An Analysis of Spanish L2 Lexical Richness
Dieter A. Waldvogel, U.S. Air Force Academy

Dieter A. Waldvogel, Ph.D., is a Latin America Regional Affairs Strategist & Foreign Language Assistant Professor at the U.S. Air Force Academy, CO.

Abstract
Reliable estimates of Spanish L2 learners' vocabulary size and richness at different proficiency levels can bring to light lexical deficiencies and may help influence decisions regarding vocabulary instruction. The focus of this paper is to explore the concepts of lexical variation and sophistication—together known as lexical richness—and to analyze the relationship between oral proficiency and lexical richness among Spanish L2 learners. This pilot study offers some insights into the process of lexical development between different levels of Spanish L2 oral proficiency.

Introduction
The importance of vocabulary research and assessment has experienced a resurgence in the past 15 years for both first language (L1) and second language (L2) proficiency (Alderson, 2005; Pearson, Hiebert, & Kamil, 2007); and yet, not much is known about native and L2 speakers' spoken vocabulary (David, 2008; Read, 2005).

Vocabulary is central to language and critical not only for the four traditionally recognized basic language skills—reading, writing, listening and speaking—but also for academic performance in general (Pikulski & Templeton, 2004; Grabe & Stoller, 2002). Nonetheless, the acquisition of L2 lexicon is constantly identified as the key problem in learning a L2 by most learners. Lexical errors are the most frequent source of confusion and the most common reason for meaning negotiations between language learners (Al-Shormani & Al-Sohbani, 2012).

Reliable estimates of learners' vocabulary size and richness at different levels of proficiency can bring to light these lexical deficiencies and help influence decisions regarding vocabulary instruction at different proficiency levels, including whether it is even worthwhile to explicitly teach vocabulary at all. Several analytical tools and measures have been designed to help researchers and educators analyze L2 learners' lexical knowledge and use in order to provide a better picture of their lexical development. The focus of this paper is to explore the concepts of lexical variation and sophistication—together known as lexical richness—and to analyze the relationship between oral proficiency and lexical richness among Spanish L2 learners during spontaneous interactions with a trained interlocutor.

Linguistic researchers have used a range of lexical variety measures to analyze and assess vocabulary knowledge and usage, most of which are based on the ratio of different words (types) to the total number of words (tokens), known as the Type/Token Ratio or TTR (Cimino,

Dell'Orletta, Venturi, & Montemagni, 2013; Malvern & Richards, 2012; McKee, et al., 2000; Read, 2005). This measure is a ratio of different word types to the total number of words or tokens (Cimino, et al., 2013; Malvern & Richards, 2012). The TTR has been traditionally used to assess whether writers and speakers can use a variety of different words rather than a limited number of high-frequency words used repeatedly—a high TTR means the learner has produced a high ratio of total words to different word types, whereas learners with a low TTR tend to use a smaller number of words repeatedly (Cimino, et al., 2013; Read, 2005). Studies show that more proficient writers and speakers can use a wider range of vocabulary to express their ideas and to avoid repetitiousness what Read (2000) refers to as "range of expression" (p. 200). The TTR offers a macro-picture of the learner's lexicon and are used in a number of fields such as child language development, language impairment, second language acquisition, literacy development, authorship studies, forensic linguistics, stylistics, and others (McKee, et al., 2000).

The TTR, however, has come under some scrutiny by a number of researchers who have exposed some reliability issues (Daller, Van Hout, & Treffers-Daller, 2003; Malvern and Richards, 2012; McKee, et al., 2000; Read, 2005; Vermeer, 2000). Some argue that TTRs usually fail to discriminate between learners at different developmental stages (Vermeer, 2000). Vermeer (2000) instead suggests the use of a set of new advanced quantitative measures that rely less on types-tokens ratios and more on lexical sophistication; the relationship between tokens and their difficulty, measured by their frequency. This measure uses the traditional Index of Guiraud quantitative formula, which measures lexical variation and modifies it by using low-frequency word types instead of all word types. Nation (2001) proposed a similar procedure known as Lexical Frequency Profile to circumvent the weaknesses of TTR-related measures by using low-frequency word counts (or advanced words) rather than all word types. Vermeer (2000) used this lexical sophistication measure to analyze spontaneous speech samples of Dutch (L1 and L2) kindergarten children in The Netherlands and found the reliability of this measure to be much higher than that of traditional type-token measures. Daller et al. (2003) also used the advanced measures of lexical richness to analyze the speech of German-Turkish bilingual speakers. Their results supported Vermeer's findings by concluding that measures based on more advanced, low-frequency lexical items have clear advantages over traditional measures of TTR and Guiraud. Daller et al. concluded that these "new measures [of lexical sophistication] have a greater explanatory power and lead to highly significant results even with small samples" (p. 217).

