Jeff Lu, Seth Vigil, Angela Yang

Professor Bae

LING C132 Fall 24

14 November 2024

Effects of Mutual Intelligibility on Code Switching

1. Introduction

Code switching is a phenomenon in which bilingual speakers “plug in” words from one language into another within an utterance. This works both ways, usually either to capture an expression not available in the current language being spoken, or simply because the other language’s lexical entry happens to be more activated (Zhou, Y., & Wei, M. (2007)). Since code-switching occurs spontaneously during utterances, it is a challenge to study them experimentally.

Furthermore, code switching does not occur in a vacuum. In order to code switch, lexical entries and grammatical structures must be fetched. From studies of speech errors such as substitutions, we know that first syntactic constraints are created, then a certain item in the lexicon is retrieved to populate the structure. Observing how code switching likewise tends to involve the replacement of words or contained clauses, we can also support this idea that language processing at multiple levels occurs in order to code switch.

The objective of this paper is to study if mutual intelligibility between the involved languages influences the ease of code switching. The mutual intelligibility metric that we will be utilizing are the Defense Language Institute (DLI) categories, which are used by the military to measure how difficult it is for a native English speaker to learn another language. In turn, these categories determine how long a course in the language runs.

With the concepts presented thus far, we aim to answer the following research question: Does frequency and ease of code switching decrease with more disparate languages from English? Answering this question would shed light on how exactly constraints and fetching might be different when it involves different languages.

2. Literature Review

Bilingual behaviors such as code switching rely heavily on the factors that give rise to mutual intelligibility. Reviewing existing literature on mutual intelligibility will aid in framing our research into how these factors influence cognitive load, which, in turn, affects code switching timing. Mutual intelligibility, as described in the introduction, measures the degree of similarity between two distinct languages, which allows speakers of those languages to understand each other.

In Van Heuven and Gooskens' 2021 study and paper *Mutual Intelligibility*, they explored what specific linguistic factors contribute to mutual intelligibility. They hypothesized that lexical overlap, phonological similarity, and syntactic structures are all primary dependents of mutual intelligibility. Their experiment tested the reaction times and accuracy of semantically unpredictable sentences, word recognition, cloze tests, and story-to-picture matching. Van Heuven and Gooskens found that their hypothesis was supported, concluding that "linguistic similarity is multidimensional and subsumes differences in any of the linguistic subdomains, such as lexicon (shared cognates with shared meanings), phonology (same or similar sound systems, transparent correspondences between the sound systems), morphology (same or similar word structure), and syntax (same or similar word order)." (Gooskens & van Heuven, 2021, p. 54). This research is relevant to our study as it highlights how linguistic similarities influence reaction times in bilingual processing. However, the gap in this experiment is that it does not address code-switching speeds. Our research delves into the effects of code-switching depending on the distance in mutual intelligibility between the languages.

A study performed by Wang and Van Heuven in 2003 that was presented in their paper *Mutual intelligibility of Chinese, Dutch and American speakers of English* researched the topic that mutual intelligibility—specifically, whether phonological and lexical overlap and the listener's familiarity of the speaker's accent would affect the mutual intelligibility of two languages. The experiment covered two groups: speakers and listeners. The speakers were given stimuli to record and the listeners were tested for intelligibility. The listeners were tested on four different comparisons: vowel identification, constants and C-clusters, sentences, and finally, different types of linguistic aspects. Wang and Van Heuven found that "Chinese listeners performed most poorly in all of the five tests. There were no differences in overall performance between the Dutch and American listener groups when intelligibility was measured at the phoneme level (i.e., the vowel, consonant, and cluster identification tests)." (Wang and Van Heuven, 2003, p. 221-222) Their research showed that there is a level of intelligibility depending on the distance of the languages. This is relevant in our research because it suggests that mutual intelligibility is indeed affected by phonological and lexical similarity. Wang and Van Heuven's research neglects to research the code-switch that occurs for these L2 speakers, which is where our research will delve deeper.

