

# Exploring Dependency Distance Minimization in the Development of Arabic EFL Learners' Interlanguage: Insights from a Corpus-Based Study



CHEN Yixi FANG Yu

1 Department of Linguistics, Zhejiang University, Hangzhou, China, E-mail: canjikei1030@outlook.com.

2 School of Foreign Languages, Tongji University, Shanghai, China, E-mail: fangyu\_sfl@tongji.edu.cn.

## 1 Introduction

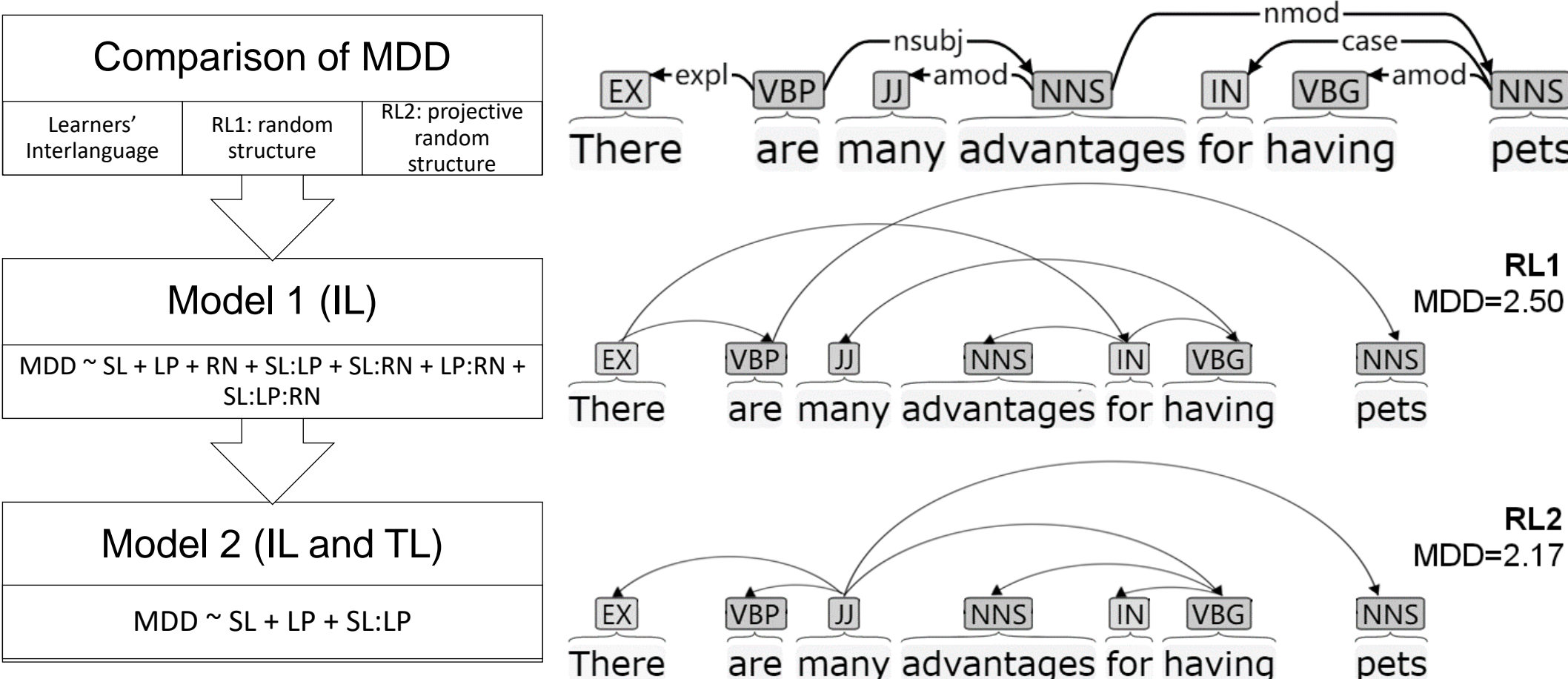
Quantifying the development of syntactic complexity in learners' interlanguage has long been an interest in language development and assessment. Under the framework of dependency grammar, mean dependency distance (MDD) has been broadly utilized as a metric of syntactic complexity in both native languages and learners' interlanguage. Whereas, dependency distance minimization (DDM), which reflects the universal tendency to reduce linguistic difficulty in human languages, has received limited attention in the context of learners' interlanguage.

