THE WRITING PROCESS IN L2 SPANISH STUDENTS AND HERITAGE

SPEAKERS: A KEYSTROKE LOGGING STUDY

An Undergraduate Research Scholars Thesis

by

JORGE VELA DE LA CRUZ

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisors:

Dr. Gabriela Zapata
Dr. Patrick Bolger

May 2018

Major: Spanish
Psychology

TABLE OF CONTENTS

ABSTRACT

The Writing Process in L2 Spanish Students and Heritage speakers: A Keystroke Logging Study

Jorge Vela de la Cruz
Department of Hispanic Studies & Psychology
Texas A&M University

Research Advisor: Dr. Gabriela Zapata
Department of Hispanic Studies
Texas A&M University

Research Advisor: Dr. Patrick Bolger
Department of Hispanic Studies & Psychology
Texas A&M University

This study belongs to the field of applied linguistics, and, in particular, the area of first and second language acquisition. The two populations of interest in this project are Spanish-English bilinguals in the US, who are referred to as heritage language learners (HLLs), and Spanish second language (L2) learners. This study uses the keystroke logging software *Inputlog* to explore the different characteristics found within the writing process of HL and L2 learners. Through *Inputlog* we gathered data on writing samples from university students in L2 Spanish classes belonging to three different levels of Spanish proficiency--low, intermediate, and advanced. To further investigate the syntactic complexity of these two groups of bilinguals we also conducted a T-unit analysis. The results show the existence of clear differences between HL and L2 learners, particularly in the complexity of their writing and the kinds of errors they made. Also, the findings point to differences among the L2 learners, despite the fact that they were all placed in the same L2 Spanish classes. The data in this study suggest that different methodologies might be needed when teaching Spanish to HL and L2 learners.

ACKNOWLEDGEMENTS

I would like to extend my most sincere gratitude to Dr. Zapata and Dr. Bolger for their guidance and support throughout the course of this research project.

I would also like to the thank the Undergraduate Research Scholars and the Hispanic Studies departments at Texas A&M University for providing the resources and assistance necessary for this research project to be successful.

Last, but not least, I would like to thank all of my friends, faculty, and staff at Texas A&M for their ongoing support.

KEY WORDS

Spanish heritage language learners

Spanish second language learners

Second Language Acquisition

Early Bilinguals

Writing Process

INTRODUCTION

The Hispanic population in the United States constitutes the largest growing minority in the country (United States Census Bureau, 2011), and the presence of Heritage Language Learners (HLLs) can be seen more and more in second language (L2) Spanish classes across the country (Montrul, 2010). Instructors in charge of these classes are often faced with an important challenge: The need to provide heritage students with instruction that can answer their needs, which are often different from those of L2 students. Unlike L2 students, most HLLs in the United States have learned Spanish through informal settings, and might have poor literacy skills (Benmamoun, Montrul, & Polinsky, 2013). Even when HLLs can use Spanish fluently, they do not have explicit knowledge of how this language functions to convey specific meanings in different kinds of written texts, which is an essential aspect of literacy. For example, in a study conducted by Potowski, Jill, and Morgan-Short (2009) HLLs responded differently than L2 learners to grammatical-based instructional treatments, which pointed to the need to engage in research endeavors to further explore these populations of students to establish different instructional methods that would be beneficial to their language development. Lynch (2003) also believes it is important to compare HL and L2 learners to shed more light onto the similarities and differences between these two groups of bilinguals and thus broaden our understanding of both, and provide effective language instruction to both.

This need is particularly important in the area of writing. To the best of the author's knowledge, only one study has investigated writing in HL populations. Mikulski and Elola (2011) conducted a study that explored Spanish HLLs' writing performances in English and Spanish. The study focused on the cognitive processes that take place during writing, more

specifically, planning time, text production time, fluency, and accuracy. The researchers used screen-capture software to record the behavior of 12 university Spanish heritage language learners, while they answered writing prompts in both English and Spanish. The study showed that HLLs took a significantly larger amount of time to produce text when writing in Spanish. Furthermore, HLLs had higher levels of fluency and accuracy in English (dominant language), compared to Spanish. The larger implications for this study suggest that teaching strategies directed at HLLs should aim to have them complete informal writing tasks at first and successively move towards more academic oriented writing tasks. This approach will help HLLs develop explicit connections between English and Spanish instruction, increasing their level of writing proficiency in Spanish.

