The Contributions of Initials and Finals to L2 Chinese Comprehensibility Based on Functional Load Principle

Rian Bao
School of Information Science
Beijing Language and Culture University
Beijing, China
boroooo@163.com

Linkai Peng
Youdao
NetEase
Beijing, China
penglinkai96@gmail.com

Jinsong Zhang
School of Information Science
Beijing Language and Culture University
Beijing, China
jinsong.zhang@blcu.edu.cn

***Abstract*—The current study set out to investigate the relative contributions of Initials and Finals to Chinese L2 comprehensibility based on Brown's functional load (FL) principle. 75 speech samples elicited from 20 Urdu-speaking learners of Chinese were subjectively rated by native speakers of Chinese for comprehensibility scores, and then the segmental errors were analyzed based on FL principle. The experimental results showed that the ratio of segment errors with high FL has stronger correlation with comprehensibility than those with low FL, and the ratio of Final errors showed stronger correlation with comprehensibility than that of Initial errors, suggesting that segmental errors with high FL inhibit comprehension more than low FL errors, and Finals are more important for successful comprehension than Initials. This study offers: (1) the adaptation of Brown's FL principle on Chinese, (2) an empirical evidence that Final is a more important constituent than Initial in speech comprehension, (3) re-examination of the stronger impact of high FL errors than low FL errors on comprehensibility.**

***Index Terms*—functional load, comprehensibility, second language learning**

I. Introduction

Comprehensibility is an important concept for second language (L2) learning, which is defined as listeners' perception of how easily and smoothly they understand L2 speech [1]. Since comprehensibility is a listener-based judgement and likely to simulate real-world interactions, it is a realistic goal for successful communication and has been consistently emphasized in prior research [2], [3]. L2 comprehensibility is most often measured based on listeners' intuitive judgements on a 9-point scale (1 = difficult to understand, 9 = easy to understand) [1], [4], [5].

To date, much work has shown that comprehensibility judgement of L2 speech is mostly linked to phonological and fluency aspects. For example, segmental errors [1], word stress [6], speech rate [7], [8], and pause errors [9], [10] were proved to be associated with comprehensibility judgements. Among the segmental features, consonant and vowel errors were proven to impede comprehension differently in different language. Consonant errors affect comprehension more severely than vowel errors in English [11], and consonant errors showed significant correlation with English L2 comprehensibility, while vowel errors didn't show any statistical significance [12]. The result is opposite in L2 Spanish that vowel errors inhibit comprehension more [13], It is noteworthy that no studies have examined the different contribution of consonant and vowel on Chinese L2 comprehensibility judgement. Chinese syllable structure is conventionally divided as Initial and Final. Initial refers to the beginning consonant, and Final refers to the ending vowel(s) or vowel(s) plus ending consonant. Thus, we focus on Initial and Final in this study, and the first goal of the current study is to examine whether Chinese Initials and Finals affect comprehension differently.

Segmental errors can be further divided as errors with high FL (an information-theoretic measure that computes contribution of phonological contrast to successful word identification, hereafter referred to as FL) errors and low FL based on Brown's FL ranking [14]: defining rank 10 to 6 as high FL, 5 to 1 as low FL, and high FL segmental errors were proven to impair comprehensibility more than low FL errors in English [12], [15]. The varying contributions of phonemic contrasts in Mandarin Chinese has been investigated using FL model based on mutual information [16], [17], and the results showed that Initials at the same articulation place with different manners, Finals sharing same onset vowel and different main vowels tend to have higher FLs. Since no studies have examined the varying importance of Chinese segmentals on listeners' comprehensibility judgement. The second goal of this study is to investigate whether the high and low FL segmental errors impede comprehensibility differently. Since Brown's FL principle is aimed at language learning [14], we attempt to apply this principle to Chinese in this study.

The present study aimed at: First, adapting Brown's FL principle to Chinese; Second, discovering whether Initial and Final errors affect native listeners' comprehensibility judge-

This study was supported by advanced Innovation Center for Language Resource and Intelligence (KYR17005), and Wutong Innovation Platform of Beijing Language and Culture University (19PT04), and Fundamental Research Funds for the Central Universities, and the Research Funds of Beijing Language and Culture University (21YCX177). Jinsong Zhang is the corresponding author.

ment differently; Third, comparing the relative contribution of high and low FL segment errors to comprehensibility judgements.

