From Context to Code-switching: Examining the Interplay of Language Proficiency and Multilingualism in Speech

Debasmita Bhattacharya[1], *Aanya Tolat*[2], *Julia Hirschberg*[1]

[1]Department of Computer Science, Columbia University, USA
[2]The Peddie School, USA

debasmita.b@cs.columbia.edu, aanyatolat@gmail.com, julia@cs.columbia.edu

Abstract

Multilingual speakers are known to code-switch across language pairs and in association with various paralinguistic aspects of conversation. We build on prior work by asking: how is a multilingual speaker's language proficiency, as shaped by their linguistic background, related to their code-switching (CSW) behavior? To answer this question, we examine the Bangor Miami corpus of spontaneous Spanish-English speech and analyze the linguistic and demographic profiles of its speakers alongside features of their conversational language production. We find that a speaker's parents' primary language, medium of secondary schooling, and self-reported ability are strongly associated with the quantity of CSW and language distribution in code-switched speech. Our work is the first to empirically show a relationship between language proficiency and multilingual speech production and calls for the inclusion of linguistic background features in future paralinguistic studies of CSW.
Index Terms: code-switching, language proficiency, speech analysis, computational paralinguistics

1. Introduction

Code-switching (CSW) occurs when speakers alternate between language varieties in speech [1]. CSW may occur within or between utterances, and has been observed across numerous language pairs. Speakers who code-switch may produce a) relatively simple *insertional* code-switches of single words or short phrases, or b) more syntactically complex *alternational* code-switches at grammatical clause boundaries [2]:

(1) a. "*Pero mi* printer *no funciona.*"
["But my printer doesn't work."]
b. "*No la puedes hacer* because you can't check it."
["You can't do it because you can't check it."]

There is growing interest in studying CSW, particularly in spontaneous conversational contexts between adult speakers, and many paralinguistic factors of speech have been shown to influence or correlate with CSW, e.g. [3, 4, 5]. However, the extent to which speaker language proficiency plays a modulating role in such multilingual speech production, and its interaction with associated paralinguistic or social outcomes, remains unclear.

Language proficiency refers to a speaker's competence and capability to use oral and/or written language accurately and appropriately in a variety of settings [6, 7]. In this work, we investigate a Spanish-English corpus and examine the link between CSW and language proficiency, as encoded by several linguistic and demographic factors of a speaker's profile. We examine language proficiency factors that span individual linguistic characteristics, e.g. years of language experience, and broader language exposure aspects, e.g. the languages primarily spoken at home. Both types of factors contribute importantly to a speaker's linguistic abilities, with the latter type having a less direct, but nonetheless meaningful, influence on the development of language proficiency. We find evidence of the language spoken by one's parents and in secondary school, along with one's self-reported higher-ability language, having the strongest positive associations with the quantity and frequency of CSW in speech, and with dominant language use within code-switched utterances. Overall, we show a significant relationship between language proficiency and multiple aspects of downstream CSW behavior. Our main contribution is the empirical investigation of a fundamental aspect of multilingual speech production that has been neglected in prior work on CSW. The impact of our contribution is enhanced by our focus on one of the most popular CSW data sets used by researchers. We recommend that future work, particularly on the paralinguistics of CSW, account for speakers' language proficiency whenever possible to add nuance to interpretation of spoken CSW behavior, and to avoid the impact of a potential confounding variable in CSW analysis.

2. Related work

Much related work on CSW has explored paralinguistic factors influencing code-switched language production. Such studies have examined demographic factors such as speaker gender [8, 9], affective aspects like the expression of emotion and empathy [3, 4], and interactive conversational features such as formality, politeness, and entrainment [10, 5, 11]. Despite the breadth of research on how CSW occurs, relatively little is known about how language ability influences CSW and its associated paralinguistic effects; few, if any, of the works cited above explicitly accounted for speaker language proficiency.