These two studies used Dutch (Vermeer, 2000), and German-Turkish bilinguals (Daller, et al., 2003) as the languages of study. The present pilot study focuses on Spanish speech samples from three different proficiency levels of university Spanish L2 learners—intermediate-low, intermediate-high, and advanced-mid—and evaluates the lexical richness of their speech during interaction using the new advanced lexical richness measures and then compares the results from each proficiency level to see if there are any significant differences between the three groups and whether an increase in sophistication will result in an increase in lexical variation and vice versa, with the assumption that L2 advanced lexical richness is a good indicator of oral language proficiency development. The goal is to determine whether the use of these advanced lexical measures could be used in the oral proficiency classification of Spanish L2 students.

The following three research questions are addressed in this study:

1) Is there a statistically significant difference in the lexical variation (as measured by Guiraud's Index) and/or sophistication (as measured by Guiraud Advanced) of Spanish L2 learners' speech samples between different Spanish L2 proficiency levels?

2) Does an increase in lexical sophistication result in an increase in lexical variation?

3) Is there evidence to support the use of these lexical richness quantitative measures as a factor in the proficiency classification of Spanish L2 learners?

Methodology

In order to address the above stated research questions, data was collected using the following research design:

Participants. Convenience speech samples from nine Spanish students were provided by the Spanish Corpus & Proficiency Level Training (SPT) website hosted by the University of Texas at Austin Liberal Arts Instructional Technology Center (http://www.laits.utexas.edu/spt/) (August, 2013). The speech samples in this website were collected by Dr. Dale Koike and are used for L2 research. At the time of this study, the website contained 252 short speech samples from 28 Spanish L2 students. Students were interviewed by Professor Koike and each was asked a series of questions to elicit interaction. Following the 2012 American Council on the Teaching of Foreign Languages (ACTFL) proficiency guidelines for speaking (ACTFL, 2012), all students were then classified into one of four proficiency levels: intermediate-low, intermediate low-mid, intermediate-high, and advanced-mid. For the current pilot study, three short speech samples were randomly selected from each of three proficiency levels: intermediate-low (n = 3), intermediate-high (n = 3) and advanced-mid (n = 3); N = 9. All of the speech samples chosen were spontaneous responses to the same statement from the researcher: "Describe the most important moment in your childhood."

Instruments. Each speech sample was fully transcribed into nine separate text transcripts to analyze lexicon use. To decrease the problems associated with text length during vocabulary richness analyses, Laufer (1991) recommended using samples of no more than 400 tokens. The short samples selected for this pilot study range in the number of tokens from 51 to 204.

Following Daller's (2003), and Vermeer's (2000) recommendations, the measures used to calculate lexical richness were the Guiraud's Index, to measure lexical variation, and Guiraud Advanced, to measure lexical sophistication. The Guiraud's Index is measured with the formula: Word Types / √ Tokes. The Guiraud Advanced is calculated using Advanced Types / √ Tokes. Types include all the different lexical words on the speech transcripts, while tokens are all the lexical words including words repeated more than once (Read, 2000). In order to determine the difference between similar lexemes (types), the present study used Read's (2000) lexeme classification: all inflected forms of verbs were used in the base form; homonyms were distinguished as separate entries; idioms were recorded as one item; and derivatives of base words were counted as separate words.

Lexical sophistication (or rareness) is a proportion of relatively unusual, low-frequency, or advanced words in the learner's text or speech (Daller, et al., 2003; Vermeer, 2000).

Researchers have used a variety of vocabulary lists to operationalize what constitutes a low-frequency or advanced word. Different studies have used anywhere from the 100 to the 10,000 most or least frequently used words (in different languages), depending on the language proficiency and academic setting. The Spanish word frequency vocabulary list used to measure lexical sophistication in this pilot study consists of the 100 most frequently used Spanish words extracted from "A Frequency Dictionary of Spanish" (Davies, 2006). When selecting the advanced word types to be used in the lexical sophistication formula, words were used only if they did not appear in the Davies' list. This low number (100) was chosen because of the size of the speech samples (51 to 204 tokens) and to give a wider margin of advanced vocabulary use to the least proficient participants, since less-proficient learners tend to produce less advanced/low-frequency words in their L2 (Daller, et al., 2003). Vermeer (2000) and Read (2000) recommend using the 1,000 or more word frequency lists for texts longer that 2,000 tokens.