Additionally, another study that contributes to the understanding of mutual intelligibility is the 2006 research conducted by Munro, Derwing, and Morton, which was covered in their paper *The mutual intelligibility of L2 speech*. Their primary research questions examined the effects of a listener's L1 on the intelligibility, comprehensibility, and accentedness of L2 speech. The study involved two groups: a speaker group that was given narratives to record, and a listener group that was given these recordings while being tested for intelligibility, comprehensibility, and accentedness. Munro, Derwing, and Morton found that "listeners from four different L1 backgrounds responded to L2 speech through dictation and rating tasks, all of these factors might have been at work. However, the most important outcome of the study was the striking similarities across listener groups in their comprehension and evaluation of nonnative utterances." (Munro, Derwing, & Morton, 2006, p. 124-125) The study concluded that linguistic factors such as lexical overlap, phonological similarity, and syntactic structures affect mutual intelligibility, comprehensibility, and accentedness ratings, while the listener's linguistic background does not. For our research, this suggests that code switching speeds may not vary significantly based on a speaker's L1, because comprehensibility is not strongly influenced on the listener's background. However, this paper neglects the reaction time factor of comprehension, leaving a gap which our research aims to fill.

These findings are further supported by research conducted by van Heuven in 2009, which was covered in their paper *Making sense of strange sounds: (Mutual) intelligibility of related language varieties*. This research conducted a similar test that concentrated more on the comprehension of the intelligibility under challenging conditions such as signal degradation and background noises. Van Heuven emphasized that "languages differ from each other not in just one dimension but in a great many respects... linguistic distance is a multidimensional phenomenon." (van Heuven, 2009, p. 39). This underscores the importance of phonological, lexical, and syntactic differences for the mutual intelligibility between languages. This research is relevant to our study because it reinforces the notion that comprehension in mutual intelligibility is primarily driven by linguistic similar factors. However, this study fails to specifically examine the effects of code-switching behaviors. Our study fills this gap by testing the reaction times of mutual intelligibility across different distances of languages.

The final piece of literature review is a study conducted by Holst in 2021 that was covered in their research paper *Code-Switching in Multilinguals: A Narrative Elicitation Study with L1 Arabic, L2 English, L3 Norwegian Speakers*. It researched whether the strength of code switching between trilinguals was more likely triggered in the presence of cognates (similar words between the languages) than non-cognates due to their interconnectedness in the mental lexicon. The procedure of their experiment was narratively driven, with participants being presented with a board with pictures, 4 cognates, and 4 non-cognate words and were told to make a verbal narrative based on the pictures. Holst found that "language dominance played a decisive role via increasing the activation level of the Norwegian mental lexicon, and thus, the Norwegian lexicon obtained stronger phonological, orthographical, and semantical representation" (Holst, 2021, p. 61) Essentially, cognates were more frequently involved in code-switching than non-cognates. This is relevant to our study because this paper emphasizes that the shared lexical words, a factor of mutual intelligibility, increases the chance of code-switching. This would suggest that a lower cognitive load is needed to code switch, which may suggest a lower time needed to code-switch. This is what our study aims to research and fill in this gap.

3. Hypothesis

There is a positive correlation between the mutual intelligibility of two languages and the ease of code-switching between them.

4. Methodology

This study investigates the impact of language mutual unintelligibility on the ease of code switching. The methodology is designed to ensure reliability, validity, and alignment with research objectives. The languages that are available to participants as bilingual L2 languages are Spanish (Category I), Indonesian (Category II), Hebrew (Category III), and Chinese (Category IV). These languages are chosen for their categorization, and to allow for specific trends to be identified, as correlation involving only a few of these might be the work of a confounding variable. The platform chosen for the experiment is Gorilla, as it provides the functionality and stimuli needed for this experiment.

Participants will first provide a language in which they are bilingual. This will constitute the operational definition of the independent variable of the experiment, with the more abstract concept being mutual intelligibility. They will then read a number of English sentences, pressing the spacebar as soon as they have done so. After each English sentence, a sentence in their indicated non-English language will appear, and the participant will be tasked with determining whether it has the same meaning as the prior English sentence as quickly as possible. These pairs will either have the same meaning or not, with those which are not correct translations involving a mix of lexical, syntactic, and semantic differences. This is to control for participants simply short-cutting the process by memorizing the words mentioned and their coinciding words in their L2. Syntactic differences might include differences in subject and object in a sentence, for example.

