part-of-Speech Specific Word Lists: Use the Word tool with POS search patterns to generate frequency lists for specific grammatical categories. In the following categories, an * is used to represent any number of characters in the word, and the Penn Treebank tagset was consulted to identify target tags. For Example, in the first pattern below "*_" allows for any word before the "_", "NN*" allows for all of the noun tags to be included, as they all *start* with NN, but some of them end with other characters.

- Search for *_NN* to display all nouns
- **Search for * _JJ* to display all adjectives** (Figure 9)
- Search for *_VB* to display all verbs
- Search for *_IN* to display all prepositions

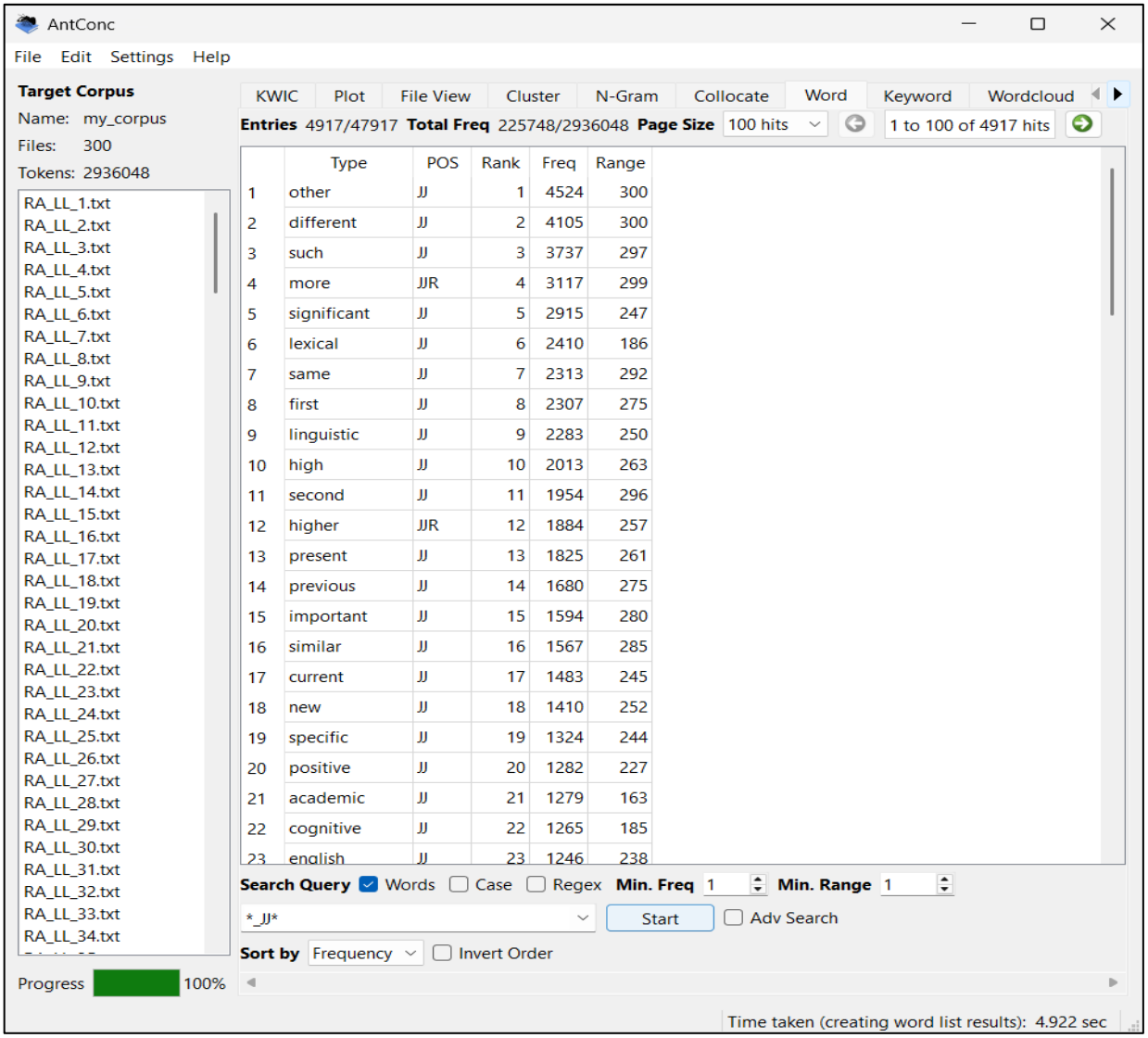
Collocate Analysis with POS Filtering: Use the Collocate tool to search for specific POS patterns around target words:

- Use *_JJ* student* to find all adjectives within the designated window of "student" (Figure 10)

Complex Pattern Searches: Use the KWIC tool to combine multiple POS tags and identify specific grammatical structures:
*_JJ * *_NN * finds all adjective-noun sequences (Figure 11)

Figure 9

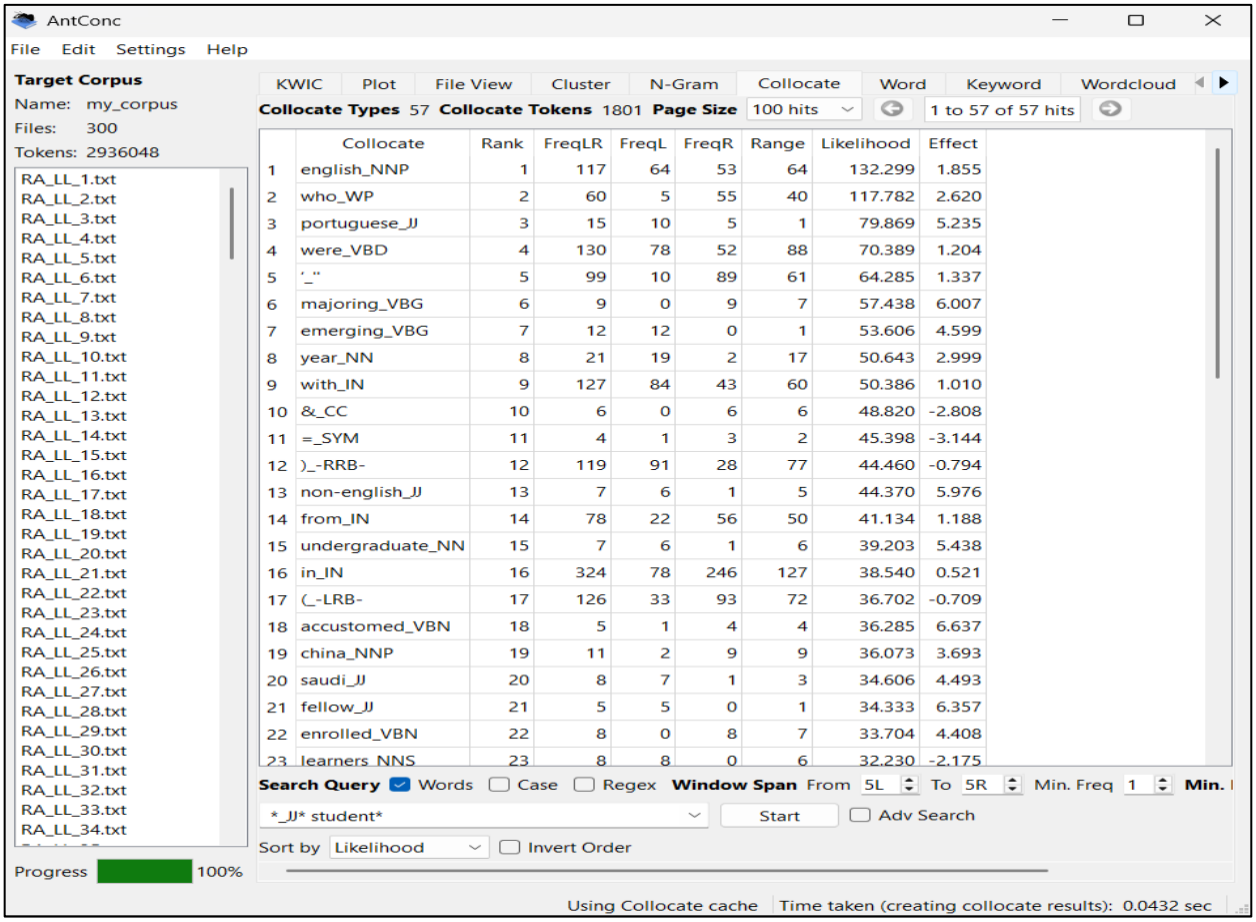
*AntConc Adjective Frequency List (*_JJ*)*



	Type	POS	Rank	Freq	Range
1	other	JJ	1	4524	300
2	different	JJ	2	4105	300
3	such	JJ	3	3737	297
4	more	JJR	4	3117	299
5	significant	JJ	5	2915	247
6	lexical	JJ	6	2410	186
7	same	JJ	7	2313	292
8	first	JJ	8	2307	275
9	linguistic	JJ	9	2283	250
10	high	JJ	10	2013	263
11	second	JJ	11	1954	296
12	higher	JJR	12	1884	257
13	present	JJ	13	1825	261
14	previous	JJ	14	1680	275
15	important	JJ	15	1594	280
16	similar	JJ	16	1567	285
17	current	JJ	17	1483	245
18	new	JJ	18	1410	252
19	specific	JJ	19	1324	244
20	positive	JJ	20	1282	227
21	academic	JJ	21	1279	163
22	cognitive	JJ	22	1265	185
23	english	JJ	23	1246	238