Procedures. Each of the nine chosen speech samples was fully transcribed into text form for analysis. In order to apply various lexical statistics to each transcript, all of the interviewer's contributions, all of the participant's non-words (i.e. expressions/sounds such as um, hehe, hmmm, ah) and all partial words were deleted (Read, 2005). Once the text was cleaned up, Microsoft Office "word count" was used to calculate the total number of tokens.

As stated in the previous section, to determine the difference between similar word types, Read's (2000) lexeme classification was used: all inflected forms of verbs were used in the base form, homonyms were distinguished as separate entries, idioms were recorded as one item, and derivatives of base words were counted as separate words. This process was done manually and with the help of the 'find and replace' tool in Microsoft Office Word 2007 to ensure each word type was not counted more than once.

In order to calculate the number of advanced types, each word type was examined to ensure it did not appear in the 100 most frequently used Spanish words (Davies, 2006) list; words that did were discarded and not used to calculate lexical sophistication. Once the numbers of tokens, types, and advanced types for each of the nine transcripts were obtained, the lexical variation and sophistication of each speech sample was calculated using the advanced lexical richness formulas described earlier. Afterwards, a two-way analysis of variance, a multiple comparisons analysis Post Hoc test, and a correlation analysis were performed to answer the research question stated in the previous section.

Results
Lexical output. The statistical analyses of the speakers' lexical output produced interesting findings. The first step was simply to calculate how many words (tokens), types and advances words each participant uttered during the short conversations with the interlocutor and calculate the means. The results show that, between intermediate-low and intermediate-high speakers, the mean number of tokens increased by 126% (from 79.33 to 179.33 tokens); the mean number of word types doubled (from 43.33 to 86); and the mean number of advanced words used increased by 138% (from 17.66 to 42). Unexpectedly, the results between intermediate-high and advanced-mid students remained almost constant. The next step was to calculate the lexical variation and sophistication scores for each speaker using the Guiraud's Index and the Guiraud Advanced formulas.

The results of the lexical richness measures, again, show significant increases in both lexical variation and sophistication between intermediate-low and intermediate-high speakers. The lexical variation scores between these two groups show an increase of 33%, from 4.81 to 6.41. While the use of advanced words between the same two groups increased by 60%, from 1.96 to 3.14. Between intermediate-high and advanced-mid speakers, however, there was only a 5% increase in the use of advanced lexicon, from 3.14 to 3.30. Possible reasons for this unexpected result are discussed later in this section.

In order to determine whether the differences found were statistically significant, a two-way analysis of variance (ANOVA) and a multiple comparisons post hoc test were used to analyze the variances. The ANOVA showed an overall statistically significant difference in both lexical variation (.039) and sophistication (.015) in the between group means at $p < .05$. However, in order to determine where exactly these significant differences in lexical variation means could be found, the data was analyzed a second time using Tukey's Post Hoc multiple comparisons test. Due to the exploratory nature of this pilot study, the significance alpha level of probability was set to .1 ($p < .1$) in order to reduce the possibility of missing a significant finding if one is present (type II error). The Post Hoc test found significant differences in lexical variation and sophistication means at $p < .1$. Both analyses showed a substantial increase is lexical richness between the intermediate-low speakers and the intermediate-high speakers. Again, no differences were found between intermediate-high and advanced-mid speakers. This post hoc analysis shows a significant difference (.019) in lexical sophistication between intermediate-low and advanced-mid Spanish L2 speakers during oral interaction. The statistical difference in lexical variation between intermediate-low and the other two groups, however, was not as strong; significant only at the $p < .06$ level.

The strength of the lexical variation and sophistication between group variances (R^2) resulted in 66% and 77% respectively. Since all the between group variance was found to be in between the intermediate-low and intermediate-high speakers, it was concluded that by using the lexical richness scores alone, we should be able to predict the correct oral proficiency level (between intermediate-low and high speakers) of Spanish L2 learner's speech with 66% accuracy for lexical variation and with 75% accuracy for lexical sophistication. The same cannot be said, however, for groups above the intermediate-high proficiency level.