In order to collect a participant's response, they will press the 'J' key if the meaning is the same, and the 'K' key if the meaning is different. Reaction time will be measured and used as a dependent variable, to get at how quickly participants are able to semantically process across their L1 and L2. Only the reaction times gathered from correctly identified trials will be used as measurements as the reaction time of an incorrect selection is meaningless to the hypothesis. Regardless of the participant's indicated language, the same 20 English sentences will be used as stimuli, with corresponding non-English probes—10 are correct translations, and 10 are semantically related but incorrect translations. Measures will be taken in order to dampen as much as possible the impact of confounding variables, including: ensuring that materials (the sentence lists and their respective translations) are the same across IV levels; randomizing the order in which stimuli are presented; recruiting participants from a variety of ages, genders, locations, and cultural backgrounds; and restricting participants to only simultaneous bilinguals so as to eliminate variations in second language acquisition.

In order to ensure statistical significance, a sample size of at least 30 participants in *each* condition is ideal. While an outcome is difficult to predict without gathering data, we predict that languages with a higher degree of mutual intelligibility will be easier to code-switch between, as their shared linguistic features and cognates will lighten the necessary cognitive load. Because we have included four languages (Spanish, Indonesian, Hebrew, and Chinese, chosen for their varying levels of linguistic similarity to English), data from this experiment would be analyzed by means of a 4-way ANOVA. This allows us to compare the difference in reaction time between languages.

5. Predicted Results

As stated previously, the expected result if a positive correlation does exist between mutual intelligibility and ease of translation would be shorter reaction times for languages more similar to English by the DLI categories.

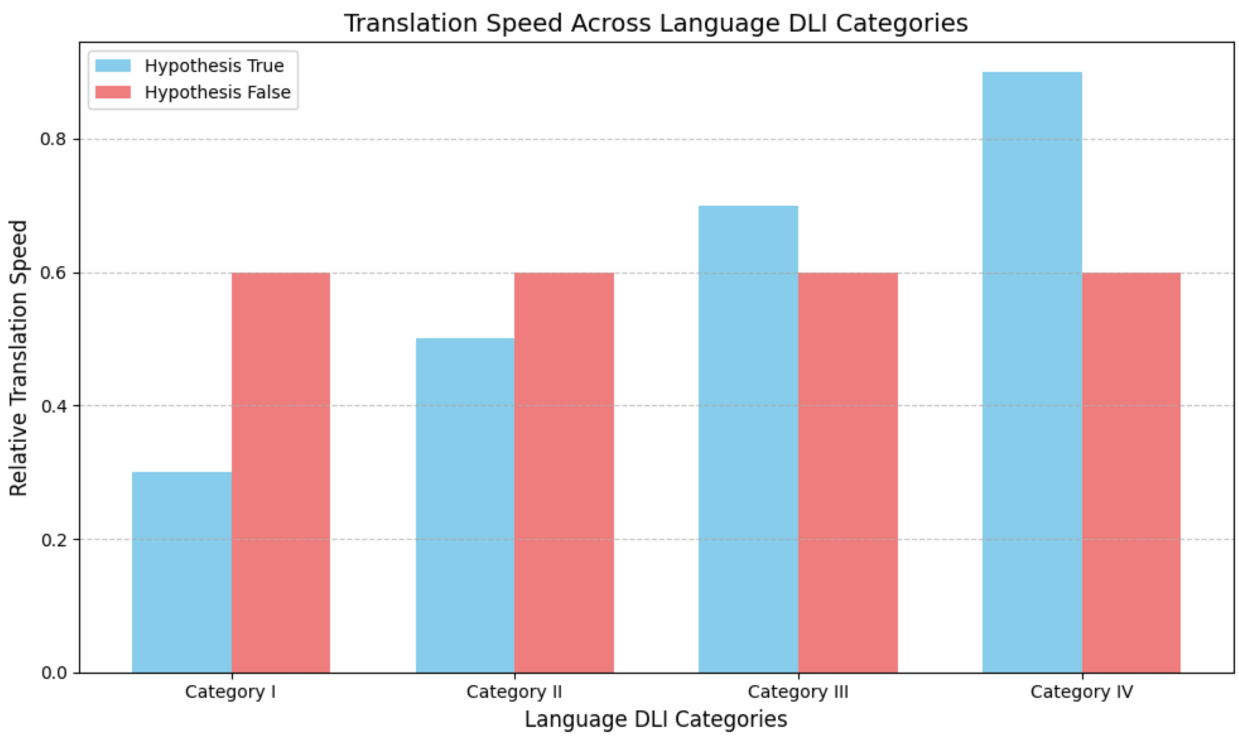

Fig. 1: a bar graph showing the average reaction time between the four categories.

