

Final Empirical Paper
Psychology of Language
Professor Madhavi Maganti

Impact of Native Language and Gender on English as Second Language Acquisition

Sk Zafar Ali & Sita Sanjivini Sawhney

Ashoka University

May 7, 2018

Abstract

The aim of the present article is to understand the roles of an individual's gender and his/her native language on his/her acquisition of English as a second language. It analyzes longitudinal corpus data of 19 children from the CHILDES database who had migrated to Canada at the average age of 3 years and 5 months. The data consists of transcribed conversations in a natural setting in 5 rounds over a period of two years, tracking the children's learning progression in English. Student's t-test and ANOVA tests were conducted on the Mean Length of Utterance (MLU) and the Lexical Diversity Indices to find correlations with respect to gender and native languages. The analysis of the results showed that female children learnt faster when compared to male children, and were also more proficient. For the second hypothesis of the impact of native language on ESL acquisition, it was found that children with East Asian native languages were better at learning English when compared to those with Indo-European native languages.

Keywords: Native Language, Gender, ESL, Mean Length of Utterance, Lexical Diversity Index

Introduction

In a constantly changing and progressing world where cultures are integrating, language has become a crucial aspect to human development and cognitive evolution. A large portion of the world population is bilingual, understanding and comprehending the various other cultures and populations that human societies have to offer. There is a significant amount of migration in the world as well. Populations, groups and people have been moving across borders in hopes for better opportunities and livelihood. In such instances, language is crucial for how they adapt to new places. Motivated by these thoughts, we decided to understand the learning of a second language in individuals who migrate from one region to another. To further narrow the aim of our study, we decided to focus on children, and their acquisition of a non-native second language. Research on second language acquisition of migrant children is interesting, since it also contributes to the understanding of the development of the upcoming generations. Since English has become a mode of communication in most countries across the world, and also a medium of cross-national and cross-cultural interactions, the learning of English is an important milestone in socio-personality development of any individual (Yip, Mathews, 2006). Thus, we decided to study the acquisition of English as a second language (ESL) in children under the age of 6 years. The major impact on learning ESL for such young children would be the influence of their native language and/or mother tongues on the L2 acquisition. Therefore, the research question for our study was how the native/mother tongue languages of immigrant children impact the learning of ESL during the critical periods of language acquisition. We also decided to examine the prospective differential impact of gender on the learning of L2. For the purpose of the study, the experiment involves the ESL proficiency of immigrant children in Canada.

Research on L2 acquisition has been conducted very extensively in the field of linguistic psychology. The Cummins' hypothesis also suggests that the level of competence in

L1 impacts the learning competence of L2 for a bilingual child (Koçak, 1987). However, this does not take away from the fact that young children, especially the ones in the critical language period are not completely proficient in L1, and the simultaneous exposure and learning of L2 is bound to impact the holistic acquisition of either of the two languages, which the Cummins' hypothesis does not address. Eric Lenneberg's research on language acquisition explains language learning as a function of the critical, biological period of language acquisition, which is between 2 years and 12 years of age (Ipek, 2009, 158). This notion was originally linked to, and explained the acquisition of L1, however, owing to advancement in research, it has now been extended to the understanding of L2 as well (Ipek, 2009, 158). This also supports the motivations behind our research question, to understand whether there is a significant relationship between the acquiring of L1, and the linguistic properties of L1, on the learning of ESL. Although, most research agrees that learning L2 is dependent on individual's proficiency in L1.

Literature Review

To further our knowledge about the patterns in language acquisition, it is important to understand the kind of research that's already been conducted in the field. Research has predicted that there are certain trends in how individuals acquire language; these are universals that are common to most languages. Krashen (1982) puts forth the Natural Order Hypothesis, according to which, learning of L2 is predictable, owing to the rules of languages, which are bound to be common for both L1 and L2. These are not the explicitly taught rules of language, but rather the patterns in which a language is taught (Ipek, 2009, 156). An example of this could be the sentence structures in particular languages, or even the universal learning of noun before verbs in almost all languages. These are highly dependent on the morphemes of the particular language being learnt. However, even though L1 and L2 seem to be following the