Even though this work has given us important information about the writing process of HLLs, it has not compared it to that of L2 learners. Considering that these two populations of students are often present in L2 classes, it is important to do so to provide both with the most effective instruction. This study seeks to address this need by exploring the characteristics of HLLs' writing as compared to that of L2 students through keystroke logging software. In the next section of the study, we discuss existing work with keystroke logging, and in particular with *Inputlog*, to highlight its advantages in the study of writing.

Keystroke Logging

In order to examine the writing processes of both HL and L2 learners, we used a keystroke logging program known as *Inputlog*. *Inputlog* is a word processor software that registers mouse movements, keystrokes, clicks, and pauses in Microsoft Word and other Window writing software (Leijten, Macken, Hoste, Van Horenbeeck, & Van Waes, 2012). This software is important in the study of writing because it provides a dynamic writer-based

perspective, as it records and dissects the writing activity as it takes places, allowing researchers to see "what the writer does instead of [presenting on] what the final product looks like" (Applebee, 1986, cited by Spelman Miller & Sullivan, 2006, p. 2). Keystroke logging has been widely used by many researchers in the field of linguistics to get information about the linguistic and cognitive processes at play during writing. For example, Leijten and his colleagues (2012) used output data from *Inputlog* to investigate the possibilities of linking logged process data from the character level to the word level by combining them with existing linguistic information and using natural language process tools. This study analyzed existing narratives and the *Inputlog* data associated with them. The results showed that enriching writing process data with linguistic annotation provides fundamental information about underlying writing processes. Furthermore, this study provided a basis for other researchers to engage in more linguistically-oriented research to explore writing processes across different languages.

In a similar study, Baaijen, Galbraith and de Glopper (2012) identified the different underlying processes that could relate to each pause, and/or revision within their participants' text. In their work, they focused on the writing processes of 80 participants recruited at the University of Groningen. Participants were asked to produce an article for the university newspaper; however, before producing the text, half the participants had to engage in synthetic planning (write a one sentence summary), and the other half, in outline planning (create an outline). They were then asked to create a well-structured and complete article in 30 minutes. The raw output produced by *Inputlog* was then analyzed to investigate the connections between keystroke-logging measures and underlying cognitive processes. The results of this study suggested that pauses between words, found outside of routine transitions between words in the collected writing samples, could presumably be attributed to higher-level thinking processes such

as planning, lexical retrieval, and/or phrase structure processing. This is important because it indicates that some writing components can be independent of one another; thus, writing needs to be investigated across different facets.

The results of these studies point to the potential of the keystroke logging software *Inputlog* to provide in-depth information about the characteristics of the writing activity of different populations of learners, and this is the main reason why it was chosen for this study. That is, by applying *Inputlog* to the analysis of the writing of both HL and L2 learners, we hoped to discover the intricacies of their writing activity as it unfolded in real time, and to find similarities and differences between these two populations of learners. In the next sections, we introduced the present study.

CHAPTER I

METHODS

Research Objectives

The main objective for this study was to investigate the similarities and differences that exist in the way in which HL and L2 learners approach and write a text. We were also interested in exploring these groups of bilinguals' syntactic complexity to further develop an understanding of the intricacies behind the texts they produced.

Participants

Our pool of participants consisted of 33 Texas A&M students recruited from three L2 Spanish classes in the Department of Hispanic Studies (two intermediate classes: Spanish 302 and 304, and one advanced class: Spanish 407). Five participants were eliminated because of problems with their data due to computer and software malfunctions. This left us with data from 28 participants. Our participants had an average age of 21, and consisted of both male (n=6) and female (n=22) students. Fifteen were HLLs, and 13 were L2 learners.

Procedures

The data collection process for this study consisted of two parts. Participants first completed a modified DELE intermediate proficiency exam in order to determine their Spanish language proficiency. This exam had been used in previous studies with HLLs (e.g., Montrul 2002, 2004, and 2006), and it had been deemed effective in the determination of HLLs' proficiency level. Based on the results of this test, all HLLs (n=15) were classified as advanced. L2 learners, however, exhibited different levels of proficiency: 4 were advanced, 6 were intermediate, and 3 were low.