Here in Figure 1, we have a simplified version of the results one might expect from the experiment, including the average reaction speed. As one can see, there is a positive correlation between the categories and the speed of translation, with higher category languages resulting in a longer average translation speed. Results that don't have a positive correlation such as the bars in red also show what a false hypothesis might look like.

We might see the data presented in a more detailed way, such as in a box plot format in which the box plot of reaction times is greater in all of its summary metrics than those with lower categories. This might help with spotting issues that the above representation might miss: for example, if one language's range is quite large and actually doesn't have conclusively longer reaction times, there might be a confounding factor that isn't being taken into account, even if the mean is in line with the hypothesis. In Figure 2, we have an example of this.

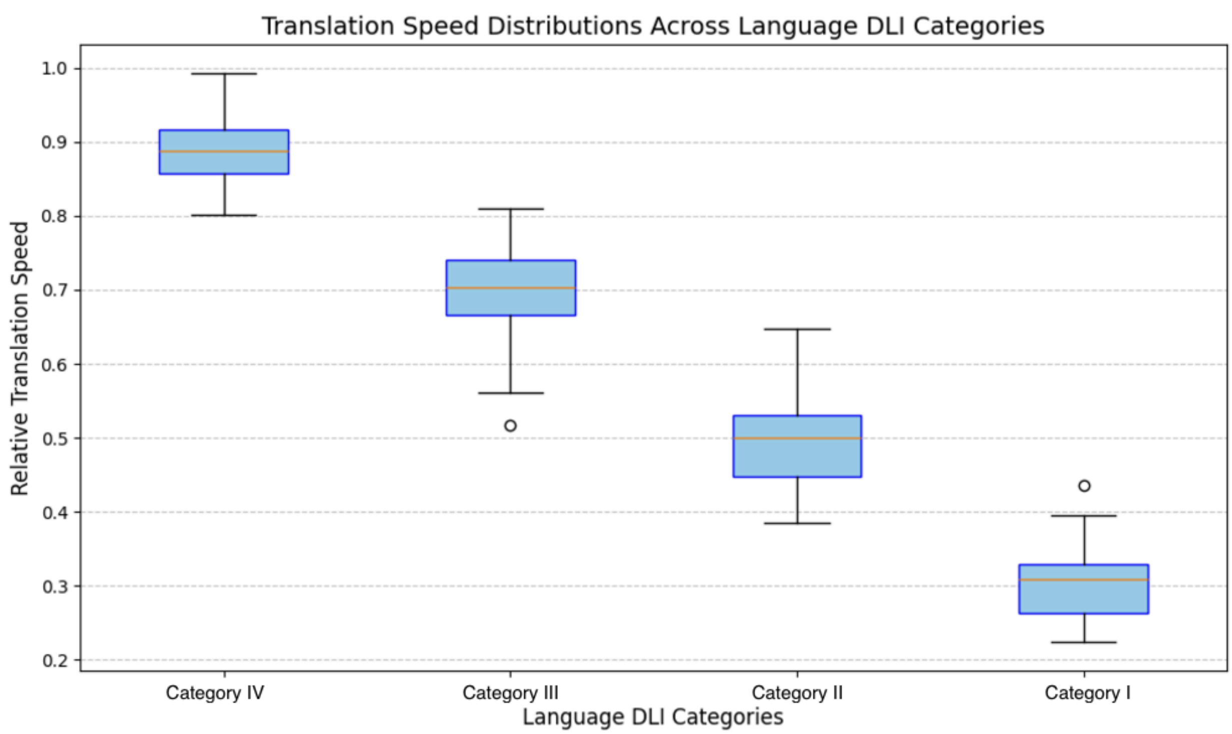

Fig. 2: a box plot comparing summary statistics of reaction time between the four categories.