The current study thus aims to investigate **whether the DDM can be observed in the development of learners' interlanguage and how it is manifested throughout the development of learners' L2 proficiency**. Understanding these issues helps shed light on learners' cognitive mechanisms to moderate the increased difficulty in language production and enrich the discussion on language development in syntax.

## 2 Data and Methods

The current study applies **corpus-based and quantitative methods**. Materials used in the current study involve

- written texts of Arabic EFL learners from the University of Pittsburgh English Language Institute Corpus (**N of sentences = 45497**) and
- texts from the Wall Street Journal portion of Penn Treebank as the sample of native English (**N of sentences = 4204**).



## 3 Results and Discussion

The Mann-Whitney U test was conducted in the comparison of MDD between IL and the two RLs. When not considering the variation of sentence length (SL), Comparisons of MDD in IL and RL1 ( $p < .0001$ , Cliff's delta = -0.6988) and MDD in IL and RL2 ( $p < .0001$ , Cliff's delta = -0.6969) both manifest significant differences.

When considering the variation of sentence length, comparison results are only significant when SL is greater than five. For short sentences between three and five words, the processing difficulty, reflected by MDD, in IL is not significantly minimized.

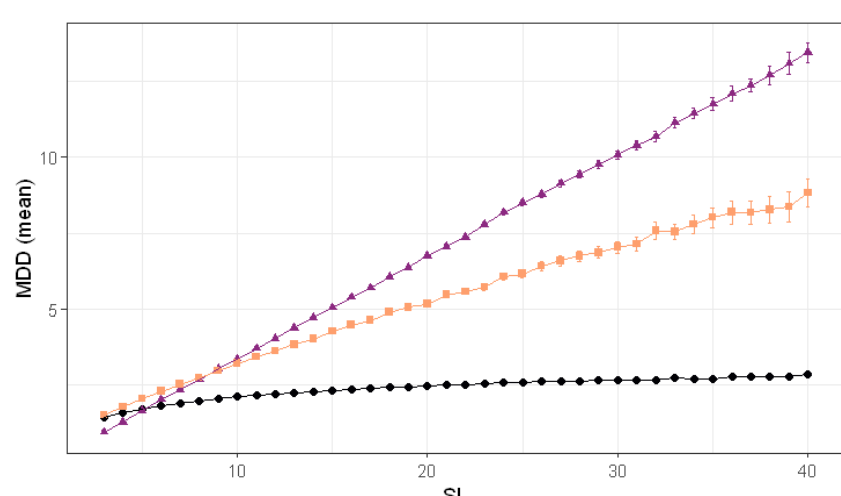


Figure 1 Comparison of MDD among IL, RL1 and RL2

SL (range)	IL, RL1		IL, RL2	
	p (M-W test)	Cliff's delta (magnitude)	p (M-W test)	Cliff's delta (magnitude)
[3, 5]	< .0001	0.1942 (small)	< .0001	-0.2771 (small)
(5, 10]	< .0001	-0.5845 (large)	< .0001	-0.6335 (large)
(10, 20]	< .0001	-0.9688 (large)	< .0001	-0.8931 (large)
(20, 30]	< .0001	-0.9999 (large)	< .0001	-0.9904 (large)
(30, 40]	< .0001	-1.0000 (large)	< .0001	-0.9979 (large)

Table 1 Results of M-W test for MDD among IL, RL1 and RL2 by sentence length (SL)

**Both SL and LP are positively correlated to MDD**, namely greater sentence length and higher language proficiency account for more processing difficulty in learners' interlanguage. RL also has a main effect on MDD, where IL can significantly minimize MDD compared to the two random languages.