Participants were also asked to self rate both their Spanish and English proficiency in reading, speaking, listening, and writing, resorting to a scale ranging from 1 to 5, with 1 being equivalent to "poor" and 5 to "native speaker command." The results of the self-ratings (Tables 1 and 2) confirmed those offered by the DELE, showing the existence of differences among participants in the two groups, particularly in the area of writing and speaking. That is, the standard deviation values suggest that, among the HLLs, the area with the highest rate of variability was writing, while in the case of L2 learners, it was speaking. These findings might reflect the exposure these two populations of students might have had to Spanish and their opportunities for language use. For example, as explained in the introductory section of this study, HLLs often learn and use Spanish in informal, family-related settings, and they do not receive formal education (schooling) in the language (Montrul, 2010). Thus, they may have little or no experience with writing and, as the self-rating data suggest, may have differing levels of experience and proficiency in this skill. Conversely, L2 students who learn Spanish through formal instruction might not have many opportunities to communicate in Spanish orally, and therefore, they may exhibit different proficiency patterns and levels of comfort in the speaking skill, which is reflected in the standard deviation values registered in Table 2. These types differences in HL and L2 learners have been reported in the existing literature (e.g., Montrul 2006, 2010, 2012).

Table 1.

HLL's Self-ratings

HLL's	Reading (Eng)	Speaking (Eng)	Listening (Eng)	Writing (Eng)	Reading (Span)	Speaking (Span	Listening (Span)	Writing (Span)
Mean	4.332	4.667	4.867	4.667	4.33	4.4	4.8	3.73
Median	5	5	5	5	5	5	5	4
SD	.834	.816	.352	.816	.899	.737	.561	1.223

Table 2.

L2 learners' Self-ratings

HLL's	Reading (Eng)	Speaking (Eng)	Listening (Eng)	Writing (Eng)	Reading (Span)	Speaking (Span	Listening (Span)	Writing (Span)
Mean	4.332	4.667	4.867	4.667	4.33	4.4	4.8	3.73
Median	5	5	5	5	5	5	5	4
SD	.834	.816	.352	.816	.899	.737	.561	1.223

1= Poor; 2=Need work; 3= Good; 4= Very good; 5= Native speaker command

In the second portion of this study, the participants wrote (typed) a narrative based on the illustrations in the book *Frog, Where Are You?* (Mayer, 1969). This story was chosen because it had been used in previous *Inputlog* studies (e.g., Leijten et al., 2012). Figure 1 shows the title page of the short picture book and its first two pages. No time or word limits were given to the participants, who were told to tell the story as comprehensively as possible. While participants

wrote their narratives, *Inputlog* recorded aspects of their writing activity: pauses, deletions, edition, planning, etc. The program then yielded this information graphically and numerically, which allowed us to characterize the different writing patterns displayed by our participants.
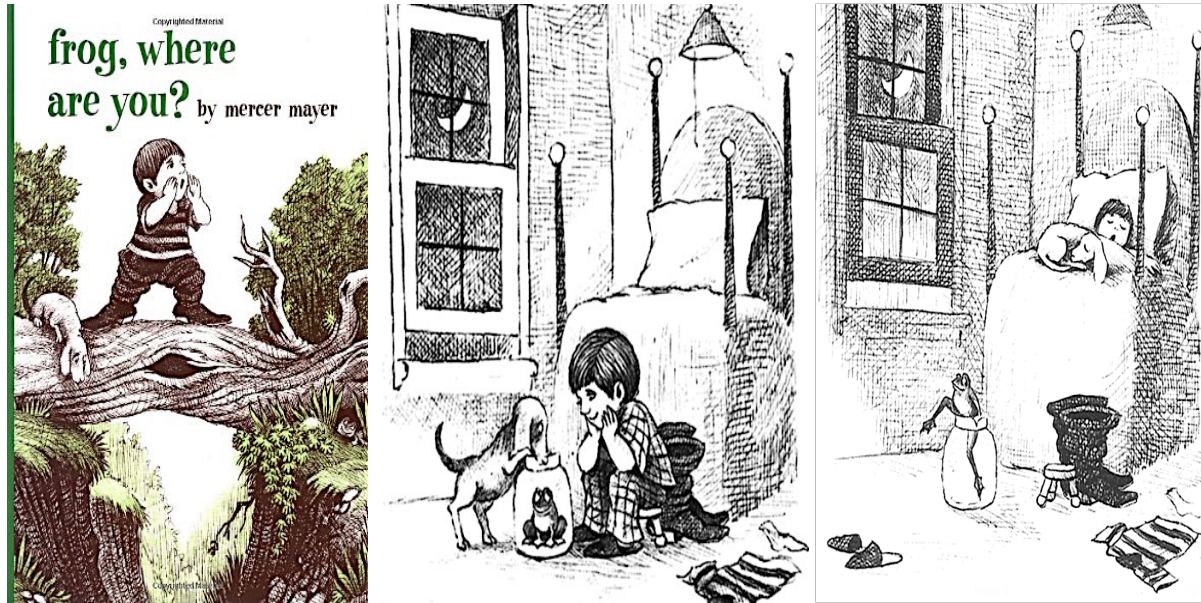

Figure 1. The title page and first two pages of the picture book *Frog Where are You?* (Mayer, 1969).