II. Method

A. Calculation of FL

We employed Brown's FL principle [14] for Chinese, and *the cumulative frequency*, *probability of occurrence*, *occurrence and stigmatization in native accents*, *phonological similarity*, *structural distribution of phonemes*, *lexical sets*, *number of minimal pairs*, *number of minimal pairs belonging to the same parts of speech*, *frequency of members of minimal pairs*, and *number of common contexts in which members of minimal pairs occur* were adopted from the original model, and the FL of Initials and Finals were calculated from a pinyin-transcribed corpus from Chinese TV show contains 300,000 sentences.

a) Cumulative frequency: The cumulative frequency of a segmental pair A and B is calculated by adding the individual frequencies for A and B in the corpus, and it is positively correlated with FL.

b) Probability of occurrence minus 0.5: To consider the fact that a high cumulative frequency might caused by one member of the pair is much more frequent than the other, the probability of occurrence of a member is calculated by dividing its individual frequency by the cumulative frequency of the pair, then subtract 0.5 from the Probability of Occurrence. The lower the value is, the greater the potential confusion is. This measure is negatively correlated with FL.

c) Occurrence and stigmatization in native accents: If a phonemic pair is conflated in native accents, it means this pair does not play a crucial role in this language. For example flat tongue Initials <z>, <c>, <s> and retroflex Initials <zh>, <ch>, <sh>, Finals with front nasal <n> and Final nasal <ng> are conflated in Southern China [18]. Phonemic pairs that contains these conflated segments are labeled as 1, and the others are labeled as 0.

d) Phonological similarity: Phonological similarity is calculated through Hamming distance of phonemic pairs [19]–[21]. First, we measure the distance between phonemes by getting the number of different phonological features, and the number is divided by the total number of phonological features, as in equation (1). Since most of the Finals are sequence of phonemes. The basic principle of similarity measurement between Finals is that medial is compared with medial, nucleus vowel is compared with nucleus vowel, and coda is compared with the coda. The similarity value of the corresponding factor is obtained first, and then the average value is taken as the similarity value of the Final. If the lengths of the two sequences are different, the nucleus vowel is filled at the positions of medial and coda [22]. The phonological features used for Hamming distance calculation can be seen in Table I and Table II. The distance between nasal codas <n> and <ng> is assigned as 2.

$$Hamming\ Dist. = \frac{\text{different features between phonemes}}{\text{total number of phonological features}} \quad (1)$$

TABLE I
Phonological Features of Vowels for Calculating Hamming Distance

	API	DOR	ROU	FRO	LOWEST	HIGHEST	HIGH	LOW	DEN	RET
a	-	+	-	+	+	-	-	+	-	-
o	-	+	+	-	-	-	+	-	-	-
e	-	+	-	+	-	-	+	-	-	-
i	-	+	-	+	-	+	+	-	-	-
u	-	+	+	-	-	+	+	-	-	-
v	-	+	+	+	-	+	+	-	-	-
i(D)	+	-	-	-	-	-	-	-	+	-
i(R)	+	-	-	-	-	-	-	-	-	+

e) Structural distribution of Initials and Finals: In Chinese syllable structure, the vowel /u/ occurs with retroflex Initials <zh, ch, sh> and cannot occur with alveolo-palatal Initials <j, q, x>. Thus, learners who conflate retroflex and alveolo-palatal will not cause misunderstanding in the context with /u/. The structural distribution is equal to the number of common pinyin contexts of a pair of Initials or a pair of Finals.

TABLE II
Phonological Features of Consonants for Calculating Hamming Distance

	SON	LAB	COR	DOR	NAS	LAT	VOIC	STRI	CONT	SPR_GL
b	-	+	-	-	-	-	-	-	-	-
p	-	+	-	-	-	-	-	-	-	+
m	+	+	-	-	+	-	+	-	-	0
f	-	+	-	-	-	-	-	+	+	0
d	-	-	+	-	-	-	-	-	-	-
t	-	-	+	-	-	-	-	-	-	+
n	+	-	+	-	+	-	+	-	-	0
l	+	-	+	-	-	+	+	-	+	0
g	-	-	-	+	-	-	-	-	-	-
k	-	-	-	+	-	-	-	-	-	+
h	-	-	-	+	-	-	-	-	-	0
j	-	-	+	-	-	-	-	+	-	-
q	-	-	+	-	-	-	-	+	-	+
x	-	-	+	-	-	-	-	+	+	0
zh	-	-	+	-	-	-	-	+	-	-
ch	-	-	+	-	-	-	-	+	-	+
sh	-	-	+	-	-	-	-	+	+	0
r	+	-	+	-	-	-	+	-	+	0
z	-	-	+	-	-	-	-	+	-	-
c	-	-	+	-	-	-	-	+	-	+
s	-	-	+	-	-	-	-	+	+	0