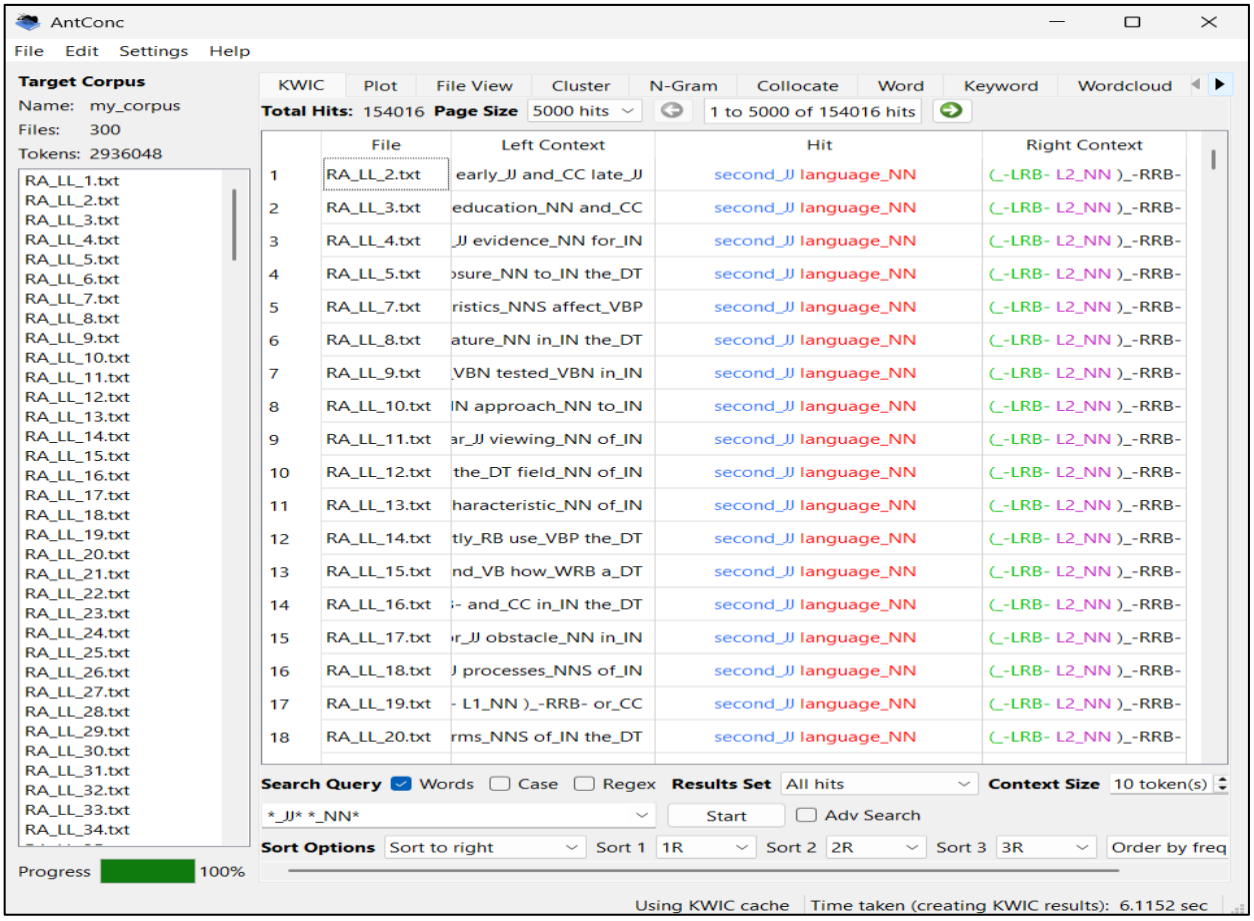
Figure 10

*AntConc Collocate Results for (*_JJ* student*)*



	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	english_NNP	1	117	64	53	64	132.299	1.855
2	who_WP	2	60	5	55	40	117.782	2.620
3	portuguese_JJ	3	15	10	5	1	79.869	5.235
4	were_VBD	4	130	78	52	88	70.389	1.204
5	'_"	5	99	10	89	61	64.285	1.337
6	majoring_VBG	6	9	0	9	7	57.438	6.007
7	emerging_VBG	7	12	12	0	1	53.606	4.599
8	year_NN	8	21	19	2	17	50.643	2.999
9	with_IN	9	127	84	43	60	50.386	1.010
10	&_CC	10	6	0	6	6	48.820	-2.808
11	=_SYM	11	4	1	3	2	45.398	-3.144
12	)_-RRB-	12	119	91	28	77	44.460	-0.794
13	non-english_JJ	13	7	6	1	5	44.370	5.976
14	from_IN	14	78	22	56	50	41.134	1.188
15	undergraduate_NN	15	7	6	1	6	39.203	5.438
16	in_IN	16	324	78	246	127	38.540	0.521
17	(_-LRB-	17	126	33	93	72	36.702	-0.709
18	accustomed_VBN	18	5	1	4	4	36.285	6.637
19	china_NNP	19	11	2	9	9	36.073	3.693
20	saudi_JJ	20	8	7	1	3	34.606	4.493
21	fellow_JJ	21	5	5	0	1	34.333	6.357
22	enrolled_VBN	22	8	0	8	7	33.704	4.408
23	learners_NNS	23	8	8	0	6	32.230	-2.175

Figure 11
*AntConc KWIC Results for (*_JJ * *_NN *) Pattern Search*



Phrase-frame Analysis with KfNgram

Another interesting kind of multiword sequences is called phrase-frames. Phrase-frames are discontinuous sequences with a variable slot that can be filled by semantically coherent variants (Fletcher, 2012). We recommend using kfNgram, a free downloadable Windows software, which you can download from http://www.kwicfinder.com/kfNgram/kfNgramStopwords.exe. More information about kfNgram and its functions can be found here: https://www.kwicfinder.com/kfNgram/kfNgramHelp.html.

First, you need to combine your text files into a single file in order to extract all word n-grams and phrase frames from one unified source. The following will show you how to do this on both Windows and Mac.

To combine files on Windows:

1. Click on the folder containing the files you want to combine.

2. Hold down Shift and right-click in the folder.
3. Click Open PowerShell window here
4. Type Get-Content *.txt | Set-Content new file.txt and press Enter to combine (Figure 12)
5. Check the folder. You will find a new file called “newfile.txt” created there with the combined contents.

To combine files on Mac:

1. Click on the folder containing the files you want to combine.
2. Hold down Shift and right-click in the folder.
3. Click New Terminal at Folder
4. Type cat *.txt > combined.txt (Figure 13)
5. Check the folder. You will find a new file called “combined.txt” created there with the combined contents.

To extract phrase-frames:

1. Open kfNgram
2. Set your parameters: decide on the N-gram size (usually 4-6 words), floor 1, not case-sensitive, replace .,-’ with space, frequency sort, retain numerals, separate
3. Add the source file (the combined text file)
4. Go to Tools, and select Get Wordgrams (Figure 14)
5. Open a new kfNgram window, use the same settings: floor 1, not case-sensitive, replace .,-’ with space, frequency sort, retain numerals, separate
6. Add the source file (the Wordgram file generated by kfNgram in the previous step), if it does not show you the file (No items match your research), check the file type and change it from Alpha n-Gram files to frequency nGram files
7. Go to tools and select Get Phrase-Frames from Wordgram file (Figure 14)
8. From the output text file, copy only the phrase frames that match your minimum frequency threshold (e.g., those which have 3 or 5 occurrences) and paste them into an Excel file. Note that there are two numbers next to each phrase frame. The first number indicates how many times the phrase frame appears in the corpus. This is the one you need to check for minimum frequency. The second number tells you how many types of variants fill this phrase frame. Then, you have the number of occurrences for each variant.
9. In the Excel file, go through the list and clean the frames and variants. When cleaning them, remove those that contain repeated words, incomplete patterns, or nonsensical word combinations. Focus on frames that are grammatically sound, meaningful, and meet the predetermined frequency and range threshold. Also, make sure to identify and manage overlapping sequences. Start by cleaning the frames, then clean the variants, and create a separate sheet for each round of cleaning. As you clean the phrases and the variants, use AntConc along with all the individual text files (the ones you combined to extract the phrase frames) to check the context and the range (number of files in which the phrase occurs). As you clean, update the frames, variants, and their frequencies.
10. Analyze the results based on their structure, function, or both, and use the concordance function to examine phrase frames in context

Figure 12
Windows PowerShell: Combine Files

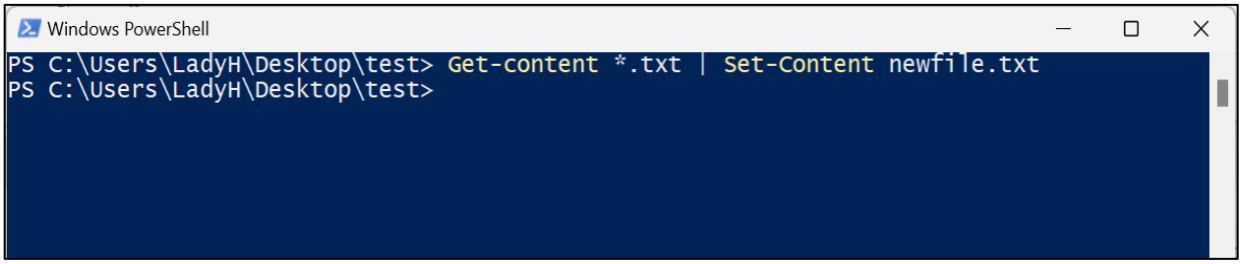


Figure 13
Mac Terminal: Combine Files

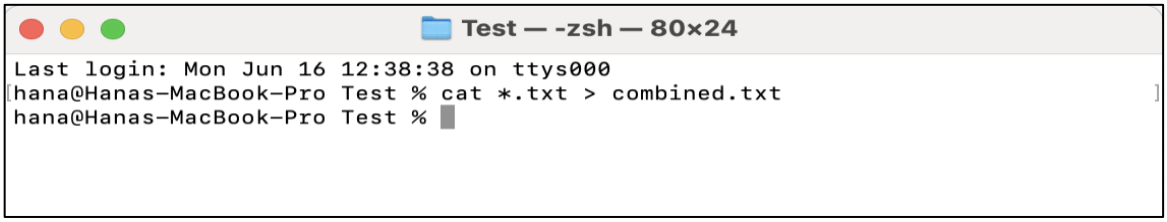
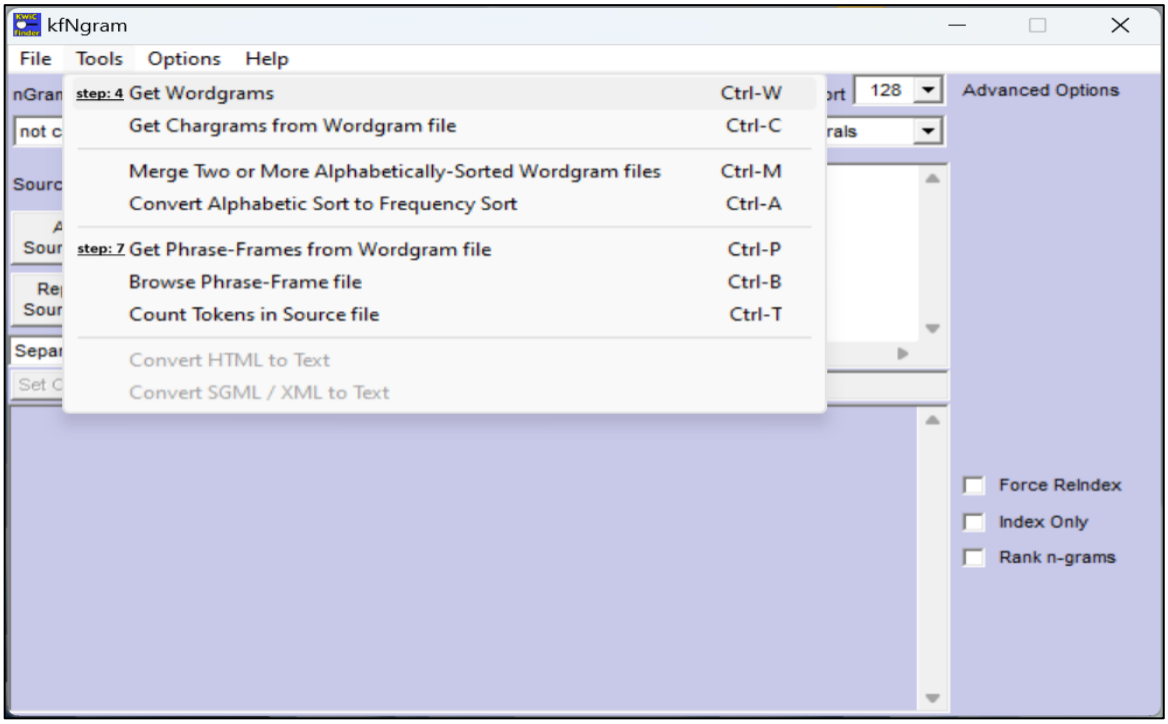