Prior work on language proficiency in CSW contexts has studied the speech of children, directly via observation or indirectly through parent assessments, in language learning and development settings. [12] found that linguistic competence influences CSW direction and function in German-Turkish, German-Italian, French-Turkish, and French-Italian child speech, while [13] showed that language proficiency is only associated with Spanish-English CSW frequency among children with developmental language disorder. Similar work has been done on groups of university students learning foreign languages. In Arabic-English, language proficiency is unrelated to elicited CSW among a small group of women [14], and there is no significant effect of English language proficiency among EFL speakers in an Indonesian-English reading task [15]. Separate from language acquisition, some work, such as [16], has studied task-oriented code-switched adult speech and shown that while language proficiency is not associated with CSW frequency, the two linguistic factors have a combined, interactive effect on

cognitive control. However, the age, gender, non-native, and/or task-restricted nature of these niche speaker populations studied limits the generalizability of these findings.

Overall, findings from prior work on the relationship between CSW and language proficiency are mixed, making it difficult to draw definitive conclusions to inform current CSW research, which tends to focus on spontaneous, open-domain conversations between adults across ages. Though some studies, such as [17], have investigated how language ability is influenced by exposure to CSW, there is little work asking what the impact of language proficiency is on CSW in the kinds of contexts that are of most interest and broadest applicability to speech researchers and downstream technology. In order to address this lack, we pose the following research question: **RQ:** How is language proficiency, as encoded by demographic and linguistic factors, related to the quantity, language distribution, and syntactic complexity of CSW in a conversational domain?

3. Corpus

We examine the Bangor Miami (BM) corpus of spontaneous Spanish-English speech [18]. Of BM's 56 conversations, we use the subset of 39 dyadic conversations spanning 20 hours between speakers aged 11 to 78 (mean: 33; median: 39). BM is popular among CSW researchers due to its mix of monolingual and multilingual utterances, word-level language identification (LID) tags, and open accessibility. Note that the corpus contains 1.88 times as many English tokens as Spanish ones. In addition to recorded and transcribed code-switched dialogue, the corpus includes response data from a questionnaire given to each speaker to solicit demographic and linguistic background information on: mother's primary language (86% Spanish, 14% English), father's primary language (83% Spanish, 17% English), medium of primary schooling (38% Spanish, 62% English), medium of secondary schooling (29% Spanish, 71% English), years of language experience (Spanish:English mean = median: 1.05), age, and self-reported ability in each language (min.: 1, max.: 10; Spanish:English mean: 0.98, median: 1.0).

4. Method

Data preprocessing. We analyze the code-switched conversation transcripts alongside the questionnaire response data. For the *conversational* features, we measure CSW quantity using previous annotations of utterance-level CSW ratio [19]. We use the provided LID tags to also compute utterance-level M- and I-indices [20, 21] to further quantify the richness and frequency of CSW in each utterance. Both indices range from 0 (monolingual) to 1 (evenly mixed between languages) and measure the extent of multilingualism within code-switched utterances. We also use LID tags to determine the language distribution within code-switched utterances; we compare counts of content tokens spoken in each language within a speaker's set of code-switched utterances to determine the speaker-level dominant CSW language (41% Spanish, 59% English). Tokens tagged as language-ambiguous are excluded at this step. We measure syntactic complexity of CSW using previous annotations of CSW strategy, labeled as insertional or alternational [11], where alternational switches are considered more complex than insertional ones. We use the *questionnaire* features as listed in the previous section. Categorical features are coded as 1 representing Spanish and 2 representing English, with 0 representing a baseline where both languages are used equally. We remove speakers with incomplete questionnaire responses, and match the remaining 73 speakers' questionnaire data across 39 conversations to conversation transcripts via speaker ID.

Statistical testing and modeling. We carry out a number of statistical tests to determine the relationship between factors of language proficiency and CSW behavior. We use chi-squared and ANOVA tests, Pearson correlation, and logistic regression to examine how language proficiency and CSW are associated across combinations of categorical and continuous feature sets of both. Statistical testing is complemented by analysis of predictive models of CSW behavior. We implement a logistic regression and SVM with `scikit-learn`, using three-fold cross-validation over an 80/20 train/test split of the combined continuous *and* categorical language proficiency data.

Table 1: *F-statistics from one-way ANOVA tests for quantity of CSW. All values are statistically significant with* $p < 0.001$.