f) Lexical sets: Lexical set is defined as the actual words of the Chinese lexicon that contain the phoneme in the corpus, and the number of the lexical sets of a phonemic pair is defined as the sum of the number of lexical set of each phoneme.

g) Number of minimal pairs: Minimal pair is the most intuitive expression of the functional load of a phonemic contrast that serves to distinguish meanings. The total numbers of the minimal pairs of the Initial and Final contrasts in the corpus are calculated.

h) Number of minimal pairs belonging to the same part of speech: Number of the minimal pairs that belong to the same part of speech are calculated from the corpus.

i) Difference of the frequencies of members of minimal pairs: If an infrequent phoneme pairs with a very frequent phoneme such as <d> and <l> (which are used in function

words), the minimal pair can hardly be said to have any importance. Thus, the difference of the frequencies of members of minimal pairs is calculated. The lower the value is, the higher the FL is.

j) Number of common contexts in which members of minimal pairs occur: This was defined as the number of the minimal pairs that preceded and followed by exactly same words.

The values of *Probability of Occurrence Minus 0.5*, *Occurrence and Stigmatization in Native Accents*, *Phonological Similarity*, *Difference of the Frequencies of Members of Minimal Pairs Occur* were multiplied by -1, since they are negatively correlated with FL. Then all the measures were weighted equally by calculating the standardized score (T-score), and the average of the standardized score is defined as FL score in the present study. Initials and Finals were calculated separately since there is no minimal pair between an Initial and a Final. After the calculation, we ranked the FL scores on the 10-point scale. Using this FL rank-ordering, we further divided the contrasts into high and low FL contrasts: 1 to 5 as low FL, and 6-10 as high FL.

B. Materials

The speech materials used in present study were selected from BLCU-SAIT corpus [23], which is an interlanguage speech corpus of L2 learners of Chinese. Speech samples with lower than 40% tone error ratio were selected in order to reduce the influence of tone errors, and 75 read speech of simple sentences from 20 Urdu-speaking learners (12 male and 8 female) from Pakistan were selected to analyze comprehensibility-related segmental features. All the speakers were students from Beijing Language and Culture University. Initial, Final and tone errors and their substitutions were annotated by a professional linguistic student.

C. Raters and comprehensibility judgements

We recruited six graduate students to participate in the comprehensibility rating experiment. All the raters are native speakers of Chinese, were all born in China and raised by monolingual parents, and none of them reported hearing disorder. All of them have no experience of teaching Chinese to speakers of other languages. They were all aged between 22 and 26 (M=23.5, 3 females and 3 males).

The comprehensibility rating tasks were conducted individually in a quiet room using the Praat's ExperimentMFC [24], and all the rating results can be automatically recorded in the software. Each rater listened to the audio through a set of headphones on the researcher's laptop. Before the data collection, the investigator trained all the raters. First, the raters familiarized themselves with the listening materials (9 sentences with standard comprehensibility ratings). Then, each rater practiced the rating procedure through three trails of experiments. In the formal experiment, they were asked to pay attention on the effort it takes to understand the sentences. If they can understand the sentence very easily, then this sentence is highly comprehensible, and vice versa.

In the formal rating experiment, 180 speech samples were rated (75 of them are analyzed in the current study, the other 105 are outside the scope of this study). They were divided into two groups following the principle of non-repetition of the sentence content. Each group was rated by three raters. 102 sentences were played for each rater in a randomized order. 90 of them are from our research materials, and the other 12 sentences from other L2 speakers (outside the 20 speakers in II-B) were included in both groups and used as computing inter-rater agreement. Each sentence can be played only one time. After hearing a sample, they made an intuitive judgement using a 9-point scale (1 = hard to understand, 9 = easy to understand). The whole session took around 30 minutes.