Figure 14
KfNgram: Phrase Frames Extraction_ Steps 4 & 7



CONCLUSION

The step-by-step instructions provided in this guide for selecting existing corpora or building custom ones and then analyzing them linguistically enable language teachers to enhance their pedagogical practice in multiple ways (O'Keeffe & McCarthy, 2022). In materials and curriculum development, teachers can use corpus data to develop vocabulary and grammar content based on frequency patterns and authentic usage rather than intuition-based approaches. Teachers can apply the same techniques to collect and analyze their students' output, e.g., writing samples, to identify common errors and discover what grammatical structures students have mastered and which areas need attention. This analysis enables them to design lessons that address actual needs rather than relying on traditional sequencing. In classroom applications, teachers can implement Data-driven learning (DDL) tasks where students use concordancing software such as AntConc to learn about language use by analyzing teacher-prepared concordances or by independently exploring corpora. The typical DDL tasks present concordances with guiding questions to recognize patterns and draw conclusions. Teachers can design tasks to teach specific linguistic features or raise awareness of how certain words or phrases function in context. In test development, teachers can use corpus data to write test items appropriate for students' proficiency levels and based on real language use. For example, by analyzing learner corpora, teachers can determine which vocabulary, grammar, and language patterns are typical of each proficiency level and select appropriate ones for their tests. These applications demonstrate that corpus analysis can inform and develop various aspects of language teaching.

In this guide, we have shown that building, cleaning, and analyzing corpora is systematic and manageable. The step-by-step instructions we provided make corpus work accessible to language teachers and novice researchers. The instructions in this guide can be used as a starting point, but expect to adapt them based on any teaching and/or research needs.

We encourage new analysts to start with small projects to build confidence before working with larger corpora. Corpus linguistics is an iterative process. The first corpus teaches lessons that improve the second project. Each analysis helps build understanding for what works best for specific research questions. Any challenges or problems that arise should not be seen as failures, but as a part of the learning process. Over time, it becomes clear that some cleaning decisions need changes, that certain analytical approaches work better than others, and that corpus work often leads to unexpected findings.

ACKNOWLEDGMENTS

We would like to thank Dr. J. Elliott Casal for his guidance and mentorship in learning how to do corpus-based research.

Authors

Ghadi Matouq is a Doctoral Candidate in Applied Linguistics at the University of Memphis and Lecturer in the English Department at Taif University. She holds a Master's in Applied Linguistics and Teaching English in International Contexts (TEIC) certificate from Texas Tech University. Her research interests include corpus linguistics, professional discourse, and computer-assisted language learning.

Hana Alqabba is a Doctoral Candidate in Applied Linguistics at the University of Memphis and Lecturer in the Department of English Language and Translation at Qassim University. Her research interests include corpus linguistics, genre analysis, and academic writing, with a focus on discipline- and paradigm-specific writing practices and literacies.

REFERENCES

Anthony, L. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software/AntConc

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*, 30–49.

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Fletcher, W. H. (2012). *KfNgram*. Annapolis MD: USA.

Foster, M. E., & Oberlander, J. (2007). Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Language Resources and Evaluation*, *41*, 305–323.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, *20*(3), 341–367.

Hyland, K. (2008). 'Small bits of textual material': A discourse analysis of Swales' writing. *English for Specific Purposes*, *27*(2), 143-160.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July). The TenTen corpus family. In 7th international corpus linguistics conference CL (Vol. 2013, pp. 125-127).

Lu, X. (2014). *Computational methods for corpus annotation and analysis*. Springer.

McCarthy, M., & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3–13). Routledge.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Newman, J., & Cox, C. (2020). Corpus annotation. In M. Lefer, M. Paquot, & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 25-48). Springer.

O'Keeffe, A., & McCarthy, M.J. (Eds.). (2022). *The Routledge handbook of corpus linguistics* (2nd ed.). Routledge. https://doi.org/10.4324/9780367076399

Reppen, R., & Simpson, R. (2002). Corpus linguistics. In N. Schmitt (Ed.), *An Introduction to applied linguistics* (pp.89–106). Arnold.

Reppen, R. (2022). Building a corpus: What are key considerations?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 13-20). Routledge. https://doi.org/10.4324/9780367076399

Sinclair, J. (2004). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*. Oxford. https://doi.org/20.500.14106/dlc

Sinclair, J., Jones, S., & Daley, R. (2004). *English collocation studies: The OSTI report*. Bloomsbury Publishing.

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice* (1st ed.). Routledge. https://doi.org/10.4324/9781315715537

Thompson, P. (2004). Spoken language corpora. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice*. Oxford. https://doi.org/20.500.14106/dlc

APPENDIX

Importance of Text Preparation and Cleaning

Converting raw textual data into a clean, standardized, and analysis-ready format is a crucial task before analyzing spoken or written language (Reppen, 2022). The following problems demonstrate why proper text preparation and cleaning are essential:

- **File organization failures:** As a first attempt you may want to put multiple texts in a large file, but this limits your analysis options because you cannot analyze texts individually (e.g., students' essays), group them by different criteria (such as gender, proficiency level, or time period), or combine them as needed for different research questions. This forces you into the time-consuming process of later dividing large files and resaving them with new names or rebuilding the corpus.
- **Poor naming conventions:** File names that do not clearly reflect the content make finding and organizing texts extremely difficult, while longer names can create compatibility issues across analytical software and backup systems.
- **PDF conversion issues:** Converting PDF files to text format can disrupt paragraph sequencing and scramble words or characters. Tools like Wondershare (https://shorturl.at/3IBqp) allow PDF editing and conversion to TXT format but may sometimes disrupt paragraph sequencing. Voyant Tools (https://shorturl.at/SSXZM) preserves paragraph sequencing during conversion, and TextFixer (https://shorturl.at/pe8jE) along with Capitalize My Title (https://shorturl.at/saRm8) help remove unwanted line breaks from texts. Generally, you need to compare with the original PDF file to check for any conversion errors or formatting issues.
- **Data loss risks:** Without backup copies of texts in multiple secure locations, you risk losing files from computer crashes, fires, or theft. These things do happen.
- **Missing metadata:** Lost headers or metadata means you lose valuable contextual information for future analysis, and if headers are not properly enclosed with angle brackets and separators, this information could be accidentally included in text analysis, which could inflate frequency counts.
- **Format incompatibility:** Files saved in incompatible formats create significant problems during analysis and require additional processing time, as plain text or UTF8 formats work best with corpus analysis tools.

- **Non-standard language handling:** When working with written texts that contain non-standard language, such as learner language, novice writing, or children's writing, you may want to create two versions of each text. An original version preserves all spelling and grammar structures because they may be valuable for certain types of analysis, and a clean version with standardized spelling for other analyses.
- **Poor recording quality:** When building spoken corpora, failing to ensure high-quality recordings using digital devices such as phones and tablets results in unclear sound that does not transfer easily to computers and creates major transcription difficulties.
- **Inconsistent transcription decisions:** You must make important decisions about how to handle reduced speech forms, unclear audio segments, overlapping speech, and prosodic elements consistently throughout the corpus.

Applied Language Learning

Research

Spanish Pronunciation Modeling by Presenters of Various Language Learning Backgrounds

Capt M. Ryan Richey
United States Air Force Academy

Jean W. LeLoup
United States Air Force Academy

The purpose of this study is to determine if heritage speakers can effectively instruct pronunciation to their non-native speaker peers to support faculty. The researcher of this study assessed non-native speaker participants' pronunciation of Spanish using pronunciation instruction and instructor background as variables (N=111: 91 underwent instruction, 20 in control group). Participants underwent a pretest. The experimental group then received pronunciation instruction addressing phonemes that do not exist in American English or that manifest differently in the target language, such as the trilled "r" and consonant combinations such as "br," focusing especially on those with a relatively higher functional load. The control group did not receive concentrated pronunciation instruction. Posttest results showed improvements in accuracy of phoneme pronunciation. Results from the delayed posttest showed a slight decline in accuracy among students of native speaker teachers and non-native speaker teachers, however, a slight rise among those of heritage speaker student peer presenters and the control group. This quasi-experimental research has implications for teachers of all languages in that they could potentially be able to leverage heritage- and native-speaker students as resources in the classroom while providing them with instructional leadership opportunities without a sacrifice in instructional quality.

Keywords: *Pronunciation, Spanish, Peer Teaching, Heritage Speakers, Native vs. Non-Native Speaker Teachers*

INTRODUCTION

There is debate in the world language instruction community regarding the pronunciation instruction effectiveness of native speaker teachers (NSTs) and non-native speaker teachers (NNSTs); traditionally-held views suggest that NSTs are superior to NNSTs in this area; however, given the occasional absence of NSTs, heritage speaker students may be able to help close this perceived gap. Conventional wisdom may suggest that NSTs are preferable because their accent may be more authentic, and they might have had their foundational education in the target language. Curiously, results from other studies suggest that instructor background may not influence the pronunciation results of their students (Levis, 2016). It seems, then, that NNSTs may inspire less confidence among their students than their NST colleagues, regardless of results. Therefore, NNSTs may find themselves with a need to influence their students' perceptions to build language learners' confidence in their instructional abilities.

Heritage speakers are those who have grown up speaking the target language in the home with their families but did not receive a formal education in that language – many times, heritage speakers are less strong in reading, writing, grammar and spelling, but thrive in listening, speaking, and pronunciation (Correa, 2011). Heritage speaker students in the classroom could serve as a valuable resource for NNSTs to impart pronunciation contents to the rest of the class in a way that increases the confidence of students, allowing NNSTs to overcome doubts that their students may have in their teacher's abilities. Surveys from the physics field suggest that students perceive peer instructors to be acceptable instructors (Braun & Kirkup, 2016). The question therefore arises as to whether heritage speaking peers can serve as presenters for their classmates, particularly in the area of pronunciation. The central idea of this study is to incorporate heritage speaking students into the lesson by allowing them to present contents, while at the same time, supplementing pronunciation instruction to generate confidence among students in refining their own pronunciation. The focus of the pronunciation for this project was on phonemes (the individual sounds that a language is composed of), namely because of the relative simplicity for heritage speakers to instruct others with specific sounds and also the ability of the researchers to clearly isolate the students' accurate or inaccurate production on the tests.