T-Unit Analysis

The narratives produced by the participants were also analyzed using T-unit analysis, a common methodology in the study of L2 writing (Bardovi-Harlig, 1992; Gaies, 1980). Bardovi-Harlig establishes that "a sentence has two (or more) T units when independent clauses (with subjects and finite verbs) are conjoined, but a single T unit when one or more clauses are embedded in an independent clause" (p. 390). Hunt (1970) suggested that longer T units reflect more syntactic complexity, which could also point to higher proficiency in the language. Our objective was, thus, to compare T units in the writing produced by our two populations of participants to establish similarities and differences in the complexity of their products.

To determine how many T units each writing sample from each participant contained, T units where counted following the guidelines established by Hunt (1965, 1970). We identified a T unit as a clause with a subject and a finite verb. If a dependent clause was attached to a T-unit, the whole statement was classified as one T-unit. The next step consisted in counting the number of words per T unit. The total number of words per T-unit for each group of bilinguals was then obtained by dividing the average word count for all the writing samples in each group by the total average of T-units found within the writing samples of that same group. However, only L2 advanced learners were compared to HLLs.

CHAPTER II

RESULTS

The results of the study show that overall HLLs produced narratives with more words, compared to the L2 learners. HLLs wrote between 218 and 686 words with an average writing time of 31.63 minutes. On the other hand, L2 learners wrote between 118 words (low proficiency) and 560 words (intermediate proficiency) with an average writing time of 31.26 minutes. The average times for both groups to complete the writing portion of the study were very similar; nonetheless, overall the HLLs wrote more words.

L2 Learners

L2 learners: Low proficiency

As seen in previous studies (e.g., Strömqvist et al., 2006), the participating L2 learners with a low proficiency level wrote fewer words and had narratives that were more linear compared to the L2 learners with intermediate and advanced proficiencies. That is, the similarity in characters produced and characters in the final product recorded by *Inputlog* showed that these students did not make major revisions to their text and they wrote in a linear fashion (as opposed to the more recurring pattern of writing and revision). Also, they had longer pauses within the text and more inaccuracies in language usage compared to the other proficiency groups. For example, their writing exhibited errors in verb conjugation and agreement as well as simple vocabulary and the presence of English. Overall, these participants had simple stories with basic cohesive devices. Figure 2 shows the data provided by *Inputlog* for the writing sample of one of the participants at the low proficiency level. The graph depicts the linearity of the text, the pauses made, and the differences between the initial text (blue line) and the final one (green line)
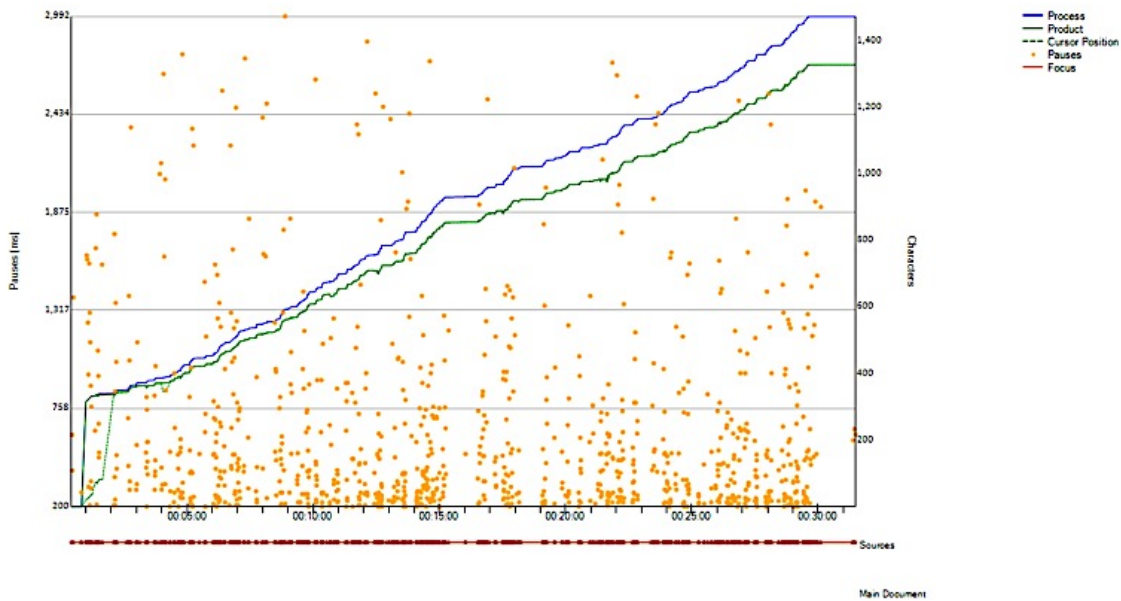
Figure 2. L2 student with low proficiency (179 words in 29.36 minutes)