Discussion
The statistical differences in overall vocabulary richness between the intermediate-low and intermediate-high oral proficiency groups were not unexpected. According to the 2012 ACTFL Proficiency Guidelines for Speaking (ACTFL, 2012),

> Speakers at the Intermediate Low sublevel are able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations… Intermediate Low speakers express personal meaning by combining and recombining what they know and what they hear from their interlocutors into short statements and discrete sentences (pg. 8.).

Intermediate-High speakers, on the other hand, "are able to converse with ease and confidence when dealing with most routine tasks and social situations of the Intermediate level" (p. 7). This difference in communicative competence suggests a higher degree of lexical and syntactic

mastery by the intermediate-high speakers, which is reflected in the speakers' lexical variation and sophistication scores. The issue of lexical variation in intermediate-low speakers is clearly addressed in the proficiency guidelines when stating that these speakers communicate "by combining and recombining into short statements what they know…" (p. 8), demonstrating a limited repertoire of advanced vocabulary. Speakers with intermediate-high oral proficiency, on the other hand, are able to "narrate and describe in major time frames using connected discourse of paragraph length" (p. 7) which arguably demands a higher level of lexical sophistication. The obvious question then emerges: Does an increase in oral lexical sophistication equates to an increase in oral lexical variation in Spanish L2 learners, and vice versa? In order to address this question, a correlation analysis was performed which found a high correlation of .922 between speakers' lexical variation and sophistication scores. From this high correlation value we can safely infer that speakers who have acquired a higher number of advanced words will tend to use more word types in oral interaction and thus, have higher lexical variation scores.

The results of this study show clear improvement in lexical diversity and sophistication between intermediate-low and intermediate-high speakers. It could be said, however, based on this study, that there is a 'slow down' after reaching the intermediate-high proficiency level. This 'slow down' is supported by the findings observed by Milton (2006), in which he found a vocabulary acquisition slowdown once the French-speaking students reached a higher proficiency level.

Studies show that L2 learners with higher oral proficiency levels who can function well in their L2 environments may not exhibit the same pattern of progress as those in the lower proficiency levels. One theory proposed by Laufer (1991) is the 'active vocabulary threshold hypothesis' which states that once a learner's vocabulary has grown to reach the average vocabulary level of the group in which he/she is required to perform, the absence of more advanced comprehensible input (i + 1) will prevent the learner from further expanding his/her lexical repertoire. It is possible, as proposed by Laufer (1991), Milton (2006) and David (2008) that the lexical proficiency state of the advanced-mid speakers in this study was sufficient enough to allow them to succeed in their Spanish L2 studies and, as a result, reached a lexical growth threshold. Therefore, if this lexical threshold theory holds true in most cases, then we can conclude that the lexical richness measures discussed in this study are only effective when used to measure or to classify the oral proficiency of Spanish L2 students rated below the advanced level. This conclusion is supported by the findings of this pilot study.

So now we can return to our original research questions: 1) is there a statistically significant difference in the lexical variation and/or sophistication (as measured by Guiraud's Index and Guiraud Advanced) of spontaneous speech during oral interaction between intermediate-low, intermediate-high and advanced-mid Spanish L2 speakers? The results of this pilot study do show statistically significant differences in the oral lexical richness scores of Spanish L2 learners' speech during interaction. These differences, however, are only significant between intermediate-low and intermediate-high speakers. Therefore, these results cannot be generalized to Spanish L2 speakers rated above intermediate-high in oral proficiency, nor can they be generalized to speakers rated as 'novice' since this proficiency level was not part of this study. Future studies should include all proficiency levels to see if these findings can be generalized to all levels.

2) Does an increase in lexical sophistication result in an increase in lexical variation? Yes. A strong correlation of .922 was found between the lexical variation and lexical sophistication scores of each of the nine speakers used in this study. From this result we may infer the pedagogical benefits of helping L2 students increase the size of their advanced vocabulary to improve their overall lexical richness.

3) Is there evidence to support the use of these lexical richness quantitative measures as a factor in the proficiency classification of Spanish L2 learners? Based on this and other studies reviewed (Daller, et al., 2003; McKee, et al., 2000; Tidball & Treffers-Daller, 2008; Vermeer, 2000), we find strong evidence to support the use of advanced lexical richness measures that focus on the degree of difficulty and frequency of the words used by L2 learners. The use of low-frequency vocabulary is an indicator of lexicon acquisition and a sign of higher proficiency in the L2 (Daller, et al., 2003; Vermeer, 2000). As stated earlier, studies show that more proficient L2 learners can use a wider range of vocabulary to express their ideas and to avoid repetitiousness (Read, 2000), what Read refers to as "range of expression" (p. 200). However, if the lexical threshold theory holds true in most cases, then we must conclude that the lexical richness measures discussed in this study may not be as effective with advanced Spanish L2 students. This conclusion is supported by the findings of this study. Future studies should include participants rated as 'novice' in oral proficiency to see whether the conclusions reached in this study can also be extended to oral proficiency novices.