LITERATURE REVIEW

Determining the effect of having heritage speaker peers present pronunciation topics on the actual pronunciation of language learners embodies many layers of complexity. First, it is necessary to define a heritage speaker or heritage language learner (HLL). A general definition indicates that HLLs have been exposed to a language (e.g., Spanish) outside the regular educational system, be it at home, through contact with relatives, friends, neighbors, or others. These students are in classrooms where English is the language of instruction. They are therefore bilingual to some degree or another (Correa, 2011). Here, then, is the first distinction to be noted when working with this population of students. HLLs comprise a wide variety of linguistic abilities and knowledge of their heritage language (Martin-Beltran et al., 2020).

A second area needing attention is the perception that HLLs have of their first language. By default, English has the hegemonic position in U.S. schools as the main language of instruction. Whether explicitly stated or tangentially implied, this depreciation or disparaging of HLLs' mother tongue often has a negative sociolinguistic effect; HLLs often prefer to use English to the detriment and disregard of their heritage language. This subtractive bilingualism can have negative consequences for the learner in terms of self-esteem, relationships within the family, and a deficit in learning the complexities of either language (Babino & Stewart, 2017; Lightbown & Spada, 2021). Employing HLLs to aid in the language classroom shows their language abilities in a positive and even esteemed light, instilling pride in their prowess in that language and contributing to their sense of belonging and who they are (Babino & Stewart, 2017; Correa, 2011; Martin-Beltran et al., 2020; Reznicek-Parrado, 2024). This new role for HLLs can also engender language comparisons between dialects. These discussions provide a fitting example of the 5 Cs (communication, cultures, connections, comparisons, and communities) of ACTFL's Comparisons Standard of the World-Readiness Standards for Learning Languages (González-Bueno, 2021; The National Standards Collaborative Board, 2015), which "stress the application of learning a language beyond the instructional setting" (ACTFL, 2022).

Another issue at play in this study is the concept of NSTs and NNSTs. While this duality has become the standard paradigm for studies investigating language teacher identify, it is anything but precise. How native is native? Native from where? Native in language production but not in cultural knowledge? Such queries indicate the complexity of this comparison (Aneja, 2016). Indeed, the NST idea frequently disadvantages NNSTs, even though they may have attained a high level of L2 ability and demonstrated themselves as perfectly proficient multilinguals (Calafato, 2019). This in no way denies there are hurdles that NNSTs must surmount when teaching language. There is a general lack of training in language teacher training programs on how to teach pronunciation, no matter the language. In addition, many NNSTs have a certain amount of insecurity vis-à-vis their own pronunciation due to not being native speakers of the language. Oftentimes students indicate a preference for NSTs but are hard-pressed to define why. Interestingly, research has shown that despite these hurdles, the teacher's language background does not significantly affect students' learning overall (Levis at al., 2016; Olson, 2014).

Acknowledging the above caveats, the present study seeks to expand the research database by adding the heretofore uninvestigated variable of HLLs as peer teachers of pronunciation. There is some evidence that peer teachers do obtain positive results (Braun & Kirkup, 2016), but this was in an undergraduate science lab course. With attention paid to the complexities of the topic, results of this study will establish a research-based foundation to the use of HLLs in the Spanish classroom. The understood premise is, of course, that target language pronunciation instruction is effective (Thomson & Derwing, 2015). Thus, the research question is: *Is there a difference in production gains between three groups of instructors: native speaking teachers, non-native speaking teachers, and heritage-speaking peer presenters?*

METHODS

Participants

The personnel included in the study, to include both experimental groups and control groups, consisted of participants (*N*=111; 91 underwent instruction: 27 with NNSTs, 27 with NSTs, 37 with heritage speaker student presenters; 20 in control group), course instructors, and peer presenters. Participants ranged between 18 and 23 years of age with the majority (60%) self-reporting to be 19 years old at the time of the study. Almost all (97%) of the participants reported their first language to be English, with 2% reporting having Spanish as a heritage language and 1% reporting having Tagalog as a heritage language.

With respect to the seven Spanish instructors, three were native speakers from Mexico, Spain, and Peru. The remaining three were non-native speakers of Spanish who had spent considerable time living in Spanish-speaking countries.

The process to select peers to instruct pronunciation lessons included a request for recommendations from Spanish instructors from the previous semester. A subsequent call for volunteers was then made among that identified group. Then, the researcher evaluated the pronunciation of the potential heritage speaker student presenters (HSSPs) using the study's pretest to evaluate their ability to produce the targeted phonemes: standard manifestations from a majority of Spanish-speaking regions (/a/, /e/, /i/, /o/, /u/, /r/, and /r/). The researcher also determined their instructional ability (interpersonal skills, ability to explain targeted concepts) through a character interview. Thirteen students were initially recommended and five were finally selected.

The five peer instructors had very similar backgrounds. All were attending the same university-level military service academy, all reported that Spanish was spoken at the home during their childhood, but that their primary language was English, all reported their ethnicity to be Latino or Hispanic, and all reported their age to be between eighteen and twenty-three. The did differ in gender however, with four reporting male and one reporting female.

Conceptual Framework

In her 2021 article, Manuela González-Bueno combines VanPatten's Processing Instruction (1996) and Adair-Hauck and Donato's PACE Model (Presentation, Attention, Co-construct, and Extension) (2002) to produce S-PACE: an instructional model used to usher students through a logical learning process that facilitates pronunciation instruction. The S-PACE method serves as the conceptual framework for the instruction in this study (see Table 1).

Table 1
S-PACE Instruction Model

Phase	Description
Presentation	Students are given an example of discourse containing the targeted sound.
Attention	Students' focus is drawn to the targeted sound within the discourse.
Co-construction	Instructor guides students to develop an explanation as to how the targeted sound is pronounced and why it is difficult for learners.
Structured Input	Students perform exercises in which they answer questions depending on the accurate perception of controlled input.
Structured Output	Students respond to prompts using words containing the targeted sound.
Extension	Students are asked to generate spontaneous speech and accurately produce the targeted sound.

This framework is utilized in this study during the instructional stage in which phonemes are introduced with the experimental group. The framework provided the structure and methodology for those specific lessons. See the Appendix for an example of a specific lesson.

Procedure

Following IRB approval for human subjects testing, the study took place during the spring semester of the 2023-2024 academic year at the United States Air Force Academy (USAFA) and followed a progression that allowed participating instructors and students to incorporate evaluations and instructional contents into the normal pace of the course of study. To begin, the researcher chose the first year of the Spanish sequence because of novice-level ability of the students, the idea being that these may be students whose pronunciation ability would be limited enough that any improvements realized during the study would be readily apparent and easily measurable. Then, within the group of first-year classes, class sections were chosen based on the following criteria: presence of heritage speaker students with adequate pronunciation in class, NSTs with adequate pronunciation, and NNSTs.

The researcher began by identifying phonemes to be targeted according to their uniqueness in Spanish versus English (such as the trilled /r/) or their differing manifestation in Spanish versus English (such as the flap /ɾ/), focusing especially on phonemes with relatively higher functional load, meaning those phonemes whose variation could drastically change the meaning of the word in question (e.g., *carro* "car" versus *caro* "expensive"). These phonemes are frequently manifested with non-nativelike production among speakers of American English (see Table 2).

Table 2
Targeted Phonemes

Word	Elicited Phoneme	Frequent American English Speaker Manifestation	Position
Asturias	/ɑ/	/ə/	Initial Unstressed
N**a**varra	/ɑ/	/ə/	Unstressed
Zamor**a**	/ɑ/	/ə/	Final Unstressed
Alicant**e**	/e/	/ei/	Final unstressed
Al**i**cante	/i/	/ɪ/	Unstressed
Madr**i**d	/i/	/ɪ/	Stressed
Toled**o**	/o/	/ou/	Final Unstressed
Universidad	/u/	/juː/	Initial Unstressed
Za**r**agoza	Flap /ɾ/	/ɹ/	Intervocalic
G**r**anada	Flap /ɾ/	/ɹ/	Consonant Combination
Cantab**r**ia	Flap /ɾ/	/ɹ/	Consonant Combination
Pontevedr**a**	Flap /ɾ/	/ɹ/	Consonant Combination
Rioja	Trilled /r/	/ɹ/	Initial
Navar**r**a	Trilled /r/	/ɹ/	Intervocalic

Best practices from previous pronunciation studies were included in the current study, such as understanding student demographics, documenting procedures to allow for replication, recruitment for statistically significant sample sizes, the inclusion of a control group, and the implementation of a delayed posttest (Thomson & Derwing, 2015).

The study began at the beginning of the spring semester with participants (*N*=111) receiving instructions by email to use *Audacity* (Mazzoni & Dannenberg, 2000) audio recording software on their school-issued laptop computers. They were instructed to record themselves using headphones while pronouncing 15 Spanish words (names of locations in Spain) leaving a three-second pause between each. The pretest took place in the language lab at the USAFA Department of Languages and Cultures. The participants were asked to record themselves saying these 15 utterances with phonemes at varying positions within words in stressed or unstressed syllables.

The instruction phase of the study consisted of 11 five-minute modules that HSSPs, NSTs and NNSTs presented to the experimental groups during the first segment of class over a three-week period. The first module provided the participants with an introduction to the study (e.g., the purposes, structure). The subsequent 10 modules each focused on a different targeted phoneme. The structure of these modules adhered to the S-PACE model with the intention of providing the participants with the purpose of the lesson, calling their attention to the differences between nativelike and non-nativelike pronunciation, and leading them through practice in pronouncing the targeted phonemes (Gonzalez-Bueno, 2019).

The researcher provided training to all presenters (NST, NNST, and HSSP) on the contents of each module prior to instruction. Within the S-PACE construct, all instructors and presenters executed each lesson as they deemed fit, which gave each of them the ability to leverage their individual strengths as necessary to achieve the study's overall goal of facilitating the participants' progress in pronunciation. Following the instructional phase, a posttest was administered with the identical phonemes and format as the pretest to be able to compare results between these two tests. Results from the posttest provided immediate results that gauged participants' learning following the three-week instructional period.

Afterward, participants underwent a two-month period in which pronunciation instruction was limited to isolated points during vocabulary lessons, periodic corrections and emphasis on the part of individual instructors. This period was intended to serve as a continuation of the extension phase of S-PACE in which participants were able to apply their pronunciation learning to scripted and non-scripted production activities (Gonzalez Bueno, 2019). Moreover, this period was designed to allow participants to further strengthen muscles required to produce more muscle-intensive sounds such as the trilled /r/.

Finally, the study concluded with a delayed posttest that instructors administered at the end of the semester. Once again, this delayed posttest had the identical contents and format as the pretest and posttest to be able to compare results. Results from the delayed posttest showed the participants' retention of the contents of the three weeks of S-PACE pronunciation instruction and strengthening of muscles in the mouth associated with non-American English sounds following two months of extension phase.

For data analysis purposes, the researcher compiled all audio recordings of all participants in order of pretest, posttest, and delayed posttest. He annotated their accurate or inaccurate execution of each targeted phoneme in an Excel spreadsheet – pronunciation was rated for accuracy in comparison to pronunciation of phonemes specifically selected for their standardized manifestations across the majority of Spanish-speaking regions (/a/, /e/, /i/, /o/, /u/, /r/, and /r/). This manual assessment proceeded one participant at a time from pretest to posttest to delayed posttest to most accurately assess individual progress throughout the course of instruction. Four additional raters screened recordings, yielding an interrater reliability of 81%, or "good" (Koo, 2016). Additional raters were instructors of the same Spanish course.