L2 learners: Intermediate proficiency

The results for the L2 learners at the intermediate level were more varied than those for low proficiency students. Some intermediate L2 learners wrote more words; however, that was not always the case. Also, intermediate students had less linear writing than their least advanced counterparts, but this was not the case for all of them. In addition, they had more revisions at the text level and a combination of short and long pauses. Overall, the students in this group wrote more sophisticated stories and were more accurate. Errors, however, were still present, but not at the vocabulary level. Figure 3 shows data provided by Inputlog for one of the writing samples from an L2 intermediate student.
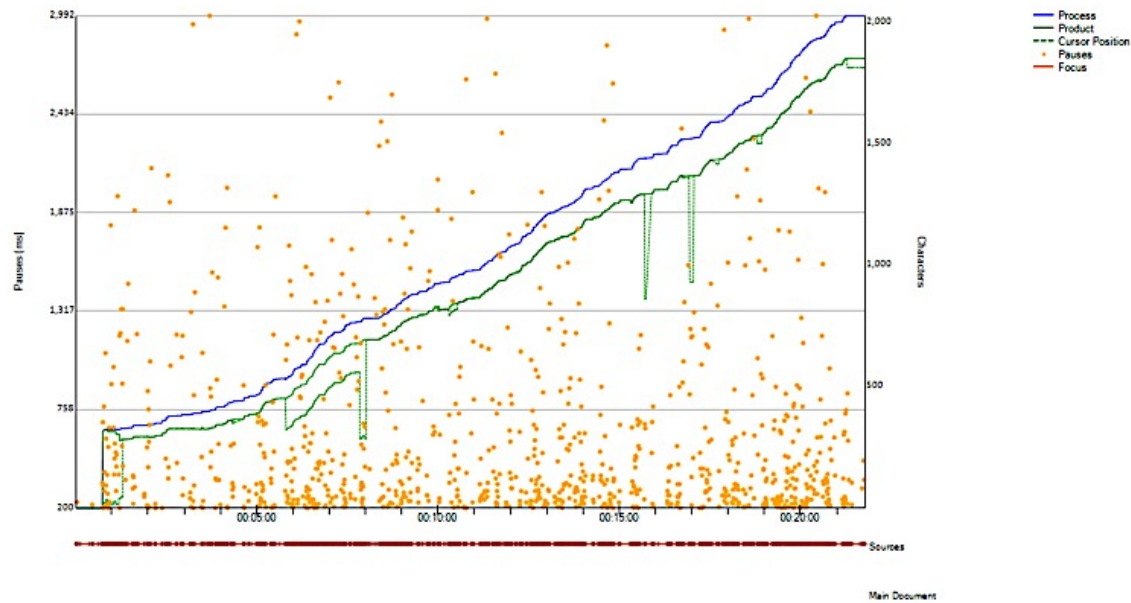
Figure 3. L2 student with low intermediate proficiency (score of 21/40) (276 words in 21.18 minutes)