Conclusion

This study examined empirical evidence to support the use of lexical richness measures in the oral proficiency level classification for Spanish L2 learners. The findings suggest that helping learners improve their vocabulary size by helping them increase their knowledge of less-commonly-used words will have a positive impact on learners' lexical variation and overall lexical richness. One particular quality of this study and the methodology utilized is that, rather than focusing on the errors the L2 learner makes during oral communication, this study spotlights the learner's lexical production and his/her ability to use more low-frequency or advanced lexicon during oral interactions.

Even though the speech samples used in this pilot study were fairly small, the advanced lexical richness measures successfully showed significant results in distinguishing the lexical proficiency of Spanish L2 learners' spontaneous speech at the intermediate level. This study also highlights the potential for using the two advanced lexical richness measures as factors in the learners' oral language proficiency classification within the ACTFL oral proficiency guidelines at or below the intermediate-high level.

Limitations and Recommendations. Sample sizes (N = 9) is an obvious limitation of this pilot study and a reason for caution when interpreting the results and making inferences based on these results. Even though the results of this study showed significant differences in the lexical variation and sophistication scores among the different levels of Spanish intermediate oral proficiency speakers, the findings were not as strong after being evaluated with a post hoc multiple comparisons test; barely significant at the $p < .1$ level. However, these results show the potential for more significant findings in future studies using larger sample sizes and perhaps including more—if not all—of the ACTFL proficiency levels for speaking to see if the 'active

vocabulary threshold hypothesis' holds true when examining all of the proficiency levels at the same time. Results may also differ if this study could be recreated in a more natural or authentic environment using casual conversations between students without the influence the *evaluator factor* casting a shadow over the participants to see if lexical richness improves.

References

ACTFL Proficiency Guidelines, 3rd Edition, 2012. American Council on the Teaching of Foreign Languages, Alexandria, VA.

Al-Shormani, M., & Al-Sohbani, Y. (2012). Semantic Errors Committed by Yemeni University Learners: Classifications and Sources. International Journal of English Linguistics, 2(6).

Alderson, J. (2005). Diagnosing foreign language proficiency: The interface between learning and assessment: Burns & Oates.

Cimino, A., Dell'Orletta, F., Venturi, G., & Montemagni, S. (2013). Linguistic Profiling based on General-purpose Features and Native Language Identification. Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications, 207–215.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. Applied Linguistics, 24(2), 197-222.

David, A. (2008). A developmental perspective on productive lexical knowledge in L2 oral interlanguage. Journal of French Language Studies, 18(03), 315-331.

Davies, M. (2006). A Frequency Dictionary of Spanish: Routledge.

Grabe, W., & Stoller, F. (2002). Teaching and researching reading: Pearson Education.

Laufer, B. (1991). The Development of the L2 Lexis in the Expression of the Advanced Learner. Modern Language Journal, 75(4), 440-448.

Malvern, D. & Richards, B. (2012). Measures of Lexical Richness. The Encyclopedia of Applied Linguistics.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. Literary and Linguistic Computing, 15(3), 323-338.

Milton, J. (2006). Language lite? Learning French vocabulary in school. Journal of French Language Studies, 16(02), 187-205.

Nation, I. (2001). Learning vocabulary in another language: Cambridge University Press, New York.

Pearson, P., Hiebert, E., & Kamil, M. (2007). Vocabulary assessment: What we know and what we need to learn. Reading research quarterly, 42(2), 282-296.

Pikulski, J., & Templeton, S. (2004). *Teaching and developing vocabulary: Key to long-term reading success.* Boston: Houghton Mifflin.

Read, J. (2000). Assessing Vocabulary. Cambridge University Press.

Read, J. (2005). Applying lexical statistics to the IELTS speaking test. *Research Notes*, *20*, 12-16

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: what frequency lists and teacher judgements can tell us about basic and advanced words. Journal of French Language Studies, 18(03), 299-313.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. Language Testing, 17(1), 65.