Finally, participants were administered a questionnaire on *Google Forms* to gather participant demographics (see Table 3).

Table 3
Participants' Self-Reported Demographics

Participant Demographic Responses	Percentage
Male	69%
Female	31%
White	76%
Latino or Hispanic	7%
Multiracial	6%
Asian	5%
African American	4%
Native Hawaiian / Pacific Islander	2%

RESULTS

Given that the purpose of this study is to determine if there is a difference in production gains between three groups of instructors (NSTs, NNSTs, and HSSPs), the researcher compiled the audio recordings and analyzed the execution of the phonemes for accuracy one participant at a time and in chronological order of recording (pretest, the posttest, then delayed posttest) and categorized them by presenter background. Change is shown using posttest score minus pretest score in one column and delayed posttest score minus pretest score in the other (see Table 4).

Table 4
Results by Presenter Background, Comparing Posttest and Delayed Posttest Improvement Over Pretest

Presenter Background	Number of Participants	Participant Posttest Improvement	Participant Delayed Posttest Improvement
Experimental—Received Instruction			
Heritage Speaker Student Presenter (HSSP)	37	14.33%	15.91%
Native Speaker Instructor (NST)	27	11.38%	10.13%
Non-Native Speaker Instructor (NNST)	27	11.53%	09.23%
Control—No Pronunciation Instruction	20	00.40%	02.18%

With respect to the research questions, if there is a difference in production gains between three groups of instructors (NSTs, NNSTs, and HSSPs), findings indicate that participants in the classes of each instructor or presenter backgrounds have a similar percent improvement at the posttest and a slightly differing percent improvement at the delayed posttest. The researcher compiled pretest results from each group as a baseline, measured the percentage of improvement to the posttest and then likewise to the delayed posttest. Improvement from pretest to posttest was similar for experimental groups. After three weeks of instruction, participants of experimental groups demonstrated an improvement of between 11% and 15%. Improvement from pretest to delayed posttest was slightly more varied, with students of HSSPs retaining and improvement of more than 15%, those of NSTs around 10%, and those of NNSTs around 9%.

Regarding the apparent relative rankings of the participants of the various instructor backgrounds at the time of posttest, there is a significant difference that separates the control group from the experimental groups, but less so between each experimental group. The difference between the control group and the HSSP group, for example, was more than 10%, suggesting existence of learning given the population size (*N*=111; 91 underwent instruction, 20 in control group) and the instructional period of three weeks. However, the difference of improvement between the participants of HSSPs and NNSTs, for example is less than 3%, which is far smaller than the difference between the experimental and control groups.

Citing the phonemes that experienced the most improvement (see Table 5), improvement could have been driven by gains in a combination of the following: the final /o/ (as in “Toled**o**”, with 18% improvement), the intervocalic trilled /r/ (as in “Nava**rr**a”, with 25% improvement), the initial trilled /r/ (as in “**R**ioja”, with 16% improvement), or the initial unstressed /ɑ/ (as in “**A**sturias”, with 13% improvement).

Table 5

Results by Phoneme, Comparing Posttest and Delayed Posttest Improvement Over Pretest

Word	Elicited Phoneme	Position	Phoneme Posttest Improvement	Phoneme Delayed Posttest Improvement
Asturias	/ɑ/	Initial Unstressed	13%	18%
N**a**varra	/ɑ/	Unstressed	11%	22%
Zamor**a**	/ɑ/	Final Unstressed	13%	14%
Alicant**e**	/e/	Final Unstressed	7%	3%
Al**i**cante	/i/	Unstressed	6%	9%
Madr**i**d	/i/	Stressed	12%	13%
Toled**o**	/o/	Final Unstressed	18%	20%
Universidad	/u/	Initial Unstressed	7%	11%
Za**r**agoza	/ɾ/	Intervocalic	13%	14%
G**r**anada	/ɾ/	Consonant Combination	9%	8%
Cantab**r**ia	/ɾ/	Consonant Combination	2%	5%
Rioja	/r/	Initial	16%	15%
Nava**rr**a	/r/	Intervocalic	25%	25%

In the examination of delayed posttest results, it appears that students of HSSPs retained more (15.91%) as opposed to those of NSTs (10.13%) and NNSTs (9.23%). Since the instruction between the posttest and the delayed posttest did not vary between these groups, the reason for this difference is not immediately clear. More analysis regarding demographics, motivation, and perception would be needed to determine this variation.

DISCUSSION

The results of this study produced findings pertinent to the central research question regarding the efficacy of heritage speaker students as presenters, findings that suggest that the data is coherent, and may contribute to the progress of related areas of interest.

Regarding the participants' results by instructor background, there are several salient points. First, there is a notable difference between the experimental groups and the control group. The control group did experience an improvement in pronunciation which may be attributed to receiving aural input, speaking, receiving periodic instruction about the pronunciation of vocabulary, and occasionally being corrected by the instructor. The experimental groups, on the other hand, experienced improvements of between 11% and 13% more than the control group, which likely confirms previous findings that formal pronunciation instruction such as that of the S-PACE construct yields superior results over informal pronunciation instruction, as posited by Gonzalez-Bueno (1997).

Also notable is the proximity in results of participants in the classes of each instructor background: HSSPs with 14.33% of improvement, NSTs with 11.38%, and NNSTs with 11.53%. This may align with previous findings that instructor background may not have an influence on student pronunciation learning, as posited by Levis in 2016.

Furthermore, and most poignant to this present study in particular is that within those groups of instructor backgrounds are not just NSTs and NNSTs, but also HSSPs. Therefore, if results of HSSPs are similar to those of NSTs and NNSTs, then this background of presenter may prove to be an acceptable resource for pronunciation instruction.

There were also findings to be derived from the study that were not centrally focused on the research question, but rather that could have application in related areas. Regarding improvement from the perspective of individual phonemes, several observations become readily apparent. The intervocalic trilled /r/ (as in "Nava**rr**a") began with an average pretest accuracy score of 16% and rose to a posttest accuracy score of 41%, which represents a 25% improvement. This is the highest improvement for any phoneme measured in the study, and it is also possibly the most difficult Spanish phoneme to pronounce for native speakers of American English due to its lack of presence in their native language and the relatively high amount of dexterity required to execute it. This high level of improvement could be due to the relatively low initial score, leaving a high potential for gains during the study. Also experiencing high improvements was the initial trilled /r/ (as in "**R**ioja"), and vowels such as the initial unstressed /ɑ/ (as in "**A**sturias"), and the final /o/ (as in "Toled**o**"). The initial trilled /r/ began with an average pretest score of 39% and rose to 55% on the posttest for a gain of 16% - the initial unstressed /ɑ/ began with an average pretest score of 81% and rose to a posttest score of 94% for a gain of 13%, and the final /o/ began with an average pretest score of 69% and rose to a posttest score of 87% for a gain of 18% (the highest gain of any vowel measured in the study). It is possible that the improvements measured were due to the ease of correction: most of American English's homologous manifestations in Spanish already exist in English and could be simply a matter of calling the learner's attention to the application, as postulated with the Noticing Hypothesis (Schmidt, 2010).

Limitations

Although the general atmosphere of the institution among colleagues and students alike was positive, thoughtful and supportive of the research, the study was not without its limitations. Fortunately, none of these limitations seem to have affected the scholarly value of the findings.

First, it seemed that the participants may have been very supportive of the research effort, but they may have rushed through the pretest, posttest, or delayed posttest due to the pressure of other assignments being due the same day. At this military service academy, students' time is precious: their mornings are filled with physical training and briefings, their lunches are also frequently occupied by briefings, and their afternoons are filled with sports practices, homework, and study. It is possible that participants felt the need to rush through these recordings – some

that sounded more relaxed lasted around 50 seconds, and others that sounded more hurried only lasted around 20 seconds. In spite of this, the value of scholarly findings do not seem to have been affected in that the tests were carried out on days with similar schedules and therefore with similar time pressures – the tests were all affected approximately in an equal measure.

Also, technology doesn't seem to have advanced enough yet so that segmentals can be rated in an efficient, automated fashion by autonomous software. At this point, manual, human scoring (raters listening to audio recordings and annotating accuracy) is still required to determine the accuracy of each phoneme: eighteen phonemes times 111 participants times three tests equals 5,994 phonemes that had to be evaluated individually by human ears and rated by the human brain. Although this limitation may have affected the findings, the data did show an interrater reliability of 81%.

Finally, the three-week length of the intervention can also be seen as a limitation (Thomson & Derwing, 2015). Many other studies have longer periods of intervention with fewer targeted phonemes. The effect that this may have is that students may have not had sufficient time to effectively absorb the volume of content presented, as demonstrated by some of the decreases in scores at the time of the delayed posttest. Replications of this study may benefit from a reduction in the number of targeted phonemes, a longer intervention period, or both.

With that, the observation of these limitations could potentially be beneficial for follow-up studies and for practical application of findings at this institution or others.

Avenues for Future Research

Although the findings of the research do point to immediate solutions for language teachers to leverage heritage speaker students in the classroom to improve confidence among learners, there is also potential for much follow-up development. The results and conclusions from this study could help to identify avenues for future research.

Most importantly, further research is needed to discover if participants are more confident in contents presented by HSSPs as opposed to NNSTs, and the reason for that. Studies from the physics field suggest that participants may be more confident in instruction carried out by their peers as opposed to faculty instructors due to similarities in age, career choice, or background (Braun & Kirkup, 2016). Similar efforts do not exist for language instruction and efforts to uncover this information could serve to advance instruction methodology.

Also, a study in which the roles between HSSPs and non-native speaker students are switched could give interesting results regarding grammar instruction rather than pronunciation instruction. In the present study, peer presenters are leveraging their native-like pronunciation capabilities to help their non-native classmates improve theirs. A mirror image study would involve a similar screening of non-native students to discover the ones that possess relatively higher grammar capabilities to instruct heritage speaker students in grammar. One limitation to conducting this study at this institution in particular is that the scope of this study would be much

smaller, relatively: the researcher was only able to identify thirteen heritage speakers out of a pool of 173 students.

CONCLUSION

In summary, the central problem of this study lies in the premise that participants may prefer NSTs over NNSTs (Hertel & Sunderman, 2009). This preference could be based on the perception that NNSTs possess an imperfect or incomplete set of skills in the target language and therefore the contents they present may also have the same attributes. The result of this preference could make participants less confident in the contents that their NNSTs impart or less comfortable in imitating their NNST's pronunciation versus imitating that of a NST. What is curious about this circumstance is that other studies suggest that teacher background, be it NST or NNST, may not have an influence on participants' pronunciation results (Levis et al., 2016).