 L2 learners: Advanced proficiency

The L2 learners at the advance proficiency level did not write as many words as expected. For example, they had a lower number of words than some of the intermediate students: Their range of words was 177 to 356, while that of intermediate students was 201 to 560. Advanced students had some linear writing, but not always, and they made more revisions at the text level then the other two L2 proficiency groups. In addition, they had short and long pauses and more sophisticated stories, and though there were still errors in their writing, they did not register at the vocabulary level. The advance L2 learners also showed planning at the beginning of their writing, which had been reported in previous studies (Strömqvist et al., 2006). Figure 4 shows the data produced by Inputlog for the writing sample from an advance L2 student.
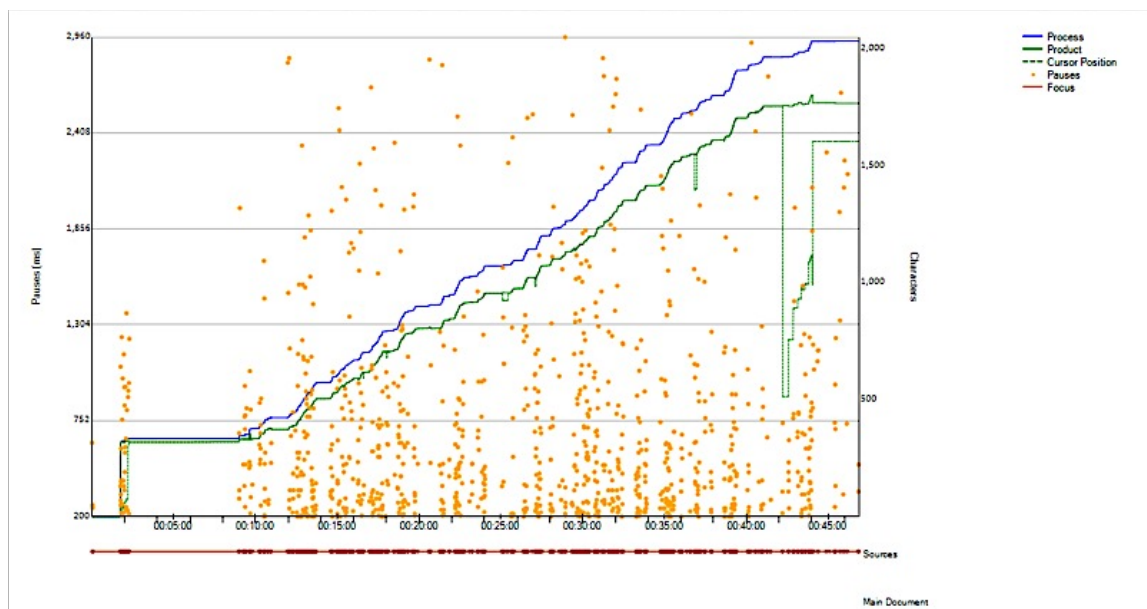
Figure 4. L2 student with a mid advanced proficiency level (score of 34/40) (279 words in 43.22 minutes).

Heritage Language Learners

The results for HLLs showed variation in the number of words within the writing samples, with a range between 216 and 686 words. For some participants, there was some linear writing with no major revisions in characters produced compared to the final product. However, some participants did make major text revisions. Overall, data for HLLs showed that participants had short pauses within the text, wrote more cohesive and sophisticated stories, and some planned their narratives before writing. Also, they were more accurate than their L2 counterparts. However, HLLs' writings exhibited problems with orthography, lexical borrowings, and semantic calques. Figure 5 shows the data produced by *Inputlog* for the writing sample of an advance HLL. The graph shows a long pause at the beginning of the text production, indicating that some form of planning took place.
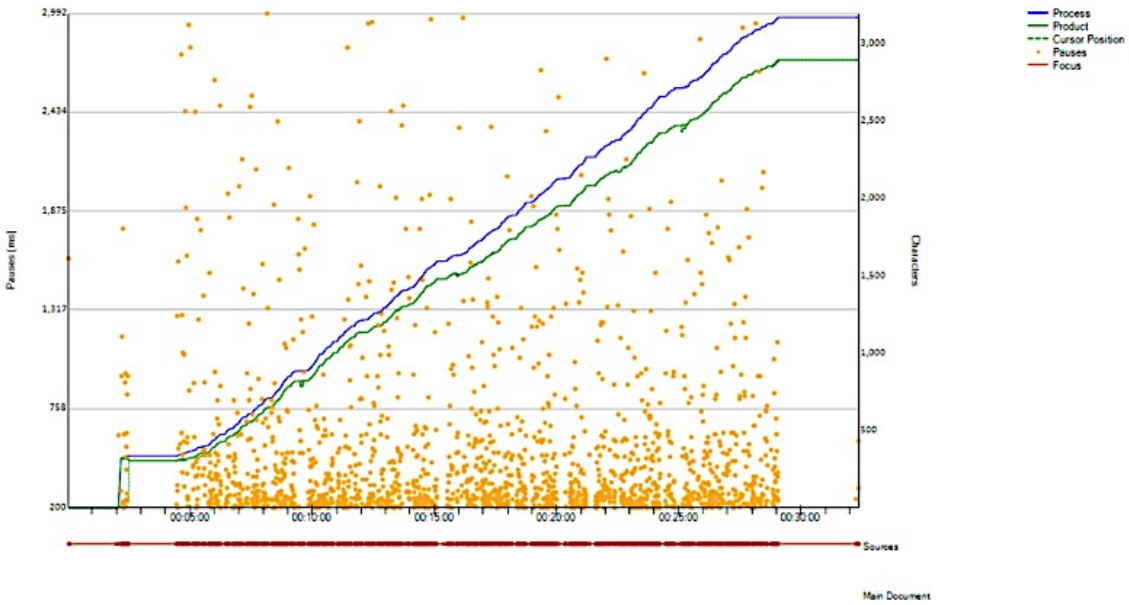
Figure 5. HLL student with a high-advanced proficiency level (score of 37/40) (489 words in 29.04 minutes).