D. Comprehensibility-related Segmental Factors

To compare the contributions of Initials and Finals, high FL and low FL segments to comprehensibility judgements, 10 variables were calculated for the correlation analysis with comprehensibility, i.e., segment error ratio, FL rank of segment errors, FL rank of Initial errors, FL rank of Final errors, high FL error ratio, low FL error ratio, high FL Initial error ratio, low FL Initial error ratio, high FL Final error ratio, low Final error ratio. To eliminate the effect of different number of Initial and Final errors in a sample, we further selected the samples with same number of Initial and Final errors and calculated the correlation between these variables the comprehensibility scores.

- Segment error ratio: was defined as the total number of substitutions (Initials and Finals) divided by the total number of Initials and Finals.
- FL rank of segment errors: was defined as the sum of the FL ranks of all the segmental substitutions divided by the total number of Initials and Finals.
- FL rank of Initial or Final errors: was defined as the sum of the FL ranks of segmental errors divided by the total number of Initials and Finals.
- High or low FL error ratio: was defined as the number of high or low FL segmental substitutions divided by the sum of total number of Initials and Finals.
- High or low FL Initial error ratio: was defined as the number of high or low FL Initial substitutions divided by total number of Initials and Finals.
- High or low FL Final error ratio: was defined as the number of high or low FL Final substitutions divided by total number of Initials and Finals.

III. Result

A. FL distribution

The overall distributions of Initials and Finals FL can be seen in Fig. 1 . Y-axis is the average of normalized value of all the FL measures, and X-axis is the rank of the average value. It is clear that both Initials' and Finals' FL distributions show a steep descending trend at the beginning of the curve line, indicating that few numbers of phonemic pairs are carrying the majority of information transmission. This distribution is similar to previous research conducted on English [25]. The

Initial pair with the highest FL is <j, x> followed by <zh, sh>, <d, l> and <j, q>, and the Final pair with the highest FL is <i, ing> followed by <i, u>, <i, ian> and <i, ie>, which is in line with the previous conclusion that Initials at the same articulation place with different manners, Finals sharing same onset vowel and different main vowels tend to have higher FLs [16], [17]. Since there are too many pairs of Initials and Finals, we selected the pairs that are conflated in the speech materials used in current study and assigned a simplified FL ranks to them on a scale of 1-10 based on the raw FL calculation (see in Table III).

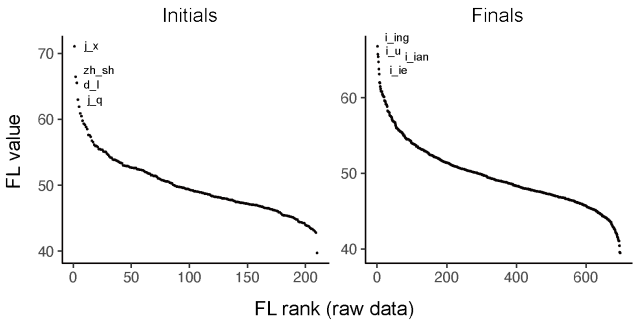

Fig. 1. FL distribution of Chinese Initials and Finals.

TABLE III
FL RANK ORDERING OF CHINESE INITIAL AND FINAL PAIRS CONFLATED BY L2 LEARNERS IN THE CURRENT STUDY

FL rank	*Initials*	*Finals*
10	sh-zh, q-j, q-x	i-ing, i-u, i-ie, u-ai, u-e, ian-ing, u-a, i-in, u-uei, i-ai, eng-ang, i-a, u-ou, i-e, ian-in, ao-e, i-uei
9	h-g, t-d, ch-zh, ch-sh	i-ia, ao-ou, an-ai, ang-ong, ao-uo, ian-an, an-a, an-en, ian-iang, uan-an, uei-uo, u-uen
8	z-d, q-b	e-en, ang-a, uo-ong, ing-iang, a-e, eng-e, u-iou, ai-iao, ao-iao, ü-iou, ang-ing, ü-u
7	k-g, h-k	ua-e, ou-uo, ai-ei, uan-uen, ie-e, eng-ing, uan-ian, ian-uen, iou-üe, ong-uang, iao-uo
6	p-b, c-s, s-z, l-r	uang-ang, ei-ie, ü-ou, uen-uo, iang-eng, u-o, en-in, a-o, iong-iang
5	c-sh, c-z, z-zh	o-ao, uo-ua, uai-iao, an-ang, ou-o, u-uo
4	g-q, c-ch	in-ing
3	zh-j	en-eng
2	x-sh	-
1	ch-q, s-x	ang-iang

B. Comprehensibility judgement

In terms of inter-rater reliability of comprehensibility judgements, Pearson's r was computed among three raters' scores from same group. The strength of correlations is relatively high, which varied from r = .658 to r = .828, and the inter-group reliability is r = .805.