With that in mind, one resource that NNSTs have at their disposal that could inspire confidence among their students with the contents and emphasize accurate pronunciation in to produce advanced-level speakers is heritage speaker students. Studies from the physics field suggest that participants may view their peers as acceptable instructors in comparison to faculty instructors (Braun & Kirkup, 2016). The question then becomes whether the same could be true for the language classroom: initially, one must determine the efficacy of heritage speaker students. The present study suggests that heritage speaker students can in fact achieve similar results in pronunciation presentation as NSTs and NNSTs. Furthermore, one must consider the level of speaker that his or her language program seeks to produce. If the goal is to produce an intermediate mid speaker, then an emphasis on communicative objectives at the expense of form allows for the program to make efficient use of time. However, if the program's goal is to produce speakers at the intermediate high level or above, accurate pronunciation is a requirement for concise communication without the necessity of circumlocution (ACTFL, 2024), and it could be useful for the students to be confident in the nativelike nature of the contents being presented. Studies have suggested that the average language instructor spends eight minutes per week on pronunciation, and that instruction tends to be ad-hoc and impromptu in nature (Olson, 2014). This lack of attention to pronunciation instruction could be due to what studies suggest is a lack of confidence in the ability to effectively provide pronunciation instruction and a skepticism of its usefulness due to recent emphasis on communicative methodologies versus a focus on form (Nagle et al., 2018). However, a body of studies suggests the effectiveness of including deliberate pronunciation instruction into curriculum (Lord, 2010). Many language programs do seek to develop students to the intermediate high level or above, and therefore, intelligible pronunciation could be a valuable asset for students to strive to attain – heritage speaking students can help NNSTs get their students to that level.

Assigning heritage speaker students to present pronunciation contents is important for middle schools that want to start their students out with good speaking habits, high schools that want to get heritage speakers on board with the lesson, and universities that want to produce advanced-level speakers. More research is required in this area to determine how NNSTs can

inspire more confidence in their students in the contents they present and how to create individualized plans for heritage and native speaker students so that their foundations are also solidified.

Acknowledgements

This study is the result of the teamwork of language instructors in the Department of Languages and Cultures at the United States Air Force Academy, concept development with Dr. Barbara Schmidt-Rinehardt, refinement with Dr. Manuela Gonzalez-Bueno, literature review by Dr. Jean LeLoup, and the encouragement and guidance of Dr. Pete Swanson – the researcher gratefully acknowledges their valuable contributions.

Disclaimer

The views expressed in this article, book, or presentation are those of the author and do not necessarily reflect the official policy or position of the United States Air Force Academy, the Air Force, the Department of Defense, or the U.S. Government.

Authors

Capt M. Ryan Richey is an instructor of Spanish in the Department of Languages and Cultures at the United States Air Force Academy. His research explores how heritage speaker students can contribute in the L2 classroom, language and ethnicity demographics within the armed services.

Jean W. LeLoup, Ph.D. (The Ohio State University), is Professor Emerita of Spanish in the Department of Languages and Cultures at the U.S. Air Force Academy, where she directed the Intermediate Spanish curriculum. Her areas of interest include the immersion classroom, the voseo and the linguistic landscape of Costa Rica, and proficiency benchmarking in the Spanish curriculum.

REFERENCES

ACTFL (2022). World-readiness standards for learning languages. *ACTFL Language Connects*. http://www.actfl.org/

ACTFL (2024).Proficiency Guidelines. *ACTFL Language Connects*. http://www.actfl.org/

Adair-Hauck, B., & Donato, R. (2002). The PACE model: A story-based approach to meaning and form for standards-based language learning. *The French Review, 76*(2), 265-276.

Aneja, G. (2016). (Non) native speakered: Rethinking (non) nativeness and teacher identity in TESIL teacher education. *TESOL Quarterly, 50*(3), 572–596.

Babino, A., & Stewart, M. (2017). "I Like English Better": Latin Dual Language Students' Investment in Spanish, English and Bilingualism. *Journal of Latinos and Education, 16*(1), 18–29. https://doi.org/10.1080/15348431.2016.1179186

Braun, M., & Kirkup, L. (2016). Non-physics peer demonstrators in undergraduate laboratories; A study of students' perceptions. *European Journal of Physics, 37*(1).

Calafato, R. (2019). The non-native speaker teacher as proficient multilingual: A critical review research from 2009–2018. *Lingua, 227*.

Correa, M. (2011, Summer). Advocating for critical pedagogical approaches to teaching Spanish as a heritage language: Some considerations. *Foreign Language Annals, 44*(2), 308–320.

Diaz, M., & Simonet, M. (2015). Second language acquisition of Spanish /e/ and /ei/ by native English Speakers. *Hispania, 98*(4), 750-761. https://muse.jhu.edu/article/605433

Gonzalez-Bueno, M. (1997). The effects of formal instruction on Spanish stop consonants. In W. Glass & A. Pérez-Leroux (Eds.), *Contemporary Perspectives on the Acquisition of Spanish* (pp. 57–75). Cascadilla Press.

Gonzalez-Bueno, M. (2019). Suggestions for teaching Spanish voiced stops/b, d, g/and their lenited allophones [β̞, ð̞, ɣ̞]. In R. Rao (Ed.), *Key Issues in the Teaching of Spanish Pronunciation* (1st ed., pp. 60–83). Routledge.

González-Bueno, M. (2021). S-PACE to teach L2 grammar: Adding structure to the PACE model. *Applied Language Learning, 31*(1-2), 29–49.

Hertel, T., & Sunderman, G. (2009, Fall). Student attitudes toward native and non-native language instructors. *Foreign Language Annals, 42*(3).

Koo, T. (2016, June). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Levis, J., Sonsaat, S., Link, S., & Barriuso, T. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly, 50*(4), 894–931.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*(3), 488–503.

Martin-Beltran, M., García, A., & Montoya-Ávila, A. (2020). "I know there's something like that in Spanish": Heritage language learners' multifaceted interactions with linguistically diverse peers. *International Journal of Applied Linguistics, 30*(3), 530–552.

Mazzoni, D., & Dannenberg, R. (2000). Audacity [software].

Nagle, C., Sachs, R., & Zarate-Sandez, G. (2018). Exploring the intersection between teachers' beliefs and research findings in pronunciation instruction. *The Modern Language Journal, 10*(3), 512–532. https://doi.org/10.1111/modl.12493

Olson, D. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*(1), 47-68.

Reznicek-Parrado, L. (2024). *The peer-effect: Non-traditional models of instruction in Spanish as a heritage language.* Routledge.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. *CLaSIC 2010* (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

The National Standards Collaborative Board. (2015). World Readiness Standards for Learning Languages. *4th Ed.* Alexandria, VA.

Thomson, R., & Derwing, T. (2015). The effectiveness of L2 pronunciation instruction: a narrative review. *Applied Linguistics, 36*(3), 326-344.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition.* Norwood, NJ: Ablex Publishing Co.

APPENDIX

Example Lesson: "Mis amigos excéntricos"

Objective: Students will accurately produce /o/ sound in Spanish

Procedure:

1) Presentation: Students listen to a recording of two friends talking about current fashion: clothes, piercings, tattoos, etc. This recording includes unit vocabulary: avergonzado, conozco, sombrero.
2) Attention: Presenter calls students' attention to the /o/ - how the /o/ never changes from one position to another and how it is a monothong, not a diphthong /ou/, and how it is always the /o/ and not the schwa /ə/ regardless of position, using example vocabulary from the unit: Avergonzado, conozco, sombrero. Note: presenter does not explain these rules, but rather gives these examples and asks students to repeat, correcting pronunciation of individuals and of the group.
3) Co-construction: Presenter asks students to describe the rules of what they have learned, annotating them on the board as they generate them: /o/ never changes according to position: monothong and not diphthong, /o/ and not schwa.
4) Structured Input: Students perform an activity in which they listen to the same recording as in the presentation phase and answer content question regarding that recording.
5) Structured Output: Students perform an exercise in which the presenter asks them to guess a unit vocabulary word containing the /o/. Presenter corrects pronunciation as required.
6) Extension: Students must record a 2-minute conversation in which they describe their own eccentric friends using the unit vocabulary.

Applied Language Learning

Book Review

Review of *Developing Writing Competence in L2 Chinese Classrooms: Research and Application*

Yan Wang
Defense Language Institute Foreign Language Center

Learners' interactions in and through text influence their L2 acquisition (Lam, 2000). Specifically, writing facilitates the attainment of L2 competence through reinforcing other language skills. Learners can focus more easily on forms in writing than in speaking, allowing them to draw on their explicit knowledge while composing (e.g., Polio 2017; William 2012). In the past decades, writing's instrumental role in the acquisition of a second language has garnered significant research attention; however, the majority of the research on L2 writing development has been done in the field of teaching English as a second or foreign language (Polio, 2017). Studies about how learners develop writing skills in Chinese, a character-based language, are still in an emerging stage. More work needs to be done to advance the theoretical knowledge and pedagogical practices related to writing in L2 Chinese classrooms. *Developing Writing Competence in L2 Chinese Classrooms: Research and Application,* a book edited by Li Yang and Laura Valentín-Rivera and published by Multilingual Matters (2023), is one such effort intended to further the understanding on L2 Chinese writing. It aims to provide readers with "cutting-edge empirical research and insightful teaching methods and strategies for effectively developing L2 writing competence in L2 Chinese classroom contexts" (per the book cover).

Developing Writing Competence in L2 Chinese Classrooms: Research and Application comprises eight chapters encompassing the following main areas: understanding the impact of learners' linguistic upbringing (heritage vs. foreign language learner) on their writing; comprehending writing as a social endeavor (collaborative writing vs. individual writing); identifying the strategies used in L2 Chinese writing; investigating learner development in terms of nominal structure and connective usage in writing; assessing the effectiveness of teachers' corrective feedback; and exploring the role of social media in L2 writing as well as how instructors adapt to online learning under pandemic conditions. All the chapters follow a similar organizational style. Each chapter starts with an introduction section, providing an overview of the research on the topic of focus and identifying the research gap that serves as the purpose of the study. The introduction section

is followed by a literature review of previous studies on the topic, research design, data analysis, results, discussion, and pedagogical implications.

Chapter 1, "Writing processes and products of Chinese as heritage and foreign language learners" by S. Yang, reports on a comparison between the writing process and products of low-intermediate-level Chinese Heritage language learners (CHLLs) and foreign language learners (CFLLs). The study found that CHLLs and CFLLs went through similar writing stages, with both devoting most time to the formulation stage, but with differences between the groups in terms of time spent on macro- vs. micro-level issues. For writing products, CHLLs' texts were overall syntactically more complex; however, CHLLs did not outperform CFLLs in using conjunctions per clause. Regarding lexical complexity, CHLLs' texts exhibited higher sophistication, containing more low-frequency words. The findings of the study shed light on how CHLLs and CFLLs differ in terms of their writing process and writing products, providing abundant resources for instructors to adjust their instruction to better suit the needs of CHLLs and CFLLs. Researchers examining writing process and product will also find this study helpful, as the study employed several instruments to measure learners' syntactical and lexical complexity, and clear descriptions of these instruments were provided. Although the pedagogical strategies proposed are practical and appear to be straightforward for instructors to incorporate into the classroom to address learner deficiencies in writing, it would have been useful if the authors had proposed specific strategies for developing students' ability in connecting sentences with appropriate connective devices, a skill found to be lacking in both CHLLs and CFLLs.