6. Results

Now that we've seen *examples* of what data might look like, here we have some proof-of-concept data. Surveying a smaller convenience sample of 13 participants (3 Spanish, 3 Indonesian, 4 Hebrew, and 3 Chinese), we got a better sense of what confounding variables existed and the strength of the trend we were looking for. Fig 3 shows our set of box-plot data capturing the collected reaction times:

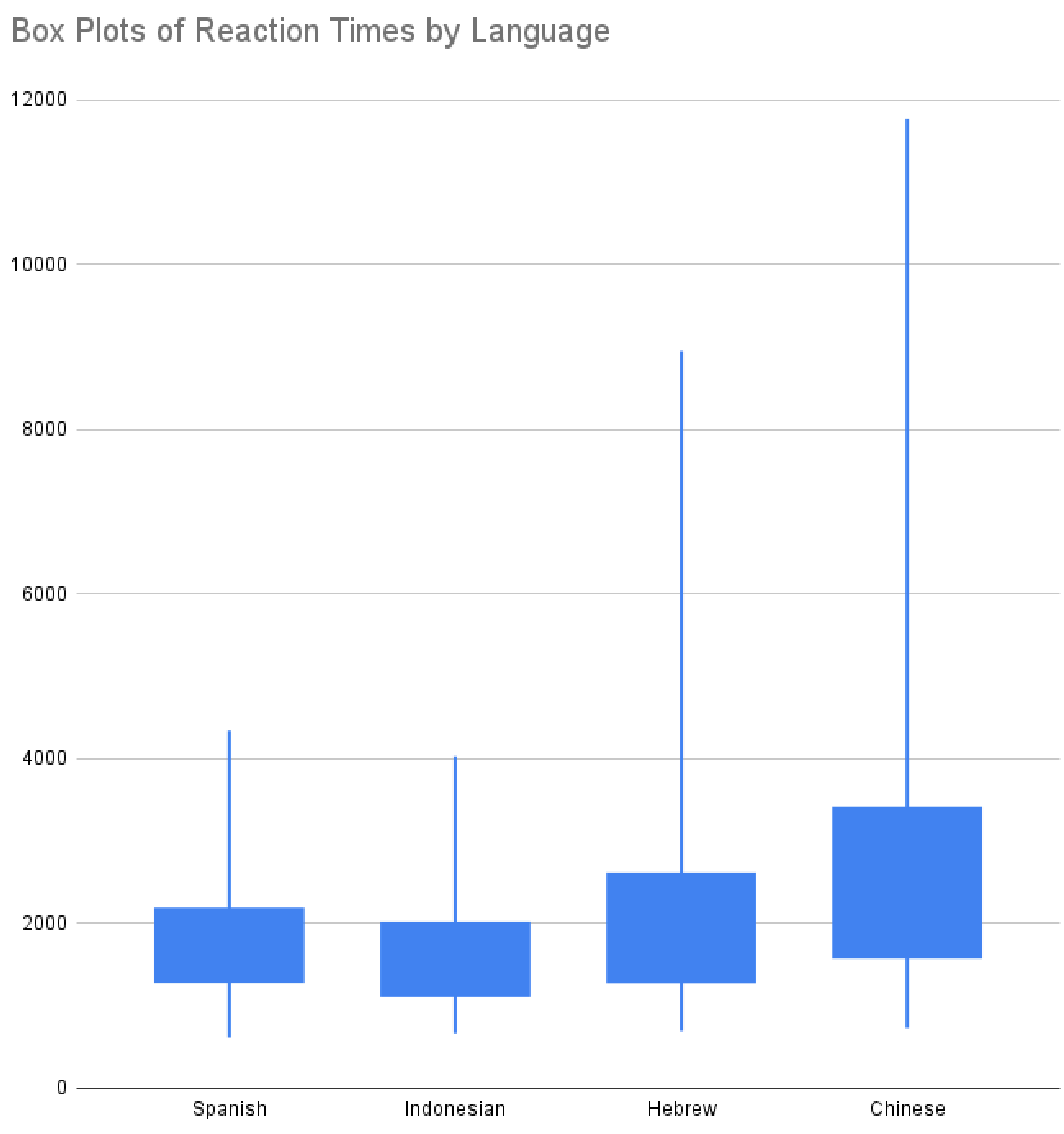

Fig. 3: a box plot made of real collected data as an example

As previously disclaimed, because this is not a statistically significant amount of participants, and participants were chosen by convenience (specifically, friends and family members), there will be less control in this sample than in the proposed experiment design. For example, though the majority of the data seems to follow the hypothesis well, with a mostly positive correlation being observed, the Indonesian data was significantly faster than the Spanish data, likely due to the majority of the Indonesian sub-sample including Indonesian-born students, whereas the Spanish bilinguals were all American-born.