C. The correlation between FL measures and comprehensibility judgments

In order to identify the relative importance of Initials and Finals, a set of correlation analysis was computed. 10 segmental measures and comprehensibility scores of 75 speech samples were submitted to Pearson's correlation coefficient (see in Table IV). According to the results, significant negative correlations with comprehensibility judgments were found in FL rank of segment errors (r = -0.539, p < 0.001), segment error ratio (r = -0.523, p < 0.001), high FL error ratio (r = -0.484, p < 0.001), FL rank of Final errors (r = -0.455, p < 0.001), high FL Final error ratio (r = -0.405, p < 0.001), high FL Initial error ratio (r = -0.321, p < 0.01), FL rank of Initial errors (r = -0.306, p < 0.01), and low FL error ratio (r = -0.275, p < 0.05). However, low FL Final error ratio low FL Initial error ratio did not show statistical significance (p > 0.05). The data with balanced Initial and Final errors showed the similar results. The significant correlation was found in FL rank of Final errors (r = -0.700, p < 0.001), high FL Final error ratio (r = -0.655, p < 0.001), high FL Initial error ratio (r = -0.627, p < 0.001), and FL rank of Initial errors (r = -0.600, p < 0.001). Low FL Final and Initial error ratio did not show statistical significance (p > 0.05).

TABLE IV
PEARSON CORRELATION COEFFICIENTS BETWEEN L2 SEGMENTAL FEATURES AND LISTENER-BASED JUDGEMENT OF L2 COMPREHENSIBILITY ($*** \leq 0.001, ** \leq 0.01, * \leq 0.05$)

Features	*Corr.*	*Sig.*
FL rank of segment errors	-0.539	***
Segment error ratio	-0.523	***
High FL error ratio	-0.484	***
FL rank of Final errors	-0.455	***
High FL Final error ratio	-0.405	***
High FL Initial error ratio	-0.321	**
FL rank of Initial errors	-0.306	**
Low FL error ratio	-0.275	*
Low FL Final error ratio	-0.145	
Low FL Initial error ratio	-0.055	
FL rank of Final errors (balanced)	-0.700	***
High FL Final error ratio (balanced)	-0.655	***
High FL Initial error ratio (balanced)	-0.627	***
FL rank of Initial errors (balanced)	-0.600	***
Low FL Final error ratio (balanced)	-0.088	
Low FL Initial error ratio (balanced)	0	

IV. DISCUSSION AND CONCLUSIONS

The current study examined how high FL and low FL, Initial and Final segmental errors differently impact Chinese L2 comprehensibility with using 75 speech samples articulated by 20 Urdu-speaking learners. First, Brown's FL principle [14] was adapted for Chinese to calculate the relative importance of Initial and Final pairs. Second, the correlation analyses between 10 segmental variables and the comprehensibility scores showed that high FL segmental errors impede understanding more than low FL segmental errors, which is consistent with previous research [12], [15]. More specifically, our results revealed that Finals play more important role than Initials in speech understanding. Lastly, the strong correlation between FL rank of segmental errors and L2 comprehensibility provided insights that FL model can probably be used as a measure for automatic assessment of L2 comprehensibility. The current model assigns a fixed FL value for a phonemic pairs, but the importance of a phonemic pairs varies in different context. Future work should calculate FL in a more accurate way considering context effect to provide a more reliable predictor for comprehensibility judgements.

REFERENCES

[1] M. J. Munro and T. M. Derwing, "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," *Language learning*, vol. 45, no. 1, pp. 73–97, 1995.

[2] J. M. Levis, "Changing contexts and shifting paradigms in pronunciation teaching," *TESOL quarterly*, vol. 39, no. 3, pp. 369–377, 2005.