Chapter 2, "Collaborative writing in a tertiary Chinese as a foreign language classroom: Processes and products" by B. Olovson and S. Huang, delves into the realm of collaborative writing by examining the impact of collaborative writing in a college-level writing course. Specifically, the researchers assessed the types and frequency of language-related episodes (LREs), instances where the participants negotiated their linguistic choices while writing collaboratively. Additionally, a comparison was made between collaboratively written texts and individually written texts in terms of complexity, accuracy, and fluency. The results of the study revealed that lexical LREs represented the most frequent type, while syntactic and discourse LREs were the least frequent. Compared to the individually produced texts, the collaborative texts were found to be longer, lexically more accurate, and syntactically of greater quality in terms of mean sentence length. Chapter 2 provides useful scaffolding activities, such as including a video to model successful collaboration that resulted in a high number of resolved LREs. It would be interesting to replicate the study with students at the beginning level to assess whether collaborative writing impacts them in a similar way.

Chapter 3, "Learners' writing strategies in L2 Chinese: A cross-sectional study" by L. Yang and Z. Zhao, reports on a cross-sectional approach to examine CFL learners' writing strategies and the relationship between their strategy frequency and text quality. The findings showed that the while-writing strategies were more frequently used than the pre- and post-writing ones. No statistically significant relationship was found between the strategy frequency and text quality. The findings of Chapter 3 enhance understanding of how learners use strategies when composing Chinese texts and how teachers can facilitate the writing process through strategy training. Data

for Chapter 3 was collected using a questionnaire, which, as the authors pointed out, may impact the validity of the findings due to limitations inherent to self-reporting. It is recommended that future studies supplement questionnaires with the think-aloud protocol. Additionally, future researchers should consider allocating more time for participants to complete the writing task. In Chapter 3, participants were given only 30 minutes for the writing task, possibly contributing to reporting not using pre- or post- strategies.

Chapter 4, "Investigating nominal structures in L2 Chinese writing: A systemic functional linguistics perspectives" by X. Pan, examines how L2 Chinese learners expand the forms of nominal structures (NSs) to construe meaning in writing, as well as the functions and the development of NSs in writing. The study found that, overall, participants used more NSs in the expository essay. Most NSs were formed through Epithet/Classifier, followed by an act clause as Head and Qualifier. The study also found that participants' use of NSs increased with their proficiency. Interestingly, the difference between Intermediate High and Advanced Low was not statistically significant, indicating that nominal structures likely remain stable when learners' proficiency reaches a certain level. Informed by Systematic Functional Linguistics, Chapter 4 illuminated the relationship of nominal structures to second language development in writing. The pedagogical suggestions proposed by the researchers are particularly informative for classroom instruction. Despite the aforementioned strengths of the study, readers without much research experience and a background in linguistics may find the article challenging to follow. Additionally, the examples provided in the article are in Pinyin without tone marks, making it difficult to understand a sentence without referring to the English translation. Including the Chinese characters to aid readers in processing the information would have been helpful.

Chapter 5, "Exploring L2 Chinese learners' connective usage in writing: An error analysis approach" by J. Lin and G. Gao, explores how instructors can help Chinese learners develop their skill in using appropriate connective devices to create coherent texts. The study found that learners had difficulty mastering the use of conditional and successive connectives. Another interesting finding is that the unpaired connectives within casual, progressive, and adversative categories were used more accurately than those in pairs. Chapter 5 provided in-depth knowledge on Chinese learners' acquisition of connectives in writing. Instructors are likely to find the pedagogical implications proposed by the researchers particularly helpful in designing activities to develop learners' skill of using appropriate connective devices to create coherent texts. It needs to be pointed out that the participants in the study are advanced-level students. It would be interesting to see whether the findings from the study hold true for both intermediate- or beginning-level students, which necessitates further examination of this topic by including participants of different proficiency levels. The study can be easily replicated, thanks to the detailed and clear explanation of data analysis process.

Both Chapter 6 and 7 explored learners' perceptions. Chapter 6, "Facebook as a mediator for L2 Chinese writing: Practices and perceptions" by L. Ye, investigates the facilitative role of Facebook-mediated social communication in L2 Chinese writing and the students' perception on the usefulness of this social platform. The findings of the study showed that Facebook played a positive role in improving CFL students' writings. Interestingly, the study did not find Facebook

to be an interactive platform as the earlier studies did (i.e., Wang & Vásquez, 2014; Paul & Friginal, 2019). This is mostly likely due to the writing task design that did not specifically require participants to comment on each other's posts. One limitation of the study is its small sample size. Out of nine participants, only five viewed Facebook writing tasks favorably. Additionally, even though seven out of nine participants produced more sentences at the end of the semester, it is possible that the improvement is a result of repeated writing practice rather than the facilitative role of Facebook. It would be interesting for future studies to include a control group that engages in the same type of writing but without using Facebook for comparison purposes. Despite its constraints in the scope of the study, as Yang and Valentín-Rivera summarized, this chapter "deepens our understanding of the application of Facebook in the CFL context and sheds light on the efficient incorporation of public and virtual spheres as an academic platform to promote written production" (Yang & Valentín -Rivera, 2023, p.14)

Chapter 7, "The efficacy of teachers' written corrective feedback in the L2 Chinese classroom: Learner perceptions and preferences" by L. Valentín-Rivera, examines learners' perceptions and expectations of teachers' written corrective feedback (WCF) in the L2 Chinese classroom, as well as their level of familiarity and preferences with different types of WCF. The results showed that all participants expected feedback from their teachers, and they generally perceived teachers' comments to be helpful. Although all participants reported being familiar with direct WCF, the majority preferred indirect comments because they believed that indirect WCF strengthened their linguistic connections. Chapter 7 provided further empirical evidence supporting the writing-to-learn approach (Manchón, 2011). Therefore, as the author of Chapter 7 recommended, writing should be emphasized as a vital part of students' learning experience. Even in a language program where writing is not assessed, regular writing practice should be utilized to facilitate learners' language development and develop their critical thinking skills. Other recommendations worth mentioning include providing metalinguistic feedback accompanied by examples to help learners achieve a greater level of processing, as well as exposing learners to excellent writing examples produced by their peers for peer modeling.

Chapter 8, "L2 writing under pandemic conditions: How do Chinese and Spanish instructors adapt" by D. Román-Zúñiga, I. Elola, and R. Vasseur, examines how Spanish and Chinese L2 writing instruction was affected by the transition from a face-to-face to an online modality as a result of the COVID-19 pandemic. The findings revealed that the instructors in this study in general were not familiar with digital tools that foster multiliteracies and there was a lack of training on developing writing tasks fitting online modalities. Distinctive approaches were found between the two language programs with the Spanish program incorporating a wider variety of writing assignments while the Chinese program offering limited types of assignments. Chapter 8 provides important considerations for L2 online writing pedagogy and serves as a guide for the design of future training on developing online L2 writing courses.

Overall, *Developing Writing Competence in L2 Chinese Classrooms* is an excellent resource for both researchers and language educators. It offers a solid foundation of empirical knowledge and pedagogical recommendations to further the researching and teaching of writing in L2 Chinese classrooms. It is most likely to be beneficial for graduate students and scholars interested in

conducting research in L2 Chinese writing. All the studies selected in the book followed a rigorous research process with a clear description of the research design, making it possible for future researchers to replicate the study in a different context. Moreover, the detailed introduction and explanation of the instruments used for assessing writing complexity, accuracy, and fluency provide future researchers much-needed tools to expand the scope of the research area and deepen the understanding of L2 Chinese writing. Language educators are also likely to find the information in the book informative. The pedagogical suggestions grounded in empirical data can help them make well-informed decisions on writing instruction in L2 Chinese classrooms and facilitate the implementation of writing-focused activities.

Author

Yan Wang, Ed.D. in Learning and Instruction, serves as Dean of the Chinese School at the Defense Language Institute Foreign Language Center. Her research interests include second language acquisition, learning strategies, and performance-based assessments. She has presented at conferences such as LEARN and ACTFL and has published articles on language teaching pedagogies and students' strategies in learning the Chinese language.

REFERENCES

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Lam, W. S. E. (2000). L2 literacy and the design of the self: A case study of a teenager writing on the internet. *TESOL Quarterly*, *34*(3), 457–482. https://doi.org/10.2307/3587739

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. Longman.

Manchón, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-Learn in an additional language* (pp. 61–82). John Benjamins.

Yang, L., & Valentín-Rivera, L. (Eds.). (2023). *Developing writing competence in L2 Chinese classrooms: Research and application.* Multilingual Matters.

Paul, J., & Friginal, E. (2019). The effects of symmetric and asymmetric social networks on second language communication. *Computer Assisted Language Learning, 32*(5-6), 587–618. https://doi.org/10.1080/09588221.2018.1527364

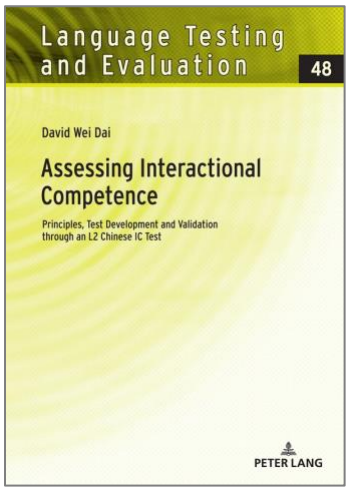
Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching, 50*(02), 261–275. https://doi.org/10.1017/S0261444817000015

Wang, S., & Vásquez C. (2014). The effect of target language use in social media on intermediate-level Chinese language learners' writing performance. *CALICO Journal, 31*(1), 78–102. https://doi.org/10.1558/cj.31.1.78-102

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21*(4), 321–331. https://doi.org/10.1016/j.jslw.2012.09.007

Book Review

Review of *Assessing Interactional Competence: Principles, Test Development and Validation through an L2 Chinese IC Test*



Myoyoung Kim, Mirtha Justiniano Galley, Ruiqi Ma, Aleksey Novikov, and Jongkwan Lee[2]
Defense Language Institute Foreign Language Center

The book *Assessing Interactional Competence: Principles, Test Development and Validation through an L2 Chinese IC Test*[3] (Dai, 2024) is grounded on the author's experience of teaching English to medical graduate students. Dai observed that while in school, students were taught and assessed based on linguistic criteria; in intercultural professional settings, they were evaluated by entirely different standards. This discrepancy led him to explore whether interactional competence (IC) constitutes a distinct test construct.

With the increasing adoption of the communicative approach in language learning and teaching, there has been growing interest in assessing IC as it pertains to situationally-appropriate language use in speaking (Galaczi & Taylor, 2018). Even before Kramsch (1986) introduced the term *interactional competence*, the interactive nature of spoken language ability had been recognized through earlier concepts such as the interactional features of speaking in context (Hymes, 1974) and communicative competence (Canale & Swain, 1980). He and Young (1998) further expanded the IC construct beyond individual ability, framing it as a co-constructed process that involves language processing within interaction.