T-Unit Results for L2 Learners and HLLs

Overall, advanced L2 students had an average of 22.75 T units, with an average length of 8.43 words; on the other hand, the average T-units for HLLs participants was 36.92, with an average of 9.96 words. This shows that the writing produced by HLLs was more syntactically complex than that of their L2 counterparts.

CHAPTER III

DISCUSSION

The results for this study showed differences among L2 students in writing across individuals and different levels of proficiency. Intermediate and advanced students seemed to have made more revisions at the text level. Advanced students took more time to finish writing and sometimes wrote less, which indicates that this group of L2 learners were perhaps more metalinguistically aware than their counterparts. That is, advanced learners probably had more knowledge of grammatical rules, and thus, could have limited their writing and could have reflected more on their product in order to avoid errors. This is supported by previous research (e.g., Spellman Miller, 2005), which has shown that L2 learners can become cognizant of their own approaches to the writing task, and they can come to recognize how their attention to certain details might have an impact on the overall result of their product.

Another difference that manifested among the three groups of L2 learners was connected to the way in which the story was told. For example, low proficiency students wrote simple narratives, which might have been related to their lack of vocabulary. That is, in order to tell Mayer's (1969) story, participants had to resort to the low frequency words (e.g., frog, jar, pond, hole, etc.), which could have posited extra difficulties, and could also explain the presence of English in some of the low-proficiency narratives. In addition, there were different patterns of accuracy among the three groups of L2 participants. Overall, these findings suggest that, in spite of the fact that the L2 learners participating in the study were all part of the same L2 university classes, it was clear that they exhibited different levels of attainment in Spanish. The results have pedagogical implications because they imply that instructors cannot assume that all students are

the same even if they are in the same class, and that instructed L2 acquisition does not always result in the same proficiency level for students that have followed the same instructional path.

The study also shows differences among advanced L2 students and HLLs. The writing produced by HLLs was more cohesive and comprehensive: The content was more sophisticated, and the T-unit results suggest that it was more syntactically complex. Also, HLLs wrote more words in a shorter period of time, and exhibited more grammatical accuracy. However, HLLs' writing also pointed to their difficulties with literacy-related aspects such as orthography, and language transfer (e.g., borrowings and semantic calques), which has been reported in previous studies (e.g., Benmamoun, Montrul, & Polinsky, 2013; Montrul 2010, 2012).

Even though the results point to clear differences among L2 learners and between L2 learners and HLLs due to proficiency level and also the nature of their Spanish acquisition (i.e., instructed vs. naturalistic), the data also suggest that some of the findings could be related to individual differences. For example, not all HLLs had shorter pauses within the text. Also, some advanced L2 and HL learners did not make many revisions at the text level, nor did they plan their writing beforehand. There were also some L2 students within the same proficiency level that had differences in the amount of text they produced. These differences could be attributed to the different experiences with the language, or to individual preferences. However, there are limitations to this study, and the nature of its data cannot provide information to elucidate what aspects of the participant writings was indeed related to the personal nature of the process.

HLL's and L2 learners, as shown by the results in this study, have different linguistic needs that are not always addressed through language instruction. These needs may result due to differences in experience with language or informal language acquisition, and they should be addressed accordingly if the desired outcome is for students to develop competence in the target

language. Based on our results, HLLs seem to be more apt to benefit from language instruction concerning literacy skills. On the other hand, L2 learners would benefit more from instruction that allows them develop larger sets of vocabulary, and better language fluency. Conversely, both groups would benefit from instruction that addresses dominant language transfer. Pedagogical implications that can be derived from our results, regarding individual differences among participants, are that instruction directed towards HLL's and L2 learners cannot be simply seen as dichotomous. Although overall L2 learners may have the same problems in language usage when producing text, approaches of instruction should be modified to address the individual differences among students with different proficiency levels that belong to the same group of bilinguals.