[3] K. Saito, S. Webb, P. Trofimovich, and T. Isaacs, "Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations," *Studies in Second Language Acquisition*, vol. 38, no. 4, pp. 677–701, 2016.

[4] T. M. Derwing and M. J. Munro, "Accent, intelligibility, and comprehensibility: Evidence from four l1s," *Studies in second language acquisition*, vol. 19, no. 1, pp. 1–16, 1997.

[5] P. Trofimovich and T. Isaacs, "Disentangling accent from comprehensibility," *Bilingualism: Language and Cognition*, vol. 15, no. 4, pp. 905–916, 2012.

[6] J. Field, "Intelligibility and the listener: The role of lexical stress," *TESOL quarterly*, vol. 39, no. 3, pp. 399–423, 2005.

[7] J. Anderson-Hsieh and K. Koehler, "The effect of foreign accent and speaking rate on native speaker comprehension," *Language learning*, vol. 38, no. 4, pp. 561–613, 1988.

[8] M. J. Munro and T. M. Derwing, "Modeling perceptions of the accentedness and comprehensibility of l2 speech the role of speaking rate," *Studies in second language acquisition*, vol. 23, no. 4, pp. 451–468, 2001.

[9] O. Kang, D. Rubin, and L. Pickering, "Suprasegmental measures of accentedness and judgments of language learner proficiency in oral english," *The Modern Language Journal*, vol. 94, no. 4, pp. 554–566, 2010.

[10] D. Crowther, P. Trofimovich, and T. Isaacs, "Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives," *Journal of Second Language Pronunciation*, vol. 2, no. 2, pp. 160–182, 2016.

[11] J. Walz, "Studies of error gravity. native reactions to errors produced by swedish learners of english," 1980.

[12] Y. Suzukida and K. Saito, "Which segmental features matter for successful l2 comprehensibility? revisiting and generalizing the pedagogical value of the functional load principle," *Language Teaching Research*, vol. 25, no. 3, pp. 431–450, 2021.

[13] K. E. Schairer, "Native speaker reaction to non-native speech," *The Modern Language Journal*, vol. 76, no. 3, pp. 309–319, 1992.

[14] A. Brown, "Functional load and the teaching of pronunciation," *TESOL quarterly*, vol. 22, no. 4, pp. 593–606, 1988.

[15] M. J. Munro and T. M. Derwing, "The functional load principle in esl pronunciation instruction: An exploratory study," *System*, vol. 34, no. 4, pp. 520–531, 2006.

[16] B. Wu, J. Zhang, and Y. Xie, "A clustering analysis of chinese consonants based on functional load," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.

[17] Y. Chen, Y. Xie, and J. Zhang, "A comparison study of information contributions of phonemic contrasts in mandarin," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2017, pp. 1579–1582.

[18] H. Zhao and Y. Liu, "An analysis of southern dialects speakers' pronunciation errors of initials and finals in mandarin," in *Proceedings of 2018 National Acoustic Conference*, 2018.

[19] D. Gildea and D. Jurafsky, "Learning bias and phonological-rule induction," *Computational Linguistics*, vol. 22, no. 4, pp. 497–530, 1996.

[20] J. Pierrehumbert, "Dissimilarity in the arabic verbal roots," in *Proceedings of NELS*, vol. 23. University of Massachusetts Amherst, 1993, pp. 367–381.

[21] Y. Do and R. K. Y. Lai, "Accounting for lexical tones when modeling phonological distance," *Language*, vol. 97, no. 1, pp. e39–e67, 2021.

[22] H. Chen, "Study on similarity measure of vowel system between dialects," *Zhongguo Yuwen*, no. 2, p. 7, 2000.

[23] B. Wu, Y. Xie, L. Lu, C. oCao, and J. Zhang, "The construction of a chinese interlanguage corpus," in *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*. IEEE, 2016, pp. 183–187.

[24] P. Boersma and D. Weenink. (1992) Praat: doing phonetics by computer [computer program].

[25] Y. Zhang, Z. Li, and J. Zhang, "Towards the use of pretrained language model gpt-2 for testing the hypothesis of communicative efficiency in the lexicon," in *2021 International Conference on Asian Language Processing (IALP)*. IEEE, 2021, pp. 62–66.