Regardless of this, even in a smaller sample size, we do see that there is a strong positive correlation between translation time and linguistic distance to English. This is a promising result, as it suggests that our hypothesis holds some merit. One factor to notice in this study is the fact that the lower quartile of data seems to be quite close between the languages, and was achieved by those who were more fluent in both languages. With bilinguals who had a bit less proficiency, we saw a longer tail or skew in regards to the upper bound of the reaction time with languages that were further from English. This might correlate to some specific kinds of mistranslations or correct translations being more difficult to convert and requiring more thought across languages. Overall, these test results show that a correct hypothesis is more likely to be true.

7. Further Issues & Future Research

We can further explore this idea by expanding and controlling for more factors to better answer the research question. One way that we could expand this experiment would be to reproduce the results using a different measure of linguistic similarity, to better control for factors that might've determined the DLI categories. The DLI categories were created to fit classes and bilingual testing in the defense industry, so they might not have been the best measure of cross-language similarity. Furthermore, in this experiment, there is no control for proficiency, so the relative speed of comprehension will likely be influenced by proficiency more than linguistic distance. This is a very large confounding variable as being a beginner or expert in the language has more direct influence on comprehension time.

Another significant issue with this study is the size of the participant pool upon which results are drawn. Though the intended size is large enough for statistical significance, the actual observed participant pool we acquired was far too small to draw a conclusion about the hypothesis. The particular trials mentioned in this paper are mainly for demonstration of how a full-scale execution might look.

More abstractly, the method might not address the hypothesis directly enough. Because code-switching is a phenomenon in naturalistic production, and we are indirectly making conclusions about it through the speed of translation, the relationship between the two might not be close enough. In other words, we might be making implications about how code switching is processed underlyingly that are incorrect. One possible difference might lie in the fact that code-switching usually involves an adjacently activated lexicon in the L2 when the concept is fetched during L1 speech production, rather than conscious effort to translate a piece of lexicon *from* L1 to L2. In this case, we would only be able to extend our conclusions to the frame of ease of translation, rather than the likelihood and ease of code-switching.

8. Conclusion

In short, we predict that a higher degree of mutual intelligibility between two languages will decrease the difficulty of code switching between them. In our attempts to collect preliminary data on this topic, we found weak evidence in support of our hypothesis; however, our sample size was too small to make any strong conclusions. Future expansion of this research would first and foremost involve a larger and more controlled sample of the population. Regardless of whether our predicted results hold true or not, however, this experiment would have profound implications for our understanding of the cognitive faculty necessary to code switch between one's L1 and L2, our ability to teach languages to new learners, and our sociolinguistic perspectives on bilingual communities.

9. References

https://app.gorilla.sc/admin/project/148142

Gooskens, C., & Van Heuven, V. (2021). **Mutual Intelligibility**. In M. Zampieri, & P. Nakov (Eds.), *Similar Languages, Varieties, and Dialects: A Computational Perspective* (pp. 51–95). Cambridge University Press. doi:10.1017/9781108565080.006

Van Heuven, V. (2009). **Making sense of strange sounds: (Mutual) intelligibility of related language varieties.** A Review. doi:10.1515/9780748641642-005

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). **The mutual intelligibility of L2 speech**. *Studies in Second Language Acquisition*, *28*(1), 111–131. doi:10.1017/S0272263106060049

Holst, S. (n.d.). **Code-switching in multilinguals: A narrative elicitation study with L1 Arabic, L2 English, and L3 Norwegian speakers**.

Wang, Hongyan & Van Heuven, Vincent. (2003). **Mutual intelligibility of Chinese, Dutch and American speakers of English**. P. Fikkert & L. Cornips (eds.), Linguistics in the Netherlands, 213 - 224 (2003). 20. doi: 10.1075/avt.20.22wan

Zhou, Y., & Wei, M. (2007). *Code-switching as a result of language acquisition: A case study of a 1.5 generation child from China.* Paper presented at the 2007 Annual Conference of Teaching English to Speakers of Other Languages (TESOL), Seattle, WA.