same routes of acquisition especially in the critical language learning period, studies have also shown how not all first language learners follow these trends (Ipek, 2009, 157). There is also a significant influence of individual differences, culture, background, and especially the linguistic nature of L1, that impact L2 acquisition. These affects, especially those of the native languages, according to McLaughlin (1987), can have both, beneficial or a disagreeable impact on the L2 learning (Ipek, 2009, 157). To control for most confounding variables, especially the individual differences of the speakers, the study looks largely at the Mean Length of Utterance (MLU) of the participants, to control for the grammatical, syntactical, lexical errors the young participants might commit.

Gender and Language Acquisition

Considering how language acquisition is a cognitive and biological process, the study also attempts to draw correlations between gender and the L2 acquisition. Research on L2 acquisition in adults has time and again proven to be faster and more proficient in females than in males (Feery, 2008). Van der Silk, Van Hout, and Schepens (2015) findings support this claim with evidence on how, despite their native languages, women have outperformed men in the learning and proficiency of Dutch as a second language. The question of whether these findings can be extended to children, is answered by the research by Ineta Dabašinskienė (2012), who suggests that, it is the difference in the socio-cultural environments provided to men and women as children, that impacts their language acquisition. Since men are provided with toys like trucks and cars, their development at that age is more inclined to spatial development. However, women, who are provided with dolls and house-playsets, are required to have communicative interactions, contributing to better socio-linguistic proficiencies. Thus, gender does play a role in language acquisition, and the results of our study attempt to understand their role in ESL acquisition.

Native Language Acquisition

Consequently, it is also crucial to understand the extent and the possible ways in which a native language might impact the learning of ESL. The major question here is, does the similarity between the sentence structure and language rules between L1 and L2, cause an interference of L1 in the learning of L2? Research says, for adult L2 learners, the closer the structures of the previously learnt language and the new language are, the easier and faster is the learning (Yip, Mathews, 2006). Nevertheless, the learning of L2 for children under the ages of 6 is more complicated since they hardly have proficiency in either of the languages, and the interference of learning one with the other, is highly possible. However, the learning of L1 is still crucial for the efficient learning of L2, since the active L1 creates a linguistic structure for the learning of L2 and promotes easier learning of a different language (Derakhshan, Karimi, 2015). Though the exposure and the age of L1 learning is an important aspect to consider. Thus, our study controls for the age of the participants and the age of their L2 acquisition. Since the learning of language during the critical period in childhood is a rather subconscious, informal, and implicit process, the L1 learning is easier in its impact on the plasticity of the brain (Derakhshan, Karimi, 2015). Consequently, even though L1 and L2 learning, especially in children follow similar routes of acquisition, the phenomenon of learning two separate languages is differently conducted in the mind, where the nature of L1 impacts how well L2 is learnt.

Methodology

To work on the research question, we received access to CORPUS data available on the internet. The dataset we use for our study provided a comprehensive report for 19 young participants, who had migrated into Canada below the ages of 4 (Table 1, Appendix A). We decided on this dataset from the CHILDES archive for two reasons. One, the data available on

CHILDES is preformatted in a methodological format for easy comprehension and analysis, and two, the availability of the CLAN program to analyse speech data on the CHILDES database.

The process of data collection and composition, was a longitudinal, naturalistic process performed in Edmonton, Canada in 2002-2005 by Johanne Paradis (Department of Linguistics, University of Alberta) on 25 immigrant children who were learning English as their second language. The children were recorded talking in English in a naturalistic setting and their conversation was transcribed in CHAT format using CLAN software. This was done in 5 rounds over a period of two years for each child.

We downloaded the data as 95 CHAT files (19 children, 5 rounds per child). We eliminated data for 6 children as these did not have data for all five rounds. The native languages of the participants ranged from Eastern (Mandarin, Japanese, Cantonese, and Korean) to European (Romanian, Spanish) and included Middle Eastern (Arabic, Farsi) languages. The average age at which these children started learning English was 4 years and 7 months. Their first languages had been recognized and established by this age, however, we do not have record for their proficiency in their L1. At the start of the study, the average age of the children was 5 years and 4 months. The rounds were conducted at an interval of about six months. 12 of these children were male, and 7 were female.