Factor	CSW ratio	M-index	I-index
Mother's language	304.17	35.70	281.96
Father's language	205.05	20.37	191.51
Primary school language	162.05	45.37	156.38
Secondary school language	201.17	65.24	194.54

5. Results

5.1. Language proficiency is significantly related to CSW quantity

We begin by examining the relationship between *categorical* features of language proficiency and the quantity of CSW in speakers' utterances. One-way ANOVA tests (Table 1) show a strong and statistically significant association between all the categorical variables and CSW quantity, as encoded by CSW ratio and I-index in particular. This suggests that there are significant differences in both the raw quantity of code-switches in an individual's speech and the frequency with which a speaker alternates between languages in that speech between speaker groups with differing backgrounds of language exposure, which go on to shape language proficiency. We specifically find that greater CSW quantity is associated with exposure to Spanish rather than English. Note that for M-index, the medium of primary and secondary schooling is more salient, whereas for CSW ratio and I-index, the language used by parents is more relevant, as reflected by relative F-statistic values across factors.

Factorial ANOVA tests combining the categorical variables in relation to CSW quantity reveal similar results. We find additional significant two-way interactions between mother and father's language ($F = 2.967$, $p = 0.028$) and primary and secondary school medium ($F = 3.369$, $p = 0.016$) when CSW ratio is the dependent variable, and significant two-, three-, and four-way interactions across all combinations of the linguistic environment variables when setting M-index as the dependent variable ($2.887 < F < 6.382$, $p < 0.02$). However, there is a reduction of at least an order of magnitude of the F-statistic compared to the one-way ANOVA tests, which reflects interaction effects within the set of independent variables.[1] Overall, the relationship between language proficiency, measured via the language spoken at home and in school, and the quantity of CSW performed in speech is most significant and best understood through analysis of individual factors.

Next, we examine the *continuous* variables of language proficiency in relation to CSW quantity using Pearson correlations.

[1] E.g., parents in a household always share the same primary language, the medium of schooling at both levels is always the same as the mother's primary language, and two-thirds of speakers' medium of education is the same at both levels.

We find little evidence of a linear relationship between any of the continuous questionnaire variables and any of the CSW quantity metrics in an individual's speech. A speaker's years of experience in Spanish relative to English is weakly *negatively* related to CSW ratio ($R = -0.212$, $p = 0.070$), M-index ($R = -0.045$, $p = 0.070$), and I-index ($R = -0.219$, $p = 0.062$). So, more experience in English is weakly associated with more CSW in speech, which approaches statistical significance. Speaker age is weakly, but not significantly, *positively* correlated with CSW ratio ($R = 0.161$, $p = 0.175$), M-index ($R = 0.216$, $p = 0.066$), and I-index ($R = 0.157$, $p = 0.185$). A speaker's self-reported ability in Spanish relative to English is inconclusively, but in all cases non-significantly, correlated with CSW quantity ($R = -0.102$, $p = 0.389$ and $R = 0.022$, $p = 0.854$) and frequency ($R = -0.099$, $p = 0.402$). These results across the metrics of CSW quantity are generally consistent, with no linear relationships between any of our features of interest. However, years of experience recurs across correlations with all three CSW metrics as approaching significance, suggesting that there may still be a non-linear relationship between this factor and quantity of CSW. Further testing will be required to uncover this definitively.

5.2. Language proficiency is significantly related to dominant language in CSW

We examine the *categorical* language proficiency factors in association with the dominant language used within a speaker's code-switched utterances. Chi-squared tests show that the language of schooling is significantly related to dominant language use in code-switched utterances. Both primary and secondary school media are associated with the same language being the dominant one in a speaker's code-switched productions ($\chi^2(1, N = 73) = 6.140$, $p = 0.013$ and $\chi^2(1, N = 73) = 7.785$, $p = 0.005$, respectively). The odds that a speaker is Spanish-dominant in CSW, given that their primary school medium was Spanish, are 4.97 (95% CI: 1.34-20.32) times those for a speaker with an English-medium primary school. Similarly, the odds that a speaker is Spanish-dominant in CSW, given that their secondary school medium was Spanish, are 7.32 times greater than those for an English-medium school (95% CI: 1.63-39.93). These results are especially significant given the overall skew towards English in the corpus. So, there appears to be a meaningful relationship between the medium of a speaker's education, particularly at the secondary level, and the dominant language used in their code-switched utterances.