**The interactions between SL and RN (a) and between SL and LP when RN = IL (b) are found significant**, whose coefficients are both negative. The current results imply that despite learners with higher proficiency may encounter more processing difficulty when producing long sentences, their language knowledge is able to moderate the augmented effort and restrain it under an acceptable threshold.

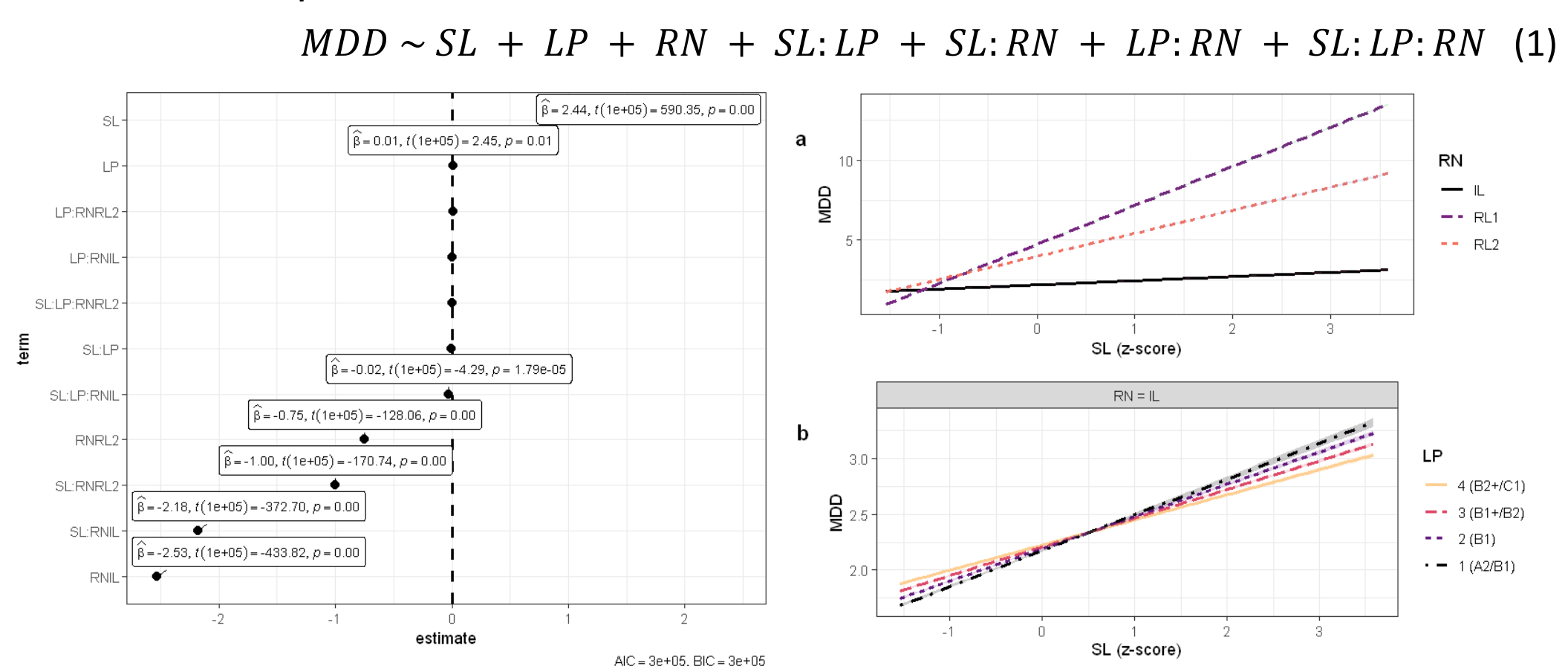


Figure 2 Multiple regression results of Model 1

Figure 3 Interact plot of SL:RN (a) and SL:LP:RN (b) in Model 1

**The interaction between SL and LP is again found significant in Model 2.** Through the development of LP, the slope of SL increases while its intercept decreases. It suggests that the development of learners' L2 proficiency can moderate the positive effect of sentence length on MDD, which may facilitate their production of long sentences.