Analysis

We analysed the CHAT files in CLAN software and copied the results in a spreadsheet to be further analysed in JASP. For the link to the analysis, see Appendix A.

In CLAN, we ran the following commands:

1. MLU – To get the mean length of word utterance (total number of words divided by total number of utterances) per child per round.
Output – Word MLU and Standard Deviation.
2. MLU (%mor) – To get the mean length of morpheme utterance (total number of morphemes divided by total number of utterances) per child per round.
Output – Morpheme MLU and Standard Deviation.
3. VOCD – To get the Type Token Ratio (TTR – The total number of different words divided by the total number of words. A high TTR indicates a high degree of lexical variation while a low TTR indicates the opposite) and the Lexical Diversity Index, which estimates the skill of language from vocabulary size and grammatical structure.
Output – Type Token Ratio and Lexical Diversity Index score.

We imported this data onto the spreadsheet and standardized the mean length of word and morpheme utterances by combining the MLU and SD into Z-scores of Mean Length of Word and Morpheme utterances. We then averaged these two Z scores to get a final MLU value. We ran tests on the Final MLU value and the Lexical Diversity Index. We ignored the Type Token Ratio as CLAN factored in the Type Token Ratio score to calculate the Lexical Diversity Index, and thus running analysis on the Type Token Ratio would be redundant.

We then ran the following tests on the Final MLU value and the Lexical Diversity Index:

1. Independent samples t-tests to check for learning differences between Gender – Gender as the grouping variable and MLU and Lexical Diversity Index as the dependent variable (Appendix B).
2. 2 ANOVAS to with MLU and Lexical Diversity Index as the dependent variables and Round, Gender and Language as the independent variables (Appendix B).

Results

We saw that females scored better than males on average when tested for MLU and Lexical Diversity Index. Results indicate that females ($M = 55.962$, $SD = 11.793$) perform significantly better than males ($M = 50.635$, $SD = 10.338$) in Lexical Diversity Index. ($t(93) = 2.299$, $p = 0.024$) (Output with Graph in Appendix B).

A 5 (Round: 1 vs. 2 vs. 3 vs. 4 vs. 5) x 2 (Gender: Male vs. Female) x 8 (Language: Romanian vs. Mandarin vs. Cantonese vs. Spanish vs. Korean vs. Arabic vs. Japanese vs. Farsi) mixed ANOVA on Final MLU Score revealed main effects of Round, $F(4, 45) = 5.109$, $p = 0.002$, and Language, $F(7, 45) = 10.069$, $p < 0.001$. However, there were no main effects of Gender, $F(1, 45) = 0.230$, $p = 0.634$. Moreover, there were no significant interactions between Round*Gender, $F(4, 45) = 0.653$, $p = 0.628$, Round*Language, $F(28, 45) = 0.698$, $p = 0.842$, Gender*Language, $F(1, 45) = 3.007$, $p = 0.090$, and Round*Gender*Language $F(4, 45) = 1.078$, $p = 0.379$. We see a steady increase in MLUs across the first four rounds with a marginal drop in the fifth round. We also see an exceptionally high value in MLU scores for Japanese children compared to other languages.

Another 5 (Round: 1 vs. 2 vs. 3 vs. 4 vs. 5) x 2 (Gender: Male vs. Female) x 8 (Language: Romanian vs. Mandarin vs. Cantonese vs. Spanish vs. Korean vs. Arabic vs. Japanese vs. Farsi) mixed ANOVA on Lexical Diversity Index revealed main effects of Round, $F(4, 45) = 6.052$, $p < 0.001$, Language, $F(7, 45) = 5.758$, $p < 0.001$, and Gender, $F(1, 45) = 7.708$, $p = 0.008$. Moreover, there were significant interactions between Gender*Language, $F(1, 45) = 7.701$, $p = 0.008$. In contrast, there were no significant interactions between Round*Gender, $F(4, 45) = 0.794$, $p = 0.536$, Round*Language, $F(28, 45) = 0.399$, $p = 0.994$, and Round*Gender*Language $F(4, 45) = 0.443$, $p = 0.777$. We notice a sharp increase in the Lexical Diversity Index between the first and second rounds and steady increases after that,

which plateau at the fourth and fifth rounds. We also notice higher Lexical Diversity Indices in Japanese, and Mandarin speakers, compared to speakers of other languages.