Turning to the *continuous* language proficiency factors and their relationship with language dominance in CSW, we implement a binary logistic regression (Table 2) distinguishing Spanish-dominant CSW from English-dominant CSW. Most notably, we find that higher self-reported ability in Spanish relative to English is associated with Spanish-dominant CSW. This finding is statistically significant and aligns with expectations. Somewhat less intuitive is the result for years of experience in Spanish relative to English, where speakers with more Spanish experience are less likely to be Spanish-dominant in CSW. This only approaches statistical significance, though, and we are inclined to discount it, but we do note that the result is consistent with our previous Pearson correlation results. Finally, we find that older speakers in this corpus are slightly more likely to use Spanish as the dominant language in their code-switches, which could reflect increasing comfort in one language over time. More information is needed to assess whether this interpretation is reasonable given a speaker's context. Overall, there appears to be an association between one continuous measure of speaker language proficiency – self-reported ability – and the dominant language used in code-switched utterances.

Table 2: *Summary of logistic regression for Spanish-dominant (coded 1) versus English-dominant (coded 0) CSW.*

Factor	coef	std err	t
Intercept	1.1875	0.993	1.196
Years of experience (es:en)	-1.5283	0.944	-1.619
Reported ability (es:en)	0.7417	0.248	2.993
Age	0.0032	0.004	0.834

5.3. Language proficiency is not directly related to CSW strategy

Finally, we consider the language proficiency factors associated with the CSW strategy most used in conversation. For the *categorical* proficiency features, we find neither parent's language is significantly associated with the CSW strategy used by speakers ($\chi^2(1, N = 73) = 0.251$, $p = 0.882$ and $\chi^2(1, N = 73) = 0.107$, $p = 0.948$). Nor does the language of schooling at either level have a significant association with CSW strategy ($\chi^2(1, N = 73) = 0.333$, $p = 0.188$ and $\chi^2(1, N = 73) = 1.289$, $p = 0.524$). Perhaps syntactic complexity of CSW, encoded by strategy, is too fine-grained an aspect of CSW behavior to be directly impacted by language proficiency, especially when measured via influential but indirect aspects of the linguistic environment.

Table 3: *Summary of logistic regression for alternational (coded 1) versus insertional (coded 0) CSW.*

Factor	coef	std err	t
Intercept	2.7343	1.020	2.681
Years of experience (es:en)	-1.7861	0.970	-1.841
Reported ability (es:en)	-0.3678	0.255	-1.444
Age	0.0037	0.004	0.919

Our final logistic regression looks at the *continuous* language proficiency variables and distinguishes speakers' use of alternational and insertional CSW (Table 3). The most notable result which approaches statistical significance is that greater years of experience in Spanish relative to English are more associated with the use of insertional CSW than alternational CSW. This might simply be a reflection of the majority of English insertions into Spanish utterances in the corpus (about 58% of insertional code-switched utterances fall into this category); this does not provide much insight into a relationship between language proficiency and CSW strategy choice. We also note that greater self-reported ability in Spanish relative to English is more associated with the use of insertional than alternational CSW. This is somewhat counter-intuitive, as alternational CSW is more syntactically complex, and we expected its use to be aided by greater proficiency in relatively grammatically complex Spanish, and thus preferred by such proficient speakers. But, this result is not statistically significant, and so we may discount it. Finally, we find that older speakers are slightly more likely to use alternational code-switches in this corpus. Overall, it is difficult to conclude that any of the language proficiency factors are directly related to CSW behavior as encoded by the CSW strategy used in conversation. However, note that alternational CSW tends to have higher M-index values than insertional CSW (median: 0.83 v.s. 0.6), while insertional CSW tends to have higher CSW ratio and I-index values

(median: 0.2 v.s. 0.1), indicating a potential indirect relationship between CSW strategy and language proficiency via CSW quantity, which is significantly associated with proficiency as encoded by language exposure factors. This aligns with early theoretical work on CSW [22] that suggested speaker ability influences the level of complexity of code-switched productions.