**Both the slope and intercept in learners' interlanguage approximate the corresponding parameters in the target language.** The development of learners' L2 proficiency enables their language use to share more similarities with the target language while remaining distinct from native speakers' language use.

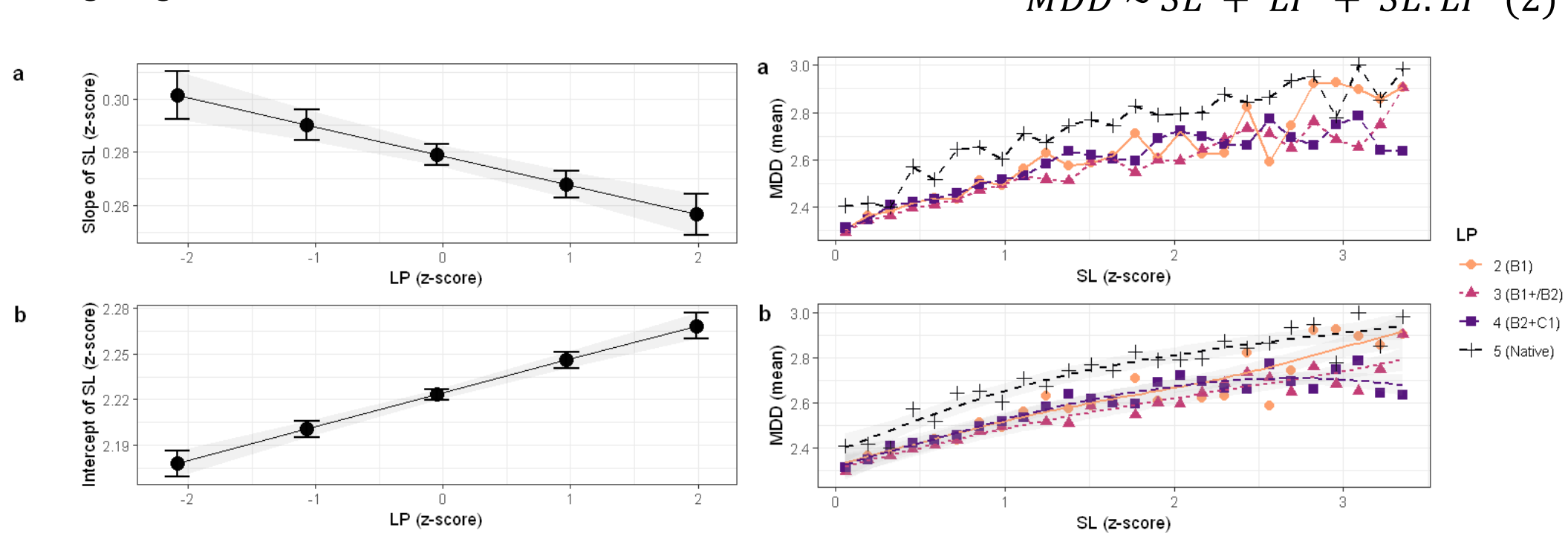


Figure 4 Slopes and intercepts of SL under different levels of LP in Model 2

Figure 5 MDDs when SL (z-score) is greater than or equal to 1.5 in different levels of LP

## 4 Conclusion

Our findings validate that MDD is significantly minimized in learners' interlanguage except in short sentences under five words. Moreover, multiple regression analysis validates the interaction effect between L2 proficiency and sentence length in learners' interlanguage, where the increase of MDD by sentence length is moderated in terms of slope as L2 proficiency develops. Last but not least, learners' interlanguage demonstrates a tendency to approach but remains distinct from the target language in terms of the variation of MDD by sentence length, which echoes the findings in previous studies.