[2] This book review is the result of a collective effort by the Proficiency Standards Division Book Reading Group at the Defense Language Institute Foreign Language Center.
[3] Assessing Interactional Competence (2024) is available as an open-access publication, making it freely accessible to researchers and educators. All research materials are also available in the appendix (https://www.peterlang.com/document/1306687).

In light of the IC context, this book seeks to address three key questions across eight chapters: First, can critical IC activities be identified for the development of an IC test? Second, can IC abilities be effectively rated using the test? Third, can the measurement process yield reliable results? These overarching questions were further refined based on different validation inferences for the IC test developed by the author. Each research question was examined separately across three studies, with a summary provided in the final chapter. Additionally, the author shares the training materials, surveys, and rating scales used in the studies in both English and Chinese in the appendix section, making them valuable resources for those interested in conducting similar studies.

Chapter 1 proposes IC assessment as an alternative to traditional linguistic assessments, focusing on the ability to engage in real-life communication. It identifies research gaps, including the limited exploration of non-English languages and computer-mediated communication (CMC). The chapter highlights the need for further research on the real-world applicability of IC assessment and its relationship with linguistic competence (LC). Lastly, it outlines the book's structure with research questions and assumptions for inferences of the validity framework.

Chapter 2 provides a comprehensive literature review on the assessment of IC. The chapter not only delves into the philosophical foundations of IC assessment in exploring various perspectives on interaction and pragmatics but also examines the role of CMC in language learning and its implications for IC assessment. This chapter also introduces the two main principles Dai adopted for the development of IC assessment: Conversation Analysis (CA) for analyzing the sequential aspect of the interaction and Membership Categorization Analysis (MCA) to investigate the categorical side of interaction. A significant portion of this chapter is dedicated to defining the IC construct, both theoretically and operationally, discussing its differentiation from LC and the importance of assessing IC in both first (L1) and second language (L2) speakers. The chapter concludes by addressing challenges in test design and rating materials development for IC assessment.

Chapter 3 highlights the significance of the interpretive argument within Kane's validity framework (2012), which necessitates supporting evidence for inferences such as domain description, evaluation, generalization, explanation, and extrapolation. The research design is structured into three studies: task-based needs analysis and test design, test construct and rating scale development, and the principal testing study. By following this staged approach, comprehensive evidence is gathered for the interpretive argument, which is later evaluated in Chapter 7.

In Chapter 4, Dai details the design of an interpersonal communication test, starting with a task-based needs analysis (TBNA) to identify challenges L2-Chinese speakers face in interpersonal interactions. The test development involved drafting items based on TBNA findings and refining them through feedback from various informants. The final IC test includes nine items covering three interactive modes, three power variables from politeness theory, and three sub-target language use domains, designed for CMC. The methodology section explains the TBNA and test development process, using triangulation methods for comprehensive data collection. The study

involved 18 participants from diverse backgrounds, using hermeneutic-Socratic interviews and longitudinal reflective diaries. To ensure quality and reliability, norming questionnaires were used to validate the items in two rounds. Data analysis for the TBNA used thematic analysis, while the test design involved qualitative and quantitative data analysis to refine and validate the test items effectively.

In Chapter 5, Dai presents his second study, in which he develops and validates an indigenous criteria-based rating scale for IC. This rating scale prioritizes real-world communicative demands over primarily academic standards. By incorporating the perspectives of domain experts, the author improves on the ecological validity of the scale. As a result, the rating instrument better captures how speakers handle the challenges of interpersonal discourse, such as managing disagreements or cultivating cooperative relationships. Building on these expert insights, the author structures the scale around six performance categories: Conflict Management, Solidarity Promotion, Reasoning Skills, Personal Qualities, Social Relations, and Linguistic/Prosodic Choices. Initial pilot testing results demonstrated the scale's effectiveness in capturing the intended elements of IC and prompted minor adjustments based on empirical data. By grounding the rating scale in the lived communicative experiences of everyday users, the author demonstrates how an indigenous, criteria-based rating scale can reliably measure authentic interaction.

Chapter 6 presents the main findings of the study, including IC test results, questionnaire responses, and the relationship with the traditional language proficiency scores from the Chinese Standard Exam, Hanyu Shuiping Kaoshi (HSK). A large-scale IC test involving 106 Chinese speakers from diverse cultural backgrounds (26 countries) and different age groups (from the teens to the forties) provided insights into IC assessment and offered detailed guidelines for those interested in replicating similar studies, particularly in material development and rater training. The study emphasized the necessity of LC for IC, as speakers require sufficient linguistic resources to manage social interactions effectively. However, LC alone does not guarantee strong IC, as evidenced by the wide dispersion of IC scores among L1 speakers, supporting the distinction between the two constructs at higher proficiency levels. While CA-informed examples and training materials improved rater consistency, test takers expressed less confidence in their suprasegmental abilities such as pronunciation and prosody, suggesting a need for targeted feedback for those features, specifically in teaching and assessment. Additionally, although the self-assessment questionnaire was found to be reliable, its correlation with IC test scores was lower than expected, indicating a gap between perceived and actual IC performance.

Chapter 7 returns to the validity argument, evaluating evidence and inferences from prior chapters to assess the IC test's overall validity. It expands beyond Kane's framework by including test practicality and its impact on teaching and learning (Chapelle, 2021). The chapter highlights the contribution of this study to advancing assessment methodologies and addressing gaps, such as the need for more studies on CMC platforms and IC assessment in diverse settings. It also emphasizes the importance of further investigation into practical applications and theoretical developments.

In Chapter 8, Dai summarizes the entire research project, highlighting its significance, innovation, and limitations, as well as discussing future research directions. It begins with a philosophical discussion on pragmatics theories and proposes a unified model of interaction to reconcile epistemological conflicts. Focusing on Chinese and using CMC as a testing platform, the book adopts a methodical approach using the Kanean argument-based validation framework, rigorously designing and validating test tasks through TBNA and feedback from various informants. The book also democratizes applied linguistics by involving everyday-life domain experts in the test construct development, as well as ensuring that the IC rating scale is more reflective of real-life language use. It emphasizes cross-disciplinary enrichment, explores the relationship between interpersonal communication and LC, and argues for teaching and assessing IC independently from LC. The author also addresses outstanding issues, limitations, and future research directions, including incorporating L2 perspectives and exploring decision and consequence inferences. Additionally, the chapter notes that the use of raters for IC assessment also requires further investigation.

In the conclusion of this book, the author recognizes the importance of linguistic abilities traditionally emphasized in standardized speaking proficiency tests, such as vocabulary, grammar, and cohesive discourse, while contextualizing them within meaningful, real-world language use rather than treating them in isolation. This study extends the construct by integrating sociocultural and identity-related dimensions of communication, reinforcing that IC is fundamentally about responding to others, accomplishing social actions, and maintaining coherence in interaction.

Methodologically, Dai adopts an innovative approach by incorporating laypeople's perceptions of effective interaction rather than relying solely on theory-driven linguistic criteria. This bottom-up perspective offers a more nuanced and contextually grounded understanding of what constitutes successful interaction. By drawing on CA and MCA, the study presents a comprehensive model that accounts for both the sequential organization of interaction and the ways speakers convey identities, social roles, and community memberships.

To conclude, the studies in this book demonstrate that interactional competence is not entirely dependent on linguistic proficiency, as even lower-proficiency L2 speakers can outperform L1 speakers in interactive effectiveness. The findings challenge traditional assessment paradigms that favor native speakers, and Dai alerted readers to the need for a more equitable framework for evaluating communication skills. This reconceptualization of IC aligns with broader shifts in language assessment as IC concepts such as turn-taking and joint development of discourse are incorporated in the revised Interagency Language Roundtable (ILR) Speaking skill level descriptions (2021) on the social interaction aspect of language function. By offering an empirically supported and theoretically robust model, this book paves the way for assessing and teaching IC in diverse language use contexts, highlighting the importance of adaptability, collaboration, and real-world communicative effectiveness.

Authors

Myoyoung Kim, Ph.D., serves as an Educational Measurement Specialist, Proficiency Standards Division, Language Proficiency Assessment Directorate, at the Defense Language Institute Foreign Language Center in Monterey, California.

Mirtha Justiniano Galley serves as an OPI Education Specialist, Proficiency Standards Division, Language Proficiency Assessment Directorate, at the Defense Language Institute Foreign Language Center in Monterey, California.

Ruiqi Ma, Ph.D., serves as an OPI Education Specialist, Proficiency Standards Division, Language Proficiency Assessment Directorate, at the Defense Language Institute Foreign Language Center in Monterey, California.

Aleksey Novikov, Ph.D., serves as an Educational Measurement Specialist, Proficiency Standards Division, Language Proficiency Assessment Directorate, at the Defense Language Institute Foreign Language Center in Monterey, California.

Jongkwan Lee serves as Chief, Proficiency Standards Division, Language Proficiency Assessment Directorate, at the Defense Language Institute Foreign Language Center in Monterey, California.

REFERENCES

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1–47. https://doi.org/10.1093/applin/I.1.1

Chapelle, C. (2021). *Argument-based validation in testing and assessment*. SAGE Publications. https://doi.org/10.4135/9781071878811

Dai, D.W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test*. Peter Lang Group. doi:10.3726/b21295

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding Questions. *Language Assessment Quarterly, 15*(3), 219–236. https://doi.org/10.1080/15434303.2018.1453816

He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young, & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1-24). John Benjamins. https://doi.org/10.1075/sibil.14.02he

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. University of Pennsylvania Press. https://doi.org/10.1017/CBO9780511611810.029

Interagency Language Roundtable. (2021). *Interagency Language Roundtable Language Skill Level Descriptions-Speaking*. https://www.govtilr.org/Skills/Speaking.htm

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing, 29*(1), 3–17. https://doi.org/10.1177/0265532211417210

Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal, 70*(4), 366–372. https://doi.org/10.1111/j.1540-4781.1986.tb05291.x

Call for Papers

Applied Language Learning, a refereed journal published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, is soliciting articles for publication.

The Journal (US ISSN 1041-679X and ISSN 2164-0912 for the online version) is to provide a forum for the exchange of ideas and information on

- Instructional methods and techniques
- Curriculum and materials development
- Assessment of needs within the profession
- Testing and evaluation
- Implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences

The journal seeks to serve the professional interest of language teachers, administrators, and researchers concerned with the teaching of foreign languages to adult learners. We welcome articles that describe innovative and successful practice and methods and/or report educational research or experimentation.

Deadline: Submissions are welcome at any point. Send your manuscript electronically to the Editor: journals@dliflc.edu.

View the **Guide for Authors** and read the recent and past issues of ***Applied Language Learning*** at http://www.dliflc.edu/resources/publications/applied-language-learning/

Send all questions, comments, and/or submissions to the Editor: journals@dliflc.edu.