Discussion

Understanding the impact of one's native language on their acquisition of L2, is crucial to comprehend the development of the upcoming generations. This research attempted, not only to explain the significance of gender differences on the acquisition of L2, but also to understand the correlation between the native languages and L2. Though the reasons for females' greater proficiency in L2 than men, is still under debate, this study reinforces the impact of gender gap on the acquisition of new languages. To further understand the why this happens, we must look at the difference in socio-biological environments of the two genders, as pointed out by Dabašinskienė (2012).

However, the results also point toward an interesting aspect of analysis where the kind of children of certain native languages performed better than the rest. One notices that, children belonging to East Asian native languages performed significantly better than those of proto-Indo-European native languages like Spanish and Farsi. A possible reason for this could be the similarity in the structures of English and the non-East Asian languages in the study. Similar linguistic structures, and parallel learning of two languages could have caused interference. There could have been interference of the L1 and L2 learnings for the child since the lexical diversity and the structures of these languages are similar (Derakhshan, Karimi, 2015). However, the East Asian languages like Mandarin, Korean, and Japanese, showed significantly higher lexical diversities. This could be a result of non-interference and differentiated learnings of the native languages and of English, separating the acquisition process for the two.

Understanding of L2 acquisition can explain the proficiencies and capabilities of an upcoming, culturally integrated generation. It is possible to further the results of this study by

considering the domestic environments of children and how they impact the language acquisition abilities. There is also a need to understand the role of gender, and its impact of language and cognitive advancement. Today's societal structures are moving toward gender equality, and such biological, cognitive, and learning differences, especially based on the research by Dabašinskienė (2012) would be a great contribution. Although, our study did not consider the impact of native languages and their structures on that of genders, and whether females are more inclined to learn certain kind of languages better than males. Such a research could be very interesting and would contribute well to understanding whether there's inherent sexism in linguistic structures. This study attempts to contribute to the larger picture of how societal structures are changing today owing to impacts of globalization and movements like feminism around the world. Language being one of the key connectors of cognitive and the social world, is bound to be impacted by these and further research to understand it is crucial in the field of linguistic psychology.

Limitations of the Dataset

As interesting as the study was, there were a couple of reasons the analysis might be ideal for application to a larger sample. The size of the sample was merely 19 of the 25 children whose data had been collected. The size for such an intricate scientific analysis is very small to make larger generalizations. Within these 19 participants also, there wasn't consistency in the native languages, thus the native language hypothesis was tested on even smaller sample sizes. Such studies are highly dependent on the nature-nurture arguments, however, the data we received hardly had any information of the parents and the domestic environment of the participants. This is an important limitation as the environment of the participant could have either inhibited or encouraged the learning of ESL.

References

- Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., . . . Foryś-Nogala, M. (2017). How Does L1 and L2 Exposure Impact L1 Performance in Bilingual Children? Evidence from Polish-English Migrants to the United Kingdom. *Frontiers in Psychology*, 8. doi:10.3389/fpsyg.2017.01444
- Kocak, A. (1987). A relation between acculturation and second-language acquisition in the classroom: A study of Turkish immigrant children born in the Netherlands. *Journal of Multilingual and Multicultural Development*, 8(5), 3-41. Retrieved from <http://arno.uvt.nl/show.cgi?fid=141784>
- Yip, V., & Matthews, S. (2006). Assessing Language Dominance in Bilingual Acquisition: A Case for Mean Length Utterance Differentials. *Language Assessment Quarterly*, 3(2), 97-116. doi:10.1207/s15434311laq0302_2
- Ipek, H. (2009). Comparing and Contrasting First and Second Language Acquisition: Implications for Language Teachers. *English Language Teaching*, 2(2). doi:10.5539/elt.v2n2p155
- Derakhshan, A., & Karimi, E. (2015). The Interference of First Language and Second Language Acquisition. *Theory and Practice in Language Studies*, 5(10), 2112. doi:10.17507/tpls.0510.19
- Dabašinskienė, I. (2012). Gender Differences in Language Acquisition: A Case Study of Lithuanian Diminutives. *Journal of Baltic Studies*, 43(2), 177-196. doi:10.1080/01629778.2012.674795
- Van der Slik, F. W. P., van Hout, R. W. N. M., & Schepens, J. J. (2015). The Gender Gap in Second Language Acquisition: Gender Differences in the Acquisition of Dutch among Immigrants from 88 Countries with 49 Mother Tongues. *PLoS ONE*, 10(11), e0142056. <http://doi.org/10.1371/journal.pone.0142056>

Feery, Karen (2008) "Current Perspectives on the Role of Gender in Second Language Acquisition (SLA) Research," The ITB Journal: Vol. 9: Iss. 1, Article 4.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with Specific Language Impairment. *Language, Speech and Hearing Services in the Schools*, 36, 172-187

Appendix A

Table 1: List of children with Gender, Age, and Native Language

Serial	Code	Primary Language	Gender	Age of Arrival	Age of Acquisition	Age at Round 1	Age at Round 5
1	CHRS	Romanian	F	5.8	5.8	6.2	8.4
2	CNDX	Mandarin	F	6.1	6.1	6.8	8.8
3	DNLN	Cantonese	M	0.1	4.0	5.2	7.2
4	DNNC	Mandarin	M	4.4	4.6	5.3	7.4
5	DNNS	Mandarin	M	3.5	3.9	4.5	6.6
6	DVDC	Spanish	M	5.6	5.6	6.3	8.3
7	GSYN	Korean	F	4.0	5.0	5.2	7.2
8	JNNH	Mandarin	F	4.4	4.4	5.9	7.9
9	LLKC08	Arabic	F	0.1	3.9	4.8	6.8
10	MRSS24	Mandarin	F	4.7	4.7	5.0	7.0
11	RMLM10	Japanese	F	2.2	3.5	4.3	6.2
12	RNL28	Cantonese	M	0.1	3.3	4.7	6.7
13	SBST17	Spanish	M	3.8	3.8	5.1	7.2
14	SHHN32	Farsi	M	5.3	5.5	6.5	8.5
15	SMNS22	Spanish	M	5.0	5.0	5.5	7.4
16	THRJ33	Farsi	M	3.1	3.3	4.2	6.2
17	TNYN20	Mandarin	M	5.7	5.8	6.4	8.6
18	TRRK11	Arabic	M	0.1	3.5	4.2	6.6
19	YSSF12	Arabic	M	0.1	4.8	4.9	6.9

Link to the dataset : <https://chilides.talkbank.org/access/Biling/Paradis.html>

The analysed files along with the Excel and JASP outputs can be found at <https://github.com/zafar21/Bilingual-CHILDES>.

Appendix B

Independent Samples T-Test

	t	df	p
Lexical Diversity Index	2.299	93.00	0.024
MLU Final	1.175	93.00	0.243

Note. Student's t-test.

Group Descriptives

	Group	N	Mean	SD	SE
Lexical Diversity Index	Female	35	55.962	11.793	1.993
	Male	60	50.635	10.338	1.335
MLU Final	Female	35	0.008	0.364	0.061
	Male	60	-0.077	0.324	0.042

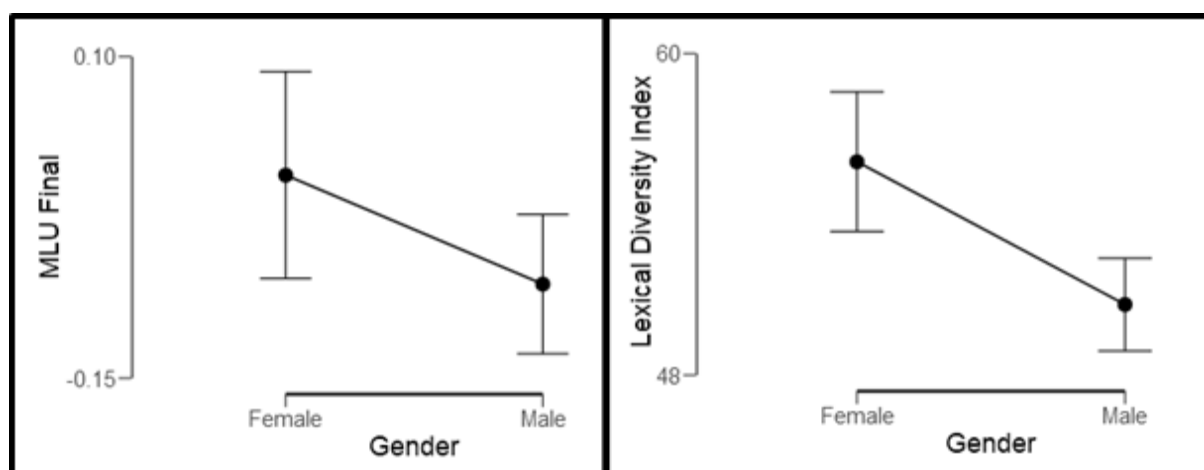


Fig 1: MLU and Lexical Diversity Index by Gender

ANOVA - MLU Final

Cases	Sum of Squares	df	Mean Square	F	p
Round	1.220	4	0.305	5.109	0.002
Gender	0.014	1	0.014	0.230	0.634
Language	4.209	7	0.601	10.069	< .001
Round * Gender	0.156	4	0.039	0.653	0.628
Round * Language	1.168	28	0.042	0.698	0.842
Gender * Language	0.180	1	0.180	3.007	0.090
Round * Gender * Language	0.258	4	0.064	1.078	0.379
Residual	2.687	45	0.060		

Note. Type III Sum of Squares

Simple Main Effects - Round

Level of Gender	Sum of Squares	df	Mean Square	F	p
Female	0.539	4	0.135	2.256	0.078
Male	1.643	4	0.411	6.876	< .001

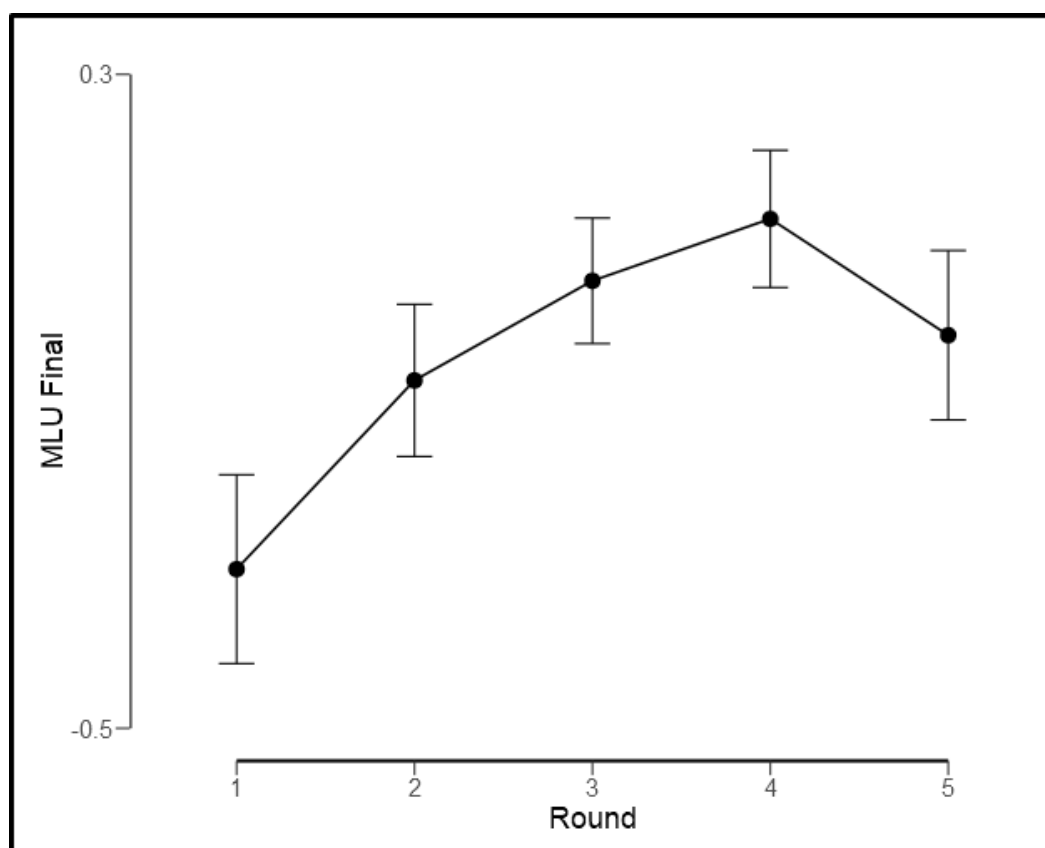


Fig 2: Standardized MLU Scores with 80% Confidence Intervals

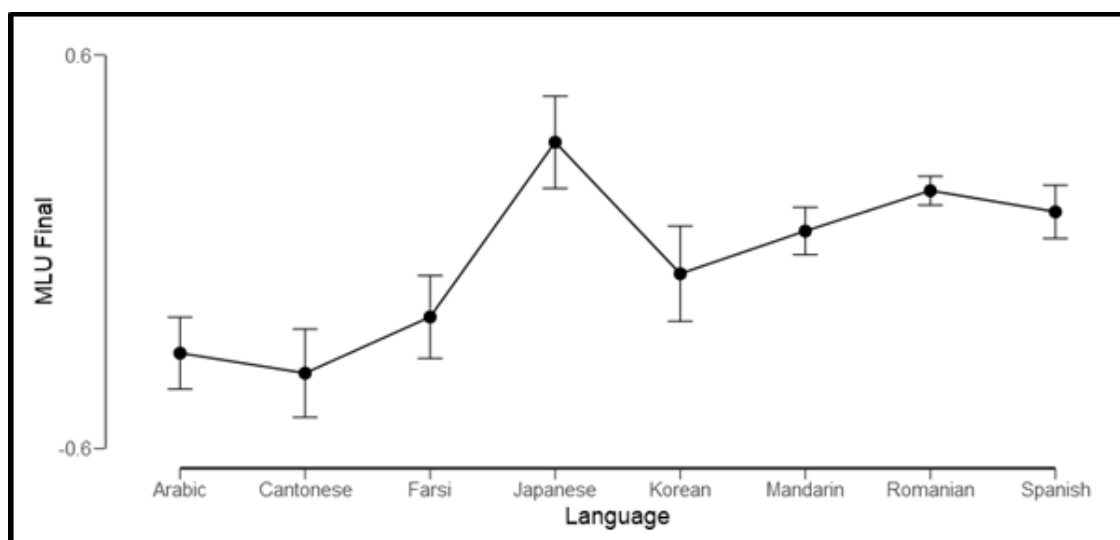


Fig 3: Standardized MLU Scores by Language

ANOVA - Lexical Diversity Index

Cases	Sum of Squares	df	Mean Square	F	p
Gender	577.4	1	577.37	7.708	0.008
Language	3019.1	7	431.30	5.758	< .001
Round	1813.4	4	453.36	6.052	< .001
Gender * Language	576.9	1	576.89	7.701	0.008
Gender * Round	237.8	4	59.45	0.794	0.536
Language * Round	837.2	28	29.90	0.399	0.994
Gender * Language * Round	132.7	4	33.17	0.443	0.777
Residual	3370.9	45	74.91		

Note. Type III Sum of Squares

Simple Main Effects - Round

Level of Gender	Sum of Squares	df	Mean Square	F	p
Female	512.6	4	128.2	1.711	0.164
Male	2898.2	4	724.6	9.673	< .001

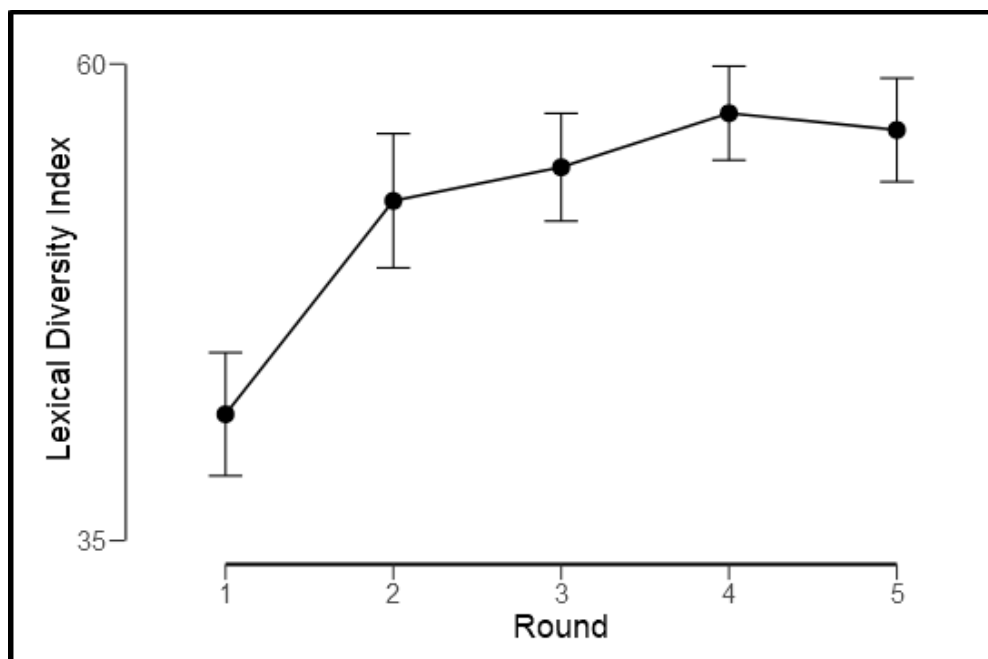


Fig 4: Lexical Diversity Index over 5 Rounds with 80% Confidence Intervals

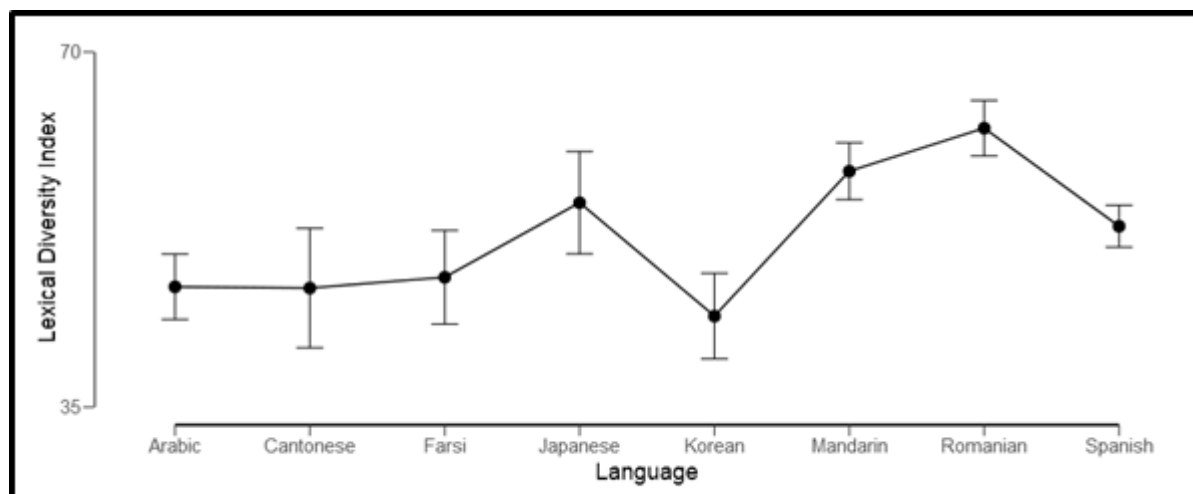


Fig 5: Lexical Diversity Index by Language with 80% Confidence Intervals