5.4. Language proficiency can predict CSW behavior

We complete our investigation by assessing whether predictive models can leverage the relationships we found between language proficiency and aspects of CSW behavior to accurately predict CSW from a speaker's linguistic profile. To make CSW quantity prediction a binary task, we define a speaker with a "high" value for a given metric as one whose CSW quantity surpasses the midpoint of values for that metric across all speakers.

Both models perform similarly and reasonably well on all CSW prediction tasks, bettering a random baseline, 0.5, in each case (Table 4). Stronger results for CSW quantity and dominant language prediction reflect the relative statistical significance of associations we found above. These results overall provide further evidence for the role of language proficiency in CSW.

Table 4: *Evaluating model performance on predicting CSW behavior from language proficiency factors.*

	Mean Accuracy (SD)	
Prediction task	**Logistic regression**	**SVM**
High CSW ratio	0.85 (0.04)	0.88 (0.00)
High M-index	0.71 (0.03)	0.73 (0.02)
High I-index	0.90 (0.02)	0.90 (0.02)
Dominant CSW language	0.75 (0.12)	0.69 (0.02)
Dominant CSW strategy	0.63 (0.06)	0.62 (0.01)

5.5. Discussion

We show that language proficiency, measured via speakers' linguistic environment and individual linguistic characteristics, has notable relationships with several aspects of CSW. These relationships can even be learned and applied by simple models of CSW behavior, lending validity to our statistical findings.

With respect to the **quantity of CSW** in Spanish-English speech, the language spoken by one's parents is strongly associated with CSW behavior and can be used to differentiate speakers who code-switch in abundance and often from speakers who do neither. This finding echoes known impacts of language exposure in the home on language proficiency [23, 24, 25, 26], and provides new empirical evidence for the downstream impact of this effect on multilingual, spontaneous speech production.

When considering the **dominant language used in CSW**, we show that the medium of speakers' education, especially at the secondary level, and speakers' self-reported higher-ability language, are strongly related to language choice in spontaneous CSW. The first of these factors further highlights the known influence of language exposure on proficiency [27, 28], while the second acts as a more direct indicator of linguistic ability. Both are significantly associated with specific downstream behaviors in code-switched speech. The salience of self-reported linguistic ability is particularly interesting, as this may be considered a subjective measure. But, the intuitive results that come about from its inclusion in our experiments follow from and generally agree with related work [12, 14], and point to the reliability of using it as a proxy for language proficiency.

With regard to **CSW strategies** used in conversation, we note the inconclusive evidence for a direct relationship between these and language proficiency as encoded by any feature of a speaker's linguistic profile. We speculate that language proficiency may still be tied to CSW strategy choice in the BM corpus, in which the majority of code-switched utterances are insertional [11], but in a more complex way than we anticipated. We plan to conduct more sophisticated tests to uncover the subtle relationship between CSW quantity, strategy, and language proficiency in an admittedly imbalanced empirical setting.

Finally, based on our weak, counterintuitive, and insignificant results involving years of language experience and age, we suggest that these individual linguistic characteristics are not necessarily useful as stand-alone measures of language proficiency. For instance, years of experience and age as bare statistics do not account for periods of inactivity in a speaker's language use, which would undoubtedly have an impact on multilingual speech production. Age in particular could be more informative as a measure of language proficiency when considered in interaction with aspects of language exposure.

6. Conclusion

We examine the relationship between language proficiency and spontaneous CSW in the BM corpus. We find that (1) the language spoken by a speaker's parents is strongly associated with the quantity and frequency of CSW in that speaker's productions; (2) the medium of instruction during a speaker's education and a speaker's self-reported higher-ability language are strongly related to the dominant language used in their CSW; and (3) there is a potential indirect relationship between language proficiency and the strategy of CSW used in conversation. We expand the scope of study of language proficiency in code-switched contexts to open-domain speech between adult bilinguals across age and gender, and our findings across new aspects of CSW are more generalizable than prior work. We conclude that certain aspects of language proficiency may well have a downstream impact on code-switched language production, and should be accounted for in any study of CSW.

We recommend that future work on modeling CSW, or assessing other paralinguistic factors' effect on CSW, incorporate language proficiency features in their analysis. For example, studies of CSW quantity should ideally include demographic information about speakers' language exposure in childhood. We acknowledge that data on certain language proficiency factors might be easier to collect than others, and suggest that, wherever possible, the inclusion of some metrics of language proficiency in CSW research is preferable to none. Our work is the first to explicitly and empirically examine language proficiency in a relatively large-scale code-switched context. We hope that our investigation of a popularly used corpus will inform future studies of CSW in speech, leading to richer understanding of the interplay between various aspects of multilingual communication and their influence on how and why people code-switch.

7. Limitations

Our work focuses on a single language pair in a single corpus, which is somewhat skewed towards English and insertional CSW. We acknowledge the need to extend our method to additional language pairs, cultural contexts, and corpora with different kinds of CSW to test the robustness of our findings. Due to lack of access to CSW datasets containing speaker metadata, our work makes use of the best currently available resources and serves as a reasonable first step towards understanding the role of language proficiency on code-switched speech production.

8. Acknowledgments

We thank Lin Ai and Run Chen for helpful discussions and feedback. This work was supported in part by the National Science Foundation under Grant IIS 2418307.

9. References

[1] S. Poplack, "Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching," *Linguistics*, vol. 18, pp. 581–618, 01 1980.

[2] P. Muysken, *Bilingual speech: a typology of code-mixing*. Cambridge University Press, 2000.

[3] A. V. A. Ferreira, "Code-switching and emotions display in Spanish/Galician bilingual conversation," *Text & Talk*, vol. 37, pp. 47 – 69, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:199060969

[4] D. Bhattacharya, E. Lin, R. Chen, and J. Hirschberg, "Switching Tongues, Sharing Hearts: Identifying the Relationship between Empathy and Code-switching in Speech," in *Interspeech 2024*, 2024, pp. 492–496.

[5] E. Lavric, "Code choice and face," *Lodz Papers in Pragmatics*, vol. 3, no. -1, pp. 23–35, 2007. [Online]. Available: https://doi.org/10.2478/v10016-007-0003-x

[6] E. Brière, "Are we really measuring proficiency with our foreign language tests?*," *Foreign Language Annals*, vol. 4, pp. 385 – 391, 12 2008.

[7] N. Cloud, F. Genesee, and E. V. Hamayan, "Dual language instruction: A handbook for enriched education." Heinle Heinle, 2000.

[8] P. Gardner-Chloros and M. Edwards, "Assumptions behind grammatical approaches to code-switching: When the blueprint is a red herring," *Transactions of the Philological Society*, vol. 102, no. 1, pp. 103–129, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0079-1636.2004.00131.x

[9] K. A. Finnis, "Variation within a Greek-Cypriot community of practice in London: Code-switching, gender, and identity," *Language in Society*, vol. 43, no. 3, pp. 287–310, 2014. [Online]. Available: http://www.jstor.org/stable/43903872

[10] D. Bhattacharya, J. Chi, J. Hirschberg, and P. Bell, "Capturing Formality in Speech Across Domains and Languages," in *Proc. INTERSPEECH 2023*, 2023, pp. 1030–1034.

[11] D. Bhattacharya, S. Ding, A. Nguyen, and J. Hirschberg, "Measuring Entrainment in Spontaneous Code-switched Speech," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, K. Duh, H. Gomez, and S. Bethard, Eds. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 2865–2876. [Online]. Available: https://aclanthology.org/2024.naacl-long.158

[12] L. T. Schächinger Tenés, J. C. Weiner-Bühler, L. Volpin, A. Grob, K. Skoruppa, and R. K. Segerer, "Language proficiency predictors of code-switching behavior in dual-language-learning children," *Bilingualism*, vol. 26, no. 5, pp. 942–958, 2023.

[13] M. Kapantzoglou, J. E. Brown, L. M. Cycyk, and G. Fergadiotis, "Code-Switching and Language Proficiency in Bilingual Children With and Without Developmental Language Disorder," *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 5, pp. 1605–1620, 2021. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/2020_JSLHR-20-00182

[14] K. K. Aldalbahy, "The Relationship between Code-switching and the Level of Language Proficiency in the Second Language of Saudi Female Students," *Arab World English Journal (AWEJ)*, vol. 13, no. 4, pp. 309–324, 2022.

[15] D. Oktavia and W. Trisnawati, "The Code Switching and English Language Proficiency Performed by Students in Learning English as a Foreign Language at STKIP-MB," in *Proceedings of the 7th International Conference on English Language and Teaching (ICOELT 2019)*. Atlantis Press, 2020, pp. 334–338. [Online]. Available: https://doi.org/10.2991/assehr.k.200306.056

[16] S. Kheder and E. Kaan, "Cognitive control in bilinguals: Proficiency and code-switching both matter," *Cognition*, vol. 209, p. 104575, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027720303942

[17] M. Kaushanskaya and K. Crespo, "Does Exposure to Code-Switching Influence Language Performance in Bilingual Children?" *Child Development*, vol. 90, no. 3, pp. 708–718, 2019. [Online]. Available: https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/cdev.13235

[18] M. Deuchar, "Miami corpus: Preliminary documentation - bangortalk," 2011. [Online]. Available: http://bangortalk.org.uk/docs/Miami_doc.pdf

[19] V. Soto, N. Cestero, and J. Hirschberg, "The Role of Cognate words, POS Tags and Entrainment in Code-Switching," in *Interspeech 2018*, 2018, pp. 1938–1942.

[20] R. Barnett, E. Codó, E. Eppler, M. Forcadell, P. Gardner-Chloros, R. van Hout, M. Moyer, M. C. Torras, M. T. Turell, M. Sebba, M. Starren, and S. Wensing, "The LIDES Coding Manual: A document for preparing and analyzing language interaction data version 1.1—July, 1999." *International Journal of Bilingualism*, vol. 4, no. 2, pp. 131–270, 2000. [Online]. Available: https://doi.org/10.1177/13670069000040020107

[21] G. A. Guzman, J. Serigos, B. E. Bullock, and A. J. Toribio, "Simple tools for exploring variation in code-switching for linguists," in *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, M. Diab, P. Fung, M. Ghoneim, J. Hirschberg, and T. Solorio, Eds. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 12–20. [Online]. Available: https://aclanthology.org/W16-5802

[22] S. Poplack, "Syntactic Structure and Social Function of Code Switching," *Latino Language and Communicative Behavior*, vol. 6, pp. 169–184, 1978.

[23] E. Hoff and C. Core, "Input and language development in bilingually developing children," *Seminars in speech and language*, vol. 34, no. 4, pp. 215–26, 2013.

[24] L. M. B. I. K. Neelima Wagley, Rebecca A. Marks, "Contributions of bilingual home environment and language proficiency on children's Spanish-English reading outcomes," *Child development*, vol. 93, no. 4, pp. 881–899, 2022.

[25] R. R. Romeo, J. A. Leonard, S. T. Robinson, M. R. West, A. P. Mackey, M. L. Rowe, and J. D. E. Gabrieli, "Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function," *Psychological Science*, vol. 29, no. 5, pp. 700–710, 2018, pMID: 29442613. [Online]. Available: https://doi.org/10.1177/0956797617742725

[26] E. Brière, "Variables Affecting Native Mexican Children's Learning Spanish as a Second Language," *Language Learning*, vol. 28, pp. 159 – 174, 10 2006.

[27] H. Umiera and M. Yunus, "English as a Second Language (ESL) Learning: Setting the Right Environment for Second Language Acquisition," *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah*, vol. 3, pp. 207–215, 12 2018.

[28] L. Harklau, "The role of writing in classroom second language acquisition," *Journal of Second Language Writing*, vol. 11, no. 4, pp. 329–350, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1060374302000917