This study makes important contributions to the existing body of knowledge regarding HL and L2 learners' writing processes; however, it is important to note some of the limitations within this investigation. The study had a very small sample size, which leaves room for error when trying to make our results applicable to the overall population of bilinguals. It is also plausible that the book *Frog, Where Are You?* (Mayer, 1969) was perhaps too difficult for our L2 learners. As stated earlier, the story required our participants to use low frequency words, which might have affected their production, limiting the comprehensiveness of their narratives and resulting in their use of English. Therefore, future research should focus on incorporating a writing task based on a simpler story. Another limitation of the study is that we did not produce any statistical data to determine statistical significance or other patterns within our results. Nonetheless, this work serves as a basis for future research involving the study of writing processes amongst HL and L2 learners using keystroke logging software.

CONCLUSION

This study examined the differences and similarities in the writing processes of HL and L2 learners in order to provide information that might result in the development of teaching strategies that could be appropriate for both populations of learners. Although results are limited, this study shows supporting evidence that different methodologies might be needed when teaching HLLs compared to L2 learners. For instance, instruction directed at L2 students should be focused on vocabulary, grammar, and in establishing ways to increment Spanish fluency. On the other hand, instruction directed at HLLs should focus on strengthening literacy skills (e.g., orthography) in their first language. Furthermore, this study shows that there are differences in the writing process among individuals, despite belonging to the same group of bilinguals. Therefore, instructors need to be aware not only of differences in language proficiency, but they also they need to consider individual differences among their students, which suggests that a one-size-fits-all approach might not be effective in the L2 classroom. Being aware of these discrepancies will allow educators to best instruct their students and aid in their language development.

REFRECNES

Baaijen, V. M., Galbraith, D, & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication, 29*(3) 246–277.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly, 26*(2), 390–395.

Benmamoun, E., Montrul, S. & Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theoretical Linguistics, 39*(3-4), 129-181.

Gaies, S. J. (1980). T-Unit analysis in second language research: Applications, problems, and limitations. *TESOL Quarterly, 14*(1), 53-60.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels: Research Report No. 3*. Urbana, Illinois: National Council of Teachers of English.

Hunt, K. W. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development, 53*(1)(Serial No. 134).

Leijten, M., Hoste, V., Van Waes, L., Macken, L., Van Horenbeeck, E. (2012). From character to word level: Enabling the linguistic analyses of Inputlog process data. In *Proceedings of the EACL 2012 Workshop on Computational Linguistics and Writing* (1–8). Avignon, France: Association for Computational Linguistics.

Lynch, A. (2003). The relationship between second and heritage language acquisition: Notes on research and theory building. *Heritage Language Journal, 1*, 26-43.

Mayer, M. (1969). *Frog, where are you?* New York: Dial.

Mikulski, A., & Elola, I. (2011). Spanish heritage language learners' allocation of time to writing processes in English and Spanish. *Hispania, 94*(4), 715-733.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition, 5*, 39-68.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition, 7*, 125-142.

Montrul, S. (2006). On the bilingual competence of Spanish heritage speakers: Syntax, lexical semantics and processing. *International Journal of Bilingualism, 10*(1), 37-69.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics, 30*, 3-23.

Montrul, S. (2012). The grammatical competence of Spanish heritage speakers. In S. M. Beaudrie & M. Fairclough (Eds.), *Spanish as a heritage language in the United States* (pp. 101-120). Washington: Georgetown University Press.

Potowski, K., Jegerski, J., & Morgan-Short, K. (2009). The effects of instruction on linguistic development in Spanish heritage language speakers. *Language Learning*, 59, 537–579

Spellam Miller, K. (2005). Second language writing research and pedagogy: A role for computer logging? *Computers and Composition, 22*(3), 297–317.

Spelman Miller, K., & Sullivan, K. P. H. (2006). Keystroke logging: An introduction. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 1-10). Amsterdam, Netherlands: Elsevier

Strömqvist, S., Holmqvist, K., Johansson, V., Karlsson, H., & Wengelin, A. (2006). What keystroke logging can reveal about writing. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: Methods and applications* (pp. 45-72). Amsterdam, Netherlands: Elsevier.

United States Census Bureau (2011). *The Hispanic population: 2010. Census briefs*. Washington: U.S. Department of Commerce Economics and Statistics Administration. http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf.