



University of Ljubljana
FACULTY OF ARTS

Acta Linguistica Asiatica

Volume 12, Issue 1, 2022

ACTA LINGUISTICA ASIATICA

Volume 12, Issue 1, 2022

Editors: Andrej Bekeš, Nina Golob, Mateja Petrovčič

Editorial Board: Alexander Alexiev (Bulgaria), Bi Yanli (China), Cao Hongquan (China), Luka Culiberg (Slovenia), Tamara Ditrich (Slovenia), Kristina Hmeljak Sangawa (Slovenia), Hsin Shih-chang (Taiwan), Ichimiya Yufuko (Japan), Terry Andrew Joyce (Japan), Jens Karlsson (Sweden), Lee Yong (Korea), Lin Ming-chang (Taiwan), Wei-lun Lu (Czech Republic), Arun Prakash Mishra (Slovenia), Nagisa Moritoki Škof (Slovenia), Nishina Kikuko (Japan), Sawada Hiroko (Japan), Chikako Shigemori Bučar (Slovenia), Irena Srdanović (Croatia).

© University of Ljubljana, Faculty of Arts, 2022
All rights reserved.

Published by: Znanstvena založba Filozofske fakultete Univerze v Ljubljani
(Ljubljana University Press, Faculty of Arts)

Issued by: Department of Asian Studies

For the publisher: Dr. Mojca Schlamberger Brezar, Dean of the Faculty of Arts

The journal is licensed under a

Creative Commons Attribution-ShareAlike 4.0 International License.

Journal's web page:

<http://revije.ff.uni-lj.si/ala/>

The journal is published in the scope of Open Journal Systems

ISSN: 2232-3317

Abstracting and Indexing Services:

Scopus, COBISS, dLib, Directory of Open Access Journals, MLA International Bibliography, Open J-Gate, Google Scholar and ERIH PLUS.

Publication is free of charge.

Address:

University of Ljubljana, Faculty of Arts
Department of Asian Studies
Aškerčeva 2, SI-1000 Ljubljana, Slovenia

E-mail: nina.golob@ff.uni-lj.si

TABLE OF CONTENTS

Foreword	5
----------------	---

RESEARCH ARTICLES

Preference for Deletion vs. Epenthesis in Japanese Phonological Adaptations: Lexical Stratification and Input Medium

SHOJI Shinichi	9
----------------------	---

Correlations Between Proposed Orthoepic Competence Descriptors and Japanese Language Ability

ITO Hideaki	19
-------------------	----

Chinese Idioms: Stepping Into L2 Student's Shoes

Mateja PETROVČIČ	37
------------------------	----

Coroneologisms and Word Formation Processes in Hindi-English Codemixed Words

Md. Tauseef QAMAR, Juhi YASMEEN, Md. Arfeen ZEESHAN, Sanket PATHAK.....	59
---	----

A Synchronic and Historical Look at Akeanon Phonology

Philip RENTILLO, Ruchie Mark D. POTOTANON	91
---	----

RESEARCH ARTICLES IN SLOVENE / ZNANSTVENI ČLANKI V SLOVENŠČINI

Vloga prevajanja pri razumevanju literature korejske manjšine v ZDA: Analiza prevodov Leejevega romana *Native Speaker*

Byoung Yoong KANG	131
-------------------------	-----

Foreword

The new calendar year brings a new winter issue. We are pleased to announce that the issue boasts a selection of articles which mainly focus on Japanese, Chinese, and Korean. This is the emerging long-term focus of our journal though, of course, we will continue to be accessible to outstanding submissions concerning the wider East Asian linguistic field in the future as with the two articles on Akeanon phonology and Hindi-English coroneologisms in this issue.

The articles in English are wrapped up with two phonological topics and further include second-language acquisition, pragmatics, and word formation. The final article deals with translation challenges and is written in Slovene.

The issue opens with the article “Preference for Deletion vs. Epenthesis in Japanese Phonological Adaptations: Lexical Stratification and Input Medium” in which **SHOJI Shinichi** investigated phonological adaptation of non-loan words in Japanese and their preference for either deletion or epenthesis.

The following is the article entitled “Correlations Between Proposed Orthoepic Competence Descriptors and Japanese Language Ability” by **ITO Hideaki**. In it, the author proved the correlation by analyzing online tests and self-assessment questionnaires and categorized learners of the Japanese language into different skill levels.

Mateja PETROVČIČ in her article “Chinese Idioms: Stepping Into L2 Student’s Shoes” overviews the status of Chinese idioms *chengyu* in foreign language education, stresses their importance, and offers practical suggestions as to which aspects of teaching idioms should be taken under consideration.

The next article “Coroneologisms and Word Formation Processes in Hindi-English Codemixed Words” was written by **Md. Tauseef QAMAR, Md. Arfeen ZEESHAN, Juhi YASMEEN**, and **Sanket PATHAK**. It investigated coroneologisms and word-formation processes in Hindi-English code-mixed words and came to the conclusion that compounding, affixation, blending, and reduplication appear mainly as a result of compounding and borrowing.

“A Synchronic and Historical Look at Akeanon Phonology” is the article by **Philip RENTILLO** and **Ruchie Mark D. POTOTANON**, which provides a review and reevaluation of a reflex of the proto-Bisayan */ and *-d- as well Akeanon phonology in general based on synchronic distribution, dialectology, historical accounts, and acoustic analysis.

Last but not least is the Slovene article “The Role of Translation in Understanding the Literature of the Korean Minority in the United States: Analysis of the Translation of Chang-rae Lee’s Novel *Native Speaker*” written by **Byoung Yoong KANG**. In the article, the author examined the role of the language in the original and translation, further revealed some problems arising from the translation, and reinterpreted the meaning of *Koreanness* in Korean literature.

Editors and Editorial board wish the regular and new readers of the ALA journal a pleasant read full of inspiration, and a rise of new research ideas inspired by these papers.

Editors

RESEARCH ARTICLES

Preference for Deletion vs. Epenthesis in Japanese Phonological Adaptations: Lexical Stratification and Input Medium

SHOJI Shinichi

Mie University, Japan
shoji.shinichi@dhier.mie-u.ac.jp

Abstract

This study investigated phonological adaptation of non-loan words in Japanese and their preference for either deletion or epenthesis. Earlier studies argue that non-loan Japanese words prefer deletion while loanwords prefer epenthesis. Studies further show that the input medium affects the adaptation; text-input leads to epenthesis while sound-input to deletion. The present study experimented with text-input of non-loan nonce words and investigated how native Japanese speakers adapt their causative, passive, and potential forms. Results showed a strong preference for deletion in causative forms, a relatively weak preference for deletion in potential forms, and no significant preference in passive forms. The outcome indicates that deletion is not present by default, and further investigation is needed to define factors that influence the selection.

Keywords: Japanese, phonological adaptation, deletion, epenthesis, verb formation

Povzetek

Ta študija je raziskala fonološko prilagoditev neizposojenih besed v japonščini in njihovo naklonjenost izbrisu oziroma vstavljanju. Prejšnje študije kažejo, da se neizposojene japonske besede nagibajo k izbrisu, izposojene pa k vstavljanju. Študije nadalje kažejo, da vhodni medij vpliva na prilagajanje; vnos besedila vodi do epenteze, medtem ko vnos glasu do brisanja. Ta študija je eksperimentirala z neizposojenimi besedami v obliki besedila in raziskala, kako materni japonski govorci prilagajajo vzročne, pasivne in potencialne oblike besed. Rezultati so pokazali močno naklonjenost izbrisu v vzročnih oblikah, razmeroma šibko naklonjenost izbrisu v potencialnih oblikah in nejasne smernice v pasivnih oblikah. V splošnem lahko sklepamo, da izbris ni privzeto prisoten, zato so za opredelitev dejavnikov, ki vplivajo na izbiro, potrebne nadaljnje raziskave.

Ključne besede: japonščina, fonološka prilagoditev, izbris, vstavljanje, tvorba glagola



1 Introduction: phonological adaptation in Japanese

Phonological adaptations are widely discussed in a great amount of linguistic research, especially in the field of studies on loanwords. In general, certain sounds and phonological patterns are often unique to a group of languages, which may or may not exist in the phonetic and phonological systems of the other languages. Thus, when a language imports (or loans) new words from another language, and if the words include some sounds or phonological patterns that the borrower language does not have, then the pronunciations of the loanwords have to be adapted to be compatible with the borrower language's sound system. Among many adaptation methods, two of them are deletion and epenthesis; non-existent sounds or phonological patterns in the borrower language are deleted, or additional sounds are epenthesized so that the non-existent phonological patterns do not violate the borrower language's system.

In the case of Japanese loanwords from other languages, the phonological adjustment relies on epenthesis by default, rather than deletion (McCawley, 1968; Smith, 2006; Shoji & Shoji, 2014), to dissolve consonant clusters and coda in most cases, due to the language not allowing consonant clusters or coda (with a few exceptions). Specifically, the vowel [u] is epenthesized in many loanwords. This is because, in principle, the adaptations should be done in a way which “do as little violence as possible” to the source language's system (Sapir, 1921, p.210). Among Japanese vowels, [u] is the most unmarked and perceptually the least salient, as supported by the facts that [u] is ‘the most readily subject to devoicing’ and to ‘weakening and elimination’ (Lovins, 1975, p. 106; Mori, 1929, p. 58 as cited in Lovins, 1975, p. 106), which allows the loanword to be minimally different from its source word (Kubozono, 2002). An exception is, [i], which is epenthesized when the first consonant in clusters or coda is [tʃ] or [dʒ] in the source words. Also, [o] is epenthesized when the first consonant in clusters or coda is [t] or [d]. These adaptations are shown in the words such as an English word ‘little [lɪtl̩]’ turning into [rit̩ou] and ‘beach [bitʃ]’ becoming [bi:ɕi] as loanwords in Japanese.¹ Shoji and Shoji (2014) experimented with native Japanese speakers utilizing English nonce words as the experimental items. Participants were given the nonce words written in the English alphabet that included consonant clusters and coda (e.g., ‘krito’), and the task of the participants was to rewrite the given words in Japanese katakana text.² Because it is impossible to write consonant clusters or coda in Japanese orthography, the participants were forced to either delete a consonant

¹ It should be noted that this tri-partite epenthesis pattern with [u], [i], and [o] may not rigidly default adaptations. In terms of which vowel is epenthesized, studies including that of Mattingley, Hall, and Hume (2019) show that the vowel that native Japanese speakers epenthesize in resolving consonant clusters vary to a great extent, loosely consistent with traditional patterns regarding [u], [i] and [o].

² In Japanese, katakana texts are used to write loanwords (except those from China) and hiragana texts are used to write non-loan words.

(e.g., 'rito') or epenthesize a vowel (e.g., 'kurito'). The results indicated that the participants significantly tended to epenthesize vowels in consonant clusters and after coda, rather than deleting consonants. The outcome confirms that epenthesis is the preferred method for the phonological adaptation of loanwords.

While phonological adaptations are widely discussed in the studies of loanwords, similar phenomena are also found within a language in its word formations. In Japanese, early studies, including that of McCawley (1968), argue that its native vocabulary exhibits its tendency for deletion rather than epenthesis. For example, in Japanese verb formation, deletion is preferred to epenthesis by default when the verb stem ends with a consonant. For example, when a verb, *yomu* 'read', becomes its causative form, passive form, and potential form, the stem of the verb *yom* has to be suffixed by *sase* (causative), *rare* (passive), and *re* (potential). The combinations of stems and suffixes, *yom-sase*, *yom-rare*, and *yom-re*, should be adjusted to dissolve the clusters with stem-ending consonants and suffix-initial consonants. Deletion is applied for these formations, i.e., the causative form is *yom-ase* (*yom-~~s~~sase*) rather than *yom-V-sase*, the passive form is *yom-are* (*yom-~~r~~are*) rather than *yom-V-rare*, and the potential form is *yom-e* (*yom-~~r~~e*) rather than *yom-V-re*.

As mentioned above, Japanese seems to have different preferences between epenthesis and deletion depending on whether the words are non-loan or loanwords. This discrepancy is compatible with the explanation by lexical stratification, as Fukazawa, Kitahara, and Ota (1998) argue. In their analyses of native Japanese words and loanwords, these words behave differently in their phonological adaptations. A part of Fukazawa et al's analyses is related to the conflict between the markedness constraint and the faithfulness constraint, which are applied, for example, to a native Japanese word, *kan-da* 'bit (past tense of 'bite')', and a loanword, *komp*i*uutaa* 'computer'. In the native Japanese word, *ta* is the past tense morpheme, but *ta* is voiced to become *da* in *kan-da*. It is not *kan-ta* but *kan-da* because the markedness constraint (i.e., Post Nasal Voicing) has priority over the faithfulness constraint (i.e., IDENT[voice]); *ta* after *kan* needs to be voiced to be *kand*a**.³ In contrast, the other loanword is not *komb*i*uutaa* but *komp*i*uutaa* because the faithfulness constraint (i.e., IDENT[voice]) has priority over the markedness constraint (i.e., Post Nasal Voicing); the *p* in *komp*i*uutaa* has to be voiceless (and thus must not be *komb*i*uutaa*) to be faithful to the source word 'computer'. These examples indicate that phonological constraints are applied in different rankings depending on whether the word is native or loaned. If we apply the idea of lexical stratification to the deletion vs. epenthesis conflict in the current study, we could suggest that a constraint DEP-IO (that applies deletion) has a priority over MAX-IO (that applies epenthesis) in non-loan Japanese words, while a

³ This word, *kan-da*, includes a consonant cluster, which is usually unacceptable in Japanese. As mentioned earlier in this paper, however, there are exceptions that can stand in the coda position, which are voiceless obstruents and nasals (Kubozono, 2002), the latter applying in this case.

constraint MAX-IO has priority over DEP-IO in loanwords. However, this account is not convincing enough because there are some instances of loanwords that include deletion, not epenthesis, e.g., ‘handkerchief [hæŋkətʃɪf] is [hankatɕi] as a loanword in Japanese, which deletes the coda [f].

Another possible account for the discrepancy between non-loan preference for deletion vs. loanword preference for epenthesis can be offered by Mathieu’s (2012) and Smith’s (2006) studies. According to their explanations, the selection between deletion and epenthesis depends on whether the source words were initially emerged/imported in text or sound. Mathieu and Smith argue that epenthesis is preferred when the source words were input in text, while deletion is preferred when the source words were input in sound. Mathieu shows an example, in which a Romanian word for ‘step’ is pronounced as [pas], which is loaned from French. The French source word is pronounced as [pa] but spelled as *pas*. Although the French pronunciation does not include the word-final [s], the Romanian loanword is pronounced as [pas], being influenced by the French text, *pas*. This is because this Romanian loanword came from the French text-based source. This argument may be supported by Altenburg and Vago (1987, as cited in Hancin-Bhatt, 2008), who maintain that a text-input new word is visually perceivable as in spelling, and thus a speaker tends to utter all the text (i.e., “spelling pronunciations”), which results in epenthesis rather than deletion. These arguments could supplement the lexical-stratification account. That is, Japanese loanwords prefer epenthesis possibly because most of them were imported as text, but a small number of loanwords that prefer deletion were imported auditorily. That could be why a small number of loanwords exhibit deletion while most loanwords show epenthesis. Also, Shoji and Shoji (2014) explain that this deletion occurred for auditorily transmitted loanwords possibly because the deleted consonant was not perceptual to the speakers of the borrower language (i.e., they did not hear the consonant).

Regarding the effect from the input medium that Mathieu (2012) and Smith (2006) argue, Shoji and Shoji’s (2014) study mentioned earlier also tested native Japanese speakers with sound-input nonce words in English as the experimental items. It was predicted that participants would rewrite the English nonce words utilizing deletion when the words were auditorily input. The results indicated that, compared with the other experiment with text-input, participants’ preference for deletion increased (i.e., less than 1% → less than 10%). However, the number of instances of deletion remained small, which may indicate that epenthesis is the default adaptation method for loanwords in both cases of text-input and sound-input. Nevertheless, Shoji and Shoji’s second experiment exhibited the effects of input mediums to be real, as the results showed increased instances of deletion for sound-input items, compared with those when the items were text-input.

Summarizing the earlier studies mentioned in this section, default phonological adaptations appear to be epenthesis for loanwords and deletion for non-loan words in Japanese, compatible with Fukazawa, et al's (1997) lexical stratification which argues that non-loan and loan words behave differently. However, the effect from input medium that Mathieu (2012) and Smith (2006) researched is also true, although the effect is weak, as shown by Shoji and Shoji (2014). A remaining issue that should be studied is whether Japanese non-loan word adaptation prefers deletion or epenthesis when source words are text-input, which is the research question of the current study, as summarized in Table 1.

Table 1: Preferences of deletion or epenthesis

Loanwords	Text-input Significant preference for epenthesis (Default adaptation) → - evidenced by existing instances, i.e., actual words - evidenced by experiment (Shoji & Shoji, 2014)	Sound-input Decreased preference for epenthesis & Increased preference for deletion - evidenced by experiment (Shoji & Shoji, 2014)
Non-loan words	Sound-input Significant preference for epenthesis (Default adaptation) → - evidenced by existing instances, i.e., actual words	Text-input - tested in the current study

Non-loan word adaptation obtained from the experiment of the current study may exhibit the preference for deletion if it is the default method for non-loan word adaptation. Also, the effect from input-medium could be found, i.e., the adaptation might tend to rely on epenthesis when the source words are text-input.

2 Experiment

As mentioned, the objective of the current study is to investigate whether non-loan Japanese word adaptation prefers deletion or epenthesis when source words are text-input. To empirically test this, an experiment in written test format was conducted. The experiment was participated in by 20 native Japanese speakers from the community, whose ages ranged from 23-61 ($M = 39.2$). As the experimental items, the participants were given 5 nonce words written in Japanese hiragana texts. Because hiragana texts

were normally used to write non-loan words in Japanese, the participants should interpret the given nonce words as non-loan, not as loanwords. The experimental items were all verbs with consonant-ending stems, and participants were tasked to create the nonce verbs' causative, passive and potential forms by adding the suffixes *sase*, *rare*, and *re*, respectively.⁴ Also, the items' stem-final consonants were avoided to be *s* or *r* because, otherwise, the formation with the stem-final *s* or *r* and suffix-initial *s* or *r* (in *sase*, *rare*, and *re*) would make a succession of two identical consonants, which tends to be avoided in general (McCawley, 1968). The items, their stems, suffixes, and expected formations with deletion and epenthesis are shown in Table 2.

Table 2: Experimental item words

Items (nonce words)	Stem	Suffix	Expected formation	
			by deletion	by epenthesis
<i>rokomu</i>	<i>rokom</i>	<i>sase</i> (causative)	<i>rokom-ase</i>	<i>rokom-V-sase</i>
		<i>rare</i> (passive)	<i>rokom-are</i>	<i>rokom-V-rare</i>
		<i>re</i> (potential)	<i>rokom-e</i>	<i>rokom-V-re</i>
<i>tomaku</i>	<i>tomak</i>	<i>sase</i> (causative)	<i>tomak-ase</i>	<i>tomak-V-sase</i>
		<i>rare</i> (passive)	<i>tomak-are</i>	<i>tomak-V-rare</i>
		<i>re</i> (potential)	<i>tomak-e</i>	<i>tomak-V-re</i>
<i>ratu</i>	<i>rat</i>	<i>sase</i> (causative)	<i>rat-ase</i>	<i>rat-V-sase</i>
		<i>rare</i> (passive)	<i>rat-are</i>	<i>rat-V-rare</i>
		<i>re</i> (potential)	<i>rat-e</i>	<i>rat-V-re</i>
<i>darinu</i>	<i>darim</i>	<i>sase</i> (causative)	<i>darim-ase</i>	<i>darim-V-sase</i>
		<i>rare</i> (passive)	<i>darim-are</i>	<i>darim-V-rare</i>
		<i>re</i> (potential)	<i>darim-e</i>	<i>darim-V-re</i>
<i>jimu</i>	<i>jim</i>	<i>sase</i> (causative)	<i>jim-ase</i>	<i>jim-V-sase</i>
		<i>rare</i> (passive)	<i>jim-are</i>	<i>jim-V-rare</i>
		<i>re</i> (potential)	<i>jim-e</i>	<i>jim-V-re</i>

The order of the items given to each participant was randomized by the investigator. The experiment was conducted in traditional paper-and-pencil format, which took approximately 10 minutes. The participants' responses should show whether native Japanese speakers prefer deletion or epenthesis when phonologically adapting non-loan words given in the text.

⁴ Although the participants were instructed to create the causative form, the passive form, and the potential form, the suffixes were not provided to the participants. The given items in the experiments were only the nonce verbs.

3 Results

For analyzing the forms that participants created, only those that consisted of the stem and the suffix (and an inserted vowel in case of the epenthesis) were subject to the analyses. Other instances such as the ones that overly deleted non-stem-ending consonants (e.g., *rokom + sase* → *rom-ase*) or the ones that overly epenthesized consonants (e.g., *rokom + sase* → *rokom-ase-sase*) were excluded from the data analyses because these instances may indicate that participants did not correctly understand the stems and suffixes. Also, the other types of excluded instances were those showing participants' misinterpretations such as writing a causative-passive form (e.g., *rokom-ase-rare*) when a passive form was needed (i.e., *rokom-are* or *rokom-V-rare*). After excluding these instances, we obtained 85 instances for the causative form, 86 instances for the passive form, and 81 instances for the potential form. In total, 252 instances were subject to data analyses.

Results with a number of instances with deletions and epenthesis are shown in Table 3.

Table 3: Results: numbers of instances

Form	Deletion	Epenthesis
Causative	64 (75.3%)	21 (24.7%)
Passive	47 (54.7%)	39 (45.3%)
Potential	58 (71.6%)	23 (28.4%)
Total	169 (67.1%)	83 (32.9%)

The repeated measure ANOVA was run for by-participant analysis and by-item analysis. The omnibus analysis for the results showed that, for all the formations, there were significant (and marginally significant) preferences for deletion over epenthesis [$F_1=11.09, p=.004; F_2=5.90, p=.072$]. In the participant analysis, there was a significant interaction between the adaptation method (deletion vs. epenthesis) and forms (causative, passive, vs. potential), although the item analysis did not find significance [$F_1=6.89, p=.007; F_2=1.36, p=.380$].

A series of individual T-test analyses were conducted for each form comparing the occurrences of deletions and epenthesis. Results of the tests indicated that the passive form showed no significant preference either for deletion or epenthesis [$t_1=.44, p=.685; t_2=.648, p=.525$] although the causative form showed significant preference for deletion [$t_1=3.90, p=.018; t_2=3.56, p=.002$] and the potential form showed a significant preference for deletion in the participant analysis and a marginally significant preference in the item analysis [$t_1=3.60, p=.002; t_2=2.75, p=.052$].

The results of all the statistical analyses are summarized in Table 4.

Table 4: Results: Statistic analyses

	Participant analyses	Item analyses
Omnibus	Significant preference for deletion	Marginally significant preference for deletion
Causative form	Significant preference for deletion	Significant preference for deletion
Passive form	No preference	No preference
Potential form	Significant preference for deletion	Marginally significant preference for deletion

4 Discussion

The results shown in the previous section appear to be mixed. The omnibus analyses imply that there is a trend that prefers deletion to epenthesis when native Japanese speakers adapt non-loan words with text input. This outcome supports the earlier studies, which suggests that phonological adaptation for non-loan Japanese prefers deletion by default. However, this preference was inconsistent, or possibly selective, depending on the type of formation. The causative formation showed the strong preference for deletion, but the potential formation showed a weaker preference for deletion, and the passive formation did not elicit any preference. The inconsistent outcome may not be a strong evidence for deletion to be the default method for non-loan word adaptation in Japanese.

The weak preference for deletion in the potential form as well as no preference in the passive form could be attributed to the effect from input medium. Following Mathieu (2012) and Smith's (2006) arguments, text-input may have led native Japanese-speaking participants to epenthesis for many instances. However, the possible effect from the text input was still random or selective because the occurrence possibilities of the epenthesizing instances greatly differed between the causative, passive, and potential forms. Thus, this outcome also could not be strong evidence of the effect of the input medium.

The unclear results from this experiment may indicate that deletion as the phonological adaptation for Japanese non-loan words is “weakly default”, but not “absolutely default”, in spite of the fact that the formations of existing Japanese verbs appear to consistently exhibit deletion. Thus, this defaultness could be subject to change depending on different forms, which could not be clarified in the current study. Also, it could be possible that the preference for deletion or epenthesis changes over time. For example, a historical anthology *Manyōshū* from the 8th century shows that, when *siru* ‘know,’ whose stem is *sir*, was attached by the potential suffix *yu*, the form was *sir-a-yu*; with *a* epenthesized. Likewise, the potential form of *naku* ‘cry’, whose

stem is *nak*, was *nak-a-yu* (Shibuya, 1993). In addition, another possibility is that the preference for deletion or epenthesis is dialectologically different. For example, in the Kansai region, the potential forms of *yomu* ‘read’, whose stem is *yom*, can be *yom-a-re*, with *a* epenthesis.⁵ All these inconsistent instances imply that the preference for deletion in non-loan words is not always true and that there is much to be further researched.

5 Conclusions

This study investigated the native Japanese speakers’ preferences for deletion and epenthesis for non-loan word adaptation. The experiment implied, but not strongly evidenced, the defaultness of deletion and/or the effect from the input medium. Further research is necessary in order to investigate such effects. Specifically, an experiment should be conducted utilizing non-loan nonce words, which could be identical to the experimental items in the current study, but with the sound input. The defaultness of deletion for non-loan word adaptation will be confirmed if the results show significant preferences for deletion to epenthesis. Also, the effect from the input medium will be ascertained if reliance on epenthesis significantly decreases in the case of sound-input items, compared to that with text-input. In addition, when utilizing sound inputs, there may be effects from accentuation, as Kubozono (1997, 2002) argues. Although Kubozono finds that loanwords tend to preserve input accent while native Japanese words do not as much, we may find some extent of effects from accentuation leading Japanese speakers to choose deletion or epenthesis, either of which would preserve the source-word accentuation. Therefore, the accents of the sound-input items should be controlled experimentally. Moreover, data from Shoji and Shoji’s (2014) experiment with sound-input loanwords may need to be reanalyzed with the consideration of the items’ accents.

Another possible future research study should investigate whether the preferences of deletion and epenthesis are selective in response to each formation. The current study exhibited that the causative formation significantly preferred deletion while the passive formation showed no preference. This inconsistency might be due to the phonological difference between suffixes. In addition, the preferences either for deletion or epenthesis are selective for the sake of assimilation or dissimilation between the suffix and stems. All these possibilities indicate that there are many issues remaining to be investigated.

⁵ This type of epenthesis in the potential form in the Kansai region appears mostly in the negative-potential form, e.g., *yom-a-re-hen* ‘cannot read’ or *nom-a-re-hen* ‘cannot drink’.

Acknowledgments

The author wishes to thank Mr. Michael Donohue, who supported this article in providing comments and proofreading.

References

- Altenburg, E. P. & Vago, R. M. (1987). Theoretical implications of an error analysis of second language phonology production. In G. Loup & S. Weinberger (Eds.), *Internlanguage phonology: The acquisition of a second language sound system* (pp. 148–164). Cambridge, MA: Newbury House.
- Fukazawa, H., Kitahara, M., & Ota, M. (1998). Lexical Stratification and Ranking Invariance in Constraint-based Grammars. *Proceedings of Chicago Linguistic Society* 34.
- Hancin-Bhatt, B. (2008). Second language phonology in optimality theory. In J. G. H. Edward & M. L. Zampini (Eds.), *Phonology and second language education* (pp. 117–146). Amsterdam, Netherlands: John Benjamins.
- Kubozono, H. (1997). Lexical markedness and variation: A non-derivational account. *West Coast Conference on Formal Linguistics* 15. Stanford, CA: CSLI, 273-287.
- Kubozono, H. (2002). Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Journal of Phonetic Society* 6, 79-97.
- Lovins, Julie. B. (1975). *Loanwords and the Phonological Structure of Japanese*. Bloomington: Indiana University Linguistic Club.
- Mathieu, L. (2012). Orthographic traces in Romanian and Japanese loanwords: Enriching phonological representations. *Journal of Language Contact* 5, 144-181.
- Mattingley, M., Hall, K. C., & Hume E. (2019). Epenthetic vowel production of unfamiliar medial consonant clusters by Japanese speakers. *Laboratory Phonology* 10, 1-35.
- Mori, Masatoshi. (1929). *The pronunciation of Japanese*. Tokyo, Japan: The Harald-sha.
- McCawley, J. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton & Co.
- Sapir, Edward. (1921). *Language: An introduction to the study of speech*. Eahway, NJ: Quinn & Boden Company.
- Shibuya, K. (1993). Nihongo kanou hyougen no shosou to hatten [Aspects and development of Japanese potential expressions] *Bulletin of School of Letters, Osaka University* 33, 1-262.
- Shoji, S. & Shoji, K. (2014). Vowel epenthesis and consonant deletion in Japanese loanwords from English. *Supplemental Proceedings of Annual Meetings on Phonology 2013*. Retrieved from: <http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/16>.
- Smith, J. L. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. In T. Vance (Ed.), *Japanese/Korean Linguistics* 14 (pp. 63–74). Stanford, CA: CSLI.

Correlations Between Proposed Orthoepic Competence Descriptors and Japanese Language Ability

ITO Hideaki

University of Tsukuba, Japan
ito.hideaki.gb@u.tsukuba.ac.jp

Abstract

In this paper, three types of Japanese online tests, and self-assessment questionnaires comprised of 13 descriptor categories, including one category on Japanese orthoepic competence, were issued to 15 Japanese language learners attending language schools in Japan. As a result, we confirmed a more than moderate positive correlation between the orthoepic competence descriptors and test scores, both concerning the individual scores on the three tests and the aggregate total of those scores. Based on these test results, learners were categorized into different skill levels, such as novice, intermediate, and advanced. Learners who scored at the intermediate level with their grammar test or scored over 170 total points across all tests tended to evaluate themselves at a B-level or higher competency level.

Keywords: Japanese language education; orthoepic competence; CEFR; descriptors; self-assessment

Povzetek

V tem prispevku so bili 15 učencem japonskega jezika, ki obiskujejo jezikovne šole na Japonskem, izdani tri vrste japonskih spletnih testov in vprašalniki za samoocenjevanje, sestavljeni iz 13 kategorij deskriptorjev, vključno z eno kategorijo o japonski ortoepski kompetenci. Rezultati testov so potrdili več kot zmerno pozitivno korelacijo med deskriptorji ortoepske kompetence tako posamično kot tudi v skupnem seštevku vseh točk. Na podlagi teh rezultatov so bili učenci razvrščeni v različne ravni spretnosti, kot so začetniki, srednji in napredni. Učenci, ki so pri svojem slovničnem testu dosegli vmesno stopnjo ali dosegli več kot 170 skupnih točk na vseh testih, so bili ocenjeni na ravni B ali višji ravni kompetenc.

Ključne besede: poučevanje japonskega jezika; ortoepska kompetenca; CEFR; deskriptorji; samoocenjevanje



1 Introduction

In 2001, the Council of Europe introduced the “Common European Framework of Reference for Languages” (hereinafter referred to as CEFR) (Council of Europe, 2001). This new framework rapidly spread throughout the world’s language education circles. In Asia, including Japan, its acceptance is growing as the language education standard for various foreign languages (Chéng, 2017). In the case of Japanese language education, the Japan Foundation created its standard, named the JF Standard for Japanese-Language Education (JFS) taking into consideration the CEFR. Programs using this standard expanded in recent years, deploying class level descriptions, learning material, and other components based on common reference levels that are further divided into six stages (Ito, 2019, Majima, 2018). However, North (2014) noted problems applying CEFR to languages such as Japanese, which use large numbers of characters.

North writes:

In the context of current pedagogy for Japanese and Chinese it is not possible for a learner at A2 or B1 or B2 to read the types of text that appear in CEFR descriptors for the levels concerned, simply because they do not know enough signs. (p. 45)

North continues:

...[u]sing the CEFR for such languages implies either *profiling* proficiency, admitting that such learners are a higher level for listening and speaking than they are for reading and writing – which the CEFR scales will facilitate describing – or alternatively, developing completely new descriptors for reading and writing. (p. 45)

There have also been reports on problems conducting task-based tests in Japanese language education, where a lack of sign ability may impede carrying out tasks, or make a performance evaluation difficult (Kumano et al., 2013). Therefore, in recent years discussions have advocated the need to resolve orthoepic competence to adopt the CEFR in Japanese language education. Engaging with this issue, we have already noted (see Ito 2017, 2019, 2020) that although orthoepic competence is acknowledged in the CEFR, the framework does not outline any descriptors for this competency. From 2017 to the present year, we proposed several draft descriptors for orthoepic competence, arranged according to different proficiency levels. It should be noted, however, that these drafted descriptors for orthoepic competence were created by using text mining analysis to extract characteristic words from communication ability descriptors as given within the CEFR. For this reason, the relationship between these descriptors and actual Japanese language learners’ abilities has not been clarified. The present research, therefore, aims to clarify the relationship between Japanese language learners’ abilities,

and descriptors of Japanese language orthoepic competence, as proposed in our previous study (Ito, 2020).

The structure of this paper is as follows. Section 2 outlines the preceding research related to Japanese language orthoepic competence and clarifies the research project's goal. Section 3 explains the examination methods used, and reports on the results attained. Section 4 contains an analysis and observation of the results. Finally, Section 5 provides a summary of the details included in this paper and raises several future research directions.

2 Prior research

2.1 The definition of orthoepic competence

In the CEFR, orthoepic competence is defined in the following manner: “[U]sers required to read aloud a prepared text, or to use in speech words first encountered in their written form, need to be able to produce a correct pronunciation from the written form, need to be able to produce a correct pronunciation from the written form.” (Council of Europe, 2001, p. 117) Meanwhile, Bellassen and Zhang (2008) define orthoepic competence as: “The ability of the language user to accurately read aloud and pronounce a text or speech in a loud voice” (p. 68). We (Ito, 2019), however, take a somewhat different approach. After examining orthoepic competence entries in the CEFR as well as Bellassen and Zhang (2008), we suggest that, when a person reads, an ambiguous continuity exists between pronunciation and the understanding of a word's meaning: “The individual can pronounce characters or vocabulary, and is at least partially capable of understanding the meaning of characters or words within context, as well as the function of how they are written.” (p. 76) In this paper, orthoepic competence is understood by the definition provided in Ito (2019).

2.2 Orthoepic competence descriptors

A large volume of research relating to the CEFR has been conducted in numerous languages, often focusing on a characteristic aptitude outlined in the CEFR, known as “competence in accomplishing tasks” (Ito, 2019). However, very little research has specifically investigated the connection between the CEFR's orthoepic competence and languages that use non-alphabetic scripts. It would seem that the only examples are Bellassen and Zhang (2008), and Ito (2017, 2019, 2020).

Bellassen and Zhang (2008) experimented with the introduction of the CEFR to Chinese language education in France. They noted that the characters used in the Chinese language pose a challenge for the implementation of the CEFR. This is because there is almost no relationship between the pronunciation of Chinese characters and

how they are written. It is almost impossible for learners to read characters they are unfamiliar with. Furthermore, understanding the meaning of characters can be impeded by misreading, which stems from the fact that different characters look similar to each other. Therefore, Bellassen and Zhang (2008) proposed an evaluative standard known as a “literacy threshold.” First, Chinese characters can be grouped into different tiers of language proficiency, based on the frequency with which they are used, their frequency in everyday conversation, and the extent to which they can be combined with other characters to create new ones. Then, the language proficiency of a learner can be determined based on how many characters they can recognize and write (Table 1). According to our study (Ito, 2019, 2020), the research behind the ‘literacy threshold’ proposed by Bellassen and Zhang (2008) is valuable because of the way it links together Chinese characters and the CEFR. However, we note the proposed “literacy threshold” does not include any concrete methods for selecting Chinese characters, except based on frequency. Furthermore, it only indicates tiers of the characters, with no specific descriptors for each tier. We, therefore, doubt the applicability of this approach, given that it runs counter to how the CEFR understands language users; i.e., as ‘social agents’ who strengthen and revise their language capabilities while carrying out tasks under particular environmental conditions.

Table 1: Literacy threshold (Bellassen & Zhang, 2008, p. 69).

Levels	Number of Chinese characters (approximately)
C2	Over 3,000
C1	2,200
B2	1,500
B1	800
A2	500
A1	250

In the previous studies (Ito, 2017, 2019, 2020), we researched the formulation of Japanese language orthoepic competence descriptors, making progressive advancements. We noted that, although the CEFR lacks concrete descriptors for orthoepic competence, the framework itself is not complete, but is characterized by an orientation towards continual expansion and refinement (Ito, 2017). As the Council of Europe (2001) states: “The framework should be open and flexible, so that it can be applied, with such adaptations as prove necessary, to particular situations” (p. 7). We further assert that the description of the Japanese language orthoepic competence is a pressing task. In our view, it is necessary to establish some concrete descriptors as soon as possible, even if they are initially in a rough form that leaves room for further discussion and refinement. We therefore developed some draft descriptors for Japanese language orthoepic competence. However, to produce these drafts, we

turned to the six linguistic competences where the CEFR does provide descriptors (general linguistic range, vocabulary range, vocabulary control, grammatical competence, phonological competence, orthographic competence), and arbitrarily extracted characteristic words from each language proficiency level (Ito, 2017). Therefore, the resulting descriptors cannot be said to have a high degree of objectivity. Following this initial attempt, we asserted, “It is necessary to research [orthoepic competence] further so that it becomes a more objective measure” (Ito, 2019, p. 78). This time, a text mining method was used to extract characteristic words, and re-consider the orthoepic competence descriptors for levels A1 and A2. Finally, with Ito (2020), the work conducted in Ito (2019) is extended from levels B1 through to C2, with a more objective method deployed to create competence descriptors for Japanese language education (Table 2).

Table 2: Orthoepic competences (Ito, 2020)

C2	The learner can consistently read accurately, and can even read difficult kanji such as those found in literature.
C1	While [a learner] may make slight mistakes at the vocabulary level, they can accurately read linguistic expressions and vocabulary, such as [those] used in their field of expertise.
B2	[The individual] has a high level of orthoepic competence, and can accurately read words if they are common.
B1	While there are cases where [a learner] may make obvious mistakes, they are relatively able to accurately read characters related to a broad range of material, including everyday topics.
A2	There are many cases where a learner may need to re-read a section of text or reads incorrectly; however, if they have the necessary basic vocabulary, then they are able to read material encountered in daily situations.
A1	The learner is able to read a section of text if they have studied the material and has a basic, concrete, and limited repertoire (words and expressions, etc.) that relates to his/her personal information.

3 Examination

3.1 Examination participants

In February 2020, data were collected from a total of 15 students attending Japanese language school E in Japan. The breakdown of the 15 students’ nationalities was as follows: 11 from China, 1 from Taiwan, 1 from South Korea, 1 from Vietnam, and 1 from Indonesia. As the aim was to collect data from learners with a diverse range of skill levels, no particular restrictions were given with regards to Japanese language ability. As a result, data was collected from 3 individuals with less than a year of experience, 9

individuals with one to three years of experience, 1 individual with three to five years of experience, and 2 individuals with more than five years of experience.

3.2 Examination Methods

For the examination, participants were gathered into a single classroom, and the following two steps were carried out.

1. Data were collected on the Japanese language ability of examination participants by having them sit three online Japanese language tests, the SPOT90, Grammar90, and Kanji SPOT50. This set was provided by the Tsukuba University Center for Distance Learning of Japanese and Japanese Issues, and is known as the “Tsukuba Test-Battery of Japanese” (TTBJ).
2. Data were collected on how students evaluated their Japanese language ability. For this purpose, students were asked to complete a self-assessment questionnaire, after being shown descriptors from A1 to C2 as a random for 13 categories. These included six communicative language activities designated in the CEFR: “overall oral production” (oral production), “overall written production” (written production), “overall listening comprehension” (aural reception), “overall reading comprehension” (visual reception), “overall spoken interaction” (spoken interaction), and “overall written interaction” (written interaction); the six communicative language competences designated in the CEFR, of “general linguistic range,” “vocabulary range,” “vocabulary control,” “grammatical competence,” “phonological competence,” and “orthographic competence;” and the focus of the present research, “orthoepic competence.”

The respective aims of the three TTBJ tests used in step 1 are as follows: the SPOT90 test measures total Japanese language ability, including practical know-how, the Grammar90 test measures grammatical knowledge, and the Kanji SPOT50 test measures capability in using kanji-based vocabulary (Kobayashi, 2015; Sakai et al., 2015). The test time was between 30 to 60 minutes in total.

CEFR自己評価_eng

[OVERALL ORAL PRODUCTION]
Please choose the one closest to your Japanese language ability.

<input type="checkbox"/> Can give clear, detailed descriptions and presentations on complex subjects, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	<input type="checkbox"/> Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail. Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.
<input type="checkbox"/> Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.	<input type="checkbox"/> Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes, etc. as a short series of simple phrases and sentences linked into a list.
<input type="checkbox"/> Can produce simple mainly isolated phrases about people and places.	<input type="checkbox"/> Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.

▶ Next

◀ Back ▶ Next

2 / 14
Question List

Figure 1: An example of the English version of the display used in the self-assessment questionnaire

3.3 Examination Results

From the TTBJ tests and self-assessment questionnaires, data were collected from each of the 15 individuals. This data included the results of individual tests, the total score for all three tests combined, and the 13 category self-assessment results. All 15 individuals sat the TTBJ tests until the end, and also selected answers for all 13 categories of the self-evaluation questionnaire. Therefore, the data from all 15 individuals were used in the analysis. An overview of the examination results is provided in Tables 3 and 4 below.

Table 3: Overview of TTBJ results

Test Name	Full Marks	Average Value	Minimum Value	Maximum Value
SPOT90	90	66.2	41	79
Grammar90	90	62.5	43	82
Kanji SPOT50	50	38.9	15	47
Total	230	167.7	108	208

Table 4: Overview of self-assessment results

Self-Assessment Categories	Most Commonly Selected Proficiency	Least Commonly Selected Proficiency	Highest Selected Proficiency
Overall Oral Production	A2	A1	C1
Overall Written Ability	B1	A2	C2
Overall Listening Comprehension	B1	A2	C1
Overall Reading Comprehension	A2/B1/B2	A1	C1
Overall Spoken Interaction	A1	A1	C1
Overall Written Interaction	B2	A1	C1
General Linguistic Range	A1/A2/B1	A1	C1
Vocabulary Range	A1	A1	C2
Vocabulary Control	A2	A1	C2
Grammatical Competence	A2/B1	A1	C2
Phonological Competence	A2	A1	C1
Orthographic Competence	A1/A2	A1	C2
Orthoepic Competence	B1	A1	C2

4 Results

4.1 Method of analysis

With the data gained through the TTBJ, along with the self-assessment questionnaire, correlation analysis was conducted between the total test score and individual test results, and the descriptor self-assessment. This examination used an ordinal scale for descriptor self-assessment. Therefore, correlation analysis of the TTBJ results and the descriptor self-assessment was conducted by applying Spearman's rank correlation coefficient, which is nonparametric. The significance of the correlation coefficient was also examined by testing for non-correlations. Section 4-2 shows the correlation between the descriptor self-assessment and the test total score. Sections 4-3 to 4-5 show correlations between the respective test results and the descriptor self-assessment. In Section 4-6, some observations are made about the analysis results.

4.2 Method of analysis correlation between total test score and self-assessment

The correlation results between the total test score and the descriptor self-assessment are shown in Table 5. If we examine the correlation coefficients and *p*-values, for the 3 categories: "overall oral production," "overall listening comprehension," and "general

linguistic range,” a statistically significant moderate positive correlation at 5% was observed (overall oral production $\rho = 0.64$, $P = 0.010$, overall listening comprehension $\rho = 0.68$, $P = 0.006$, general linguistic range $\rho = 0.65$, $P = 0.009$). Furthermore, for “orthoepic competence” as well, a statistically significant strong positive correlation at 5% was observed ($\rho = 0.72$, $P = 0.003$).

Table 5: Correlation between total test score and self-assessment

Variable 1	Variable 2	Correlation Coefficients (ρ)	P-Value	Frequency
Test Total	Overall Oral Production	0.639	0.010	15
	Overall Written Production	0.192	0.493	15
	Overall Listening Comprehension	0.676	0.006	15
	Overall Reading Comprehension	0.463	0.082	15
	Overall Spoken Interaction	0.144	0.610	15
	Overall Written Interaction	0.261	0.348	15
	General Linguistic Range	0.648	0.009	15
	Vocabulary Range	0.160	0.569	15
	Vocabulary Control	0.495	0.060	15
	Grammatical Competence	0.076	0.787	15
	Phonological Competence	0.298	0.280	15
	Orthographic Competence	0.091	0.748	15
	Orthoepic Competence	0.715	0.003	15

4.3 Correlation between the SPOT90 and self-assessment

The correlation results between the SPOT90 and the descriptor self-assessment are presented in Table 6. The correlation coefficient for “orthoepic competence” was somewhat lower than for the correlation with the total test score, with a moderate correlation. However, as with the correlation with the total test score, a statistically significant moderate positive correlation at 5% was observed for four categories: “overall oral production,” “overall listening comprehension,” “general linguistic range,” and “orthoepic competence” (overall oral production $\rho = 0.65$, $P = 0.009$, overall listening comprehension $\rho = 0.64$, $P = 0.010$, general linguistic range $\rho = 0.68$, $P = 0.006$, orthoepic competence $\rho = 0.51$, $P = 0.050$).

Table 6: Correlation between the SPOT90 and self-assessment

Variable 1	Variable 2	Correlation Coefficients (ρ)	P-Value	Frequency
SPOT90	Overall Oral Production	0.649	0.009	15
	Overall Written Production	0.223	0.424	15
	Overall Listening Comprehension	0.640	0.010	15
	Overall Reading Comprehension	0.339	0.216	15
	Overall Spoken Interaction	0.232	0.405	15
	Overall Written Interaction	0.192	0.494	15
	General Linguistic Range	0.676	0.006	15
	Vocabulary Range	0.110	0.695	15
	Vocabulary Control	0.376	0.167	15
	Grammatical Competence	0.083	0.769	15
	Phonological Competence	0.186	0.507	15
	Orthographic Competence	0.068	0.810	15
	Orthoepic Competence	0.515	0.050	15

4.4 Correlation between the Grammar90 and Self-Assessment

The correlation results between the Grammar90 and the descriptor self-assessment are presented in Table 7. A statistically significant moderate positive correlation at 5% was observed for the four categories “overall oral production,” “overall listening comprehension,” “overall written interaction,” and “orthoepic competence” (overall oral production $\rho = 0.58$, $P = 0.023$, overall listening comprehension $\rho = 0.54$, $P = 0.040$, overall written interaction $\rho = 0.53$, $P = 0.044$, orthoepic competence $\rho = 0.60$, $P = 0.019$). A statistically significant strong positive correlation at 5% was also observed for ‘general linguistic range’ ($\rho = 0.79$, $P = 0.001$).

Table 7: Correlation between the Grammar90 and self-assessment

Variable 1	Variable 2	Correlation Coefficients (ρ)	P-Value	Frequency
Grammar90	Overall Oral Production	0.583	0.023	15
	Overall Written Production	0.183	0.513	15
	Overall Listening Comprehension	0.535	0.040	15
	Overall Reading Comprehension	0.414	0.125	15
	Overall Spoken Interaction	0.345	0.208	15
	Overall Written Interaction	0.525	0.044	15
	General Linguistic Range	0.786	0.001	15
	Vocabulary Range	0.338	0.218	15
	Vocabulary Control	0.434	0.106	15
	Grammatical Competence	0.136	0.630	15
	Phonological Competence	0.289	0.296	15
	Orthographic Competence	0.358	0.190	15
	Orthoepic Competence	0.595	0.019	15

4.5 Correlation between the Kanji SPOT50 and Self-Assessment

The correlation results between the Kanji SPOT50 and the descriptor self-assessment are presented in Table 8. A statistically significant moderate positive correlation at 5% was observed for the three categories “overall oral production,” “overall listening comprehension,” and “orthoepic competence” (overall oral production $\rho = 0.60$, $P = 0.018$, overall listening comprehension $\rho = 0.53$, $P = 0.044$, orthoepic competence $\rho = 0.63$, $P = 0.012$).

Table 8: Correlation between the Kanji SPOT50 and self-assessment

Variable 1	Variable 2	Correlation Coefficients (ρ)	P-Value	Frequency
Kanji SPOT50	Overall Oral Production	0.601	0.018	15
	Overall Written Production	0.295	0.286	15
	Overall Listening Comprehension	0.527	0.044	15
	Overall Reading Comprehension	0.303	0.272	15
	Overall Spoken Interaction	-0.061	0.828	15
	Overall Written Interaction	-0.092	0.743	15
	General Linguistic Range	0.332	0.226	15

Variable 1	Variable 2	Correlation Coefficients (ρ)	P-Value	Frequency
	Vocabulary Range	0.026	0.927	15
	Vocabulary Control	0.437	0.103	15
	Grammatical Competence	0.082	0.772	15
	Phonological Competence	0.148	0.597	15
	Orthographic Competence	-0.142	0.613	15
	Orthoepic Competence	0.631	0.012	15

5 Discussion

Table 9 summarizes the correlation results for the total test scores as well as the results of each test, and the descriptor self-assessment.

Table 9: Categories where a statistically significant correlation was observed in the examination

Test Name	Categories with a Moderate Correlation Observed	Categories with a Strong Correlation Observed
Total Test Score	Overall Oral Production, Overall Listening Comprehension, General Linguistic range	Orthoepic Competence
SPOT90	Overall Oral Production, Overall Listening Comprehension, General Linguistic Range, Orthoepic Competence	
Grammar90	Overall Oral Production, Overall Listening Comprehension, Overall Written Interaction, Orthoepic Competence	General Linguistic Range
Kanji SPOT50	Overall Oral Production, Overall Listening Comprehension, Orthoepic Competence	

As we can see from the results shown in Table 9, there was a statistically significant positive correlation between the Japanese orthoepic competence descriptors proposed by Ito (2020) and the total test score, as well as with the scores for the individual tests themselves; SPOT90, Grammar90, and Kanji SPOT50. Furthermore, if we examine the correlation coefficient for orthoepic competence, and each score, the total test score was the highest at $\rho = 0.72$. Then, we have Kanji SPOT50 at $\rho = 0.63$, Grammar90 at $\rho = 0.60$, and SPOT90 at $\rho = 0.51$. The total test score displayed the highest degree of correlation. The next highest score was obtained by the test that measures kanji ability, the Kanji SPOT50, which is directly related to orthoepic

competence. Follow is the Grammar90, which measures grammatical knowledge. At the end is the SPOT90, which primarily measures practical conversational ability. If we consider the definition of orthoepic competence as “the capacity to read text or characters aloud,” the necessary abilities for this skill in order are overall ability, kanji knowledge, grammatical knowledge, and conversational ability. Therefore, the above ordering of correlation coefficients suggests a general correlation between Japanese language ability and the Japanese orthoepic competence descriptors proposed by Ito (2020). Moreover, for this examination, the descriptor “overall reading comprehension,” was prepared for the self-assessment categories. However, “overall reading comprehension,” had only a weak or moderate non-significant correlation with any of the test scores. This also indicated that orthoepic competence descriptors are assessed separately from reading comprehension descriptors.

Furthermore, if we change our focus and consider the abilities the TTBJ test sets measure within the CEFR descriptors, we found that for each test, a statistically significant positive correlation was confirmed for “overall oral production,” “overall listening comprehension,” and “orthoepic competence.” In other words, the TTBJ test set used for this examination has the potential to function as a test to measure the CEFR criteria of “overall oral production,” “overall listening comprehension,” and “orthoepic competence.”

Here, as can be seen in Table 10, scores received for the TTBJ tests SPOT90, Grammar90, and Kanji SPOT50 have equivalent Japanese-Language Proficiency Test (JLPT) standards for comparison. Following these standards, an examination was conducted into the relationship between self-assessment and proficiency levels such as novice, intermediate, and advanced.

Table 10: Standards for interpreting results of individual test scores (from the TTBJ website)

Test Name	Total Score	Proficiency	JLPT Equivalent (Rough Equivalent)
SPOT90	0 - 30	Beginner	None
	31 - 55	Novice	N4, N5
	56 - 80	Intermediate	N3, N2
	81 - 90	Advanced	N1
Grammar90	0 - 20	Complete Beginner	None
	21 - 50	Novice	N4, N5
	51 - 80	Intermediate	N3, N2
	81 - 90	Advanced	N1
Kanji SPOT50	0 - 15	Beginner	None or N5
	16 - 30	Novice	N4
	31 - 40	Intermediate	N3, N2
	41 - 50	Advanced	N1

As a result, we have found correspondences presented in Table 11. As correspondence only concerns test scores, it is difficult to reach any categorical conclusions. However, a certain tendency was observed and that is if Grammar90 was at the intermediate level then self-assessment would be level B or higher. Furthermore, when the total score for SPOT90, Grammar90, and Kanji SPOT50 reached 170 marks, there was a tendency for self-assessment at the B-level or higher. Japan Foundation (2017) has also investigated the relationship between the JLPT results and JFS assessment conducted by teachers. According to their findings, many individuals who passed the N3 level or higher were assessed as a B1 level or higher. This result matches with the standard used for the intermediate level in Grammar90, which is JLPT N3 or N2.

Moving forward, we can anticipate further investigations into the relationship between various test results and self-assessment questionnaires, like those conducted in this examination. These will shed more light on the connections between the CEFR and various resources that are already happening within Japanese language education.

Table 11: Interpreted Level of Tests Taken by Participants, and Self-Assessment Equivalence

Number	SPOT90	Grammar90	Kanji SPOT50	Total Score	Most Commonly Selected Proficiency
1	Intermediate	Intermediate	Advanced	208	B2
2	Intermediate	Intermediate	Advanced	207	C1
3	Intermediate	Intermediate	Advanced	199	B2
4	Intermediate	Intermediate	Advanced	189	B1
5	Intermediate	Intermediate	Advanced	188	B1/B2
6	Intermediate	Intermediate	Advanced	180	A2
7	Intermediate	Intermediate	Novice	169	B1
8	Intermediate	Novice	Advanced	166	A2
9	Intermediate	Intermediate	Novice	166	A2/B1
10	Intermediate	Intermediate	Advanced	157	A2/B1
11	Intermediate	Novice	Advanced	156	A1
12	Intermediate	Novice	Advanced	148	A2/B2
13	Novice	Novice	Intermediate	137	A1/A2
14	Novice	Novice	Intermediate	137	A2
15	Novice	Intermediate	Beginner	108	B2/C2

6 Conclusion and Future Tasks

The goal of this research project was to clarify the correlation between the abilities of Japanese language learners and the orthoepic competence descriptors that we provided for different proficiency levels in our previous studies (Ito, 2017, 2019, 2020). For this purpose, Japanese language learners were requested to sit the three TTBJ tests: the SPOT90, Grammar90, and Kanji SPOT50. Subsequently, they were shown 13 category descriptors before completing a self-assessment questionnaire. These descriptors included the following: the six communicative language activities designated in the CEFR, of “overall oral production” (oral production), “overall written production” (written production), “overall listening comprehension” (aural reception), “overall reading comprehension” (visual reception), “overall spoken interaction” (spoken interaction), and “overall written interaction” (written interaction); the six communicative language competences designated in the CEFR, of “general linguistic range,” “vocabulary range,” “vocabulary control,” “grammatical competence,” “phonological competence,” “orthographic competence,” and finally, “orthoepic competence.” After conducting correlation analysis using the data from the tests and self-assessment, a statistically significant positive correlation was confirmed for the orthoepic competence descriptors and test scores. This included both the total score for all three tests, as well as individual test scores. Furthermore, when examining the examination results according to the TTBJ test classification, for each of the three tests a statistically significant positive correlation was confirmed for “overall oral production,” “overall listening comprehension,” and “orthoepic competence.” The TTBJ test set used for this examination displayed potential for functioning as a test measuring “overall oral production,” “overall listening comprehension,” and “orthoepic competence,” from the CEFR standard. Here, further examination was conducted into the relationship between Japanese learners’ self-assessment, and the proficiency levels associated with the scores for the TTBJ’s SPOT90, Grammar90, and Kanji SPOT50 tests (novice, intermediate, advanced, etc.) A tendency was observed among students at the intermediate level in the Grammar90 to self-assess their skill at a B-level or higher. Students who had reached the total score of 170 or higher for all three tests also tended self-assess their skill at a B-level or higher.

Although the present examination was conducted with a limited number of participants, it provided some indications to prove the relationship between orthoepic competence descriptors and Japanese language ability. Japanese language education requires the use of three types of characters: hiragana, katakana, and kanji. Studying these characters poses a significant burden for learners. For this reason, going forward, a future task will examine the relationship between self-assessment questionnaires for the CEFR descriptors, and tests offered by external organizations. I hope to thereby clarify the utility of the orthoepic competence descriptors presented by Ito (2020).

Acknowledgments

This work was supported by JSPS KAKENHI under Grant JP18K12419.

References

- Bellassen, J., & L. Zhang. (2008). The CEFR: New concept and its implications and impetus—Chinese language teaching at its critical moment. *Chinese Teaching in the World*, 85, 58–73. (「《欧洲语言共同参考框架》新理念对汉语教学的启示与推动—处于抉择关头的汉语教学」『世界汉语教学』)
- Chéng, Y. (2017). *The introduction and contextualization of CEFR in the Chinese speaking world*. Tokyo, Japan: Coco Publishing. (程遠巍『中華世界における CEFR の受容と文脈化』ココ出版)
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Ito, H. (2017). Devising tentative descriptors of orthoepic competence for extension and refinement: With reference to CEFR/JFS linguistic competence. *Journal of Japanese Language Teaching*, 168, 55–62. (伊藤秀明, 「拡張・精緻化のための読字能力の能力記述文試案作成—CEFR/JFS の言語構造的な能力を参考に—」, 『日本語教育』) https://doi.org/10.20721/nihongokyoiku.168.0_55
- Ito, H. (2019). Reconsidering can-do descriptors for orthoepic competence with a focus on the CEFR's "Basic Language User." *The Journal of the JASBEL*, 3, 72–86. (伊藤秀明, 「読字能力の評価尺度の再考—「基礎段階の言語使用者」に注目して—」, 『基礎教育保障学研究』) https://doi.org/10.32281/jasbel.3.0_72
- Ito, H. (2020). Orthoepic competence descriptors in Japanese language education: CEFR Levels B1 to C2. *Acta linguistica asiatica*, 10(1), 49–66. <https://doi.org/10.4312/ala.10.1.9-26>
- Japan Foundation (2017) *JF standard for Japanese-Language education*. <https://jfstandard.jp/pdf/jfs_jlpt_report2017.pdf> (May 26, 2021).
- JF Standard for Japanese-Language Education 2010 (Second Edition). <https://jfstandard.jp/pdf/jfs2010_all_en.pdf> (November 24, 2021)
- Kobayashi, N. (2015) SPOT (Simple Performance-Oriented Test). In J. Lee (Ed.), *The language test guide book for Japanese language education* (pp. 110-126). Tokyo, Japan: Kurosio Publishers. (小林典子「SPOT (Simple Performance-Oriented Test)」李在鎬(編)『日本語教育のための言語テストガイドブック』くろしお出版)
- Kumano, N., Ito, H., & Hachisuka, M. (2013). Level certification test (A1) development for JFS Japanese-language courses based on JFS/CEFR. *The Japan Foundation Japanese-Language Education Bulletin*, 9, 73–88. (熊野七絵・伊藤秀明・蜂須賀真希子「JFS/CEFRに基づくJFS日本語講座レベル認定試験(A1)の開発」『国際交流基金日本語教育紀要』9号) <https://doi.org/10.20649/00000107>
- Majima, J. (2018). The domestic and international impact of CEFR on Japanese language education. In H. Sensui (Ed.), *Teaching language, learning language: Language education and the Common Framework of Reference for Languages for a multi-lingual,*

multi-cultural Europe (CEFR) (pp. 249–274). Shiga, Japan: Kohrosha. (真嶋潤子「CEFRの国内外の日本語教育へのインパクト」泉水浩隆（編）『ことばを教える・ことばを学ぶ 複言語・複文化・ヨーロッパ言語共通参照枠（CEFR）と言語教育』行路社)

North, B. (2014). *The CEFR in Practice*. Cambridge, UK: Cambridge University Press.

Sakai, T., Kano, C., & Kobayashi N. (2015) TTBJ (Tsukuba-test-battery of Japanese). In J. Lee (Ed.), *The language test guide book for Japanese language education* (pp. 86-109). Tokyo, Japan: Kurosio Publishers. (酒井たか子・加納千恵子・小林典子「TTBJ (Tsukuba-test-battery of Japanese)」李在鎬（編）『日本語教育のための言語テストガイドブック』くろしお出版)

TTBJ (Tsukuba Test-Battery of Japanese). *The Center for Distance Learning of Japanese and Japanese issues, University of Tsukuba*. <<https://ttbj.cegloc.tsukuba.ac.jp/index.html>> (May 26, 2021)

Chinese Idioms: Stepping Into L2 Student's Shoes

Mateja PETROVČIČ

University of Ljubljana, Slovenia

mateja.petrovcic@ff.uni-lj.si

Abstract

Idiomatic expressions play an important role in communication. They are widely used in both spoken and written discourse and are often one of the indicators of native-like proficiency. This paper discusses Chinese idioms *chengyu* in foreign language education, presents some observations regarding the scope of idioms predicted for the advanced level of Chinese, and introduces conceptual gaps between L1 and L2 as possible reasons for learners' mistaken usages. This study argues that equipping students with skills and knowledge needed for reaching contextually appropriate use of idioms should be the main goal in the teaching/learning process and provides a set of practical suggestions as to which aspects of teaching idioms should be considered thoroughly.

Keywords: Chinese idioms; foreign language education; teaching and learning strategies; text corpora; language transfer

Povzetek

Idiomatski izrazi imajo v komunikaciji pomembno vlogo. Uporabljajo se v govornem in pisnem diskurzu, pogosto so tudi pokazatelj dobrega poznavanja jezika, ki meji na raven znanja naravnih govorcev. Članek obravnava kitajske pregovore *chengyu* pri poučevanju kitajščine kot tujega jezika in predstavi opažanja glede predvidenega izbora pregovorov, ki so glede na smernice poučevanja predvideni šele na najvišji jezikovni ravni, torej na ravni mojstrstva. Razprava o tem opozori na konceptualne vrzeli med L1 in L2 kot možnem izvoru nekaterih napak, ki jih tujci storimo. Članek poleg tega poudari pomen veščin in znanja, ki sta potrebna za doseganje kontekstualno ustrezne uporabe pregovorov, v sklepnih mislih pa ponudi še vrsto praktičnih predlogov o tem, čemu moramo pri poučevanju/učenju idiomatskih izrazov nameniti pozornost.

Ključne besede: kitajski pregovori; poučevanje tujih jezikov; strategije učenja in poučevanja; besedilni korpusi; jezikovni prenos



1 Introduction

Idiomatic expressions are a very fascinating part of a language. They make the language use more colorful, subtle, and precise. They are widely used in both spoken and written discourse and are considered as one of the hallmarks of native-like proficiency. Considering a wide range of situations in which idioms are used and the important role that they play in communication, there is little doubt that foreign language learners need to be familiar with their meaning and use (Vasiljevic, 2015).

The inclusion of idioms into the syllabus should therefore be an important part of teaching Chinese as a foreign language due to their constant presence in media, literary works, and daily conversations. As Wang and Luo claim (2021, p. 284), through learning idioms, students not only enrich their vocabulary, improve their language skills, and enhance the vividness of their language expression, but also better understand Chinese history and culture.

While in one's native language idioms are typically acquired through exposure, they are difficult nuts to crack in a foreign language for several reasons. As Vasiljevic (2015) notes, even if the learners recognize the figurative use of expressions, they often lack the knowledge and the skills to disambiguate the phrase meaning in the way the native speakers may do due to their limited linguistic proficiency and vocabulary size. Moreover, many idioms are also culturally embedded and not motivated only by their lexical components, but also by the specific cultural and historical context in which they originated (Boers et al., 2004).

Even though learners of Chinese usually come across idioms at the intermediate and advanced level, this aspect of language use still causes several problems. Learners often make mistakes because they misunderstand the structural, semantical, or syntactical features of idioms (Shi, 2008).

Several previous studies have discussed the types of mistakes foreigners make and the reasons thereof (Shi, 2008; D. Xiao, 2016; Yang, 2011), some of them examined textbooks and learner's dictionaries (Hong, 2012; S. Wang & Luo, 2021), or proposed improvements regarding teaching strategies and methods (Vasiljevic, 2015; Wang W. 王伟, 2016; D. Xiao, 2016; among others).

We agree with Vasiljevic (2015) that idiomatic language can and should be taught, whereby it is our duty as teachers to look for how the learning process can be optimized. Besides the published materials, online resources, and classroom idiom teaching activities, we hereby propose that teachers should equip students with the knowledge and skills of how to gain reliable information, encourage them to keep exploring language use to reinforce their language skills. The time available for contact hours is always limited, no matter how many classes there are scheduled in the syllabus.

To achieve this goal, we first need to understand how the present generations of students seek information, which resources they select, whether they evaluate their findings or not, and at which stage they get satisfied with answers. This study further points out the aspect which should be thoroughly considered at the advanced level of language teaching, namely the differences in the distribution of idioms across various domains.

Section 2 first reviews the terms related to idiomatic expressions in Chinese and closes with the decision to use English word *idiom* to denote Chinese *chengyu*. Section 3 focuses on idioms in foreign language education, points out some observations regarding the scope of idioms taught, and introduces some possible reasons for usage mistakes. Section 4 discusses the role of Chinese corpora and compares their strengths. Section 5 presents some major observations regarding the information-seeking process and students' attempts to use given idioms on their own. Section 6 briefly touches on the fact that occurrences of idioms vary across domains. The paper closes with Section 7 which sums up this study's findings.

2 Idiomatic expressions in Chinese

A common overarching term for idiomatic expressions in Chinese is *xiyu* 习语 or *shuyu* 熟语. In a broader sense, the term *shuyu* 熟语 is used to denote a category, and in a narrower one, it is used to refer to individual items in the category (X. Zhang, 2012, p. 20). Generally speaking, it serves as an umbrella term which includes various subcategories, i.e. common phrases *guanyongyu* 惯用语, idioms *chengyu* 成语, proverbs *yanyu* 谚语 and two-part allegorical sayings *xiehouyu* 歇后语. While some authors consider just these four groups as *shuyu* 熟语, the others include other subcategories, as well. Feng (2011) or Wang and Wang (2013), for example, also count in common sayings *suyu* 俗语, aphorisms *geyan* 格言, maxims *zhenyan* 箴言, epigrams *jingju* 警句, slang *liyu* 俚, and many other specific forms of expressions. In Xiao (2020), two of the categories are named according to the length of idiomatic expressions, namely 3-characters common expressions *sanziyu* 三字语 and 4-characters common expressions *siziyu* 四字语 as separate groups. Categories of idiomatic expressions are overlapping in many aspects, and it is impossible to draw a clear dividing line between them.

As far as the term *chengyu* 成语 is concerned, it can be understood in a broader or a narrower sense as well. In the broader sense, it is a category of set phrases that includes *suyu* 俗语 and *yanyu* 谚语, but in the narrower sense, it refers to four-character set phrases that can be traced back to written records in ancient fables, histories, and literary works (X. Zhang, 2012, p. 26). *Chengyu*, therefore, form a special part of the Chinese lexicon that has a fixed structure, conveys concise and comprehensive meanings, contains rich cultural connotations, and has profound

historical and cultural value. Nowadays, *chengyu* are still widely used in written and spoken Chinese and learners come across them at the intermediate and advanced level (Shi, 2008).¹

Most of the publications translate *chengyu* as *idioms* but other expressions are used for them as well. For example, Zhao (2015) uses the word *proverb* as the English equivalent of *chengyu*. On the other hand, even the English term *idiom* is not understood uniformly.

In the academic papers, the term *idiom* is understood in various ways, stretching from the traditional definition as a fixed multi-word phrase whose meaning cannot be predicted from the literal meanings of individual words that constitute this phrase, to more profound definitions from modern lexicographers, such as Glucksberg (2001), Grant and Bauer (2004), Moon (1998), and other linguists (X. Zhang, 2012). In this paper, we will not discuss the definitions and various interpretations of the terms *idiomatic expression*, *quadra-syllabic idiomatic expression*, *idiom*, *fixed expression*, *proverb*, and others in detail, but will use the English word *idiom* to denote Chinese *chengyu*, conventionally used set phrases, which are historically allusive in origin, often highly fixed in structure (the four-character mould), usually opaque in meaning and typically archaic in style (R. Xiao & Hu, 2016, p. 107).

3 Idioms in foreign language education

As Jiao et al. (2011) note, contextually appropriate use of idioms tends to impress hearers, especially if an aptly used idiom is coming from the mouth of a foreigner. Since idioms are frequently used in formal speech and higher-level written materials, they serve as a marker of proficiency level.

For a learner of Chinese as a foreign language, mastering idioms is very intriguing. This is mostly due to their meaning and figurative usage, but even their syntactical behavior is a hard nut to crack (D. Xiao, 2016).

As Xiao (2016) noted, idioms are an important part of Chinese language learning and not enough emphasis is put on them. For their complexity, idioms are usually taught no sooner than at the upper-intermediate or advanced level. This is also evident from the HSK 1-6 vocabulary lists, where almost all of the slightly more than one hundred idioms are predicted for level 6.² In other words, idioms are listed among the advanced words, ranking from #2500 to #5000. Furthermore, as provided in *Chinese*

¹ The author does not explicitly state what is meant by "intermediate and advanced level" (*zhong-gaoji jieduan* 中高级阶段), but since the results of this study are based on the data from the HSK Dynamic Composition Corpus of Beijing Language and Culture University, we assume that the approximate level of learners in question was around HSK 5 or HSK 6.

² See for example Idioms in New HSK Level 1-6. URL <http://jyangkul.net/HSK/HSK.html> (2020-04-15).

Proficiency Grading Standards for International Chinese Language Education (Center for Language Education and Cooperation, 2021), idioms are supposed to be learned even later, ranking from #5457 to #11092, which is a negatively surprising trend in standard setting.

Nevertheless, no matter how late the idioms are scheduled in the international language education, students of sinology or Chinese studies as their major should achieve the highest possible level of language proficiency during their studies, but due to the limited time and scope of the BA/MA programs, it is impossible to achieve an advanced level of language proficiency just within three/five years. Therefore, rather than omitting idioms from the curriculum or just integrating selected idioms into the courses, it is more appropriate to equip students with skills and knowledge that are needed to reach contextually appropriate use and necessary for their lifelong learning.

Zhang (1999) argues that certain syntactical errors are caused due to the lack of understanding of idioms' internal structure and could be clarified with a better understanding of their morphological features. Shi (2008) comes to similar observations and points out that foreigners make some mistakes because they are not aware that predicative idioms³ tend to be used in certain syntactical functions, whereas nominative idioms⁴ tend to be used in other syntactical functions. From this point of view, explaining idioms' structure and each character's meaning seems to have a positive impact on learners' understanding.

From the perspective of a non-native Chinese teacher, we need to mention another conceptual gap that arises when students encounter unknown idioms. Several studies have pointed to syntactical errors in general, regardless of the learners' L1 and problems that the language transfer might cause.

Although the intermediate learners should already have a sufficient understanding of Chinese notions of syntactical functions, the discrepancy in grammar and categories between L1 and L2 may lead to additional problems. For example, while the phrase *on the desk* in example (1) is considered a *subject* in Chinese, it is perceived as an *adverbial phrase of place* in Slovene.

- | | | |
|--------------|---------|------------|
| (1) 桌子上 | 放着 | 一本书。 |
| zhuōzi shàng | fàngzhe | yī běn shū |
| on the desk | lays | a book. |
- 'There is a book on the table.' (Liu et al., 2004, p. 452)

³ Chinese: *weicixing chengyu* 谓词性成语

⁴ Chinese: *tixing chengyu* 体词性成语

further look at the definition and top five examples provided by the Chinese-English online dictionary LINE Dict (<https://dict.naver.com/linedict/>), as one of the popular resources among students.⁵

(6) **HSK6** 得不偿失 débùchángshī

Definition: **not be worth the effort**

Usage examples: 得不偿失。

débùchángshī

The **gain is outweighed by the loss.**

勿做得得不偿失的事。

wù zuò débùchángshī de shì

Do not **pay too dear for your whistle.**

那些发现得不偿失。

nàxiē fāxiàn débùchángshī

Those discoveries **are not worth the candle.**

北半球秋季变暖得不偿失

běibànqiú qiūjì biànnuǎn débùchángshī

Losing more than we gain from autumn warming in the north.

寓意就是犯罪得不偿失。

yùyì jiùshì fànzùì débùchángshī

And the moral is that crime **doesn't pay.**

LINE Dict provides information that this idiom is classified as a HSK6 expression, with an English translation and several usage examples, whereby it does not mark idioms' syntactic functions. Readers are supposed to be able to understand and independently create the appropriate usage without explicit instructions.

Wang and Luo (2021) note that displaying example sentences with the targeted idiom is considered effective, whereby the selection of examples plays a demonstrative role in the usage of idioms. Compared to the printed materials, online resources have the advantage of having unlimited space for example sentences, and can therefore provide better insight into the language use.

Several authors have already stressed that teaching idioms should give priority to high-frequency idioms, as well as their high-frequency meanings and grammatical functions (Hong, 2012; Jiao et al., 2011; S. Wang & Luo, 2021; among others). To retrieve frequency data—not to mention other information—language corpora are indispensable tools. The following section, therefore, discusses Chinese corpora in relation to idioms.

⁵ Before 2014, this dictionary was known as Nciku.

4 Idioms in Chinese corpora

As far as idioms are concerned, not all Chinese corpora are equally appropriate to use. Moreover, to make full use of the materials and to ensure a successful application of corpora in classroom instruction, specialized training in corpus analysis for teachers and students would also be necessary first (Ištvánová, 2021, p. 132). The following subsections present the pros and cons of the BCC corpus, the CCL corpus, corpora in Sketch Engine, and the Hanku corpus.

4.1 Idioms in the BCC corpus

The BCC corpus (<http://bcc.blcu.edu.cn/>) is the ideal choice for the initial analysis of idioms because it has got a special tag just for them (Xun et al., 2016, p. 107). To get a list of the most frequently used idioms, we ran a simple query "i" in the balanced corpus domain and then sorted the results by frequency. The pop-up window provides information about how many different idioms there are in the selected section, and what are the absolute frequencies of each idiom, as shown in Figure 1.

共 4872 个结果			
下载		首页 上页 下页 末页	
实事求是	24078	莫名其妙	21615
艰苦奋斗	18177	不知不觉	18082
不可思议	15744	千方百计	14517
全心全意	14264	不由自主	13207
前所未有	12920	小心翼翼	11905
坚定不移	11757	理所当然	11065
不正之风	10887	无可奈何	10409
丰富多彩	9963	引人注目	9597
一模一样	9584	因地制宜	9569
一如既往	9545	众所周知	9283

Figure 1: List of the most frequent idioms in the multi-domain part of the BCC corpus with their absolute frequencies

After extracting the top 1000 idioms with the highest frequency, we compared this frequency list with the HSK idioms. The results in Figure 2 below show that idioms aimed at foreign language education do not fully overlap with the upper tenth of the most frequently used idioms but are rather distributed as follows. Among the top hundred idioms, 37 items are listed as HSK vocabulary, whereas 63 of them are not on part of the HSK wordlists. Among idioms ranked between #101 and #200, 29 are scheduled for the HSK level 6, whereas 71 of them are not considered relevant, and so on.

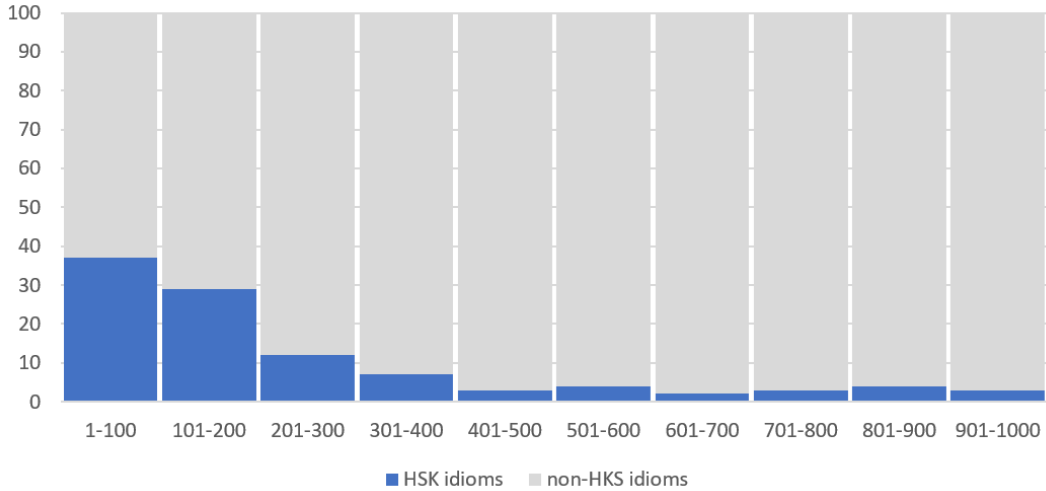


Figure 2: Distribution of HSK idioms among the top 1000 most frequently used idioms in the BCC corpus (grouping principle: 100 idioms per column)

An alternative option for idiom extraction would be the CCL corpus of Peking University (Zhan et al., 2019). In the CCL corpus, "i" stands for idioms as well, but the search is limited to the *moshi chaxun* 模式查询 (pattern query), which requires some specific context. For example, it is possible to search for a string of *zhen* 真 (really) + idiom "真(i=4)", but not for a list of idioms without any context.

Although these two corpora are very convenient for some basic information on Chinese idioms, their functions are limited to relatively simple queries and do not support the corpus query language to create flexible queries.

4.2 Idioms in SkE corpora

For more refined queries with corpus query language (CQL), Sketch Engine is a better choice. This tool returns cleaner and more precise results, but it is not ideal to extract idioms as such in the first place.

In the part-of-speech (POS) tagged corpora, there are usually no special tags for idioms, although it would be useful and convenient for Chinese since they are recognized as special fixed expressions and phraseological units. On the other hand, since idioms may be used in various syntactical functions, they are often treated as one token and are assigned the corresponding tags. For example, in corpora that are POS annotated by the Stanford Log-linear Part-Of-Speech Tagger using Chinese Penn Treebank, the majority of idioms are tagged as verbs, as shown in Table 1.

Table 1: Structure of the 1,000 idioms from BCC corpus according to their POS tags in Chinese Web 2017 (zhTenTen17) Simplified

Tag	Absolute frequency	Relative frequency	Proportion
VV	8,380,349	505.05	83.80%
NN	930,931	56.10	9.31%
AD	405,127	24.42	4.05%
VA	204,299	12.31	2.04%
CD	56,622	3.41	0.57%
Others (JJ, NR, NT, M, OD)	22,672	1.37	0.23%

Some idioms are consistently tagged with only one POS, for example, *qiánsuǒwèiyǒu* 前所未有 (hitherto unknown; such as never previously existed; unprecedented) as a verb; *sìmiànbāfāng* 四面八方 (from all sides; all around; all directions; all quarters) as a noun; *yǔcǐtóngshí* 与此同时 (at the same time; in the meantime; meanwhile) as an adjective, etc. Some idioms, on the other hand, have been assigned various POS tags. For example, the idiom *shíshìqiúshì* 实事求是 (seek the truth from facts; base on facts; be practical and realistic) is treated as a verb (VV), adverb (AD), predicative adjective (VA), or a noun (NN).

Even though 83.8% of all idioms are tagged as verbs, the query of four-character-verbs does not return only idioms. The proportion of noise increases if we expand the query to include the other tags from Table 1.

Apart from Chinese Web 2017, Sketch Engine offers several other Chinese corpora that are annotated with either Chinese Penn Treebank part-of-speech tagset, Chinese Symbols part-of-speech tagset or Chinese NEUCSP part-of-speech tagset.

Chinese Penn Treebank part-of-speech tagset,⁶ which is available in Chinese corpora annotated with Stanford taggers, does not have any special tag for idioms. This is also the case in Chinese Symbols part-of-speech tagset,⁷ which is available in Chinese corpora annotated with tagging tool developed by Institute of Information Science and CKIP group in Academia Sinica, for example, Chinese GigaWord 2 Corpus.

On the other hand, Chinese corpora annotated by the NEUCSP tagging tool developed by the Natural Language Processing Group at Northeastern University, China, use Chinese NEUCSP part-of-speech tagset,⁸ which is in many respects different

⁶ For more details, see <https://www.sketchengine.eu/chinese-penn-treebank-part-of-speech-tagset/> and Xia (2000).

⁷ For more details, see <https://www.sketchengine.eu/chinese-symbol-part-of-speech-tagset/> and (Chinese Knowledge Information Processing Group, 1993).

⁸ For more details, see <https://www.sketchengine.eu/chinese-neuscsp-part-of-speech-tagset/> and (Xiao et al. 2012).

from the former two. In this tagset, idioms are labeled as 'habitual language' (*xiyongyu* 习用语). This tag is also used for phrases, such as *méi shénme* 没什么 (it's nothing), *shì de* 是的 (that's right), *shìshí shàng* 事实上 (in fact), *méi guānxi* 没关系 (it doesn't matter; never mind), etc.

For the analysis of syntactical features of idioms, the Chinese Corpus Hanku could be used, as well, because it fully supports CQL (Gajdoš et al., 2016). Since it uses NoSketch Engine, i.e. an open-source version of the Sketch Engine corpus manager, it requires a more profound knowledge of the corpus query language to get the results that are otherwise available in the word sketch tool of Sketch Engine.

We thus suggest a combination of corpora to achieve optimal results. In our case this means the BCC corpus to extract idioms, then the Chinese Web 2017 for the syntactical analysis of the selected idioms, and the BCC corpus again for the part presented in Section 6 below.

5 Unveiling idiom's meaning and usage

We agree with the view that learning is not committing a set of facts to memory, and teachers' role is not only to transmit the required knowledge, but to guide students and give them the ability to use resources to find, evaluate, and apply information (Lujan & DiCarlo, 2006, p. 17). Developing life-long skills such as critical thinking, problem-solving, and communication is much more useful in the long term. The time available in the classes is always limited, no matter how many hours there are scheduled in the syllabus.

To achieve this goal, we first need to understand how the present generations of students are seeking information, which resources they select, whether they evaluate their findings or not, and at which stage they are satisfied with answers.

As part of their undergraduate studies, students of the sinology program at the University of Ljubljana are asked to retrieve information about given idioms and write a report describing their 'discovery path'. This is an open-ended survey conducted on a yearly basis that provides respondents with the opportunity to explore information retrieval in their personal way. Every student selects a package of five different idioms⁹ and is asked to do whatever they want to understand the meaning and usage of the given 'items'. In the concluding part of these mini-inquiries, students are asked to present a few examples that they consider useful, explain their decision, sketch a situation to be suitable for that idiom, and form a sentence with it.

⁹ The provided idioms are those with the highest frequency of occurrences, regardless of whether they are part of the HSK vocabulary or not.

The observations retrieved from these assessments provide valuable insights into the habits and preferences of current generations regarding information retrieval. This paper does not aim to discuss all the findings in detail, but rather highlights some noteworthy characteristics and common features.

First, although having easily determined that the ‘task items’ were idiomatic expressions and being fully aware that figurative meanings are not a simple sum of the meanings of the component words, all students devoted their time and energy to decompose idioms and tried to understand the meaning of individual characters. This supports the views expressed by Matlock and Heredia (2002) that non-experienced foreign language learners first establish direct connections between literal and nonliteral meanings of figurative expressions. Moreover, when processing an L2 idiom, L2 learners first attempt to make sense of it by translating it literally into L1.

Trying to understand the individual characters and relations among them seems to be a desirable thing in Chinese. Namely, the syntactical features of idioms are said to be closely related to the idiom’s internal structure. According to Shi (2008), predicative idioms tend to be used as predicates, attributives, adverbial adjuncts, and complements. On the other hand, nominative idioms would normally be used as subjects or objects.

Second, students needed fewer efforts for the so-called decomposable or compositional idioms, where the figurative meanings were relatively closely related to the literal meanings of their constituent words (e.g. *qiánsuǒwèiyǒu* 前所未有 unprecedented, lit. before-not.existing). The results have shown that even at the pre-intermediate level, students do not encounter great difficulties understanding them. Based on these observations, we claim that certain high-frequency compositional idioms may already be taught at the pre-intermediate level and do not need to be postponed to the advanced level.

Similarly, students devoted more time to non-decomposable or non-compositional idioms, probably since the meanings of individual characters seem to be unrelated to their compositional analysis (e.g. *tuōyǐng’érchū* 脱颖而出 escape.from-an.awn.of.wheat-in.such.manner-go.out; literal meaning refers to the fibers that often stick through a sack of wheat; fig. talent being fully exposed, talent revealing itself, distinguish oneself, rise above others). If they found a consistent equivalent in L1, they still managed to express the correct idea, whereas the resulting sentences often deviated from the expected usage. See example (9) below.

On the other hand, if the idiom resulted in a series of various translations, such as for example (6) above, mistakes were very frequent even if the idiom itself was a compositional one.

Third, the results have also shown that students did not give priority to the dictionaries/webpages with English translations when searching for usage examples. If

a translation was provided, they often re-checked it and compared it with other possible translations at various translation services. Surprisingly, approximately two-thirds of students seized AI-powered translation and dictionary tools, such as ReversoContext (<https://context.reverso.net/>) or Linguee (<https://www.linguee.com/>).

Forth, while focusing on the meaning of idioms themselves, students often neglected the provided context. This confirms the previous observations that learners often lack the skills to take advantage of contextual clues (Vasiljevic, 2015, p. 3). Although the selected examples shared some obvious features, students failed to notice them and to use them accordingly in their own production. For example, returning back to the idiom *tuōyǐng'érchū* 脱颖而出 'distinguish oneself', numerous examples point to the collocations with the subject *réncái* 人才 'talent', prepositional phrase *zài/cóng ... zhōng* 在/从.....中 'among ...', presence of modal verb *néng* 能 'be able to', etc., but none of these clues were applied in the created example.¹⁰ Speaking of collocates, none of the students came to the idea to make queries in language corpora. A follow-up discussion revealed that they were not even aware of language corpora in their L1.

Finally, the analysis has shown that idioms with similar syntactical behavior in L1 and L2 are easier to learn and use properly than those that differ in this respect. This aspect of idiom research has been ignored in scholarly papers.

For example, the idiom *zhòngsuǒzhōuzhī* 众所周知 is explained as *widely known; as everyone knows; as is known to all*, and it is usually used independently at the beginning of a sentence. As such, it is in our case similar to the corresponding expressions in students' native language.¹¹ It is therefore not very difficult to create well-formed sentences, as shown in examples (7)-(8) below.

(7) 众所周知，美国人口大部分居住于东半部。

Zhòngsuǒzhōuzhī, Měiguó rénkǒu dà bùfèn jūzhù yú dōng bàn bù.

'As everyone knows, the majority of the American population lives in the eastern half'
(A phrase used at the beginning of a sentence)

(8) 吸烟可能致癌是个众所周知的事实。

Xīyān kěnéng zhì'ái shì gè zhòngsuǒzhōuzhī de shìshí.

'It is a well-known fact that smoking may cause cancer.' (modifier of a noun)

¹⁰ It should be noted that there were two opposite approaches in forming sentences. The *safe-players* tended to make just minor adjustments to the existing examples, which resulted in correct use, while the *risk-takers* tried to be creative and applied idioms to their own contexts.

¹¹ Since there are no Chinese-Slovene dictionaries of idioms available, and students are used to using Chinese-English materials, we neglect the Slovene translation here. Nevertheless, some other idioms may have better Chinese-Slovene equivalents.

The reasons for such smooth application might be attributed to the compositional nature of this idiom, however, we claim that the influence of L1 grammatical features and discrepancy in grammatical categories play an important role. Even the examples with compositional idioms reflected this issue. For instance, sequences in examples (9) and (10) below show at which part of the ‘discovery path’ the usage deviated from expected.

- (9) Stage 1: Literal meaning of *tuōyǐng'érchū* 脱颖而出 is confusing but its figurative meaning sounds understandable:
talent revealing itself, to rise above others, to distinguish oneself
- Stage 2: mapping English equivalents to L1:
izkazati se, izstopati
- Stage 3: example sentences in Chinese can be perfectly translated to Slovene with the verb *izstopati*
- Stage 4: production of example in Slovene:
On vedno bolj izstopa. ‘He is increasingly rising above others.’
(perfectly acceptable)
- Stage 5: translation of *he is increasingly + rising above others* in Chinese:
他越来越 + 脱颖而出
tā yuèlái yuè + tuōyǐng'érchū
- Comments: We cannot claim that this sentence is wrong, because such instances are available in corpora but are extremely rare (relative frequency less than 0.01). The reason is that the expression *yuelaiyue* normally requires an adjective and not a verb(al phrase).¹²
- (10) Stage 1: *xìngzhìbóbó* 兴致勃勃 means:
high-spirited (adjective) (source: <https://en.bab.la/>)
alternative information: full of zest
- Stage 2: mapping English equivalents to L1:
poln žara, srečen, dobro razpoložen
- Stage 3: example sentences in Chinese can be perfectly translated to Slovene with these adjectival phrases
- Stage 4: production of example in Slovene:
Mi smo vsi srečni/polni žara, ker bo kmalu konec leta.
‘We’ll finish classes soon, so we’re full of zest.’

¹² Collocation details reveal that *yuelaiyue+VA* appears with RF 132.55, whereas *yuelaiyue+VV* appears with RF 24.76.

Stage 5: translation of *we + full of zest* in Chinese with the following explanation: this idiom is used as an adjective. When adjectives are used as predicates, they usually require *hen* (N+*hen*+Adj).
快要停课了，我们 + 很 + 兴致勃勃。
Kuàiyào tíngkè le, wǒmen + hěn + xìngzhìbóbó.

Comments: This student was not mistaken in her conclusions that adjectives can function as predicates. Moreover, based on the explicit information from the selected online dictionary and informal confirmation in L1 that this *is* an adjectival expression, her way of thinking was correct. However, placing more focus on the idiom's internal structure and near context would raise some doubts in the suggested example.

As the next step, idioms that are explained in the form of a longer phrase or sentence, are more difficult to be further used in a specific syntactical function. For example, consider the idiom *yìngyùn'éershēng* 应运而生 with the explanation *to arise or emerge as times demand; to arise according to objective demands; emerge as the times require; arise at an opportune time*. While forming sentences with this idiom as a predicate is still feasible due to the verb *arise/emerge* as the anchoring headword, it is more difficult to use it as a modifier of a noun. Consider examples (11)–(12) below.

(11) 随着网络的出现，网络语言也应运而生。

Suízhe wǎngluò de chūxiàn, wǎngluò yǔyán yě **yìngyùn'éershēng**.

'With the advent of the Internet, network language also **emerged as the times require**. >> With the advent of the Internet, network language also **came into being**.' (predicate)

(12) 慕课作为网络自主时代里应运而生的产物，是一种新型的教学资源。

Mùkè zuòwéi wǎngluò zìzhǔ shídài lǐ **yìngyùn'éershēng** de chǎnwù, shì yī zhǒng xīnxíng de jiàoxué zīyuán.

'As a **new product** (that emerged as the times require) in the era of network autonomy, MOOCs are a new type of teaching resource.' (modifier of a noun)

The most difficult idioms to grasp have proven to belong to the group of idioms that are culturally embedded. These are idioms whose meanings are not motivated only by their lexical components but also by the specific cultural and historical context in which they originated (Boers, Demecheleer & Eyckmans 2004 in Vasiljevic, 2015). For example, consider the explanation and user guidelines for the idiom *fēngshēng-hèlì* 风声鹤唳 in *A Chinese-English Dictionary of Chinese Idioms* (CEDCI).

In 383, Fu Jian's army of 800,000 strong were badly beaten by the troops of the Eastern Jin, totaling 100,000. On their way back home, the fleeing army mistook the sound of the wind and the cry of cranes as the coming of the pursuing Eastern Jin's army. (Met) To be full of suspicion and fear. Derogatory: mainly describes someone who is extremely frightened and scared. Used as 1) predicate; 2) complement (often cooperates with verbs like 闹 and 搞); 3) attributive (often modifies 程度, 地步 and 情况 as well as nouns referring to people); 4) adverbial adjunct; and 5) object (often used in cooperation with the predicate verbs 感到 and 听). May be used together with 草木皆兵 (Pan, 2002, p. 219).

Examples representing such syntactical functions are:

- (13) 中国队的攻势非常猛烈，阿根廷队门前**风声鹤唳**。

Zhōngguó duì de gōngshì fēicháng měngliè, Āgēntíng duì mén qián **fēngshēnghèlì**.

'The Chinese team's offensive was very fierce, and Argentina in front of the goal **was frightened**.'

(predicate of subject Argentina)

- (14) 整个东北被他弄得**风声鹤唳**。

Zhěnggè dōngběi bèi tā nòng dé **fēngshēnghèlì**.

'The whole northeast was **shaken** by him.'

(complement of the verb *to make*, lit. *has been made shaken*)

- (15) 资本市场有点**风声鹤唳**的感觉，股市、债市都在跌。

Zìběn shìchǎng yǒudiǎn **fēngshēnghèlì** de gǎnjué, gǔshì, zhài shì dōu zài diē.

'The capital market feels a bit **jittery**. The stock market and bond market are both falling.'

(attributive/modifier of a noun *feeling*)

- (16) 如果你总是**风声鹤唳**地看过去，生活只好让你四面楚歌。

Rúguǒ nǐ zǒngshì **fēngshēnghèlì** de kàn guòqù, shēnghuó zhǐhǎo ràng nǐ sìmiànchǔgē.

'If you always look **jittery** at the past, life will treat you surrounded by enemies on all sides.'

(adverbial adjunct to the predicate *to look*)

- (17) 犯了罪的人对每一个声音都感到**风声鹤唳**。

Fànle zuì de rén duì měi yīgè shēngyīn dōu gǎndào **fēngshēnghèlì**.

'Those who have committed a crime feel very **nervous** about every voice.'

(object of verb *to feel*)

Needless to say, for culturally embedded idioms, it is very difficult to determine appropriate contexts, even though the dictionary tells us that this idiom mainly describes someone who is extremely frightened. As observed from examples (13)-(17) retrieved from corpora, this expression can describe *fearful atmosphere* in sports, market economy, among a certain population, etc. Without detailed queries L2 learners cannot understand and use such idioms properly.

At this point, we must remind ourselves that contemporary generations first seize for online resources that return results in no time.¹³ Therefore, providing guidelines of how to refine their searches would be of great importance for their life-long learning. On the other hand, students should also be aware that specialized materials such as dictionaries of Chinese idioms capture the essence of idioms and are important supplementary reference works.

We agree with Wang and Luo (2021) that it is still a challenge for Chinese learners and teachers to answer the questions of how to effectively master the meaning and usage of idioms, however, we disagree with their opinion that there is a lack of idiom learning dictionaries for CFL learners (S. Wang & Luo, 2021, p. 285). Although compiled and published back in 2002, Pan's *A Chinese-English Dictionary of Chinese Idioms* is still a relatively good publication for learning Chinese as a foreign language. This book represents idioms relatively accurate, explains their literal and figurative meanings, which are accompanied with short but rich syntactical information, followed by up to four examples of use in various syntactical functions. In our supplementary research, we have evaluated the reliability of provided information and given examples concerning frequency of occurrences in Chinese corpora. The results were quite satisfying. Another reference work worth advising is Jiao, Kubler and Zhang's *500 Common Chinese Idioms: An Annotated Frequency Dictionary* (Jiao et al., 2011). This book is grounded in corpus research, focused on learner usage, accompanied by not numerous yet well chosen example sentences, explanations, and usage notes.

A selection of various online and printed materials in combination with the corpora would therefore be an invaluable aid to both comprehension and production.

6 Further notes on idioms across various domains

As Feng (2011) states, idioms often appear in many types of texts, from literary texts, political texts, or scientific texts, to people's daily conversations.

Therefore, we conducted further follow-up queries across various domains in the BCC corpus. Apart from the multi-domain (*duōlǐngyù* 多领域), which was designed as

¹³ One respondent in the generation 2020/2021 frankly admitted that her first intention was to dig into a comprehensive dictionary in her private library, however, she lost her nerves because of the slow pace of progress.

a balanced corpus, the BCC corpus is further composed of the following four sub-corpora of contemporary Chinese: literature (*wénxué* 文学), press (*bàokān* 报刊), Weibo (*Wēibó* 微博) and science and technology (*kē-jì* 科技). Weibo represents colloquial Chinese, whereas the section on science and technology reflects the academic writing style. All five parts are independent and do not overlap.

The analysis of idiom distribution over the five domains has shown notable differences across the domains. As demonstrated in Table 2, the most prominent difference exists between literary and academic texts. There is just a 37.5% overlap in the list of 1,000 the most frequent idioms per field. In other words, 62.5% of idioms that are frequently used in literary texts are not among the top thousand idioms in academic texts and vice versa. Table 2 below further demonstrates the ratio of idioms that are used in all registers compared.

Table 2: Ratio of common idioms across fields (range: top 1,000 most frequent idioms per field)

	Literature	Press	Colloquial Chinese
Press	401 (40.1%)	-	-
Colloquial Chinese	518 (51.8%)	449 (44.9%)	-
Academic Chinese	375 (37.5%)	676 (67.6%)	455 (45.5%)

For example, the following idioms are expected to be found in the literary texts, but not very likely in the academic papers: *yīběn-zhèngjīng* 一本正经 (be poker-faced; humorless; in all seriousness), *yīyán-bùfā* 一言不发 (not to utter a single word; be as silent as the grave), *xīnbúzàiyān* 心不在焉 (be absent-minded; absence of mind), *xīnmǎn-yìzú* 心满意足 (to one's heart's content; be fully satisfied and content) and *luànqī-bāzāo* 乱七八糟 (be out of order; all a hideous mess; be in wild disorder), among others. Vice versa, the following idioms are very likely to appear in the academic papers, but not in the literary works: *xíngzhīyǒuxiào* 行之有效 (effective), *jiānkǔfèndòu* 艰苦奋斗 (work hard and perseveringly), *yīndìzhìyí* 因地制宜 (suit one's measures to local conditions), *xiāngfǔxiāngchéng* 相辅相成 (be inseparably interconnected) and *tuōyǐng'érchū* 脱颖而出 (to stand out), among others.

We cannot claim that certain idioms appear in just one kind of text, but there is a stronger or weaker tendency to use them in various domains, as further indicated in Table 3.

Table 3: Number of occurrences in literary texts versus academic texts

	idiom	Literature	Academic Chinese
Idioms frequently used in literature, but not among the first 1,000 most frequent idioms in the academic texts.	一本正经	1,386	401
	一言不发	1,339	304
	心不在焉	1,215	364
	心满意足	1,209	491
	乱七八糟	1,204	544
Idioms frequently used in academic texts, but not among the first 1,000 most frequent idioms in the literary texts	行之有效	57	15,732
	艰苦奋斗	36	15,202
	因地制宜	16	12,965
	相辅相成	56	9,876
	脱颖而出	93	8,495

Certain idioms are therefore more typical for some domains than others. For example, data from BCC have shown that the idiom *jiānkǔfèndòu* 艰苦奋斗 (to struggle arduously; work hard and perseveringly) appeared 21,840 times per billion words in the texts from the press, but only 12 times in the literary texts. Similarly, *zhòngsuǒzhōuzhī* 众所周知 (as everyone knows) was used 6,473 times in the academic texts, but only 136 times in the literary texts. The idiom *luànqībāzāo* 乱七八糟 (be out of order; all a hideous mess) tends to be used the most in spoken language with 5,734 occurrences and the least in the academic texts with only 181 occurrences per billion words.

Therefore, when providing information about language use and instructing upper-intermediate to advanced learners what to pay attention to, we suggest it is worth considering domains, as well.

7 Conclusions

Idiomatic expressions play an important role in communication. They are widely used in both spoken and written discourse and are often one of the hallmarks of native-like proficiency. Considering the role of idioms and their vivid use, it is obvious that they should have a special place in second language acquisition, as well.

The time available for studies is always limited, therefore it is important to make the best use of it. For a learner of Chinese as a foreign language, mastering idioms is very intriguing, and according to HSK, they are usually taught after reaching upper-intermediate or advanced level. Our study has shown that there is no need to associate them with insurmountable difficulties because even at the pre-intermediate level, students are capable of proper perception and partly accurate production.

From the above, we suggest familiarizing students with idioms no later than at the intermediate level, whereby it is important to keep an eye on the main goal. That is to

say that teachers should try to equip students with the skills and knowledge needed for proper language use and independent lifelong learning. Needless to say, introducing idioms just as amazing and amusing stories (*chengyu gushi* 成语故事) is not enough to understand how these idiomatic expressions are used in the context.

As several previous studies have already pointed out, idiom teaching should pay attention to several aspects, from the role of frequency in either idiom selection, the hierarchy of meanings, grammatical functions or collocations, to the importance of their internal structure. Besides that, Hong (2012) emphasized the importance of detailed definitions and exemplifications with typical use scenarios in the textbooks. We believe that online resources should pay attention to this feature, as well, and list the good dictionary examples with extended contexts as their top results. Online resources are, unlike printed materials, unlimited in space and therefore more flexible for adding numerous additional examples. AI-powered translation and dictionary tools extract the passages from real-life contexts and cover a wide range of registers of speech. A higher degree of exposure to various examples is beneficial for developing learners' *yuyangan* (language feel/intuitive competence), and thanks to user-friendly display options, there is no need to limit results to a set of foreigners' tailored vocabulary. This, however, does not demote the importance of printed reference works.

Apart from obtaining information from the above-mentioned resources, teachers should foster students to verify their assumptions and refine their new insights in corpora, as well. As our survey has shown, this aspect of information retrieval is still very weak. As Ištvánová (2021, p. 132) noted, the direct use of corpora enhances the process of second language acquisition, however, specialized training in corpus analysis for teachers and students is necessary to ensure its successful application.

Despite all the advances of modern technologies, learners should raise their awareness of potential conceptual gaps between L1 and L2 that might lead to usage mistakes. These pitfalls are inherently related to the specifics of L1, therefore more local studies are needed to examine learning difficulties in relation to different L1 backgrounds. Moreover, getting acquainted with Chinese through English first might cause additional misunderstandings because of this additional step.

Acknowledgments

The research for this paper was supported by the Slovenian Research Agency (ARRS) in the framework of the research core funding *Asian Languages and Cultures* (P6-0243).

References

- Boers, F., Demecheleer, M., & Eyckmans, J. (2004). Etymological elaboration as a strategy for learning figurative idioms. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing* (Vol. 10, pp. 53–78). John Benjamins Publishing Company.
- Center for Language Education and Cooperation. (2021). *Chinese Proficiency Grading Standards for International Chinese Language Education*.
<https://www.blcup.com/EnPInfo/index/10865#001>
- Chinese Knowledge Information Processing Group. (1993). *Technical Report 93-05: Chinese Part-of-speech Analysis*. Academia Sinica.
https://ckip.iis.sinica.edu.tw/CKIP/tr/9305_2013%20revision.pdf
- Feng, C. 封晨. (2011). Hanyu chengyu de leixing ji keyixing yanjiu 汉语成语的类型及可译性研究. *Heihe xuekan 黑河学刊*, 08, 30–31.
- Gajdoš, L., Garabík, R., & Benická, J. (2016). The New Chinese Webcorpus Hanku – Origin, Parameters, Usage. *Studia Orientalia Slovaca*, 15(1), 21–33.
- Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphor to Idioms*. Oxford University Press.
- Grant, L., & Bauer, L. (2004). Criteria for Re-defining Idioms: Are we Barking up the Wrong Tree? *Applied Linguistics*, 25, 38–61. <https://doi.org/10.1093/applin/25.1.38>
- Hong, B. 洪波. (2012). Duiwai Hanyu chengyu jiaocai bianxie de ji ge wenti 对外汉语成语教材编写的几个问题 On the Writing of Textbooks of Chinese Idioms for Foreign Learners. *Journal of Honghe University*, 10(3), 81–84.
- Ištvánová, M. (2021). On the Use of Corpora in Second Language Acquisition – Chinese as an Example. *Acta Linguistica Asiatica*, 11(2), 131–142. <https://doi.org/10.4312/ala.11.2.131-142>
- Jiao, L., Kubler, C. C., & Zhang, W. (2011). *500 Common Chinese Idioms: An Annotated Frequency Dictionary* (1st edition). Routledge.
- Liu Y. 刘月华, Pan W. 潘文娱, & Gu W. 故韡. (2004). *Shiyong xiandai Hanyu yufa 实用现代汉语语法*. Shangwu yinshuguan 商务印书馆.
- Lujan, H. L., & DiCarlo, S. E. (2006). Too much teaching, not enough learning: What is the solution? *Advances in Physiology Education*, 30, 17–22.
<https://doi.org/10.1152/advan.00061.2005>
- Matlock, T., & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In *Bilingual sentence processing* (pp. 251–274). North-Holland/Elsevier Science Publishers. [https://doi.org/10.1016/S0166-4115\(02\)80014-0](https://doi.org/10.1016/S0166-4115(02)80014-0)
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Clarendon Press.
- Pan, W. 潘维桂. (2002). *Han-Ying Hanyu chengyu yongfa cidian 汉英汉语成语用法词典 A Chinese-English Dictionary of Chinese Idioms*. Sinolingua.
- Shi, L. 石琳. (2008). Liuxuesheng shiyong hanyu chengyu de pianwu fenxi ji jiaoxue celüe 留学生使用汉语成语的偏误分析及教学策略 (Error analysis of foreign students' chengyu use and teaching strategies). *Xinan Minzu Daxue Xuebao (Renwen Sheke Ban) 西南民族大学学报(人文社科版)*, 06, 280–283.

- Vasiljevic, Z. (2015). Teaching and Learning Idioms in L2: From Theory to Practice. *MEXTESOL Journal*, 39(4), 1–24.
- Wang, L., & Wang, S. (2013). A Study of Idiom Translation Strategies between English and Chinese. *Theory and Practice in Language Studies*, 3(9), 1691–1697. <https://doi.org/10.4304/tpls.3.9.1691-1697>
- Wang, S., & Luo, H. (2021). Exploring the meanings and grammatical functions of idioms in teaching Chinese as a second language. *International Journal of Applied Linguistics*, 31(2), 283–300. <https://doi.org/10.1111/ijal.12334>
- Wang W. 王伟. (2016). Brief Analysis of Teaching Ways of Chinese Idioms in Teaching Chinese as a Foreign Language. *Modern Linguistics*, 4(1), 1–7. <https://doi.org/10.12677/ML.2016.41001>
- Xia, F. (2000). The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0). *IRCS Technical Reports Series*, 38. https://repository.upenn.edu/ircs_reports/38
- Xiao, B. 肖冰. (2020). *Chengyu, yanyu, suyu, guanyongyu, xiehouyu, sanzhiyu, sizhiyu 成语, 谚语, 俗语, 惯用语, 歇后语, 三字语, 四字语 (Idioms, proverbs, common sayings, common phrases, allegorical sayings, common expressions (3 characters), common expressions (4 characters))*. Xin Ya chubanshe.
- Xiao, D. (2016). Errors Made by Foreign Students while Using Chinese Idioms and Corresponding Teaching Strategies. *Journal of Mianyang Teachers' College*, 35(09), 10-13+18.
- Xiao, R., & Hu, X. (2016). *Corpus-Based Studies of Translational Chinese in English-Chinese Translation*. Springer Berlin Heidelberg.
- Xun, E. 荀恩东, Rao, G. 饶高琦, Xiao, X. 肖晓悦, & Zang, J. 臧娇娇. (2016). Da shuju beijing xia BCC yuliaoku de yanzhi 大数据背景下 BCC 语料库的研制. *Yuliaoku Yuyanxue 语料库语言学*, 3(1), 93–118.
- Yang, Y. 杨玉玲. (2011). Liuxuesheng chengyu pianwu ji “Liuxuesheng duo gongneng chengyu cidian” de bianxie 留学生成语偏误及《留学生多功能成语词典》的编写. *Lexicographical Studies*, 1, 101–109. <https://doi.org/10.16134/j.cnki.cn31-1997/g2.2011.01.013>
- Zhan, W. 詹卫东, Guo, R. 郭锐, Chang, B. 常宝宝, Chen, Y. 谌贻荣, & Chen, L. 陈龙. (2019). Beijing daxue CCL yuliaoku de yanzhi 北京大学 CCL 语料库的研制 (The building of the CCL corpus: Its design and implementation). *Corpus Linguistics 语料库语言学*, 6(1), 71–86.
- Zhang, X. (2012). *Chengyu as Cultural Performances: Insights into Designing Pedagogical Materials for Four-character Chinese Idioms* [Master of Arts, Ohio State University, East Asian Languages and Literatures]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1339698374
- Zhang, Y. 张永芳. (1999). Waiguo liuxuesheng shiyong Hanyu chengyu de pianwu fenxi 外国留学生使用汉语成语的偏误分析 (How do foreign students use Chinese idioms – An error analysis). *Yuyan wenzi yingyong 语言文字应用*, 03, 25–30. <https://doi.org/10.16499/j.cnki.1003-5397.1999.03.004>
- Zhao, Z. (2015). *Stories of Chinese Proverbs: The Old Man Who Moved Mountains 中国成语故事: 愚公移山*. <https://www.goodreads.com/book/show/29585183-stories-of-chinese-proverbs>

Coroneologisms and Word Formation Processes in Hindi-English Codemixed Words

Md. Tauseef QAMAR

St. Joseph's College, Bengaluru,
Karnataka, India
tauseef.qamar@sjc.ac.in

Juhi YASMEEN

Jain (Deemed-to-be) University, Bengaluru,
Karnataka, India
juhi1421@gmail.com

Md. Arfeen ZEESHAN

Aligarh Muslim University, Aligarh,
Uttar Pradesh, India
mdarfeen65@gmail.com

Sanket PATHAK

Mihup Communications Pvt. Ltd., Kolkata,
West Bengal, India
sanketlck@gmail.com

Abstract

The COVID-19 pandemic came with a flux of new words, terminologies, and phrases, which led to the rapid coinage or neologisms in the world's different languages. These lexical innovations may take place within one language as well as with the combination of two different languages. Therefore, this paper scrutinizes coroneologisms and word-formation processes in Hindi-English code-mixed words. Such a phenomenon happened due to the acceptance of English by Indians besides their mother tongue which makes them bilingual. The data were gathered from newspapers, blogs, social media, TV news, etc. Next, the linguistic analysis of the data revealed different types of word classes in Hindi-English codemixed words such as compounding, affixation, blending, and reduplication. Out of these, compounding and borrowing were reported as the most productive types of coroneologisms in Hindi-English code-mixed words.

Keywords: COVID-19 terms; coroneologisms; Hindi-English codemixed words; word-formation processes

Povzetek

Pandemija COVID-19 je prinesla številne nove besede, terminološke izraze in besedne zveze, kar je privedlo do hitrega nastanka neologizmov v različnih jezikih sveta. Leksikalne inovacije se lahko zgodijo v enem jeziku kot tudi pri kombinaciji dveh različnih jezikov. Prispevek prouči koroneologizme in besedotvorne procese v hindujsko-angleških besedah in ugotavlja, da je do omenjenih procesov prišlo zaradi sprejemanja angleščine kot skorajšnjega maternega jezika, s čimer so govorniki postali dvojezični. Študija na podlagi podatkov iz časopisov, blogov, družbenih medijev, televizijskih novic in in drugih medijev analizira procese kot so združevanje, afiksacija, mešanje in reduplikacija, med katerimi sta se v primeru hindujsko-angleških koroneologizmov združevanje in izposoja izkazala kot najbolj produktivna.

Ključne besede: termini o COVID-19; koroneologizmi; hindujsko-angleške besede; besedotvorni procesi



1 Introduction and background

There exists a variety of literature on English word-formation processes that offers an insight into the ways English vocabularies have significantly expanded through numerous types of neologisms and over numerous domains like advertisement, internet, and mass media (Kathpalia, 2018). In this regard, we have focused particularly on code-mixed coroneologisms that are created out of the words of two linguistically different languages, i.e. Hindi and English, with special reference to COVID-19 associate terms. Such type of Hindi-English code-mixing is casually viewed as 'Hinglish' (Bhatia, 2011, p. 44), which is not merely a coincidence of English borrowing to bridge the lexical lacuna but rather a highly complex process that typical differentiates itself from the usual borrowing and getting immense popularity especially in the advertising and social media languages. In their study, Bhatia and Ritchie (2006a, p. 518) claimed that this type of codemixing should not be viewed as a corrupt form of language used by bilinguals but "as a systematic and rule-governed phenomenon which satisfies the creative needs of bilinguals", of which creativity is impossible to achieve within the boundaries of a single language. As a result of overwhelming response towards such a use of language, i.e. codemixing "English usage in day-to-day interaction, advertising and media have achieved a distinct state of fusion and hybridisation of linguistic forms, which is unprecedented in the history of human communication" (Bhatia & Ritchie, 2008, p. 11). Although codemixing operates at the discourse level, our focus in the current paper remains on bilingual words internal structure, whereby the words or their parts both from Hindi and English are fused to produce the desired level of creativity to the target audience by the content writers, news editors, and social media millennials.

Bilinguals have access to two linguistic repositories as compared to monolinguals who only use one. In Bhatt's (2008) view, to produce and interpret such combined expressions, an individual needs to have both "bilingual and bicultural competence". For instance, to produce and understand the Hindi-English blend expressions like crickshetra (cricket+ Kuruksheetra 'conquest of the Kauravas and Pandavas'), similarly agonypariksha (agony + agnipariksha 'surviving the ordeal'), one needs to have an acquaintance of both the Hindi word creation processes and the cultural nuances of Hinduism epic 'Mahabharata and Ramayana' (refer to Kathpalia & Ong, 2015). Of particular relevance here are the shadowing boundaries between English and Hindi; the boundaries of English and Hindi in the former example is easily capturable. Still, in the latter instance, the boundaries are more translucent and appear to be the case of phonetic overlapping between the English word 'agony' and the Hindi word 'agni' (fire). Recent studies (Li, 2011, p. 1223; García & Li, 2012, p. 24) show that to understand such type of linguistic mixing that leads to code-mixing, one must have a fair understanding of 'translanguaging practices' in human language because this not only stretches the boundaries between language and culture but also promotes linguistic creativity and

criticality. In this case, 'linguaging' refers to the strategy used by an individual to communicate creatively in a given context, while 'trans' means the way one tries to violate the traditionally drawn boundaries between languages (García & Li, 2014). Flexible bilingualism in this context allows us to violate grammar principles and mix two languages to create new and contextually suitable meanings.

Such a perspective encircles both creativity and criticality, and will in this paper allow us to see code-mixing through the lens of rule-governed aspect and word-formation processes fluidity perspective.

2 Literature review

Several studies have examined the language play in the English language, including several of its dialects (Crystal, 1998; Cook, 2000), however, relatively a few have focused on a language play from the multilingualism perspective. Despite the widespread use of code-mixing in a globalised era, the literature shows a dearth of such studies which have explored the relationships between code-mixing and multilingual idiosyncrasy, especially in the context of world Englishes (Bolton, 2010). This becomes more pertinent in the context of interlingual creativity and word-formation process level, which have remained largely unexplored, especially in the world's different languages (Kathpalia, 2018). Zhang (2015) is one of the exceptions in the context of non-native English stretched zone, i.e. Chinese-English code-mixed multilingual play. An official microblog was established by the municipal corporation of Shanghai to broadcast weather information. In his work, the particular interest remains on morphological innovation and the use of "X+ing" in code-switched words (viz: zhaomuing 'recruiting'), and the interlingual play in hybridised expressions created out of English weekdays (viz: mangday 'Monday') that showcase the prevailing trend of code-mixing in Chinese entertainment and social media domains, especially among multilingual communities. In addition, Lin (2011) conducted a study on multilingual creativity in Chinese-English code-mixed words among the students of China in London, others were by (Yon, 2011) on Chinese immigrants using an online forum, and Zhang (2012) examined Chinese netizens in several domains.

Another notable work was done from a stretched zone perspective, which focuses on monolingualism and multilingualism lexical innovations in the Italian linguistic landscape scenario (Vettorel & Franceschi, 2013). With a special focus on the hybridisation processes that operate at the word-creation plane that is viewed as an emblem of modernity, style, and quality in stretched zone perspectives. The hybridised data of Italian-English show creative fluidity at several linguistic levels from orthography, phonology, morphology, grammar, syntax, semantics, semantics as well as idiosyncratic constructions. These hybridised Italian English words were grouped into different word creation classes based on the derivation (viz: kissucci = kiss + ucci

'endearment's suffix'), clipping (viz: light instead of 'lighthouse'), blending (viz: ristopub = ristorante + pub), compounding (viz: isibike = isi, a shortened form of a name in Italian culture Isidoro and it is pronounced in Italian as easy + bike 'easy-bike'), and lastly idiosyncratic constructions (viz: gadget in place of 'gadget'). Authors have claimed that such hybridisation with English is relatively more frequent in bigger cities where it is used as a persuasive tool to attract customers, especially to cater to the needs of globalisation. The urban population show relatively more positive attitudes towards the hybridisation of Italian with English, especially in the advertising domain, as they relate it with the symbol of modernism, class, and prestige. Despite having positive attitudes towards such linguistic hybridity, it is viewed as less formal in several domains concerning the linguistic landscape in Italy. A similar view exists towards the use of Spanglish among the youths of Spain concerning Internet domains where they often use code-mixed words of Spanish-English. In her study, Balteiro (2012) noted that the English verb *own* has gained a special meaning and is used by Spanish when they defeat their opponent in online games. Its use has been advanced and used in several other domains in Spanish to form different hybridized words according to the given context (viz: *owneado*, *owneador*, *owneamiento*, *owneo*, and others). Moreover, several other interesting works have been done on code-switching in the European contexts (Kelly-Holmes, 2000; Schlick, 2002, 2003), in French language (Martin, 2002, 2008), and similarly in the Italian language (Pulcini, 1995; Griffin, 2004; Coluzzi, 2009; Furiassi, 2010; Vettorel, 2013) but none of them have focused primarily on lexical derivations and innovation standpoints.

Further, studies on Russian were also conducted from the stretched zone perspective. Some of them have reported the perceptual shift towards the Englishization or Romanisation of Russian from the 80s and its current traces though the effect is estimated as mild (Rivlina, 2015). The main cause for its retraction is the limited span of contact between English and Russian, fragile relations with the native English-speaking countries, and the negative attitudes among its citizens due to the linguistic resistance against the expansion of English and foreign script in these lands. As a result, the practice concerning English-Russian bilingualism have not flourished unlike others (e.g. Spanis-English, Bulgarian-English, etc.) and practised in a limited domain only like popular literature, memorable messages, media, and advertising signboards (Rivlina, 2015). Because of this, the creative experiment of English remained confined to the intra-sentential level only. Few examples are available in writing hybridisation where English writing is fused with Cyrillic writing (viz: TERRITORIA refers to 'territory'). In the current outset of Russia, English hybridisation with Russian remains very restricted though slightly gaining popularity, especially in the field of advertising. In this context, several interesting works have focused on the Englishization of Russian among which special reference to English borrowings can be found in Maximova (2002), Rivlina (2005), Eddy (2007), Yelenevskaya (2008), and Proshina (2010).

Several other studies were also conducted to examine English code-mixing in multilingual advertising signboards. Lamarre (2014, p. 132), for example, examined the linguistic landscape practices in Montreal and viewed them as “bilingual winks”, a technique of mixing French and English words to surpass the language of legislative regulating authority from the public and commercial signboards. As a product of such covert bilingual recombination that resulted in the creation of shop signboards containing ‘Chouchou’ used to refer to a shoe shop (means ‘sweetiepie’ in French language but articulated as ‘shoe-shoe’). On the contrary, a shift in the blending trend has been noticed where English appears to be more prominent in words like ‘Paw-tisserie’ = paw + patisserie, which refers to pet food shop. Such a deviation from the usual language norms makes a wider scope for English to enter into French culture overtly, is considered the creative showcasing of language dynamics in Montreal, announcing freedom from the imposed language rules on a bilingual culture. Moreover, this could lead to more vigorous recombination of linguistic hybridity, specifically for customer persuasion perspectives in the domain of advertising.

In opposition to the English stretching zone countries, the expansion and the use of English is often noticed in everyday communication in different domains, especially in non-English speaking regions of Africa and Asia. In the Indian context, where bilingual practices are seen as a natural phenomenon, English has sneaked into several Indian languages that “led to an unmarked pattern of widespread naturalistic coalescence”. This led to both inter and intra-sentential codemixing in addition to phonological and morphological recombinations (Bhatia & Ritchie, 2006b, p. 795). Of remarkable attention is linguistic hybridity or code-mixing of Hindi and English because the two languages “symbolise economic power, social mobility, and wider communicative access to the speakers of other Indian languages” (Bhatia & Ritchie, 200b, p. 796). As a result, code-mixing between these two languages appear to be more productive and frequent across several domains than any other language of the Indian continent.

Moreover, a rich literature is available on bilingual recombination of Hindi and English, especially from intra- and inter-sentential perspectives (Vishwamohan, 2004; Bhatia & Ritchie, 2006a, 2006b; Kachru, 2006b; Nair, 2008; Si, 2010; Bhatia, 2011; Kothari & Snell, 2011; Sailaja, 2011; Bhatia, 2012; Kathpalia & Ong, 2015). On the other hand, a few studies have focused on word-level code-mixing (Kathpalia, 2018) but excluded prefixes. However, only one study was done in the 1970s that have concentrated on Hindi-English code-mixing. Still, from a data-driven perspective, the discussed classes were hybrid collocation, lexical combinations, word order, and reduplication but ignored blends (Kachru, 1975).

Therefore, the main objective in the current study is to bridge such lacuna by addressing coroneologisms in Hindi-English code-mixed words in addition to prefix-through the recently gathered words related to COVID-19 terms. In particular, we will shed light on three key word-formation processes: i) affixation, which undertakes the

application of English affixes on Hindi words or viz a viz, ii) blending, a situation where the words of English and Hindi are mixed, and iii) compounding, a process of combining two words and one from both the languages, in such a way that the meaning of newly compounded words can be determined from its component word. In case of reduplicated words, where the whole or its part or syllable of the existing word is repeated to support the compounding process.

Based on the above, this paper aims to answer the following three questions:

3. What are the particular word-formation types responsible for these coroneologisms through Hindi-English code-mixing?
4. Are such word creation processes structured, patterned, productive, and appear across different domains?
5. What factors are involved in creating such coroneologisms through word-formation processes?

The existing study further expands on the plane that will display how the linguistics resources of the two linguistically different languages have been exploited to coin these coroneologisms in Hindi-English code-mixed words and how these lexical innovations will contribute to the literature of English stretched zone context.

3 Methodology

This paper focuses on coroneologisms formed through the code-mixing of Hindi and English words from a word-formation processes perspective in different domains. Hereafter, the term “Hindi-English words” will be used with special reference to coroneologisms. Examples of Hindi-English words used in this study are taken from our collected wordlists. In some cases, it is difficult to understand whether the word primarily belongs to Hindi or English language, however, most Hindi speakers across the country will recognize these Hindi words as a combination of code-mixed words.

This study will distance itself from the traditional notions of matrix and embedded language because of the disagreement shown towards the concepts of mixing English-Hindi or vice versa. Alternatively, this study will approach the items through the lens of bilingualism that allows us to analyze two linguistically different, independent, additive, and interdependent languages (García & Li, 2014) to imagine bilingualism practices concerning intra- and inter-related languages. Mixing this angle of bilingualism with that of linguistic fluidity, we reach the term “translanguaging” (Li, 2011, p. 1222; García & Li, 2014, p. 2) that will further be used to examine the internal structure of coroneologism in Hindi-English code-mixed words since it allows Indian bilinguals to cross both linguistic and cultural boundaries between the two languages to produce code-mixed words. In particular, we will use the notion of “flexible” multilingualism to examine the code-mixed words on the scale of multilingualism to check whether the

language presents clear or shadowing boundaries (Blackledge & Creese, 2010; Zhang & Chan, 2015). Therefore, the paper examines the Hindi-English words in terms of the stem, whether the given stem belongs to English (viz: vaccinewala) or Hindi (viz: sankarman-type).

Moreover, all the collected code-mixed words will be classified based on the scale of word-formation processes such as affixation (includes both inflectional and derivational one) and compounding (includes reduplication, specifically complete reduplication). The classification will include the nature of such formations, specifically from their predictability and ad hoc perspectives. Such an analysis allow us to understand coroneologisms from the sociolinguistic angle which stresses that linguistic hybridity or mixing languages is “a systematic and rule-governed phenomenon which satisfies the creative needs of bilinguals” that is impossible to achieve within the boundary wall of a single language system (Bhatia & Ritchie 2006a, p. 518). Those who make it are competent bilinguals with access to two linguistic repositories that enable “them to mix language to achieve maximum efficacy from the two linguistic systems at their disposal” (Bhatia, 2011, p. 49).

Lastly, words on the wordlist used in this research were collected from different sources such as advertisements, news, TV episodes, and social media (Facebook, Twitter, Instagram, and YouTube) during the two waves of the COVID-19. Details are shown in the Table 1 below.

4 Data analysis

In this section, we will discuss the prominent types of word-formation processes in Hindi-English codemixed COVID-19 terminologies that drove to coroneologisms. In this regard, we have noticed three broad categories of word creation processes, i.e. affixation, blending, and compounding. Moreover, these word-formation processes have been categorised into two types, coinage and affixation, as mentioned below:

Table 1: Coroneologisms category in Hindi-English words

Category	Words	Percentage (%)
<i>Affixes</i>	14	03.80
<i>Compounding</i>	323	87.77
<i>Blending</i>	9	02.44
<i>Abbreviations</i>	7	01.90
<i>Acronyms</i>	15	04.07
Total	368	100

4.1 Coinage

This word-formation type allows us to create a completely new lexical item either deliberately or accidentally (Yule, 2020). Our data strongly suggests that most of the terminologies about the coronavirus crisis are newly coined or created. In this context, the most dominating neologism is COVID-19, to which both the international community and lexicographers have unanimously accepted. They are subsequently stored in the dictionary as a new lexical entry that refers to the coronavirus. In the latest unscheduled update, the Merriam Webster Dictionary on 26 May 2020 has released a “COVID-19 crisis catalogue: A glossary of terms” given the ongoing pandemic, where the appearance of COVID-19 term was seen as a dominant one. Similar to this, the Oxford English Dictionary (OED) also witnessed, “The most striking change has been the huge increase in the frequency of the words coronavirus and COVID-19 themselves”. Before 2020 the commoner was hardly aware of the term ‘corona virus’, though the medical community was familiar with the coronavirus family. But the term COVID-19 was coined in February 2020, and nowadays international community got acquainted with it. Notably, most COVID-19 inspired terminologies have also been adopted by Hindi language orthography. However, some of them have been transliterated, such as कोविड-19 /kovid-19/.

4.2 Affixation

Affixes are generally divided into inflectional and derivational affixes. According to Bauer (1983), inflectional affixes are responsible only for grammatical alteration of words, such as play, plays, and playing. In contrast, derivational affixes are used to derive new words with a new lexical meaning, such as real, unreal, and result in word class, such as boy (noun) to boyish (adjective). We have found that Hindi-English words allow both inflectional such as selfiyaan (Hindi suffix -iyaan is added to make the plural of English word ‘selfie’) and derivational affixes such as doctorgiri (refer to someone who acts like doctor), similarly, with English such as desiness (refer to the quality of being Indian) and Hindi word such as heropanti (acting like a hero) affixes. More examples have been presented in the table below of how Hindi-English codemixed COVID-19 words allow inflectional and derivational affixes, including prefixes and suffixes and whether they are from Hindi or English. But notably, the data show scarcity of infix as both Hindi and English do not have an infix system (Kathpalia, 2018).

Table 2: Examples of English inflections in Hindi-English COVID-19 words

Suffix	Examples	Meaning
-s	yodhas/tikaas/jawans	warriors/vaccines/warriors

Table 3: Examples of Hindi inflections in Hindi-English COVID-19 words

Suffix	Examples	Meaning
-yaan	entriyaan	-yaan= a direct feminine, plural suffix
-on	vaccinon/bedon	-on = plural suffix
-ein	filein	-ein = feminine plural suffix marker

Interestingly, the above Table 2 shows examples from both Hindi and English languages and presents the ways inflections are attached to the stems in each other language's words. While in Table 2, English suffix -s was attached to Hindi stems, the subsequent Table 3 shows examples of Hindi suffix -yaan, -on, and -e attached to English stems.

Nair (2008) claims that using English verb endings (e.g., -ing, -s, and -ed) and plural morpheme (-s) are frequently used with Hindi Nouns and Verbs. Similarly, Hindi suffixes are also affixed to English stems responsible for marking number, gender, and case into English words, as exemplified in the above tables.

Kathpalia (2018) argued that the percentage of derivational (53%) affix is slightly higher than the inflectional one (46%). Further, she noted that out of the total derivational affixes from both English and Hindi languages, the number of English affixes were (24%). In comparison, Hindi affixes were (76%), even more pertinent that the majority of them were suffixed (95%) as compared to prefixes, i.e. (5%). In this study, we have also considered the Hindi prefix a- which is frequently attached with English stems, especially with nouns. The data of inflectional (prefixes) and derivational suffixes of both English and Hindi languages are exemplified in Tables 4, 5, and 6 below:

Table 4: Examples of Hindi derivation in Hindi-English COVID-19 words

Prefix	Examples	Meaning
a-	a-sankarmit	un-infected
	a-symptomatic	without-symptom

Table 5: Examples of English derivations in Hindi-English COVID-19 words

Suffix	Examples	Meaning
-type	sankarman-type	infection-type

Table 6: Examples of Hindi derivations in Hindi-English COVID-19 words

Suffix	Examples	Meaning
-garast	coronagarast	patient = refer to suffering of disease 'occurs with noun'
-giri	doctorgiri	giri = refer to '-ism'
-vaalaa (M)*	coronavaalaa	vaalaa = occur with noun and indicate a possessor or owner
-vaalii (F)	maskvaalii	
-vaale (P)	policevaale	
-vaalon(OP)	hospitalvaalon	
-sb (H)	Dr sb	sb = as an honorific marker

* M= Masculine, F= Feminine, P= Plural, OP= Oblique Plural, H= Honorific

Our data suggest that the frequently used English suffix was -type and Hindi suffix -waala/waalii/waale/waalon, -giri, and -sb. Next, the data further revealed that inflectional affixes are slightly less frequent than derivational affixes in both English and Hindi. Therefore, it is pertinent to mention that the data show a wider scope of using derivational affixes in both languages. Bauer (1983) noted that the inflection affixes is fragile because it belongs to smaller and closed classes. They are also known as blockers. However, the scope of derivational affixation is much wider than inflectional one, even though it invites some language-specific restrictions. The English suffix -able shows semantic consistency and can be attached to any transitive verb to form an adjective. For example, workable (Bauer, 1983), and interestingly this feature is loaned in the Hindi language to perform the same function as in the word 'chaleable', meaning workable. Typically, the Hindi-English words do not contain English suffix only but also prefix -un, as in unjhelable. Similarly, the Hindi prefix a- is also frequently attached to English stems to perform the same function means used to make negative of any word (see table 4). The data presented in the tables above mainly shows the suffixation of English in Hindi-English words because the use of prefix is slightly less common, comparatively. However, we could not encounter words which would simultaneously contain prefixes and suffixes. We strongly recommend checking such a phenomenon with more Hindi-English words.

Interestingly, the Hindi suffix morphemes are frequently attached to English stems. Some of these trending suffixes are: -vaalaa/vaalii, -baazi, -giri, -panti, sb, and -yaan, etc., used by bilingual Indian speech community in their daily conversation as well as in Hindi cinemas, advertisements, TV series, and dramas. Even they appear in Hindi cinema names, for example, a movie released in 2015 named 'Meeruthiya Gangsters', 'womeniya' (women+duniya) a famous song in a Hindi movie named 'Gangs of Wasseypur'. The trace of these expressions are also found in the works of several scholars like Kachru (1975): educated-type, school-vaalaa, police-vaalaa; Gargesh

(2006) and Sinha (2011): milk-vaalaa; Kachru (2006b), heropanti. It is noteworthy to mention that inflectional morpheme/suffix -valaa can appear in different forms depending on the number and gender of the noun, as for example valaa (male), vaalii (female), and vale (singular), valon (oblique plural). In this context, Singha (2011) noted that all -valaa constructions form multiword expressions in Hindi language. The suffix -valaa construction appears to be productive in terms of Hindi-English bilingual COVID-19 terminologies. Further, the suffix -baaz is originally from Persian, borrowed into the Hindi language, and used frequently with English stems. For example, as in Hindi TV dramas and serials; drame**baaz**. In a work Kathpalia (2018) noted such examples: country**baaz**, daring**baaz**, design**baaz**, fight**baaz**, flirt**baaz**, fraud**baaz**, help**baaz**, smart**baaz**, etc. The other type of Hindi suffix -giri, and -panti were in use but gained recognition recently only through political and Bollywood discourse titles, for example, Political**giri** (its Hindi equivalent is Netagiri), Police**giri**, and Heropanti. Finally, the use of Hindi honorific -sb is also frequently used with English words, for example, Drs**sb**, to perform dual function; to show honour and to show geographical affiliation, i.e. north-India (Nair, 2008).

Despite several language-specific grammatical constraints and differences between inflectional and derivational affixes, our data revealed that both affixes were used frequently with Hindi and English words to coin new Hindi-English expressions of different domains given the COVID-19 pandemic. The data we have presented reveal that the process of affixations is quite productive between the two languages (Hindi and English) and allows us (Indian bilinguals) to create new terms/words related to the COVID-19 crisis, eventually more than what is feasible within the same language as per the existing grammar rules. In the following sections, we have focused on another potential type of word formation process that allows Indian bilinguals to coin new Hindi-English words, that is blending.

4.3 Blending

Blending refers to amalgamation or fusion of two words to create a new word. The literature suggests different names of this process (blending) like: coalesced words, portmanteau words, and telescoped words. Despite different expressions, all refer to the notion of creating new words by conjoining parts of existing words in human languages across the globe. In this work, we have examined Hindi-English blends as an “extra grammatical phenomenon” (Mattiello, 2013, p. 127) but with consistency in patterns. The data were analyzed based on Mattiello’s (2013) proposed classification. Mattiello’s classification is an improved and revised version of different previous taxonomies. In the following Table 7, some examples of blending patterns in Hindi-English COVID-19 words have been presented.

The examples of Hindi-English blends revealed one pattern, i.e. complete blend (beginning+end). Mattiello (2013, p. 118) divided blends into three types: ‘morphotactical’ (complete and partial reduplicated blends), ‘morphological’, and ‘graphical’ (overlapping vs. non-overlapping blends) but we have found only one type of blend in our data, i.e. complete blend. Moreover, the COVID-19 related blends were made within the same language, which means either in Hindi or English. Furthermore, only a few examples of Hindi blends concerning pandemics were reported. In contrast, blends in the English language were more frequent, and most of them has been adopted in the Hindi language through transliteration. In this study, we have found examples for morphotactical blends, which means the data showed only complete blends. Referring to complete blends, this is formed by reducing the source words of both languages to the splinter. Notably, the structure of complete blends contains its sub-pattern, which is as beginning+end, as exemplified below.

Table 7: Examples of COVID-19 blends

Type	Illustration	Hindi	Combination	Pattern	English	Combination
Total blend	All source words reduced to splinter	tikotsav	tika ‘vaccine’ + ustav ‘festival’	Beginning + End	Cipremi	Cipla + Remdesivir
					Covaxin	Corona + vaccine
					Covishield	Covid + shield
					Cowin	Corona + win

4.4 Compounding

The concept of compounding is perceived as a “lexeme containing two or more potential stems that have not subsequently been subjected to a derivational process” (Bauer, 1983, p. 29) (as cited in Kathpalia, 2018). Further, we have classified compounds on the scale of semantics into endocentric and exocentric types, as exemplified in the Table 8 below.

These compounds have been classified on syntactic accounts based on the formation patterns/structure of the whole or component part of compound constructions. Based on the existing types of compounds (noun+noun, noun+verb, adjective+noun, and phrasal compound, and others) found in the English language, the analysis was further extended to examine the existing patterns/structures of compounds in Hindi-English, as presented in the Table 9 below.

Table 8: Types of COVID-19 specific compounds in Hindi-English
(Adopted from Kathpalia, 2018)

Type	Illustration	Examples
Endocentric compounds	A hyponym of a grammatical head as in <i>mahawave</i> means 'big covid second wave'	mahawave = maha 'big' + wave 'relating to corona surge'; mahavaccinatedrive = maha 'biggest' + vaccine drive 'relating to vaccination in view of corona'
Exocentric compounds	A hyponym of an unexpressed semantic head as in <i>sampurn lockdown</i> means 'complete lockdown'	sampurnlockdown = sampurn 'complete' + lockdown

Table 9: Showing formation patterns/structures of COVID-19 associated compounds
(Adopted from Kathpalia, 2018)

Type	Examples	Meaning
Noun+noun	sankramandar	infected rate
Adjective+noun	atmanirbharbharat	atmnirbharbharat, (Independent India) a slogan given during first wave of covid-19 by GOI.
Noun+verb	tika khoj	vaccine (re)search
Phrasal compound	parvasi mazdoor	migrant labor

Our data revealed that most of the Hind-English bilingual compounds were reported in the noun compounds class belonging to both noun+noun and adjective+noun types. Most of the COVID-19 related words were formed of noun+noun combination. See an interesting case of compounding; tandoorii 'physical/social distancing' (Bhatia & Edmonds, 2021). The second trending type of compounding was formed of adjective+noun English-Hind pair (see appendix). Additionally, a limited number of phrasal compounds have also been reported.

Yet another important type of compounding is formed through reduplication, i.e., called 'eco-words' or 'rhyming words' formation. Such words are created through the exact replication of sounds, syllables, or words or else by altering the first part or syllable of the second word (Mattiello, 2013, p. 141). The alterations could be performed with vowels, consonants, or both vowels and consonants. Notably, Hindi speakers form eco-words by altering the first part or syllable of the second word with 'va' or 'u'. In contrast, the second part or syllable remains the same according to the first word, for example, taxi-**v**axi or taxi-**u**xi 'taxi and like'. Further, the first word can stand alone and bear its semantic value, while the second word serves a

communicative or rhyming function. The instance of complete/full reduplication was not reported. Table 10 below shows such examples of reduplication.

Table 10: Types of COVID-19 reduplication (adopted from Kathpalia, 2018, originally from Mattiello, 2013)

Type	Illustration	English-Hindi
Partial reduplication	Replicating only part of word or sound	corona-vorona = corona and the like vaccine-phaiksin-uksin = vaccine and like

The eco-word formation presence in our data was found to have a higher frequency. The data also revealed specific patterning of making eco-word in both the languages, i.e. in English, the second word starts with /p, b, or w/ (such as lockdown-phockdown) and in Hindi, the second word starts with /v or u/ (as in, sankarman-vankarman/unkarman 'infection and the like'). Kachru (2006a) notes that the first word has its meaning while the second word bears no meaning or communicative meaning and cannot occur independently. The case of complete reduplication in Hindi-English was not present. However, the examples in partial reduplication are more productive than complete reduplication in Hindi-English scenario. Lastly, the other types of reduplication were also not found in the data, like semantic reduplication (such as lathi-stick) in the case of Hindi-English scenario.

4.5 Borrowing

It is a very common practice whereby a lexical item of one language travels to another language, and the process is called adaption or borrowing. In Crystal's (2010) opinion, in modern times, English is considered as the biggest donor of lexical items to the world's different languages remains a great receiver as well. In this context, it has borrowed lexical items from the world's 120 different languages, including Sanskrit, Hindi, Urdu, French, and Arabic, etc. (Nordfuist, 2019). In the context of borrowing from the coronavirus perspective, 'lockdown' is widely accepted by the world's other languages, including Hindi. However, Hindi has loan translated the word 'lockdown' 'process of imposing restrictions to stop the spread of the virus' as 'talaabandi' and used interchangeably by Hindi speakers. Roig-Marín (2020, p. 2) further notes that "Covid has been borrowed as an Anglicism and users of languages with grammatical gender like French, Catalan, Spanish, and Italian have tended to prefer the masculine gender because of its associations with (corona)virus, masculine in those languages". Considering this notion, it can be said that Hindi is also a gender-sensitive language, and speakers use the word 'Covid' as a masculine gender.

Further, in this study, borrowings have been analyzed according to their sub-types, i.e. loan translation or calque and loan blend. However, we have not encountered any examples of loan shifts in Hindi-English COVID-19 associated words. In the following Table 11, the examples of loan translations and loan blends have been represented.

Table 11: Showing the examples of COVID-19 words

Type	Illustration	English	Hindi
Loan translation	Equivalent translation of each word in the target language	social distancing	samajikduri
Loan blend	Either first or second ward is translated into its equivalent word in the target language	corona test partial lockdown corona warrior	corona jaanch anshik lockdown corona yodhdha

Table 12: Examples of English inflections in English-Hindi COVID-19 words

Suffix	Examples	Meaning
-s	corona yodhas corona jawans	corona warriors; -s = English plural marker corona warriors

Table 13: Showing the examples of Hindi inflections in Hindi-English COVID-19 words

Suffix	Examples	Meaning
-on	corona wardon isolation centeron	corona wards; -on = Oblique plural corona incharges; -on = Oblique plural isolation centers; -on = Oblique plural
-yaan	corona entriyaan	Corona enteries; -iyaa = a direct, feminine, plural suffix

5 Findings and discussion

This study addressed three questions: i) What are the commonly used types of Hindi-English code-mixed neologisms that have emerged due to the COVID-19 outbreak?, ii) What are the dominant word-formation classes and are these neologisms rule-governed code-mixing products, including productivity and domain dependant?, and iii) What are the factors that trigger such code-mixing in Hindi-English and concern COVID-19 neologisms?. The five word-creation processes in the Hindi-English codemixed words that we have examined related to COVID-19 terms, namely coinage, affixation, blending, compounding, and borrowing, revealed that compounding is the most

productive type of neologisms. As far as language boundaries are concerned, it was evident that the words of both Hindi and English can easily be identified concerning inflections (viz: yodhas), derivation (viz: Drsb and Drgiri), and compounding (viz: mahawave), respectively. However, in the case of blending the demarcation of boundaries between Hindi-English words and English-Hindi words were opaque and difficult. Moreover, a limited number of words have been reported (see Appendix-I below). In this regard, Kathpalia (2018) claimed that the code-mixing of English words with the words of world's other languages is also frequent and equally fluid as well as productive in terms of word-formation processes. There are instances of code-switched words that show clear demarcation of language boundaries (e. g. Chinese-English 'zhaomuing', and Italian-English 'kissucci'), and there are examples that appear to be opaque concerning language boundaries, (e.g. Chinese-English 'mangday', Russian-English 'БуGOODu', and French-English 'Chouchou').

The paper also attempts to answer the question whether or not the Hindi-English code-mixed word-formation processes are systematic and whether they show any patterns. It further attempts to establish the difference between grammatical and extra-grammatical constructions on account of morphological knowledge of both Hindi and English. In this context, words formed through regular rules were kept under grammatical class, while unpredictable words were kept under extra-grammatical category. For example, words formed through compounding and affixation processes were transparent and predictable from the existing rule of Hindi and English, whereas blends were extra-grammatical constructions. Therefore they were unpredictable as they "are generally not transparently analyzable into morphemes" (Mattiello, 2013, p. 250). Our data further revealed that the words formed through affixation in Hindi-English were easily breakable into their component words or morphemes or parts of stem and affixes, for example, doctaron = doctor + -on (Hindi plural suffix). However, the scenario is relatively less transparent in terms of blends, as they are formed by combining the words of two different languages. In conclusion, they are partially or fully unpredictable. Moreover, our data show sparseness in blend formation as compared to the affixation and compounding, but their formation shows some regularity (see the Appendix- I).

The formation of Hindi-English codemixed COVID-19 terms are both productive and rule-governed, as they are formed through the regular word creation processes of both languages, specifically compounding and suffixation. Additionally, we have also devised such words that are formed of extra-creativity and, through the extension of regular rules of the language, kept under extra-grammatical or non-rule governed constructions, as they are unpredictable from their component parts. The data of this study revealed both types of lexical innovations concerning Hindi-English codemixed COVID-19 terminologies; one created out of productivity (coining new terms using the existing word-formation processes), and the second formed out of creativity (creation of new words/terms through manipulating the existing rules). Based on the

aforementioned discussion, it can be noted that the productivity of rule-governed coinage is significantly higher than non-rule governed coinage in Hindi-English words. This may be because the coinage through the rule-governed process is easier. Bauer (2001) noted that both rules and analogy contribute to morphological innovations. Still, the product of rule-governed coinage is quite predictable from its components, while extra-grammatical construction is only partially predictable.

Moreover, Mattiello (2013) argued that the motivation for word creation of different classes, including grammatical and extra-grammatical categories, could be context-dependent. According to this, the current study's data reveals that coinage through the affixation process in Hindi-English words is a part of Indian bilinguals' daily language, as they created it through the rule-governed process of Hindi-English. However, blends are not the case; rather, such constructions are created out of extra-grammatical process/morphology. Therefore, mostly we encounter them in advertisements, TV shows, and other media because in them the focus remains on customer persuasion (refer to Tables 2-7 above).

The third question of this study deals with the motivation for the creation of neologisms through the code-mixing of Hindi-English words. Examples analyzed in this study seem to be a product of social, political, cultural, and linguistic interactions among Indian bilinguals. Primary observation from these creations appear to be inspired mainly by the bilingual culture, where monolingualism appears to be an abstract idea (viz: coronayodha = corona + yodha 'survivor'). Consequently, Hindi-English codemixed COVID-19 terms also started floating frequently on digital and mass media platforms (viz: mahawave = maha 'big' + wave 'relating to corona surge'). Moreover, word-creation through Hindi-English code-mixing is mainly due to its increasing popularity among Indian youths. They view it as a new way of life or "new lifestyle mantra" (Pal & Mishra, 2011, p. 175). This gives rise to the sense that the creation of code-mixed words is due to the overwhelmed use of digital media in India, especially the dominant use of Hindi-English mixed words on popular social media websites like Facebook, Twitter, Instagram, e-Encyclopaedia, and lifestyle information related websites like health, beautification, etc. In this context, some trending words that are the product of Hindi-English code-mixing like Facehook (Facebook + bhook 'hunger'), similarly, Filmygyan (film + Hindi suffix -y, + gyan 'knowledge'), etc. So this creativity is extended and used to create Hindi-English code-mixed COVID-19 terminologies, especially by the government and TV shows to spread awareness among the Indian masses on how to fight the coronavirus. In addition to such a noble cause, given customer persuasion, the advertising agencies and experts started mixing Hindi-English words to persuade the customers, e.g., 'maha vaccine drive' (largest vaccine drive) used to spread awareness among the Indian citizens towards vaccination. Therefore, concerning COVID-19, the government and social organisations, advertisers, and educationists have also made frequent use of Hindi-English codemixing to create new names, terms, and slogans to show their awareness of their target audience in the

fight against COVID-19. For example, a famous Indian government slogan, 'do ghaz duri mask hai zaruri' means 'two-meter physical distance and mask is a must'.

The lists we have prepared in this study clearly show the speed and enthusiasm for coinage and lexical innovations in Hindi-English code-mixed COVID-19 terminologies. Crystal (2020, p. 1) has noted that there are collections that compiled neologisms concerning COVID-19 in the English language but numbers are still increasing. In this regard, the authors of the current study attempted to compile a list of such code-mixed neologisms in Hindi-English that have been created after the outbreak of the COVID-19 pandemic. Ro (2020, p. 1) claims that the only coinage is 'COVID-19', and all other terms related to the COVID-19 were created with the pre-existing words and revised and redefined given the coronavirus pandemic.

The present study's findings strongly support the notion of linguistic change, as this is an inevitable phenomenon in any human language. This study was directed to shed light on the coroneologisms of Hindi-English codemixed words, and it was noted that out of all the new coinage, COVID-19 was the only term that has topped the list of lexical innovations in 2020 after the coronavirus outbreak. Notably, the term COVID-19 was coined in February 2020 and subsequently accepted by the international community. The authors of this study also believe that there are COVID-19 inspired neologisms that are yet to be entered in the dictionary due to the frequent mutation of the virus and rounds of waves. Consequently, governments of different countries are implementing several policies and strategies to curb the virus. Therefore, the creation of COVID-19 specific terms is still underway, and social media is a good reservoir of such terminologies at a global level.

One of the significant observations of this study is that most of the COVID-19 terms went under pragmatic change. In this context, the concept of 'social distancing' (Hindi equivalent 'samajik duri') has completely changed, now this refers to the several related health behavior concepts like 'maintaining minimum social distance', 'wearing a mask', 'no physical contact', 'avoid handshaking', and 'hand sanitization', etc. may lead to saving human lives from the deadly virus of covid. Most of these terms were pre-existing in the dictionary but re-introduced and became part of our daily lives after the pandemic. Moreover, the change in social attributes due to COVID-19 will leave a long-term impact on human life, especially in terms of social behavior in everyday life.

The significance of Hindi-English code-mixed COVID-19 related terms is crucial from an Indian bilingualism perspective, as these terminologies help people in getting updates/information related to COVID-19. These terminologies may also be used as standard terminologies that may unite the Hindi-English bilingual audience in the fight against the coronavirus. Somehow, the unified terms concerning COVID-19 have already mobilised the people of the world against the virus. Moreover, such coinage and lexical innovations concerning the coronavirus needs to be collected and stored in the form of a dictionary so that in real-time, these terms can be shared with all

concerned authorities in a minute or two. In this direction, seeing and discovering these terms through the lens of the word creation processes in Hindi-English language is of paramount importance. This will lead to creating a Hindi-English bilingual database of such terms.

Lastly, the types of word creation processes outlined in this study (see section 4 above) clearly state the processes that how COVID-19 specific terms were formed through the code-mixing of Hindi-English words. This confirms that coroneologism is inevitable and contributes significantly to the development and lexical innovation in a Hindi-English scenario. Ample evidence of neologisms and lexical innovations have been presented in the current data concerning compounding and affixation in Hindi-English code-mixed words (see Appendix-I). Based on the investigation of the word-formation patterns of Hindi-English code-mixed COVID-19 terminologies, this study reports five types of word-formation processes concerning coroneologisms. This satisfies the notion of lexical innovations that took place due to pandemics in Hindi-English words. These word formation classes were: coinage, compounding, affixation, blending, and borrowing (see Appendix-I).

6 Conclusion

This study examined three key issues related to Hindi-English word-formation processes, their productivity, and motivations from a translanguaging point of view. The attention was also given to the fluidity and productivity of these lexical innovations from a linguistic perspective, particularly from a morphological viewpoint. The morphological analysis revealed that the boundaries between Hindi and English are very fluid and allow bilinguals to create new words out of their knowledge of linguistic creativity. Linguistic creativity is significant from a morphological perspective and vital to the social hybridization that further opens the way for better interaction between the two socially different languages. Moreover, the data analysis also revealed that Hindi has borrowed a significant amount of lexical items from the English language related to the COVID-19 pandemic terminologies that drove to Hindi-English code-mixed coroneologisms. Out of all the word-formation processes, compounding was reported as the most productive type. This allows Indian bilinguals to creatively combine the words of two linguistically different languages, which is less possible to achieve such a level of lexical innovations within a single language system.

References

- Aitchison, J. (2001). *Language change: Progress or decay?* Cambridge: Cambridge University Press.
- Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). *Linguistics: An introduction to language and communication*. Cambridge: MIT Press.
- Balteiro, I. (2012). When Spanish owns English words. *English Today*, 28(1), 9-14.
- Bauer, L. (1983). *English Word-Formation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139165846
- Bauer, L. (2007). *Linguistics Student's Handbook*. Edinburgh: Edinburgh University Press.
- Bhatia, T. K. (1992). Discourse functions and pragmatics of mixing: Advertising across cultures. *World Englishes*, 11(2-3), 195-215.
- Bhatia, T. K. (2011). The Multilingual Mind, Optimization Theory, and Hinglish." In R. Kothari & R. Snell (Eds.), *Chutnefying English: The Phenomenon of Hinglish* (pp. 37–52). Gurgaon: Penguin Books.
- Bhatia, T. K. (2012). Advertising and Branding in India. In J. S. Lee & A. Moody (Eds.), *English in Asian Popular Culture* (pp. 231–254). Hong Kong University Press.
- Bhatia, T. K., & Ritchie, W. C. (2006b). Bilingualism in South Asia. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of Bilingualism* (pp. 780–807). Oxford: Blackwell. DOI: 10.1002/9780470756997.ch29
- Bhatia, T. K., & Ritchie, W. C. (2008). The Bilingual Mind and Linguistic Creativity. *Journal of Creative Communication*, 3, 5–21. Doi: 10.1177/097325860800300102
- Bhatia, T. K., & Edmonds, M. (2021). COVID-19 and social distancing: Impact on Communication, Learnability and Mental Health. *Journal of exclusion studies*, 11(1), 1-16.
- Bhatia, T. K., & Ritchie, W. C. (2006a). Bilingualism in the Global Media and Advertising. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of Bilingualism* (pp. 513-546). Oxford: Blackwell. DOI:10.1002/9780470756997.ch20
- Bizhkenova, A. E., Sultanbekova, S., & Koshekova, A. (2017). Neologisms in present day german: investigation into productivity of word formation types. *Journal of Fundamental Applied Sciences*, 9(7S), 982–96.
- Burridge, K., & Bergs, A. (2016). *Understanding language change*. USA: Taylor & Francis.
- CoWIN. (2021). Updates on Corona Vaccine. <https://www.cowin.gov.in/home> (accessed: 18 May 2021).
- Crystal, D. (2008). *A Dictionary of linguistics and phonetics* (6edn). Blackwell Publishing.
- Crystal, D. (2010). *The Cambridge encyclopedia of language a dictionary of language*. Cambridge: Cambridge University Press.
- Crystal, D. (2020). Cambridge Reflections-COVID-19: Vocabulary. *Cambridgeblog*. URL: <http://www.cambridgeblog.org/2020/05/covocabulary/> (accessed: 18 May 2021).
- Durham, G. C. (2018). Back Formations and Neologisms. *3 Quarks Daily*. URL: <https://www.3quarksdaily.com/3quarksdaily/2018/09/back-formations-and-neologisms.html> (accessed: 5 April 2021).

- Gargesh, R. (2006). South Indian Englishes. In B. B. Kachru, Y. Kachru & N. L. Cecil (Eds.), *Handbook of World Englishes* (pp. 90–113). Oxford: Blackwell. doi: 10.1002/9780470757598.ch6
- Haider, A. S., & Al-Salman, S. (2020). Dataset of Jordanian university students' psychological health impacted by using e-learning tools during COVID-19. *Data in brief*, 32, 106104. DOI: 10.1016/j.dib.2020.106104
- Hickey, R. (2001). Language change. In Jan-Ola Östman & Jef Verschueren (Eds.), *Handbook of pragmatics*, 1–33. Amsterdam: John Benjamins.
- Hollett, V. (2020). New English words and language change with COVID-19. *Simple English Videos*. URL: <https://www.simpleenglishvideos.com/new-english-words-and-language-change-with-covid-19/> (accessed: 12 April 2021).
- Kachru, B. B. (1975). Lexical Innovations in South Asian English. *International Journal of Sociology of Language*, 4, 55–74.
- Kachru, Y. (2006a). *Hindi*. Amsterdam: Benjamins. DOI: 10.1075/loall.12
- Kachru, Y. (2006b). Mixers Lyricing in Hinglish: Blending and Fusion in Indian Pop Culture. *World Englishes*, 25, 223–233. DOI: 10.1111/j.0083-2919.2006.00461.x
- Kathpalia, S. S. (2018). Neologisms: Word creation processes in Hindi-English code-mixed words. *English World-Wide*, 39(1), 34–59.
- Lalić-Krstin, G., & Silaški, N. (2018). From Brexit to Bregret: An account of some Brexit-induced neologisms in English. *English today*, 34(2), 3–8. DOI: 10.1017/S0266078417000530
- Lawson, R. (2020). Coronavirus has led to an explosion of new words and phrases—and that helps us cope. *The Conversation, journalistic flair*, 28. URL: <https://theconversation.com/coronavirus-has-led-to-an-explosion-of-new-words-and-phrases-and-that-helps-us-cope-136909> (accessed: 18 May 2021).
- Liu, W., & Liu, W. (2014). Analysis on the Word-formation of English Netspeak Neologism. *Journal of Arts Humanities*, 3(12), 22–30. DOI: 10.18533/journal.v3i12.624
- Lu, N., Cheng, K. W., Qamar, N., Huang, K. C., & Johnson, J. A. (2020). Weathering COVID-19 storm: Successful control measures of five Asian countries. *American journal of infection control*, 48(7), 851–852. DOI: 10.1016/j.ajic.2020.04.021.
- Mattiello, E. (2013). *Extra-Grammatical Morphology in English: Abbreviations, Blends, Reduplicatives and Related Phenomena*. Berlin: DeGruyter Mouton. DOI:10.1515/9783110295399
- McMahon, A., & McMahon, R. (2013). *Evolutionary linguistics*. Cambridge: Cambridge University Press.
- Merriam-Webster. 2020. Coronavirus and the New Words We Added to the Dictionary in March 2020. *Merriam-webster*. URL: <https://www.merriam-webster.com/words-at-play/new-dictionary-words-coronavirus-covid-19> (accessed: 5 April 2021).
- Muhvić-Dimanovski, V. (2004). New Concepts and New Words—How Do Languages Cope With the Problem of Neology? *Collegium antropologicum*, 28(1), 139–46.
- Nair, R. B. (2008). Language and Youth Culture. In B. B. Kachru, Y. Kachru & S. N. Sridhar (Eds.), *Language in South Asia* (pp. 466–494). Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511619069.027

- Oxford English Dictionary. (2020). *Updates to the OED*. Oxford English Dictionary. <https://public.oed.com/updates/> (accessed: 5 April 2021).
- Pal, S., & Mishra, S. (2011). Hinglish and Youth. In R. Kothari & R. Snell (Eds.), *Chutnefying English: The Phenomenon of Hinglish* (pp. 161–175). Gurgaon: Penguin Books.
- Rao, P. S. (2019). The Role of English as a Global Language. *Research Journal of English*, 4(1), 65–79.
- Ratih, E., & Gusdian, R. I. (2018). Word Formation Processes in English New Words of Oxford English Dictionary (OED) Online. *Celtic: A Journal of Culture, English Language Teaching, Literature Linguistics*, 5(2), 24–35. DOI: 10.22219/celticumm.vol5.no2.24-35
- Romaine, S. (1983). Historical linguistics and language change: Progress or decay? *Language in society*, 12(2), 223–37. DOI: 10.1017/s0047404500009829
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, 14(9), 411–417. DOI: 10.1016/j.tics.2010.06.006
- Sinha, R. M. (2011, June). Stepwise mining of multi-word expressions in Hindi. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World* (pp. 110-115).
- Ten Hacken, P., & Thomas, C. (2013). *The semantics of word formation and lexicalisation*. Edinburgh University Press.
- Thorne, T. (2020). #CORONASPEAK – the language of Covid-19 goes viral – 2. *Language and Innovation*. URL: <https://language-and-innovation.com/2020/04/15/coronaspeak- part-2- the-language-of-covid-19-goes-viral/> (accessed: 18 May 2021).

Appendix-I: Hindi-English codemixed COVID-19 associated word lists

Hindi words	English words	Hindi-English words
Word formation type: Compounding		
कोरोनामरीज़ (korona mariz)	Symptomatic	राष्ट्रीयलाकडाउन (rashtriya lockdown)
कोरोनासंकट (korona sankat)	Asymptomatic	स्थानीयलाकडाउन (asthaniye lockdown)
कोरोनाकाल (korona kaal)	Coronavirus	घातकवायरस (ghatak virus)
कोरोनाविस्फोट (korona visphot)	Quarantine	वैक्सीनसंकट (vaccine sankat)
कोरोनाजिहाद (korona jihad)	Isolation	वायरससंक्रमण (virus sankraman)
कोरोनाटीका (korona tika)	Self-isolation	घातकवायरस (ghatak virus)
वैक्सीननिर्माण (korona nirman)	Covishield	फ़िज़िकलसंपर्क (physical sampark)
कोरोनामहामारी (korona mahamari)	Contactless	लाकडाउनउलंघन (lockdown ulanghan)
कोरोनालहर (korona lahar)	Antibodies	पाज़िटीवदर (positive dar)
कोरोनारोकथाम (korona roktham)	Webinar	संपूर्णलाकडाउन (sampurn lockdown)
सामाजिकदूरी (samajik duri)	Cowin	जनताकरफ्यू (janta curfew)
अदृश्यवायरस (adrishya virus)	Corona Vaccine	पाज़िटिविटीदर (positivity dar)
कोविडप्रबंधन (kovid prabandhan)	Corona Bulletin	वैक्सीननीति (vaccine niti)
कोविडनियमावली (kovid niyamawali)	Corona Positive	वैक्सीनगुरू (vaccine guru)
मुफ्तटीकाकरण (muft tikakaran)	Corona Negative	पहलीवेव (pahli wave)
संक्रमणदर (sankraman dar)	Covid Caseload	दूसरीवेव (dusri wave)
कोविडकेंद्र (kovid Kendra)	Corona Report	कोविडइलाज (covid ilaj)
कोरोनायोद्धा (korona yodha)	Corona Case	कोविडमुक्ति (covid mukti)
संक्रमणसंख्या (sankraman sankhya)	Community Transmission	जेनोमसमूह (genome samuh)

Hindi words	English words	Hindi-English words
आक्सीजनसंकट (oxygen sankat)	Quarantine Center	नयास्ट्रेन (naya strain)
कोरोनाडर (korona dar)	Positivity rate	आक्सीजनमाफिया (oxygen mafia)
कोरोनासंक्रमित (korona sankramit)	Covid return	आक्सीजनचोरी (oxygen chori)
कोरोनातांडव (korona tandav)	Corona Update	आंशिकलाकडाउन (anshik lockdown)
कातिलकोरोना (qatil corona)	Self-isolation	जीनगुरू (gene guru)
कोरोनाखतरा (corona khatra)	Self-Quarantine	देसीवैरिएंट (desi variant)
कोरोनाहाहाकार (corona hahakar)	Vaccination Programme	विदेशीवैरिएंट (videshi variant)
कोरोनापीड़ित (corona pidit)	Vaccination Scheme	कोविडचुनौती (covid chunouti)
कोरोनाव्यापार (corona vyapar)	Vaccination Center	टीकाडोज़ (tika dose)
कोरोनासंक्रमण (corona sankraman)	Vaccine Production	कोविड१९मरीज़ (covid 19 mariz)
कोरोनामामलों (corona mamlon)	Covid infection	जाँचरिज़ल्ट (janch result)
कोरोनामामले (corona mamle)	Corona case	सुपरस्प्रेडरघटना (super spreader ghatna)
सक्रियमामलों (sakriya mamlon)	Clinical Trial	महावैक्सीनेशनड्राइव (mahavaccine drive)
दूसरीलहर (dusri lahar)	Clinical Phase	कोविडगाइडलाइन (covid guideline)
तीसरीलहर (tisri lahar)	Clinical Stage	कोरोनावायरस (corona virus)
टीकाउत्पादन (tika utpadan)	Corona Patient	संपूर्णलाकडाउन(sampurn lockdown)
टीकानिर्यात (tika niryat)	Face Mask	
आत्मनिर्भरभारत (atmnirbhar bharat)	Covid Patient	
कोविडटीकाकरण (covid tikakaran)	Corona Mutants	
कोरोनाज़ख़म (corona zakhm)	Corona Variants	
वैश्विकमहामारी (vaishvik mahamari)	Complete Lockdown	

Hindi words	English words	Hindi-English words
टीकाउत्सव (tika utsav)	Covid Warrior	
टीकाअभियान (tika abhiyan)	Corona Warrior	
टीकाकरणअभियान (tikakaran abhiyan)	Frontline worker	
कोरोनाकहर (corona kahar)	Containment Zone	
कोरोनाविशेषग (corona visheshag)	Contact Tracing	
कोरोनाविषलेशन (corona vishleshan)	Social Distancing	
कोरोनामाहिरीन (corona mahirin)	Essential Business	
कोरोनारोगी (corona rogi)	Essential Shop	
कोरोनामुनाफा (corona munafa)	Night Curfew	
कोरोनादस्तक (corona dustak)	Vax Drive	
टीकाकरणमुहिम (tikakaran muhim)	Physical Distancing	
कोरोनाउछाल (corona uchal)	Super Spreader	
सरकारीआँकडे (sarkari ankde)	Community Spreader	
टीकाआयात (tika aayat)	Vaccine war	
टीकानिर्यात (tika niryat)	Human Trial	
कोरोनाशव (corona shav)	New strain	
कोरोनामंत्र (corona mantra)	Corona Expert	
कोरोनादिव्यांग (corona divyang)	Corona Testing	
कोविडशव (covid shav)	Testing kit	
कोरोनासंक्रमणरोकथाम (corona sankraman roktham)	Covid wards	
कोरोनादवा (corona dava)	Vaccine drive	

Hindi words	English words	Hindi-English words
कोरोनाअस्पताल (corona aspatal)	Corona time	
प्रवासीमज़दूर (pravasi mazdur)	Corona Period	
साप्ताहिकतालाबंदी (saptahik talabandi)	Zoom Meeting	
आरोग्यसेतु (arogya setu)	Zoom Calling	
आरोग्यसेतुएप (arogya setu app)	Oxygen Concentrator	
जमातीमरीज़ (jamati mariz)	Oxygen Cylinder	
कोरोनासेवा (corona seva)	Corona Phobia	
कोविडदौर (covid daur)	Oxygen Plant	
कोरोनाएहतियात (corona ehtiyat)	Oxygen Generator	
कोरोनासमाचार (corona samachar)	Oxygen Supply	
कोरोनासेवाभाव (corona sevabhav)	Vaccination drive	
कोरोनापरिक्षा (corona pariksha)	Community transmission	
स्वास्थ्यसुविधा (savasthya suvidha)	Respiratory droplets	
संकटकाल (sankat kaal)	Crisis time	
प्रतिरोधकक्षमता (pratirodhak chamta)	Water droplets	
कोविडवायरसमहामारी (covid virus mahamari)	Crematory Space	
कोविडमुक्तभारत (covid mukt bharat)	Covid Hospital	
कोविडसेवादल (covid sewa dal)	Covid ward	
कोविडइंतज़ामियाकमिटी (covid intezamiya committee)	Covid area	
कोरोनावायरससंक्रमण (corona virus sankraman)	Covid incharge	
आपदामेंअवसर (aapda mein avsar)	Covid-19 symptoms	

Hindi words	English words	Hindi-English words
कोरोनावायरससंकट (corona virus sankat)	Red Volunteers	
कोरोनावायरसमहामारी (coronavirus mahamari)	Cowin Platform	
आक्सीजनटास्कफोर्स (oxygen task force)	New variants	
प्रधानमंत्रीकेयरफंड (pradhanmantri care fund)	Future waves	
माइग्रेंटवरकरक्राइसिस (migrant worker crisis)	Lockdown measure	
कोरोनावैश्विकमहामारी (corona veshvik mahamari)	Peak cases	
कोरोनावायरससमाचार (corona virus samachar)	Mask guidelines	
श्रमिक रेलगाड़ी (shramik railgadi)	Fully vaccinated	
दोगज़दूरी (do gaz duri)	Mask up	
धवस्तस्वास्थ्यवस्था (dhwast swastha vyavastha)	Covid fatalities	
राष्ट्रीयवैग्यानिककार्यबल (rashtriya vaigyanik karyabal)	Massive Virus	
कोविडइलाजप्रोटोकौल (covid ilaj protocol)	Anti-covid drug	
कोरोनाटीकाकिल्लत (corona tika qillat)	Genome Committee	
कोरोनाटीकाउत्पादन (corona tika utpadan)	Corona Data	
वायरसशोधकमिटी (virus shodh committee)	Vaccine Regime	
कोविडहिफाज़तीदस्ता (covid hifazati dusta)	Healthcare infrastructure	
निःशुल्कएंबुलेंससेवा (nishulk ambulance sewa)	Genome sequencing	
पूर्वकोविडयुग (purva covid yug)	Covid Protocols	
कोरोनासंक्रमणग्राफ (corona sankraman graph)	Covid Challenge	
कोरोनाटीकाकेंद्र (corona tika Kendra)	Covid Care	
सौसालहमहामारी (sau salah mahamari)	Corona report	

Hindi words	English words	Hindi-English words
असंक्रमित (asankramit)	RT-PCR test	
संक्रमित (sankramit)	Rapid Antigen	
संक्रमण (sankraman)	Corona hotspot	
टीका (tika)	Home isolation	
टीकाकरण (tikkaran)	Get Vaccinated	
तालाबंदी (talabandi)	Community Spread	
कोरोना (corona)	Corona Therapy	
कोरोनायुग (corona yug)	Vaccine Dose	
	Plasma Treatment	
	Viral Load	
	Congregate Setting	
	Incubation Period	
	Covid shot	
	N95 respiratory	
	Swab test	
	Viral shedding	
	Curve Flattening	
	Mass testing	
	Case loads	
	Oxygen bank	
	N95 Mask	
	Anti-Virus	
	Triple Mutant	
	Covid Range	
	Covid Shelter	
	Covid Manual	
	Covid Cell	
	Covid Meal	
	Covid technician	
	Covid Block	
	Covid Fund	
	Covid Guidelines	

Hindi words	English words	Hindi-English words
	Covid helpline	
	Corona helpline	
	Novel Corona Virus	
	Covid Home Testing	
	Corona Second Wave	
	Covid Second Wave	
	Corona third phase	
	Home isolation tracking	
	Covid-19 National Emergency	
	Person to Person Transmission	
	Confirmed Positive Case	
	PM Cares Ventilator	
	Vaccination Blue Print	
	Covid-19 Ground Report	
	Corona Virus Outbreak	
	Acute Respiratory Illness	
	Covid-19 task force	
	Covid Positivity Ratio	
	Suspected covid19 patients	
	Red Volunteers Helpline	
	Antibody testing kit	
	Superspreader events	
	Vicious third wave	
	Covid shot production	
	Evidence based policy making	
	New mask guidelines	
	Test positivity rate	
	Public health measure	
	Covid Special train	
	Migrant worker crisis	
	Vaccination on call	
	Presumptive Positive Cases	
	Prevent Spreading illness	
	Emergency use authorisation	
	Flattening the curve	
	Covid Victim Body	
	Corona virus hotspot	
	Testing and tracing	

Hindi words	English words	Hindi-English words
	Morbid and comorbid	
	Super Spreader Kumbh	
	Isolation tracking app	
	Highly contagious disease	
	Covid relief fund	
	Vaccination Help Desk	
	Drive through Vaccination Center	
	Drive in Oxygen Center	

Word formation type: Blending

टीकोत्सव टीका + उत्सव (tikotsav; teeka + utsav)	Cipremi
	Covaxin
	Coronil
	Covishield
	Immunocompromised
	Immunodterefericent
	Coronials
	Cowin

Word formation type: Abbreviation & Acronyms

	PPE
	PUI
	WFH
	PPM
	COVID
	SARS
	SARS-CoV-2

Medicine and Vaccine Terms

	Covaxin
	Covishield
	Cipremi
	Steroid
	Covifor
	Coronil
	Faviflu
	Remdesivir
	Sputnik
	Pfizer- BioNTech

Hindi words	English words	Hindi-English words
	Moderna	
	Johnson & Johnson's Janssen	
	AstraZeneca	
	Novavax	
	Convidecia	
Suffix		
कोरोनावाद (coronavad)		
कोरोनाहार (coronahar)		
कोरोनाग्रस्त (coronagrast)		
Prefix		
असंक्रमित (asankarmit)		
Hind Inflections in Hindi-English Words		
		कोविडवार्डों (covid wardon)
		कोरोनाइनचार्जों (corona inchargon)
		कोविडहोस्पिटलों (covid hospitalon)
		कोविडवारियरों (covid warrioron)
		कोविडबेडों (covid bedon)
		कोविडसेंटरों (covid centaron)

A Synchronic and Historical Look at Akeanon Phonology

Philip RENTILLO

De La Salle University, Philippines
philip.rentillo@dlsu.edu.ph

Ruchie Mark D. POTOTANON

University of the Philippines Visayas, Philippines
rmpototanon@gmail.com

Abstract

Akeanon features a phoneme unique relative to many Philippine-type languages, which is a reflex of the proto-Bisayan **l* and **-d-*. This was initially described as a voiced velar fricative [ɣ], and later repositioned as both a consonant of an onset and a semivowel of a coda. Half a century later it was reaffirmed as mainly a semivowel. Based on these descriptions of Akeanon, more questions arise as per the true nature of the phoneme in focus. This paper hence provides both a review and reevaluation of Akeanon phonology based on synchronic distribution, dialectology, historical accounts, and acoustic analysis. Results point to the phoneme as a velar approximant [ɰ]. Further recommendations on both descriptive and applied contexts are provided.

Keywords: Akeanon; language documentation; phonology; Philippine languages; Bisayan

Povzetek

Aklanonski jezik se ponaša s fonemom, ki je edinstveno povezan s filipinskimi jeziki preko vizajskih **l* in **-d-*. Ta je bil sprva opisan kot zveneči mehkonebni pripornik [ɣ], pozneje pa je bil definiran njegov položaj v soglasniškem onsetu oziroma polglasniški kodi zloga. Pol stoletja pozneje so ga ponovno potrdili predvsem kot slednjega. Na podlagi teh opisov aklanonščine ostajajo neraziskana vprašanja glede resnične narave omenjenega fonema. Ta članek ponuja pregled in ponovno oceno aklanonske fonologije, ki temelji na sinhroni distribuciji, dialektologiji, zgodovinskih poročilih in akustični analizi. Na osnovi teh rezultatov opišemo omenjeni fonem kot mehkonebni drsnik [ɰ]. Na voljo so dodatna priporočila tako za opisni kot uporabni kontekst.

Ključne besede: aklanonščina; jezikovna dokumentacija; fonologija; filipinski jeziki; vizajski jeziki



1 Introduction

Mention of Akeanon in the historical literature began quite recently relative to its neighboring varieties (and languages), which have been reported or analyzed nearly two hundred years earlier. The first known published reference of the language dates back only to the mid-19th century (de Méntrida, 1841). What makes this case peculiar is that historians and philologists at that time should have long noticed the language, given a presumably sizeable speaker base alongside distinguishable features. One notable innovation in Akeanon—a correspondence with /r-l/ of the Bisayan (henceforth, Bis) group—is what was initially described as a voiced velar fricative [ɣ] (Scheerer, 1920). This was later argued as both a consonant and a semivowel (de la Cruz & Zorc 1968), and then half a century later was reaffirmed as a semivowel (Zorc, 1995, 2005) or more specifically a velar approximant [ɰ] (Zorc, personal communication, September 27, 2019). This innovation has long been represented orthographically with ⟨Ee⟩. Based on current descriptions of Akeanon alongside those of related and adjacent languages, more questions arise as to the true nature of the mentioned phonological feature. This paper attempts to revisit these facets of the Akeanon phonological system through evidence all accumulated from synchronic, dialectological, historical, and acoustic evidence.

1.1 The Akeanon language

1.1.1 Status

Akeanon (ISO 639-3: akl), also called *Aklanon*, *Inakeanon*, *Binisaya*, or *Bisaya (nga Akean)*, is a West Bisayan (henceforth, WBis) language of the Central Philippine group (Zorc, 1977; Eberhard et al., 2022). It is close to Kinaraya-a, the most widely spoken WBis, and distantly related to Hiligaynon, Waray, and Cebuano. Its speakers are concentrated within the northwestern tip of the island of Panay in the Philippines, particularly within the province of Aklan. Since the last 2000 household population census by the National Statistics Authority¹ (2002) there could be more than 500,000 speakers of the language today. It should be noted as well that this estimate was only based on ethnicity-based self-reports, and not on intelligibility or on language proficiency.

Akeanon according to the Expanded Graded Intergenerational Disruption Scale (EGIDS) is considered an educational language or Level 4, well within the range of languages classified as INSTITUTIONAL, or those used and sustained by institutions from and beyond the native-speaking community. As such, Akeanon is predominantly spoken in radio broadcasts² across the province of Aklan, and apart from use on known

¹ Officially named the National Statistics Office when the 2002 census was released.

² DYCF 88.5 FM Radyo Todo Aklan, DYYK 89.3 FM Brigada News FM Kalibo, DYRU 92.9 FM Barangay RU Super Radyo Kalibo, DYDJ 101.7 FM Energy FM Kalibo

social media platforms such as Facebook³ and Twitter, it is available in online news content particularly on the Philippine Information Agency⁴. From time to time, it is used in public signages of both commercial establishments and the local government. Its written presence is also seen to grow with the emergence of academic texts due to the recent inclusion of Akeanon into the Mother Tongue-based-Multilingual Education (MTB-MLE) scheme as a medium of instruction and as a mother tongue subject for primary schools in Aklan as mandated by the Department of Education.

1.1.2 Phonology

Akeanon has a phonological system typical of a Central Philippine language. Its canonical vowel inventory is limited to three, [a~e], [u~o], [i~ɪ], a common 3-vowel inventory among Bis languages, while [ɔ] and [ɛ] are included to accommodate common nouns, and loanwords historically from Spanish, and more recently from Tagalog and English. It features seventeen (17) native consonants, with an additional seven (7) due to both loanwords, and phonological processes. Tables 1 and 2 below illustrate the current phonological system of Akeanon based on the findings of de la Cruz and Zorc (1968) and Zorc (1995).

Table 1: Akeanon vowel inventory

	Front	Central	Back
Close	i ~ ɪ		u ~ o
Open-Mid	(ɛ)		(ɔ)
Open	a ~ e		

We consider Table 2 as the most detailed inventory on the language as to date and thus serves as the benchmark of our current analysis since it combines descriptions from de la Cruz and Zorc (1968) and Zorc (1995).

³ As of January 2022, existing groups include *Akeanon Language - Panghinambaeang Akeanon*, *Akeanon Lines*, and *Ro Akeanon*.

⁴ The PIA website as of January 2022 does not organize articles by language but only by region.

Table 2: Akeanon consonant inventory

	Bilabial		Alveolar		Post-Alveolar		Palatal	Velar		Labiovelar	Glottal
Stop	p	b	t	d				k	g		ʔ
Nasal	m		n					ŋ			
Affricate			(ts)	(dz)	(tʃ)	(dʒ)					
Fricative	(f)	(v)	s	(z)	(ʃ)			ɣ?			h
Approximant							j	ɰ?		w	
Tap			r								
Lateral			l								

A distinctive feature of Akeanon roughly described as velarized or guttural is not present in any other languages within its geographic locus. The reflex is described varyingly throughout Akeanon scholarship as a VELAR APPROXIMANT [ɰ]⁵ or as a VOICED VELAR FRICATIVE [ɣ], which will thus be tentatively labelled as the DISTINCT AKEANON REFLEX (henceforth, DAR). What is very certain, though, is that as per Zorc (1977) DAR is a reflection⁶ of the proto-Bis **l* (e.g. **lanjaw* > *Ranjaw* ‘housefly,’ **balu* > *baRu* ‘widow,’ **katul* > *katuR* ‘itch’) and intervocalic **-d-* (e.g. **wadaʔ* > *waRaʔ* ‘none, lose’).

1.2 The current study

Although Akeanon phonology has been described by de la Cruz and Zorc (1968) in detail, questions remain as to the true nature of its distinct “guttural” phoneme (i.e. DAR), orthographically (Ee). The earliest evidence of the use of the letter for the phoneme could be traced back as early as the late 20th century. It should likewise be noted that the same grapheme is used for proper nouns (e.g. *Teresa*, *Balete*), and for loanwords (e.g. *eroplano*, *puwede*) intended to be pronounced with an open-mid front unrounded vowel [ɛ]. While this already suggests a host of orthographic issues, perhaps a more pressing concern is the lingering notion among common Akeanon folk, in schools, and even among literary intellectuals, that the distinct phoneme is a vowel. This writing convention has been observed as early as the turn of the 20th century by Scheerer (1920), whereby its origin as an orthographic convention remains unclear up to this day. In terms of formal linguistic analysis, the most recent mention of DAR is by Zorc, who

⁵ Martínez-Celdrán (2004) raised the issue on the use of this symbol in Spanish, Catalan, and Galician phonetic contexts whereby [ɣ] is said to be more appropriate. Likewise, the alternative symbol is also preferred in the transcription tradition of other European languages with a similar phoneme (e.g. Danish, Icelandic, Swedish).

⁶ /l/ and prevocalic /d/ are also phonemic in Akeanon (e.g. *bulag* ‘blind’ vs *buRag* ‘[to] separate,’ *madajaw* ‘elegant,’ *pulgada* ‘inch’) while pre- and postvocalic **d-*, **-d* coalesced into /d/ (e.g. *daRan* ‘road,’ *bukid* ‘mountain’).

in footnote argued it not to be a voiced velar fricative—the recognized sound class membership for the longest time—but instead an unrounded back semivowel. Zorc (2005, 128) claims such because “...it [the consonant in question] lacks friction” (p. 128).

The points above provide an impetus for this study to take a three-pronged approach to surrounding the DISTINCT AKEANON REFLEX (DAR). Specifically, we problematize on whether DAR is a velar approximant [ɰ] or a voiced velar fricative [ɣ] through the following.

6. A review of recent and past discussions on Akeanon phonology;
7. Historical records relating to the Akeanon language; and
8. Articulatory and acoustic descriptions of DAR.

2 A review of approaches to Akeanon phonology

2.1 Descriptions

The first known published linguistic inquiry on Akeanon dates to Otto Scheerer’s (1920) paper entitled *Über einen bemerkenswerten L-Stellvertreter im Dialekt von Aklan auf der Insel Panay (Filipinen)*, lit. ‘On the remarkable L-variant in the dialect of Aklan in the island of Panay (Philippines).’ Scheerer, then working for the American colonial National Museum of the Philippines, was interested in what he calls “einen leicht stimmhaften Gaumen-Reibelaut” (p. 249), lit. ‘a slightly voiced velar (palatal) fricative.’ He reports that this sound is attested among speakers of a variety of Bisaya found within and around the Aklan River valley, once only a recognized cultural area within the province⁷ of Capiz.

Transcribing the phoneme as ê, he compares its distribution patterns to Bis cognates through the RLD (and RGH) laws described by Conant (1911, 1912) among “Indonesian”-type languages. This analysis shows that DAR is the Akeanon non-velar or /l/ reflex found in other Bis varieties such as Hiligaynon. Scheerer further enumerates lexical items with distinct /l/-reflexes in Akeanon in the same phonological environments showing that DAR is phonemic in the language. He also presents morphophonological evidence such as retention of /l/ when a velarized root is inflected (e.g. *eopad* ‘fly’ > *linupad* ‘flown away’). This phonological pattern further substantiates DAR’s consonantal yet non-velar origins. From here on the phonological description of DAR as a velar (fricative) persisted in most of Akeanon linguistic scholarship.

Almost fifty years would pass before an extensive description of Akeanon phonology would be published in English. The work in question is part of a two-book series with Zorc as a lead scholar for the United States Peace Corps. The first book is a

⁷ Aklan was once considered a mere cultural area within the Capiz Province before it seceded as a separate province on April 25, 1956, through Republic Act No. 1414.

grammatical description, which includes a phonological sketch (de la Cruz & Zorc, 1968). The second book a year later is a dictionary (Salas Reyes, Zorc, & Prado, 1969). In both books, DAR is represented with a minuscule Latin *g* with a diagonal slash (<g>) and described as a voiced velar fricative functioning as a consonant and a word-final semivowel. Less than a decade later Zorc (1977) mentions DAR once more and describes it as a “voiced velar spirant (with only some friction)” (p. xvii) transcribed as /t/ in his landmark work on the subgrouping of Bis and reconstruction of proto-Bis. This paper would be followed by Paz (1981), who reconstructed proto-Philippine phonemes and morphemes. Included in her work was a look into velarized features sparsely found throughout genetically and geographically scattered Philippine languages. This was where DAR, described as a “voiced velar fricative” (ibid, p. 23) and represented similarly in Zorc’s two-volume series, was among those used as basis for a proposed proto-velar as an additional reflex of her hypothesized Philippine proto */.

Zorc (2005) later on passingly mentions the decision not to consider the voiced velar fricative entirely as a consonant, but as a semivowel⁸. In his footnote, he says that contrary to his initial interpretation of a voiced velar fricative it is apparently an “unrounded back semivowel” (ibid, p. 128) due to its lack of turbulence. He does not mention any work expounding on such lack of “friction,” or acoustic data for substantiation. Later through another correspondence, he reaffirms his stance (Zorc, April 5, 2021, personal communication).

2.2 Historical accounts

While Scheerer’s 1920 article is the first known published linguistic inquiry on DAR, there are older written accounts on Akeanon with scant mentions of the reflex in focus. One earlier known text is by Alonso de Méntrida dating back to 1841, where he briefly mentioned such “peculiar feature,” which may be surmised to refer to DAR in comparison to other Bis varieties known during that time. There is no older document or artifact within this period that provides a detailed description of DAR let alone a mention of a distinct variety of Bis spoken within what is today the province of Aklan.

Akeanon folk history points to Borneo as the progeny of the Bisayan people, hence including Akeanons. It was believed that their ethnolinguistic ancestry traces back to two datu who were said to have had speech defects and were among the ten Bornean royalties who sailed with their consorts and servants northwards to Panay. This is recounted in various texts from the late 19th century to the early 20th century. The most

⁸ He transcribed this as a semivowel in his 1995 chapter with the lateral fricative /t/.

notable is a Hiligaynon⁹ narration by Monteclaro in his historical¹⁰ book widely referred as *Maragtas*¹¹:

Ang Pulong nga Ila Ginagamit

Ang ila pulong amo gid man ang binisayâ nga nasaktan sang hiligaynon kag hiniraya, kay ang mga taga-Bornay nga nag-alabut diri sini nga pulô, gumikan man sa nagasarisari nga pungsod sang Bornay, ugaling sang olihe nagpininigpinig ang nagahambal sang hiligaynon kag hiniraya kay ang nagapuyô sa Aklan nga mga pihit kay mga kaliwat ni Bangkaya kag ni Balinganga nga pihit man, dili matigdà sa pulong nga hiniraya nga sa m[a]sunsun ginagamit ang R kag ang L nga dili mamitlang sing ma-ayo sang mga pihit; gani man gumikan sa aklananon ang pulong nga lunsay hiligaynon nga sang olihe naglapnag sa Irong-irong tungud nga maramù nga tagadiri ang nakapangasawa sa Aklan, subong nga madamò man ang taga Aklan nga nakapangasawa sa taga Irong-[i]rong (Monteclaro, 1957, p. 42).

The Language They Use

Their [Akeanon people's] language is the same as the Binisaya such as those from where the river flows (cf. modern Hiligaynon) and those from the hinterlands (cf. modern Kinaray-a). When the people of Borneo, hailing from various Bornean nations, arrived at the island (of Panay) those who spoke Hiligaynon and Kinaray-a later on grouped together. Meanwhile, those who settled in Aklan had speech difficulties since they were descendants of (Datu) Bangkaya and (Datu) Balinganga, who had speech defects themselves. They were not fluent in Kinaray-a, which usually uses R and L, which in turn cannot be pronounced correctly by people with such defects. From the people of Aklan came words that were essentially Hiligaynon that later spread to Iloilo since many of those who migrated here intermarried with those from Aklan, and the same way many Akeanon people intermarried with those from Iloilo.

An earlier narrative comes from Spanish friar Tomas de Santaren (1856 in Pérez, 1902, p. 392) explaining how ancient Akeanon speakers appropriated the supposed speech defect of their figure head as a linguistic feature:

Como Bang-caya era zazoso en el pronunciar, así salió el hijo, y hasta en la actualidad lo sen todos los naturales de aquel partido, (no porque en realidad lo

⁹ In 20th century non-standard orthography.

¹⁰ Although Monteclaro intended his work to be considered valid history, linguistic and archeological (counter-) evidence up to this day relegate the narrative on the Ten Bornean Datus as pure myth.

¹¹ Full title is *Maragtas: Kon Sayuron (historia) sg pulô nga Panay kag sang mga pumuloyo, tubtub sang pag abut sang mga taga Borneo nga amô ang quin halinan sg mga bisayâ, kag kutob sang pag abut sg mga kastilâ.*

sean, pero si, por el dejo ó estribillo quen han tomado y siguen de sus antepasados.)

Because Bangkaya stutters in speech and so his son, it has become common to everyone native to that place [Aklan] (not because they all stutter as well, which indeed cannot be the case, but because they took it from their ancestors).

de Santaren (1856 in Pérez, 1902, p. 396) further explains the linguistic origins and diffusion of Bis in accordance with the myth of the ten datus from Borneo. His text could be considered as the earliest mention of a certain Akeanon lect distinct from other Bis varieties:

De este reparto que hizo el Dato Somacuel, entre loes restantes Datos, procedieron todas las poblaciones, así como de estas poblaciones, y ancianos proceden los que hoy existen, y de aquí la variación ó diferencia en algun tanto del idioma Bisaya, y la variación do los nombres como: Aniinjanon, Ilanodnon, Irajaynon ó Buquidnon, Aclanon, Subuanon, Buluanon, Cagayanon, Coyun-òn etc, etc.

The divisions created by Datu Sumakwel among the remaining datus resulted in these settlements, which are both ancient and continue to exist until today. Here they have variations and differences in their respective Bisayan language, which are named as ¹² Aminjanon, Ilanodnon, Irajaynon or Buquidnon, Aclanon, Subuanon, Buluanon, Cagayanon, Coyun-on and many others.

We then raise two main issues as regards the relatively late report of such feature, which has nonetheless given rise to the subsequent recognition of Akeanon as a distinct Bis variety. First, DAR is linguistically unique based on genetic and areal considerations within and beyond the Bisayan linguistic ecology. As of Paz' (1981) reconstruction of a proto-Philippine¹³ phonology, there have only been four geographically and genetically incongruous Philippine-type languages to feature a distinct reflex phonetically similar to that of Akeanon: Itbayat, Kalinga, Virac Bikol, and Bahi Barubu Manobo. Anderson (1958 in McFarland, 1975) also cites Buhinon to feature a voiced velar fricative. Likewise, one can assume that there could have been a sizeable proportion of speakers from and around the Aklan River valley out of approximately 140,000 people living in the province of Capiz during or before 1916, the year when Beyer's (1917 in Scheerer, 1920) census of the Philippine colony was conducted before the split of Aklan in 1956 into a separate province. It would be impossible not to encounter a speaker having this phonemic repertoire even if it could have been considered as an idiolect of low prestige relative to nearby statistically dominant varieties such as Kinaray-a and the much more

¹² Bisayan varieties spoken in present-day coastal Capiz, Jalaur (Halawod) River basin, Panay uplands, Aklan, Cebu, Butuan/Bohol (?), Cagayan de Oro River, and the Cuyo Archipelago, respectively, which may not accurately correspond to documented languages in these respective areas.

¹³ Issues on a unified Philippine group under Malayo-Polynesian (cf. Blust, 2005, 2019; Reid, 1982, 2018) are beyond the scope of this paper, and thus merely echo the subgrouping claimed by Paz.

politically and economically dominant Hiligaynon. Linguistically, de la Cruz and Zorc (1968) who considered¹⁴ DAR as a voiced velar fricative, argue that words in what could have been proto-Bis */ and *-d- have been reflexed to /ɣ/ in Akeanon. They deduce that “this change is also a phenomenon which has occurred more recently” (p. 15).

2.3 Distribution

We now discuss both distribution patterns of DAR by taking advantage of initial data from Scheerer (1920), and lexical entries from Salas et al.’s (1969) Akeanon dictionary. Within regular patterns, the phoneme either as onset or coda forms a syllable with either of the two canonical vowels, the open front unrounded [a], the close back unrounded [u] for root words. Only in certain morphological conditions does it emerge with the close front unrounded [i], but specifically within an onset + nucleus patter (i.e. /Ri/). These syllables vary in position: Initial, medial, and ultima. These are compared to cognates with what we call the COMMON BISAYAN REFLEX (CBR). Sound patterns within this section are represented in broad vocal transcription (i.e. /a, u/) for consistency and clarity, while Akeanon cognates featuring DAR are spelled with ⟨Ee⟩.

Table 3: Regular CBR versus DAR correspondences

	Onset + /a/	/a/ + Coda	Onset + /u/	/u/ + Coda	Onset + /i/
CBR	/la/	/al/	/lu/	/ul/	/li/
DAR	/Ra/	/aR/	/Ru/	/uR/	/Ri/

Note. The *R* in all tables do not represent a proto-R phoneme from any reconstruction, but rather for brevity to represent “reflex” in the “DAR” abbreviation, which represents the modern Akeanon reflex.

2.3.1 /a/ nucleus

This first part provides examples of correspondences between DAR (as either an onset or coda) and cognates with CBRs containing /a/ as nucleus. Syllable patterns with onset DAR and /a/ nucleus occur in lexical items across initial, medial, and ultima positions as provided in Tables 4 and 5, respectively.

¹⁴ Zorc’s stand has changed since then into a voiced velar approximant.

Table 4: Onset DAR + /a/ nucleus

Position	CBR cognate	Akeanon cognate	Gloss
Initial	<i>lawud</i>	<i>Rawud</i>	ocean, sea
	<i>landuŋ</i> <small>[CEB]</small>	<i>Randuŋ</i>	shadow, shade
	<i>lapuk</i> <small>[CEB; WAR]</small>	<i>Rapuk</i>	mud
	<i>laʔin</i>	<i>Raʔin</i>	other/s, different
Medial	<i>kalaju</i>	<i>kaRaju</i>	fire
	<i>dalagan</i> <small>[HIL; CEB]</small>	<i>daRagan</i>	(to) run
	<i>balanʔaw</i>	<i>baRanʔaw</i>	rainbow
	<i>hulas</i>	<i>huRaguk</i>	(to) sweat
Ultima	<i>ŋalan</i>	<i>ŋaRan</i>	name
	<i>wala</i>	<i>waRa</i>	left (deixis)
	<i>pula</i>	<i>puRa</i>	red
	<i>bulak</i>	<i>buRak</i>	flower

Note. CEB – Cebuano, HIL– Hiligaynon, WAR – Waray, KRJ– Kinaray-a

Table 5: /a/ nucleus + DAR coda

Position	CBR cognate	Akeanon cognate	Gloss
Initial	<i>halʔu</i> <small>[HIL; KRJ]</small>	<i>haRʔu</i>	mortar
	<i>galʔum</i>	<i>gaRʔum</i>	overcast, nimbus cloud
	<i>balʔag</i>	<i>baRʔag</i>	file (tool)
Medial	<i>kasalʔanan</i>	<i>kaSaRʔanan</i>	sin, culpability
Ultima	<i>hambal</i> <small>[HIL]</small>	<i>hambaR</i>	(to) speak, tell, speech

2.3.2 /u/ nucleus

This second part provides examples of correspondences between DAR (as either onset or coda) and cognates with CBRs containing /u/ as nucleus. Syllable patterns with /u/ and an onset DAR occur in lexical items across initial, medial, and ultima positions.

Table 6: Onset DAR + /u/ nucleus

Position	CBR cognate	Akeanon cognate	Gloss
Initial	<i>lubung</i>	<i>Rubun</i>	(to) sink
	<i>luŋib</i>	<i>Ruŋib</i>	cave, cavern
	<i>lukus</i> <small>[HIL; KRJ]</small>	<i>Rukus</i>	squid
	<i>luhaʔ</i>	<i>Ruhaʔ</i>	tear (eyes)

Position	CBR cognate	Akeanon cognate	Gloss
Medial	<i>dulungan</i>	<i>duRunḡan</i>	ear
	<i>hulubaton</i>	<i>huRubaton</i>	proverb, aphorism
	<i>ʔulunlan</i>	<i>ʔuRunlan</i>	pillow
	<i>pumulujuʔ</i>	<i>pumuRujuʔ</i>	citizen
Ultima	<i>buluḡ</i>	<i>buRuḡ</i>	medicine
	<i>talum</i>	<i>taRuḡ</i>	blade, sharp object
	<i>ʔulu</i>	<i>ʔuRu</i>	head
	<i>(na)puluʔ</i>	<i>(na)puRuʔ</i>	ten

Table 7 below shows that similar to /a/, syllables with /u/ + DAR coda only occur within ultima positions.

Table 7: /u/ nucleus + DAR coda

CBR cognate	Akeanon cognate	Gloss
<i>datʔul</i>	<i>datʔuR</i>	(to) put on a surface
<i>Habul</i>	<i>habuR</i>	cloth, blanket, (to) weave
<i>ʔumul</i> <small>[CEB; WAR]</small>	<i>ʔumuR</i>	(to) form, shape
<i>Buḡul</i>	<i>buḡuR</i>	deaf

2.4 Other patterns

2.4.1 Morphophonotactics

Inflected roots with DAR reveal their proto-Bis correspondences and are seemingly regular in occurrence as shown in Table 8 below. Zorc (1995) notes how /y/¹⁵ typically changes based on three phonological conditions (see Tables 8 and 9).

Table 8: DAR morphophonemic conditions

Type	Change	Condition
1	/R/ > /l/	If DAR is in a syllable with an apical consonant (i.e. /t, d, n, s/) and undergoes hyperthesis
2	/R/ > /l/	If DAR is followed by a close front vowel /i/ notably via infixing (e.g. <i>-in-</i>)
3	/R/ > /d+ʔ/	If onset DAR undergoes hyperthesis with a coda glottal stop /ʔ/ or \emptyset coda

¹⁵ y is the symbol Zorc used in that particular paper.

Conditions 1 and 2 reveal DAR as a reflex of the proto-Bis **l*. Condition 3 involving an epenthesis permits DAR to revert into proto-Bis **d*.

Table 9: DAR morphophonemic examples

Type	Root	Root gloss	Inflected	Inflected gloss
1	<i>putuR</i>	(to) cut	<i>putla</i>	cut (imp.)
	<i>suRud</i>	(to) enter	<i>sudlan</i>	to be entered
2	<i>Ragaʔ</i>	(to) boil	<i>linagaʔ</i>	boiled
	<i>Rubuŋ</i>	(to) bury	<i>linubuŋ</i>	buried
3	<i>waRaʔ</i>	none	<i>nawadʔan</i>	left without
	<i>maRaʔ</i>	dry	<i>madʔan</i>	dried out
	<i>pasipaRa</i>	(to) cuss, curse	<i>pasipadʔan</i>	to be cussed (at)
	<i>RaRaʔ</i>	smarting pain	<i>linadʔan</i> ¹⁶	of great pain

2.4.2 Rule exceptions

It was previously mentioned that roots with /a/ or /u/ as collocates feature DAR. The common Bis /-l+V₁-/ infix encoding the abilitative/potentive or emphatic mood is reflexed with the nucleus of the initial syllable /-DAR+V₁-/ in Akeanon. Due to possible regularization tendencies, this applies when inflecting syllables with /i/ as a nucleus.

Table 10: DAR regularization in inflection

Root	DAR	CBR	Gloss
Tipun	<i>pagtiRipun</i>	<i>pagtilipun</i>	Gathering
Hilu	<i>makahiRilu</i>	<i>makahililu</i>	Poisonous

2.5 Variation

Distribution patterns however only represent the common form. In a preliminary dialectological survey, Rentillo (2018) so far identified that this reflex is evident in almost all Akeanon speaking areas, especially in the provincial capital Kalibo and its peripheries along the Aklan River valley, where much of the population is concentrated. One notable outlier would be the northeastern variety in Nabas, which as per folk accounts, is known to possess the /l~r/ reflex of DAR. Enumerated in Table 11 below are examples of lexical items with the Nabas reflex alongside the more common DAR.

¹⁶ Combination of type 2 and 3 conditions.

Table 11: Nabas /l~r/ correspondences

Form	Position	Nabas	DAR form	Gloss
l-form	C-	<i>lambat</i>	<i>Rambat</i>	Net
		<i>lujʔa</i>	<i>Rujʔa</i>	Ginger
	-C-	<i>bulan</i>	<i>buRan</i>	moon, month
		<i>buluŋ</i>	<i>buRuŋ</i>	medicine
	-C	<i>hambal</i>	<i>hambaR</i>	(to) speak
		<i>bahul</i>	<i>bahuR</i>	big, large
r-form	-C-	<i>baras</i>	<i>baRas</i>	sand
		<i>ʔuran</i>	<i>ʔuRan</i>	rain
		<i>turug</i>	<i>tuRug</i>	(to) sleep
l~r-form	-C-	<i>talamnan, taramnan</i>	<i>taRamnan</i>	garden
		<i>dalagan, daragan</i>	<i>daRagan</i>	(to) run
		<i>(ma)hulug, (ma)hurug</i>	<i>(ma)huRug</i>	(to) fall

Based on initial data, Nabas /l/ occurs pre-, post-, and intervocalically while /r/ is confined within intervocalic positions. This /l~r/ reflex suggests an extra-Akeanon (henceforth, EA) influence possibly of Kinarayan or proto-WBis origins. There are also lexical items where both intervocalic reflexes are in complementary distribution (e.g. *talamnan, taramnan*) but patterns are inconsistent possibly due to lexicalization brought by said EA variety. For example, ‘road, pathway’¹⁷ is *daRan* < proto-Austronesian **zalan* (Blust, 1999) and *ʔayagjan* < proto-Bis **qagi* (Zorc, 1977). Compare this with *dalan* and *ʔaragan* in Nabas.

**Figure 1:** Location of Northwestern Panay in the Philippines

Historical anecdotes point to cross-generational contact among communities in northwestern Panay (see Figure 1) particularly those in Pandan (Antique), a Kinarayan speaking area, and nearby areas of Aklan (e.g. Nabas, Ibajay). This part of Panay is a key transit route linking northern Antique and the rest of Aklan. The latter serves as a more

¹⁷ Loaned variant is *karsada* from Spanish *calzada* ‘paved road’ (cf. Tagalog *kalsada*).

important economic and logistical hub for residents of bordering towns of Antique since they are geographically more distant to the provincial center in San Jose de Buenavista, which is 110 km away or nearly a 3-hour drive south. Malay and Kalibo, two major economic centers of Aklan, are just 30 km and 55 km away, respectively.

2.6 Crosslinguistic comparison

As a counterevidence against proposals of a single Philippine subgroup under Malayo-Polynesian, Zorc (2021) argues for an alternative explanation for the current linguistic macro-ecology of the Philippines through what he calls *axes*, or relationships of different languages and subgroups based on areal innovations, but which do not lead to genetic subgroupings. They are rather more related to Sprachbunds and linkages (Pawley & Ross, 1995; Ross, 1988). This may be an important point of discussion on the nature of DAR through its crosslinguistic position amid WBis and neighboring groups.

Akeanon according to Zorc belongs to the North Bisayan¹⁸ Axis (NBAXIS). Two NBAXIS members with a /j/ reflex of proto-Bis **l*, **-d-* are Romblomanon and Asi¹⁹ (see Table 12). Romblomanon is grouped under Central Bisayan (CBis), while Asi also called Banton is its own first order branch directly under Bis.

Table 12: Romblomanon, Asi /j/ correspondences

Common Akeanon	Romblomanon, Asi	Other Bisayan	Gloss
<i>waRu</i>	<i>waju</i>	<i>Walu</i>	eight
<i>duRaw</i>	<i>dujaw</i> ^[ROM] , <i>rujaw</i> ^[ASI]	<i>dulaw</i> ^[WAR]	yellow
<i>kuRanj</i>	<i>kujanj</i>	<i>Kulanj</i>	insufficient
<i>daRan</i>	<i>dajan</i> ^[ROM] , <i>rajan</i> ^[ASI]	<i>Dalan</i>	road

Note. ROM – Romblomanon, ASI – Asi, WAR – Waray

Both DAR and /j/ in ASI reflect proto-Bis **l*, while ROM /j/ is a reflex of intervocalic and coda **l*. Likewise, DAR and /j/ in both ROM and ASI are reflexes of proto-Bis intervocalic **-d-*. Compare these with reflexes of other Bis varieties within the NBAXIS in Table 13 (Zorc, 1977, p. 203).

¹⁸ The North Bisayan Axis covers a geographic perimeter much wider than WBis spanning across the southern coasts of Mindoro and Calamian Islands across Sibuyan Sea all the way to the Bikol region. NBAXIS includes WBis, CBis (excluding Warayan), Asi, Bikol languages, Hanunuo (*Southern Mangyan*), and Kagayanen (*Manobo*).

¹⁹ Data for comparison is from the Bantoanon (BAN), Odiongonon (ODG), and Sibalenhon (SIB) dialects.

Table 13: Select NBAxis correspondences reflecting proto-Bis

	Bis Subgroup	*-/	*-/	*-/	*-d-
Akeanon ²⁰	WBis	R	R	R	R
Asi (ODG, BAN, SIB)	Asi	J	j	j	j
Romblomanon	CBis	l~j ²¹	j	j	j
Hiligaynon, Capiznon	CBis	L	l	l	l
Kinaray-a	WBis	L	l	l	r
Inonhan	WBis	L	l	l	r
Cuyonon	WBis	L	l	l	r
Ratagnon	WBis	L	l	l	r
Minasbate	CBis	L	l	l	r
Northern Sorsogon	CBis	L	l	l	r
Southern Sorsogon	CBis	L	l	l	r

Based on the above proto-Bis correspondences, two Akeanon reflexes, both of which velarized as per existing documentations, will be historically and articulatorily analyzed: **l* > *R* and **-d-* > *-R-*. This velarization process in totality could have been facilitated by hypercorrection (Ohala, 1993 in Bybee, 2015) or automatization factors (Bybee & Easterday, 2019). There are two possible intermediate processes leading both proto forms into modern DAR. One is via palatalization and the other via relateralization.

2.6.1 Palatalization to velarization

The first scenario is an intermediary palatalization prior to velarization, thus trajectories of:

1. **l* > */j/* > *R*
2. **-d-* > */-j-/* > *-R-*

Palatalization is crosslinguistically a very common phenomenon as attested in Indo-European (e.g. Romance, Slavic), Chinese, Bantu, and Semitic (Bateman, 2011; Bybee, 2015). Lateral approximants undergoing palatalization are among the most common of this process. An earlier *l*-form Akeanon could have independently undergone **l* > */j/* alongside earlier Asi and ROM that recently split from CBis. All three within NBAxis could therefore be relics after other neighboring varieties switched to an */l/* reflex. Alternatively, this earlier *l*-form may have been influenced by a nearby NBAxis variety (e.g. ROM, Asi) that palatalized much even earlier. Greater linguistic diversity

²⁰ Contrastive with */l/* and intervocalic */-d-/* (e.g. *limpjo*, *ʔalima*, *ʔulij*, *sutil*, *baril*, *pulgada*, *sida*) in most of which are recently-introduced words or loans.

²¹ See French (1979).

and internal variation in northwestern Visayas surrounding Akeanon may suggest these varieties underwent sound change much earlier and that DAR is a relatively recent development (de la Cruz & Zorc, 1968).

There are sufficient crosslinguistic evidence (cf. Bhat, 1978, Bateman, 2011) pointing to how coronals especially dentals and alveolars (e.g. /d/) are inclined to undergo palatalization, a process which Parrell and Narayanan (2018) refer as *coronal reduction*. This could explain *-d- > /-j-/ in ROM, ASI, and Surigaonon, and to an extent *-d-²² > Ø²³ in Surigaonon and Cebuano varieties (e.g. proto-Bis *ʔudan > ʔujan, ʔu(w)an ‘rain’). Weakening of stops into approximants is an aerodynamically natural process based on the concept of markedness (Nagle, 2014). Voiced stops (e.g. /d/) for example are articulatorily complex to produce due to the required effort in the glottis for vibration and airflow in coordination with oral closure and release. Lenition into, say, an approximant (e.g. /j/) is articulatorily less effortful but is perceptually less marked as a trade-off.

But how could palatalization result in velarization? The interaction of an introduced [DORSAL] /j/ and pre-DAR [CORONAL] /l/ could have led to assimilation based on the need to perceptually and articulatorily “recalibrate” or reanalyze (see Bybee, 2015) both approximants as they started to overlap in the inventory during the initial period of sound change. Keating (1988 in Jagers, 2018) argues that [j] is *both* [DORSAL] and [CORONAL], which can explain for this tendency to velarize. On another note, Browman and Goldstein (1995) argue that the tip (coronal) and back (dorsal) of the tongue are mechanically interconnected such that “in initial position, the tongue tip and the tongue dorsum gesture are roughly synchronous, whereas word-finally, the wider tongue dorsum constriction precedes the narrower tongue-tip closure” (ibid, p. 2). In the case of Akeanon, coronal articulation might have dwindled due to palatal strengthening until it stabilized with an emphasized velar gesture. This strengthening is analogical to a morphophonemic process in Kimatumbi (*Bantu*) where postnasal glides [w, j] and liquids [l] undergo “hardening” (Odden, 2015, p. 12) into [d, d³, g^w], respectively (e.g. [júba] > [n-d³ɔbɛ́] ‘hide [1sg subjunctive]’). Other Bis languages surrounding eastern Bohol Sea are known for a dorsal-to-coronal shift via fortition. Boholano, Leyteño Cebuano, Surigaonon varieties (Zorc, 1977), and Baybayanon (Rubino, 2006) feature a *j- > /dʒ/ reflex while Porohanon has /z/ (Santiago, 2018; Wolff, 1967) whereby (proto-)Bis *jawaʔ > dzawaʔ, zawaʔ ‘devil.’

2.6.2 Relateralization to velarization

This second scenario posits that Akeanon first palatalized regardless of its split from proto-Bis and regardless of the influence by a neighboring *j*-form variety as in §2.6.1. It

²² Including *-l-, *-l

²³ Then resulting to vowel lengthening V: in some lexical items

then relateralized which motivated /j/ reflexes of */ and *-d- to move farther back from the hard palate to the velum, hence trajectories of:

1. */ > /j/ > /j~l/ > R
2. *-d- > /-j-/ > /-j~-l-/ > -R

A palatal-to-velar maneuver in Akeanon could have been prompted by a second phase of EA contact in the nearer past by a different prestige variety, possibly Hiligaynon as it only features the /l/ reflex for both */ and *-d-, hence a relateralization stage. ROM and Asi /j/ (or including Akeanon) could then be relics of post-Hiligaynon expansion in Western Visayas. French (1979) reports a recent shift in proto-Bis */- reflex in Romblomanon. Townsfolk would tend to add, although irregularly, a prevocalic /l-/ whereby *jaki* ‘male’ > *ljaki*. This is suggested to be influenced by Hiligaynon, the regional trade language (or Spanish, Tagalog, and English due to loaning). Meanwhile “barrio” speakers (more rural?) maintain the /j-/ reflex, which implies that the original ROM prevocalic reflex of */- is /j/. Zorc (1977) on the other hand describes this only as a /l-/ (recall Table 13) which may only speak for the representative dialect he sampled in his study. Alternatively, lateralization could have been transmitted via Capiznon. Aklan for much of its known history was an administrative component of Capiz until only in 1956 when it gained provincehood. This might have been a crucial element in shaping the community life area²⁴ of Akeanon speakers (cf. Nabas variety in northwestern Panay) and thus their linguistic repertoire. This is assuming that Capiznon is truly a CBis variety in a very close relationship with Hiligaynon akin to a mini-axis, or it could rather be an old WBis variety with a Hiligaynon superstratum. Either way, its current status remains highly controversial and necessitates further documentation.

Velarization of liquids has been attested in a wide range of languages such as Italian, Dutch, Portuguese, Puerto Rican Spanish, Sasak (cf. Flynn, 2012), languages of New Guinea, and East Chadic (cf. Lagdefoged & Maddieson, 1996). This may also explain why the velar alveolar lateral [ɫ] widely known as “dark L” is crosslinguistically typical. In one sociophonetic experiment, for example, Moosmüller and colleagues (2015) found an ongoing shift from alveolar lateral [l] to velarized lateral [ɫ] in Viennese German (e.g. [læ:ɖɛr] > [ɫæ:ɖɔ] ‘unfortunately’).

2.6.3 Nabas Akeanon reflex

The Nabas variety meanwhile could be a relic after exposure to the /l-form EA variety. It is not clear though whether Akeanon speaking inhabitants of Nabas were influenced by an EA variety, or their lineage traces to migrants speaking an EA variety

²⁴ Dialectologists and sociolinguists have extensively documented how human movement and relations (e.g. administrative jurisdictions, transport routes, marriage practices, and cultural events) are important in shaping languages and language attitudes (e.g. Mase, 1999; Gooskens, 2005; Montgomery & Stoeckle, 2013; Jeszenszky et al., 2019).

and had to assimilate with original Akeanon settlers. It is likely however that close contact with Pandan is a plausible explanation for its /l̥r/ reflex which aligns with Kinarayan reflexes (cf. Zorc, 1977).

3 Acoustic description

This section now provides a description on the acoustic properties of DAR to help characterize and confirm of whether the velarized reflex is an approximant or fricative.

3.1 Method

Data was gathered through the help of four (4) native Akeanon speakers (male = 2; female = 2). With a mean age of 29.25 years old all subjects have lived either in the towns of Kalibo or Banga in Aklan, and have been speaking Akeanon at home and within the community for no less than 12 consecutive years. Spoken data was captured with a Sony Digital Voice Recorder ICD-UX560 with a 44.1 kHz sampling rate within an unattenuated room. Audacity v. 2.3.0. was used for further background noise reduction. Data were then processed using Praat v. 6.0.49 (Boersma & Weenik, 1992-2022) for waveform and spectrogram segmentation and analyses through spectrogram with the assistance of waveform.

Two (2) tasks were conducted in this study, both of which all participants underwent. The first task was an indirect elicitation, while the second was direct. The first task involved a 22-item questionnaire (see Appendix A) eliciting responses in Akeanon. Sixteen (16) questions—anticipated to be answered with an Akeanon input featuring a DAR—were the primary experimental stimuli (e.g. *eoha* ‘tear,’ *bakae* ‘(to) buy’), while an additional set of six distractions lead to an answer without a DAR (e.g. *kulintas* ‘necklace,’ *guwa* ‘outside’). The second task prompted the subjects to read out loud 14 words (see Appendix B) flashed on a computer screen while being voice recorded. Nine items featured DAR (e.g. *sueod* ‘inside, contents,’ *eapok* ‘mud, earth’), while an additional 5 (e.g. *ayam* ‘dog,’ *panit* ‘skin [anat.]’) served as distractors. A total of 99 tokens were elicited from the four informants. 51 were used for analysis of the intervocalic DAR and another 48 for the pre- and postvocalic DAR. 8 tokens involving the words *baeay* and *kaeayo* were used to compare /j/ with DAR and vowels.

To describe DAR, this study focused on formant frequency, vocalic duration, and acoustic intensity (cf. Jagers, 2018). Vocalic context was taken into consideration to see variation in behavior. Based on typical distribution patterns of DAR (§2.3), sequences with /a/ and /u/ were prioritized. To test the argument that it is an approximant (Zorc, 2019, 2005), acoustic data of the palatal approximant /j/ were also compared. Formant frequencies specifically F1 and F2 were used to measure frontness and height, respectively. It is expected that regardless of whether DAR is a fricative or

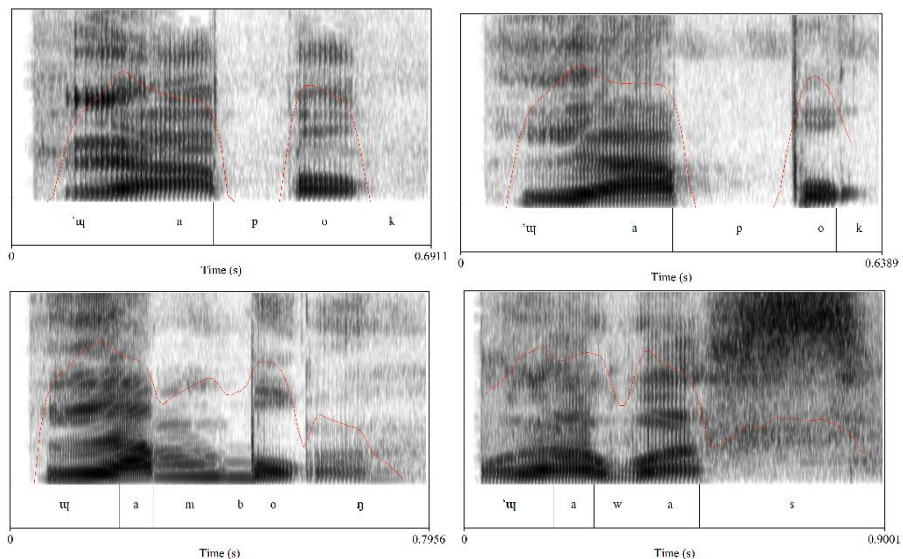
an approximant, its velar placement formant will exhibit a tighter and less anterior articulation compared to vowels, thus respectively lower F1 and F2. Acoustic intensity will be matched with formant to confirm articulatory profiles. Vocalic duration meanwhile was measured using the average absolute length of each realized DAR and relevant phoneme after onset. However, it should be noted that duration is relative due to various factors such as speech rate, pausing. Since this study is descriptive, future experiments sensitive to interaction with formant, sonority, and other (extra-)prosodic factors against a variety of phonemes are very much necessary to explore this phonological facet. Formant, duration, and intensity data between DAR and /a/, /u/, and /j/ were compared through various non-parametric tests in SPSS v. 26.

3.2 Results

Data below is divided into five major parts: The first are spectrograms of examples followed by acoustic analysis of formant, duration, and intensity. The last is a comparison of acoustic properties of velar fricatives.

3.2.1 Spectrogram

Spectrogram of Akeanon words with DAR are shown in Figures 2-5 below. Intensity is also marked (in red line).



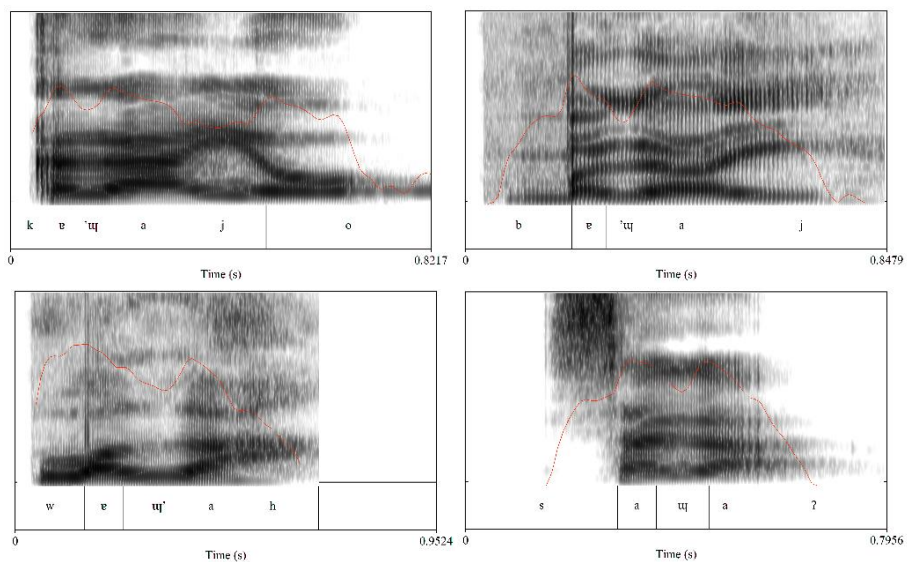


Figure 2: DAR onset + /a/ nucleus²⁵

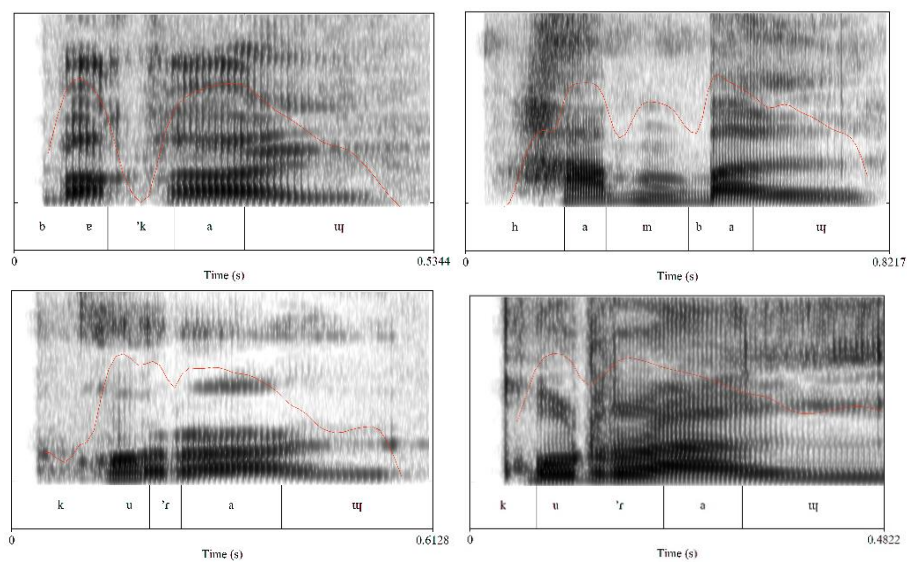


Figure 3: /a/ nucleus + DAR coda²⁶

²⁵ From top L-R: *eapok* 'mud, earth' (M1, F1) *eambong* 'upper garment, clothing' (F1), *eawas* 'body' (M1), *kaeayo* 'fire' (F2), *baeay* 'house' (M1), *waea* 'left (dir.)' (F1), *saea* 'wrong, mistake' (F2)

²⁶ From top L-R: *bakae* '(to) buy' (M1), *hambae* '(to) speak' (F1), *kurae* 'fence, enclosure' (M2, F1)

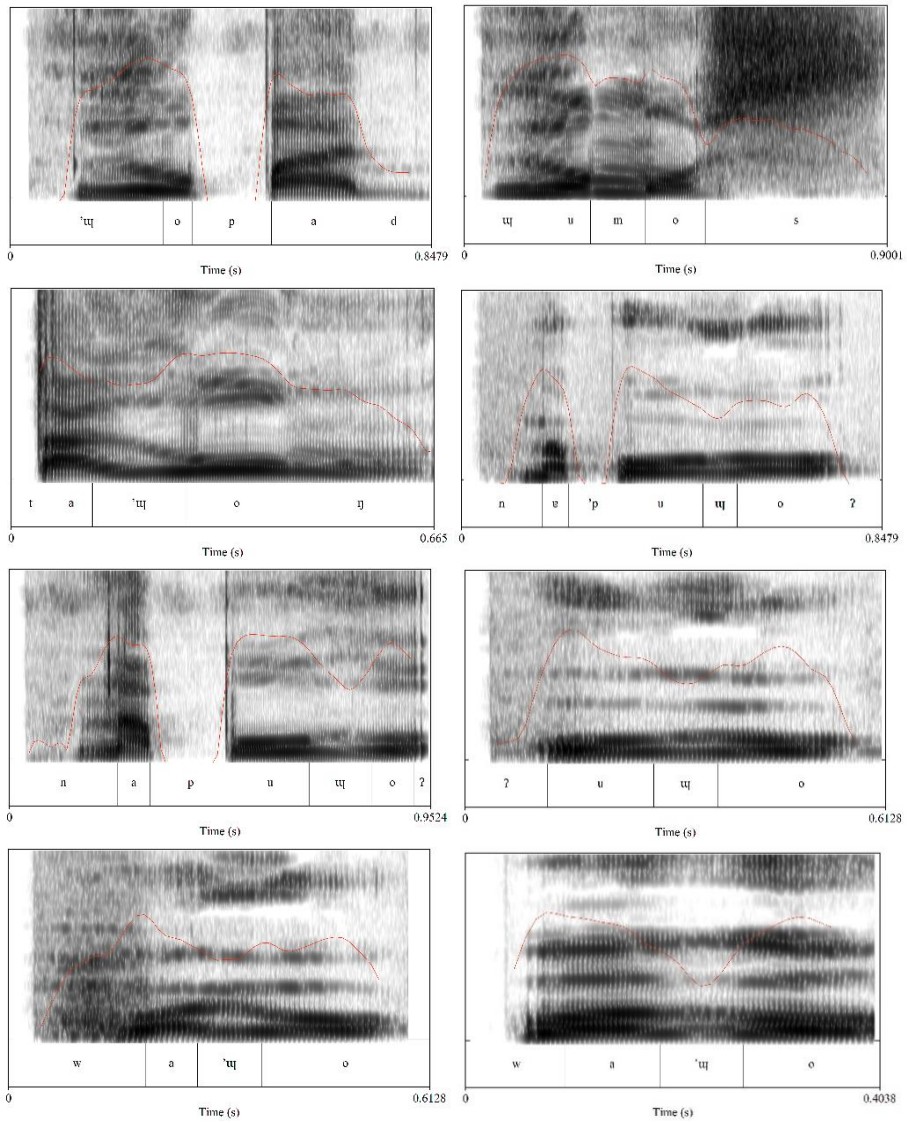


Figure 4: DAR onset + /u/ nucleus²⁷

²⁷ From top L-R: *eopad* ‘(to) fly’ (F1), *eomos* ‘(to) drown’ (F1), *taeong* ‘eggplant’ (F1), *napueo* ‘ten’ (M2, F1), *ueo* ‘head (anat.)’ (M2), *waeo* ‘eight’ (M2, F2)

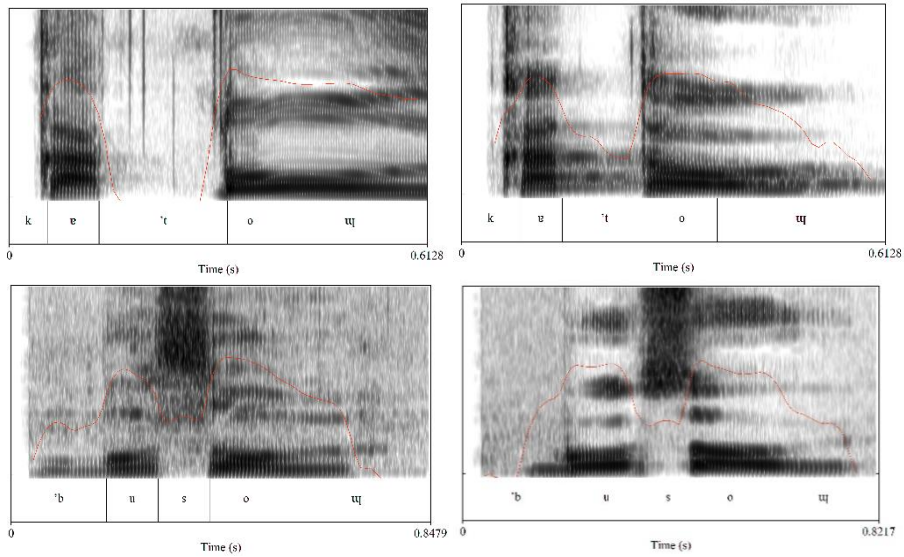


Figure 5: /u/ nucleus + DAR coda²⁸

3.2.2 Formant frequencies

Below presents formant frequencies of pre-, post-, and intervocalic DAR with consideration for phonological context (i.e. word position, adjacent phonemes). Figures 7-9 illustrate a boxplot of F1 and F2 values.

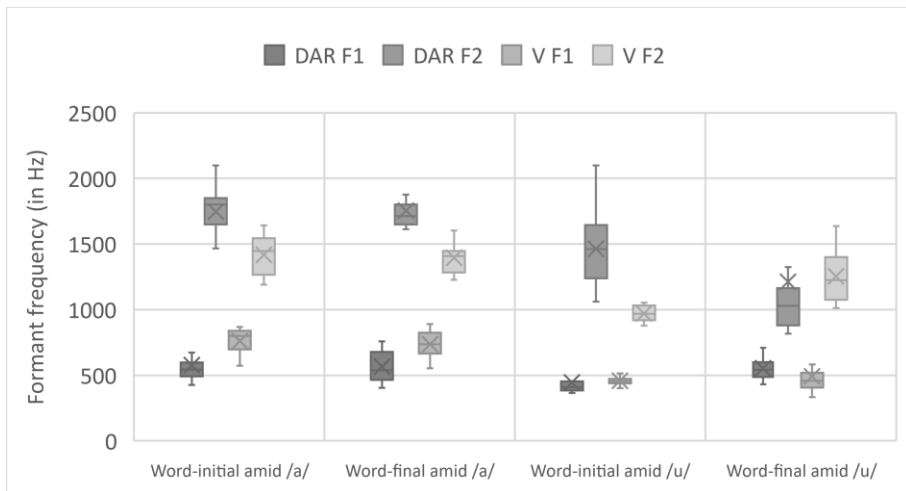


Figure 7: Formant frequencies of pre- and postvocalic DAR with adjacent vowels

²⁸ From top L-R: *katoe* '(to) itch' (F1, F2), *busoe* 'seed' (M1, M2)

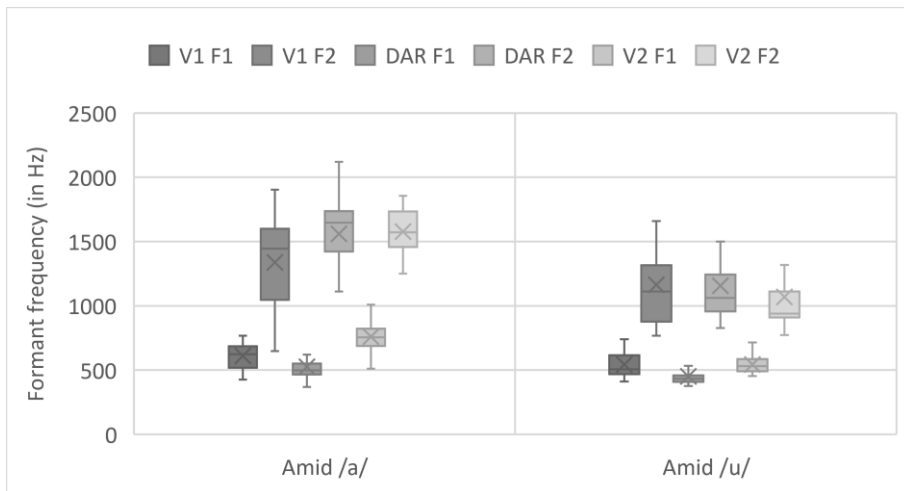


Figure 8: Formant frequencies of intervocalic DAR and adjacent vowels

Within each phoneme, there is a marked F1-F2 difference in pre- and postvocalic DAR compared to vowels. This is seen to be more modest in sequences with /u/. Across phonemes, it can be noted that DAR generally has a lower F1 compared to vowels. However, word-final DAR within /u/ contexts has a higher F1 but lower F2. While the lower F2 can be due to energy shift away from the velar, the higher vocalic F1 could imply tighter constriction compared to vowels. The F2 profiles of vowels in most other sequences are also peculiar since it is expected that higher F2 corresponds to more interior energy concentration which should be associated with vowels more than velars. The syllabic structure and stress-timing of the language could be a factor which is not thoroughly explored in this data. Generally, formant values are significantly lower when DAR and the vowel are positioned in word-initial or word-final syllables containing a close vowel (Table 14).

Table 14: Mean formants and *p*-values of pre- and postvocalic DAR and adjacent vowels

		DAR		V		
		F1	F2	F1	F2	
Kruskal-Wallis <i>p</i> -value		0.01	0.001	2e-6	7e-6	
Mean (in Hz)	Word-initial	Amid /a/	577	1747	761	1417
		Amid /u/	446	1463	457	971
	Word-final	Amid /a/	565	1755	737	1395
		Amid /u/	553	1213	492	1254

A Mann-Whitney test (Table 15) on intervocalic data reveals that there is a significant difference in formant frequency between phonological contexts for the F1 and F2 of both DAR and V2. For example, the mean F1 of DAR when next to an open

vowel is 1559 Hz versus a mean F2 of 1156 Hz if next to a close vowel. The same for the mean F1 of a V2 at 1578 Hz amidst /a/, while a mean F2 of 1069 Hz amidst /u/. This implies that the position of either DAR and V2 and/or phoneme adjacent to them has an effect. It is indeed much less perceptually clear whenever a word with intervocalic DAR is next to a close vowel compared being next to front (e.g. *sueod* vs *baeas*).

Table 15: Mean formants and *p*-values of intervocalic DAR and adjacent vowels

		V1		DAR		V2	
		F1	F2	F1	F2	F1	F2
Mann-Whitney	<i>p</i> -value	0.06	0.08	0.01	4e-5	7.79e-7	6e-6
Mean	Amid /a/	609	1339	525	1559	759	1578
(in Hz)	Amid /u/	543	1163	451	1156	543	1069

Figure 9 below compares formant frequency between DAR and /j/ in words with the palatal approximant (i.e. *baeay*, *kaeayo*; from 8 tokens). There is a marginally higher F1 for DAR, while on the other hand F2 is marginally higher for /j/ even when compared to vowels. Mann-Whitney test (Table 16) confirms a significant difference in F2 ($p = 0.001$) between DAR and /j/, but not for F1.

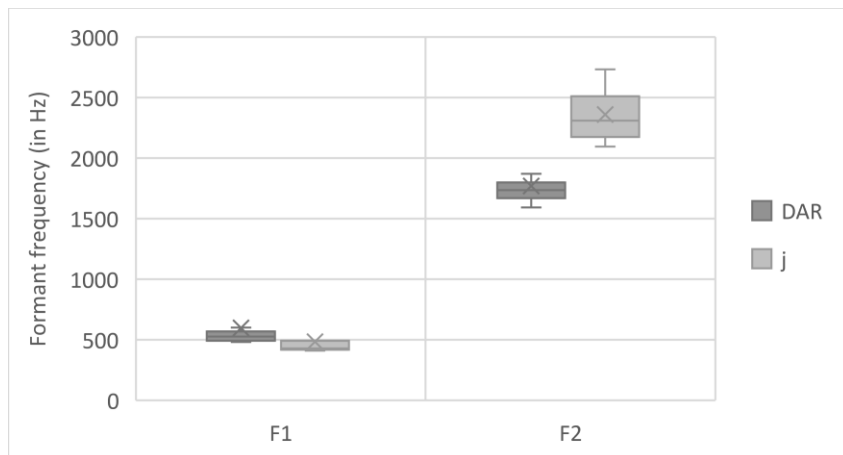


Figure 9: Formant frequencies of DAR and /j/

Table 16: Mean formant frequencies and *p*-values of DAR and /j/

		F1	F2
Mann-Whitney	<i>p</i> -value	0.07	0.001
Mean (in Hz)	DAR	598	1769
	/j/	484	2360

3.2.3 Acoustic intensity

Below presents acoustic intensity of pre-, post-, and intervocalic DAR with consideration for phonological context (i.e. word position, adjacent phonemes). Figures 10-12 illustrate a boxplot of intensity values.

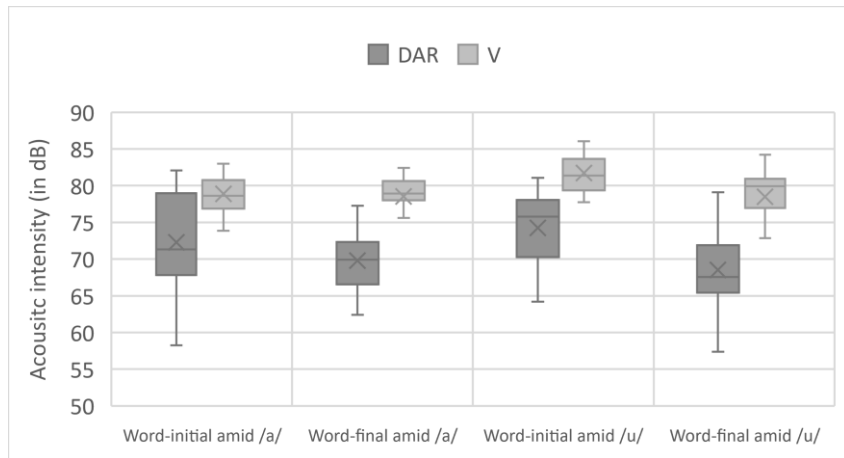


Figure 10: Acoustic intensities of pre- and postvocalic DAR and adjacent vowels

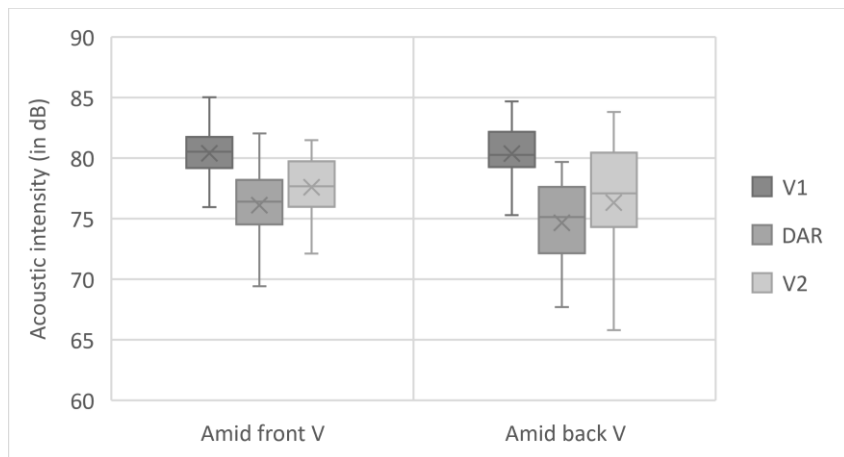


Figure 11: Acoustic intensities of intervocalic DAR and adjacent vowels

There is a minimally greater intensity in vowels compared to DAR as expected. This is a pattern evident in pre-, post-, and intervocalic positions. A Kruskal-Wallis test on pre- and postvocalic data in Table 17 shows that there is no statistically significant difference among DAR and vowels. This is the same in Table 18 where intervocalic data yields no significant difference in distribution of intensity values in both phonological contexts.

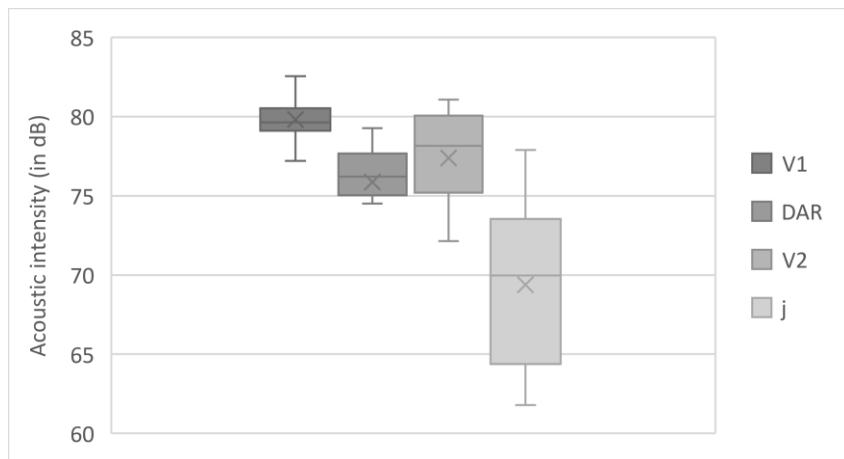
Table 17: Mean intensities and *p*-values of pre-and postvocalic DAR and adjacent vowels

		DAR	V
Kruskal-Wallis <i>p</i> -value		0.17	0.12
Mean (in dB)	Word-initial	Amid /a/	72
		Amid /u/	70
	Word-final	Amid /a/	74
		Amid /u/	69

Table 18: Mean intensities and *p*-values of intervocalic DAR and adjacent vowels

		V1	DAR	V2
Kruskal-Wallis <i>p</i> -value		0.83	0.26	0.51
Mean (in dB)	Amid /a/	80	76	78
	Amid /u/	80	75	76

Meanwhile, Figure 12 below shows intensity values of DAR, /j/, and adjacent vowels in words with the palatal approximant (from 8 tokens). Data shows /j/ has a more notable dip (compared to DAR against vowels), a tendency also found by Shaw et al. (2020).

**Figure 12:** Acoustic intensities of DAR, adjacent vowels, and /j/**Table 19:** Mean acoustic intensities of DAR, adjacent vowels, and /j/

	V1	DAR	V2	/j/
Mean (in dB)	80	76	77	69

Through a Kruskal-Wallis test, it was confirmed that there is a significant difference ($p = 0.001$) in intensity values across phonemes involved. Compared with a significantly higher formant, this demonstrates how loudness does not always go with articulatory energy. A Bonferroni-adjusted post hoc test (Table 20) confirms that /j/ has a significantly lower intensity than both vowel types, but not compared to DAR.

Table 20: Post hoc test on intensities of DAR, adjacent vowels, and /j/

	V1	DAR	V2	/j/
V1	--	0.141	1.000	0.0005
DAR		--	1.000	0.56
V2			--	0.04
/j/				--

3.2.4 Duration

This section now discusses duration of pre-, post-, and intervocalic DAR with consideration for phonological context (i.e. word position, adjacent phonemes). Figures 13-15 illustrate a boxplot of duration values.

Generally, pre- and postvocalic DAR is articulated much longer compared to neighboring vowels. This is in contrast to what Martínez-Celdrán and Reguera (2008) found in Galician, where spirant approximants show shorter duration than stops and fricatives. Through a Kruskal-Wallis test (Table 21) it was confirmed that only DAR is significantly different across phonological contexts ($p = 8.6e-5$) where it tends to be articulated much longer when adjacent to close vowels. This could be due to stress-timing of words used for the elicitation, or cross-articulatory effects of a vowel type.

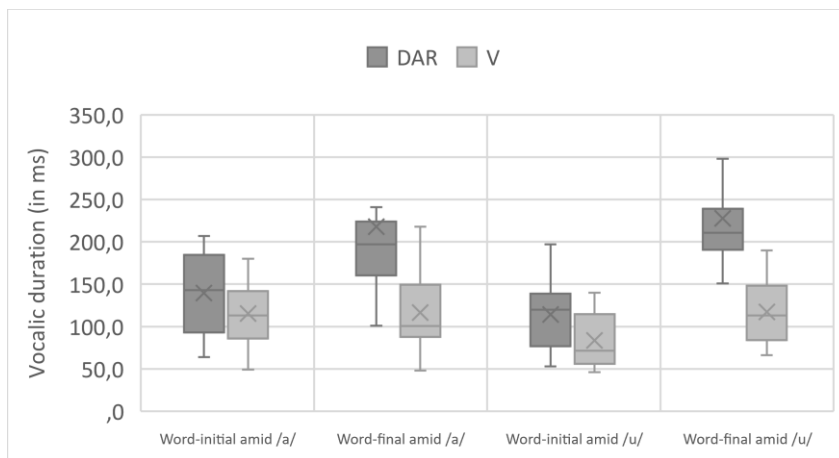


Figure 13: Vocalic duration of pre- and postvocalic DAR and adjacent vowels

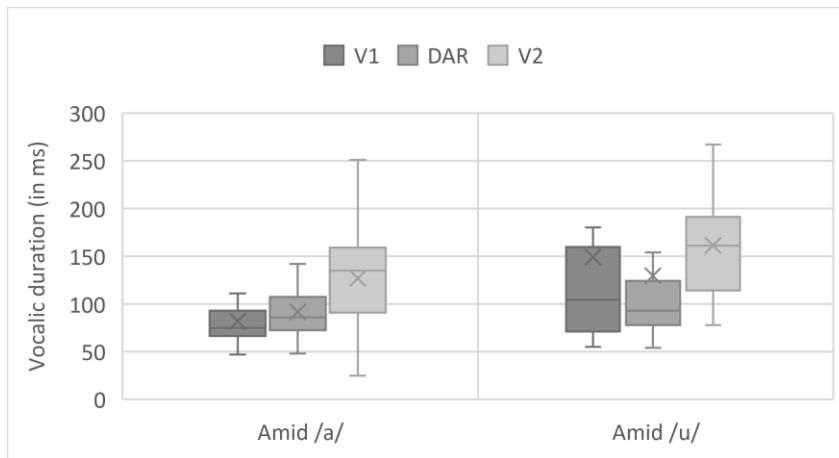


Figure 14: Vocalic duration of intervocalic DAR and adjacent vowels

Table 21: Mean duration and p -values of pre- and postvocalic DAR and adjacent vowels

		DAR	V
Kruskal-Wallis p -value		8.6e-5	0.11
Mean (in ms)	Word-initial	Amid /a/	139
		Amid /u/	218
	Word-final	Amid /a/	114
		Amid /u/	228

On the one hand, it seems that V2 has longer duration compared to V1 and DAR in intervocalic positions. Compared to pre- and postvocalic positions, intervocalic DAR also seems shorter. This could be attributed to the mechanical demands of articulating word-medial phonemes or vowels. A Mann-Whitney test (Table 22) meanwhile reveals a significant difference in distribution of V1 ($p = 0.01$) and V2 ($p = 0.03$) durations amid both an open and a close vowel. This could imply that an intervocalic DAR is stabler in duration regardless of phonological context.

Table 22: Mean duration and p -values of intervocalic DAR and adjacent vowels

		V1	DAR	V2
Mann-Whitney p -value		0.01	0.23	0.03
Mean (in ms)	Amid /a/	82	92	127
	Amid /u/	149	129	161

Figure 15 below presents the duration of DAR, /j/, and adjacent vowels in words with the palatal approximant (from same 8 tokens). It can be seen that /j/ is articulated much longer than DAR and the vowels.

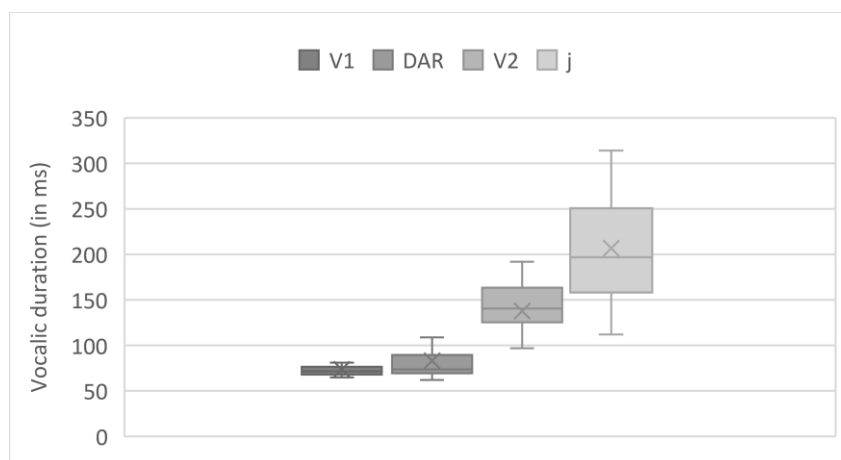


Figure 15: Vocalic duration of DAR, adjacent vowels, and /j/

Table 23: Mean duration of DAR, adjacent vowels, and /j/

	V1	DAR	V2	/j/
Mean (in ms)	74	83	138	207

A Kruskal-Wallis test reveals that there is a significant difference ($p = 0.0003$) in duration across all phonemes. This is confirmed through a Bonferroni-adjusted post hoc test (Table 24) showing that /j/ has significantly longer duration than V1 and DAR.

Table 24: Post hoc test on intensities of DAR, adjacent vowels, and /j/

	V1	DAR	V2	/j/
V1	--	1.000	0.09	0.001
DAR		--	0.27	0.005
V2			--	1.000
/j/				--

3.2.5 Compared with velar fricatives

When reviewing spectrograms, it seems that the spectral shape of the approximant is identical to the fricative counterpart. Martínez-Celdrán and Reguera (2008, pp. 57-58) present examples of open and closed variants of the “spirant approximant [ɣ̥]” in Spanish [laβo'ðeɣa] and Galician ['loɣo]. The “close” variant has F1 and F2 concentrated within low frequencies indicating a less anterior, more constricted

articulation characteristic of velars as opposed to the “open” variant. See also Figure 16 below for contrasts between [i^huó] ‘money’ (left) and [èyó] ‘darkness’ (right) in Urhobo (Rolle, 2013, p. 304). F1 and F2 in both approximant and fricative are quite low which are typical of a velar (Baart, 2010). Our Akeanon samples meanwhile have much wider gaps between F1 and F2 within DAR articulation. Turbulence is also concentrated in higher formant frequencies with less noise at low frequencies.

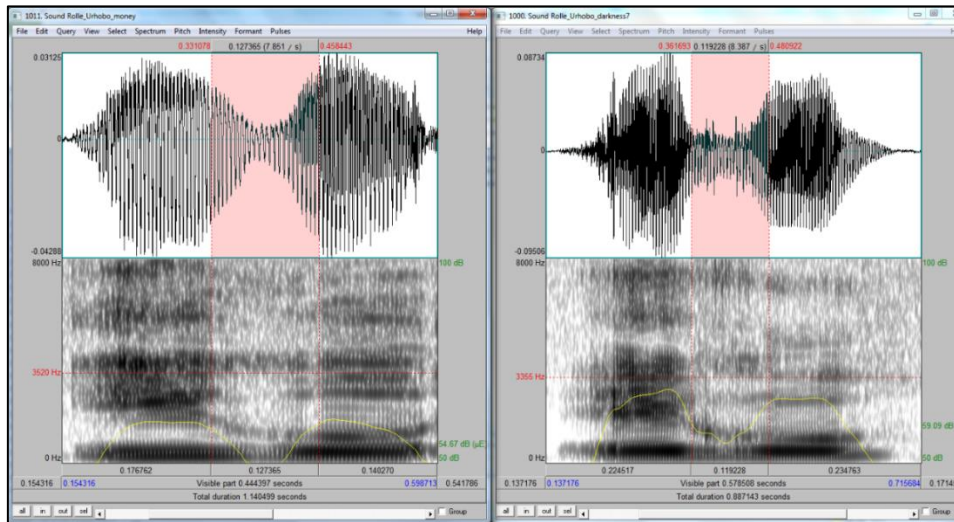


Figure 16: Spectrogram of [u] and [ɣ] variants in Urhobo

Upper formants from our data have less “noisy” spectral shapes in contexts with /a/ compared to those with /u/. They also have slightly higher F2 band than the latter most particularly within word-medial positions. From a perceptual standpoint, DAR within /u/ contexts also have more roundedness. Baart (2010) mentions that F2 behavior varies depending on quality of the adjacent vowel. It is known that non-vocalic variability is largely affected by the vowel context which has time and time been observed (Luce & Charles-Luce, 1985; Nearey & Rochet, 1994; Weglarski et al., 2000; Staroverov & Tebay, 2021) including semivowels in a notable crosslinguistic observation by Maddieson and Emmorey (1985).

Speakers note that DAR does not have a “raspy” or “coarse” guttural sound. Zorc (2005) also argues how the phoneme in focus lacks “friction” which can rather be observed in languages featuring a voiced velar fricative [ɣ] such as Modern Hebrew, and certain varieties of German and Dutch. All these articulatory properties corroborate the distinction set by Ball and Rahilly (1991 in Martínez-Celdrán, 2004) for both phonemes of contention in Spanish. Romero (1995 in *ibid*) also argues that there is no reliable difference in degree of constriction between fricatives and approximants. Rather, the basis of distinction should be the “lack of articulatory tension” (Martinet 1980-1981 in *ibid*, p. 204) possessed by the latter. From these premises, we adhere to

Martínez-Celdrán's (2004) proposed definition which effectively fits the properties attested in DAR whereby APPROXIMANTS are:

“...segments that, having a certain degree of constriction, lack a turbulent airstream, either due to the non-existence of the necessary articulatory precision required to produce it, or because the vocal tract is not narrow enough, or because these conditions occur simultaneously” (p. 208).

It can be presumed that this approximant quality identical to Spanish semivowels prompted its early codification into ⟨Ee⟩. It is known that much of modern written traditions in the Philippines trace their roots to the Spanish language and Spanish-based education²⁹ of the colonial period. It is undoubted that the earliest writing practice of modern Akeanon was started by those who received formal education through a Spanish orthographic perspective.

4 Conclusion

In this paper we provide an extended description of Akeanon phonology by describing the distinct Akeanon reflex (DAR). Acoustic data reveal that DAR has a distinct property compared to vowels and the approximant /j/. DAR has a generally lower F1 than that of vowels but greater than the glide /j/. Its F2 is significantly greater than vowels but less than /j/. Phonemic context may also play a role since DAR formants are lower if amidst /u/ than amidst /a/ in any position. Duration-wise, pre- and postvocalic DAR is also articulated longer than vowels especially if it is within the context of /u/ although this is not seen intervocalically. Meanwhile, /j/ has a significantly longer duration than V1 and DAR. In terms of intensity, DAR and vowels have no significant difference, but /j/ has a significantly lower intensity than DAR and V2.

Based on historical accounts and descriptions, it is clear that DAR is a reflex of the proto-Bis *l and *-d-. Its distribution appears as an onset and a coda with /a/ or /u/ as a nucleus. Few examples, possibly lexically motivated, permit its operations alongside /i/. We posit that its velarization (whether from a palatal or a relateralized form) may have been triggered either independently or by another prestige variety. Physiological and perceptual restrictions nonetheless should have played a crucial facilitative role. Meanwhile, its relative infrequency in phonological inventories of Philippine-type languages may be explained by the Obligatory Contour Principle (McCarthy, 1986 in Bailey, 2020), which argues that a sequence of two identical features is susceptible to a high perceptual variability. DAR is a reflex of a coronal and dorsal, and their articulatory mechanisms are relatively close if not intertwined. This may explain the

²⁹ Prior to American introduction of public education in the 20th century, an extremely small portion of the population made of landed elites gained access to formal education.

diachronic instability (e.g. coalescence) or inability to stabilize, which may explain the **-d- > /j/ > Ø/V*: reflex in other Bis varieties such as Cebuan.

Spectrogram shapes of phonemes with nearly similar articulation such as Urhobo velars, and Spanish and Galician spirant approximants somewhat resemble those of DAR. However, its perceived reduced turbulence and other acoustic qualities (e.g. spectral shape) further provide evidence that the reflex is not a fricative. We concur with Zorc's reevaluation that DAR is rather a velar approximant [ɰ] and therefore propose an updated chart of Akeanon consonants as seen below.

Table 25: Akeanon consonant inventory

	Bilabial		Alveolar		Post-Alveolar		Palatal	Velar		Labiovelar	Glottal
Stop	P	b	t	d				k	g		ʔ
Nasal		m		n					ŋ		
Affricate			(ts)	(dz)	(tʃ)	(dʒ)					
Fricative	(f)	(v)	s	(z)	(ʃ)						h
Approximant							j	ɰ		w	
Tap				r							
Lateral				l							

From this realization, there are two major issues that could be addressed in the future. First, the analysis relies on acoustic documentation. More thorough extrapolation of phonological environment with consideration to contrast, syllabicity, stress, pausing, and perceptual distance (e.g. Padgett, 2008), experimental or otherwise, is advisable to provide more comprehensive documentation. Ultrasound analysis of constriction and positioning is also ideal.

Second is to acknowledge the pedagogical implications of a revisited Akeanon phonology in orthographic and grammatical instruction. As far as orthography is concerned, there is no need to reappropriate the use of ⟨Ee⟩ which represents both the front vowel and the approximant. Looking for an alternative in computers would foremost be an added challenge; most widely available keyboards in the Philippines do not have accessible special characters or accents that would effectively and conveniently exemplify its phonetic property. The closest could be ⟨Jj⟩ whereby *baeay* would be written as *bajay*. However, this grapheme has long been used in many loanwords and proper nouns representing varying phonemes. This defeats the purpose if the goal is orthographic transparency. Nevertheless, there is a need to rectify existing texts describing the language, which would normally emphasize the phoneme as an important feature in distinguishing Akeanon from other Bisayan varieties and as something unique apart from Philippine languages in general. Current public Akeanon understanding treats [ɰ] as a vowel and is taught as such in many local schools. We

recommend that pedagogical approaches and instructional materials be reevaluated towards the more accurate phonetic character. This is not mere pedantry on phonology. Better awareness may have positive implications to grammatical instruction given that Akeanon, like other Philippine languages being strongly agglutinative, has a complex and diverse affix system.

References

- Baart, J. (2010). *A field manual of acoustic phonetics*. SIL International.
- Bateman, N. (2011). On the typology of palatalization. *Language and Linguistics Compass*, 5(8), 588-602. <https://dx.doi.org/10.1111/j.1749-818x.2011.00294.x>
- Bhat, D. N. S. (1978). A general study of palatalization. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 2, pp. 47-92). Stanford University Press.
- Blust, R. (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In E. Zeitoun & P. J.-K. Li (Eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics* (pp. 1-54). Academia Sinica.
- Blust, R. A. (2005). The linguistic microhistory of the Philippines. In H.-C. Liao, & C. R. G. Rubino (Eds.), *Current issues in Philippine linguistics: Parangal kay Lawrence A. Reid* (pp. 31-68). Linguistic Society of the Philippines & SIL Philippines.
- Blust, R. A. (2019). The resurrection of proto-Philippines. *Oceanic Linguistics*, 58(2), 153-256. <https://doi.org/10.1353%2Fol.2019.0008>
- Browman, C. P., & Goldstein, L. (1995). Gestural syllable position effects in American English. In F. Bell-Berti, & L. J. Raphael (Eds.), *Producing speech: Contemporary issues* (pp. 19-34). American Institute of Physics.
- Bybee, J. (2015). *Language change*. Cambridge University Press.
- Bybee, J., & Easterday, S. (2019). Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology*, 23(2), 263-302. <https://doi.org/10.1515/lingty-2019-0015>
- Conant, C. E. (1911). The RGH law in Philippine languages. *Journal of the American Oriental Society*, 31(1), 70-85. <https://doi.org/10.2307/3087491>
- Conant, C. E. (1912). The Pepet law in Philippine languages. *Anthropos*, 7(4), 920-956.
- de la Cruz, B. A., & Zorc, R. D. (1968). *A study of the Aklanon dialect. Volume one: Grammar*. Peace Corps.
- de Méntrida, A. (1841). *Diccionario de la lengua bisaya hiligueina y haraya de la isla de Panay*. Imprenta de D. Manuel y de D. Felis S. Dayot.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2022). *Ethnologue: Languages of the world* (24th ed.) SIL International.
- Flynn, D. M. (2012). Vocalic dorsality in Articulator Theory. *School of Languages, Linguistics, Literatures and Cultures*. University of Calgary.

- French, R. A. (1979). The phonology of Romblomanon [Manuscript]. *SIL International*.
<https://www.sil.org/resources/archives/43456>
- Gooskens, C. (2005). Travel time as a predictor of linguistic distance. *Dialectologia et Geolinguistica*, 13, 38-62. <http://dx.doi.org/10.1515/dig.2005.2005.13.38>
- Jaggers, Z. S. (2018). Evidence and characterization of a glide-vowel distinction in American English. *Laboratory Phonology*, 9(1), 1-27. <https://doi.org/10.5334/labphon.36>
- Jeszenszky, P., Hikosaka, Y., Imamura, S., & Yano, K. (2019). Japanese lexical variation explained by spatial contact patterns. *International Journal of Geo-Information*, 8. <http://dx.doi.org/10.3390/ijgi8090400>
- Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustic Society of America*, 78(6), 1949-1957.
- McFarland, C. D. (1975). *The dialects of the Bikol area* [Doctoral dissertation]. Yale University.
- Maddieson, I., & Emmorey, K. (1985). Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica*, 42, 163-174. <https://doi.org/10.1159/000261748>
- Martínez-Celdrán, E. (2004). Problems in the classification of approximants. *Journal of the International Phonetic Association*, 34(2), 201-210. <https://doi.org/10.1017%2FS0025100304001732>
- Mase, Y. (1999). On dialect consciousness: Dialect characteristics given by speakers. In D. R. Preston (Ed.), *Handbook of perceptual dialectology* (Vol. 1, pp. 101-114). John Benjamins.
- Monteclaro, P. A. (1907). *Maragtas kon sayuron (historia) sg pulô nga Panay kag sang mga pumuloyo, tubtub sg pag-abut sg mga taga Borneo nga amo ang ginhalinan sg mga bisaya, kag sg pag-abut sg mga katsila / Ginsulat ni gino-ong Pedro Alcantara Monteclaro pumuluyo sa Miag-ao, Ilong-ilong, siling sg iya nabasahan sa mga tolon-an nga iya na-ayop sa mga tigulang*.
- Montgomery, C., & Stoeckle, P. (2013). Geographic information systems and perceptual dialectology: A method for processing draw-a-map data. *Journal of Linguistic Geography*, 1, 52-85. <http://dx.doi.org/10.1017/jlg.2013.4>
- Moosmüller, S., Schmid, C., & Kasess, C. H. (2015). Alveolar and velarized laterals in Albanian and in the Viennese dialect. *Language and Speech*, 59(4), 488-515. <https://dx.doi.org/10.1177/0023830915615375>
- Nagle, C. (2014). *A longitudinal study on the role of lexical stress and motivation in the perception and production of L2 Spanish stop consonants* [Doctoral dissertation, Georgetown University]. DigitalGeorgetown. <https://repository.library.georgetown.edu/handle/10822/761009>
- National Statistics Office. (2002). *Aklan: Population growth rate up to 2.05 percent in 2000 (Results from the 2000 census of population and Housing, NSO)*. Philippine Statistics Authority. <http://web0.psa.gov.ph/content/aklanpopulation-growth-rate-205-percent-2000-results-2000-census-population-andhousing-nso>

- Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception of French and English stops. *Journal of the International Phonetic Association*, 24(1), 1-18.
- Odden, D. (2015). Bantu phonology. *Oxford handbooks online*. Oxford University Press. <https://dx.doi.org/10.1093/oxfordhb/9780199935345.013.59>
- Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237-278). Longman.
- Padgett, J. (2008). Glides, vowels, and features. *Lingua*, 118, 1937-1955.
- Parrell, B., & Narayanan, S. (2018). Explaining coronal reduction: Prosodic structure and articulatory posture. *Phonetica*, 75(2), 151-181. <https://dx.doi.org/10.1159%2F000481099>
- Pawley, A., & Ross, M. D. (1995). The prehistory of Oceanic languages: A current view. In P. Bellwood, J. Fox, & D. Tryon (Eds.), *The Austronesians: Historical and comparative perspectives* (pp. 39-74). Australian National University.
- Paz, C. J. (1981). *A reconstruction of proto-Philippine phonemes and morphemes*. University of the Philippines.
- Pérez, A. (1902). *Igorrotes: Estudio geográfico y etnográfico sobre algunos distritos del norte de Luzon*. El Mercantil.
- Reid, L. A. (1982). The demise of proto-Philippines. In A. Halim, L. Carrington, & S. Wurm (Eds.), *Papers from the Third International Conference on Austronesian Linguistics, volume 2: Tracking the travellers* (pp. 201-216). Australian National University.
- Reid, L. A. (2018). Modeling the linguistic situation in the Philippines. *Senri Ethnological Studies*, 98, 91-105. <http://doi.org/10.15021/00009006>
- Rentillo, P. (2018). Preliminary dialectology of Aklanon [Conference presentation]. 2018 Linguistic Society of the Philippines - National Conference and General Meeting, Balanga, Philippines.
- Rolle, N. (2013). Phonetics and phonology of Urhobo. *UC Berkeley Phonology Lab Annual Report*, 9(9), 281-326. <https://escholarship.org/content/qt2q6685k5/qt2q6685k5.pdf>
- Ross, M. D. (1988). *Proto Oceanic and the Austronesian languages of western Melanesia*. Pacific Linguistics.
- Rubino, C. (2006). Utudnon, an undescribed language of Lyete. In H.-C. Liao, & C. R. G. Rubino (Eds.), *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid* (pp. 306-337). Linguistic Society of the Philippines & SIL Philippines.
- Salas Reyes, V., Zorc, R. D., & Prado, N. L. (1969). *A Study of the Aklanon dialect. Volume two: Dictionary (of root words and derivations)*. Peace Corps.
- Santiago, V. (2018, October 25). Some notes on Porohanon, a Bisayan variety spoken in the Camotes islands, Cebu [Conference presentation]. 1st College of Social Sciences and Philosophy Graduate Students Association (CSSP GSA) Brown Bag Session, University of the Philippines Diliman, Quezon City, the Philippines.
- Scheerer, O. (1920). Über Einen Bemerkenswerten L-Stellvertreter im Dialekt von Aklan auf der Insel Panay (Filipinen) [On the remarkable L-variant in the dialect of Aklan in the

- island of Panay (Philippines)]. *Zeitschrift für Eingeborenen-Sprachen* [Journal of Native Languages], 11, 241-259.
- Shaw, J. A., Caringan, C., Agostini, T. G., Mailhammer, R., Harvey, M., & Derrick, D. (2020). Phonological contrast and phonetic variation: The case of velars in Iwaidja. *Language*, 96(3), 578-617. <https://doi.org/10.1353/lan.2020.0042>
- Staroverov, P., & Tebay, S. E. (2021). Velar lateral allophony in Mee (Ekari). *Journal of the International Phonetic Association*, 1-31.
- Weglarski, A., Sewall, A., Schiavetti, N., Metz, D. E., & Whitehead, R. L. (2000). Effect of vowel environment on consonant duration: An extension of normative data to adult contextual speech. *Journal of Communication Disorders*, 1(1), 1-10. [https://doi.org/10.1016/s0021-9924\(99\)00022-2](https://doi.org/10.1016/s0021-9924(99)00022-2)
- Wolfenden, E. (1975). A description of Hiligaynon syntax. In I. Davis (ed.), *Linguistics and related fields*, 46. Summer Institute of Linguistics.
- Wolff, J. U. (1967). History of the dialect of the Camotes Islands, Philippines, and the spread of Cebuano Bisayan. *Oceanic Linguistics*, 6(2), 63-79.
- Zorc, R. D. (2021). Axis relationships in the Philippines – When traditional subgrouping falls short [Conference presentaton]. 14th Philippine Linguistics Congress, Facebook Live.
- Zorc, R. D. (2005). Aklanon tag- and extra-systemic linguistic phenomena. In H.-C. Liao & C. Rubino (Eds.), *Current issues in Philippine linguistics and anthropology: Parangal kay Lawrence A. Reid* (pp. 124-135). Summer Institute of Linguistics & Linguistic Society of the Philippines.
- Zorc, R. D. (1995). Aklanon. In D. T. Tryon (Ed.), *Comparative Austronesian dictionary: An introduction to Austronesian studies* (pp. 343-350). De Gruyter Mouton.
- Zorc, D. P. (1977). *Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Australian National University.

Appendix A

List of words for the indirect elicitation task

- | | |
|-------------|------------|
| 1. eawas | 12. puea |
| 2. kulintas | 13. saea |
| 3. eambong | 14. kuring |
| 4. eoha | 15. ueo |
| 5. guwa | 16. waeo |
| 6. eomos | 17. bakae |
| 7. bueak | 18. lima |
| 8. baeas | 19. kurae |
| 9. tueog | 20. katoe |
| 10. Malay | 21. relo |
| 11. taeong | 22. busoe |

Appendix B

List of words for the direct elicitation task

1. pispis
2. baeay
3. hambae
4. eapok
5. igkampod
6. kaeayo
7. waea
8. panit
9. haboe
10. eopad
11. ayam
12. sueod
13. napueo
14. suba

RESEARCH ARTICLES IN SLOVENE
ZNANSTVENI ČLANKI V SLOVENŠČINI

Vloga prevajanja pri razumevanju literature korejske manjšine v ZDA: Analiza prevodov Leejevega romana *Native Speaker*

Byoung Yoong KANG

University of Ljubljana, Slovenia

ByoungYoong.Kang@ff.uni-lj.si

Abstract

The novel *Native Speaker* written by Chang-rae Lee who is a member of the Korean minority in the United States, was translated and published in Korean twice, in 1995 and 2003. The novel's main theme is exploring (Korean American's) identity and the most important key to understanding it is the language. In other words, depending on how one understands the identity of its protagonist Henry, who is, as the writer Chang-rae Lee himself, a member of the first-and-a-half generation Korean-Americans, the interpretation, evaluation, and even translation of the work can differ greatly. This study examined the role of the language used in the novel and analyzed its second translation into Korean. It reveals the problems arising from translation and emphasizes the role of translation when it comes to understanding the novel. Ultimately, the study explains and reinterprets the meaning of *Koreanness* in Korean literature.

Keywords: Chang-rae Lee; *Native Speaker*; diaspora literature; translation; Korean diaspora; American literature; Korean literature

Povzetek

Roman *Native Speaker*, čigar avtor je Chang-rae Lee, pripadnik korejske manjšine v ZDA, je bil v korejščino preveden in objavljen dvakrat, prvič leta 1995 in zatem leta 2003. Glavna tema omenjenega romana je iskanje identitete (Američana korejskega rodu), ključ za razumevanje dela pa jezik. Glede na to, kako razumemo identiteto poldruge generacije korejskih priseljencev, ki ji pripadata glavni junak romana, Henry, in pisatelj sam, se lahko interpretacija, ovrednotenje in celo prevod dela precej razlikujejo. Članek proučuje vlogo jezika v omenjenem romanu in analizira novejši prevod v korejščino. Izpostavljene so težave, ki se pojavijo pri prevajanju Leejevih literarnih del, in vloga, ki jo pri razumevanju dela igra prevod. Poleg tega članek obravnava tudi pomen *korejskosti*, ki ga vsebuje korejska literatura, in ga poskuša ponovno pojasniti.

Ključne besede: Chang-rae Lee; naravni govorec; literatura diaspore; prevajanje; korejska diaspora; ameriška književnost; korejska književnost



1 Uvod

Za razumevanje literarnega dela moramo najprej poznati avtorja dela in literarne osebe, ki se v delu pojavijo. Ko govorimo o literaturi korejske manjšine v ZDA ali literaturi Američanov korejskega rodu (*jaemi hanin munhak* 재미 한인 문학), moramo poleg tega poznati še zgodovinsko ozadje korejske manjšine v ZDA.

Z izrazom "prva generacija" označujemo izseljence, ki so se rodili v Južni Koreji, se tam izšolali, nato pa se v odrasli dobi preselili v ZDA. Zanje velja, da jim je bližje korejščina kot angleščina, običajno je tudi njihov življenjski slog bližje korejskemu. Izraz "poldruga generacija" označuje potomce prve generacije, ki so bili sicer rojeni v Koreji, a so v Ameriko odšli kot mladostniki ter se tam tudi izšolali. Za razliko od prve generacije niso bili del korejskega izobraževalnega sistema, zato nimajo težav pri sporazumevanju v angleškem jeziku. Druga generacija izseljencev se je rodila, izšolala in odrasla v ZDA, zato nima težav pri sporazumevanju v angleščini in popolnoma razume ameriško kulturo ter ameriški način razmišljanja. Stik s korejsko kulturo ohranjajo preko staršev, vendar je njihov življenjski slog bližje ameriškemu.

Literatura prve generacije korejske manjšine v obdobju pred koncem druge svetovne vojne je bila napisana v korejščini, a ker so bila ta dela objavljena v ZDA, je v njih moč najti ameriške elemente. Pisatelji poldruge generacije so dela objavljali v angleščini, a je v njih moč zaslediti korejske elemente in značilnosti korejskega jezika, kot so na primer tujke, prevzete iz angleškega jezika (*konglish*), ki niso nujno poznane ameriškemu bralcu.

Za ustrezno razumevanje literature korejske diaspore moramo torej poznati obe kulturi in oba jezika. Dobri prevajalci, ki ta dela prevajajo iz angleščine v korejščino, poznajo obe strani. Korejski prevod je kot filter, s katerim lahko raziskovalci iščejo korejske elemente v literarnih delih korejske manjšine v ZDA.

Pisatelj Chang-rae Lee¹ pripada poldrugi generaciji korejske manjšine. V članku se bom osredotočil na njegov prvenec *Native Speaker*, kjer so prisotne vse zgoraj navedene značilnosti. Objavljen je bil v angleščini, a vsebuje veliko elementov korejskega jezika in korejske kulture, hkrati pa je bil dobro sprejet med ameriški literarnimi krogi. Pozornost je kasneje vzbudil tudi v Koreji, zato je v korejščino preveden dvakrat.

Članek sem zasnoval na podlagi analize novejšega korejskega prevoda. Moj namen je proučiti, kako so v prevodu interpretirani in prikazani elementi, ki se nanašajo na korejski jezik in kulturo. Izhajal bom iz predpostavke, da dober prevod pomaga pri globljem razumevanju literarnega dela, saj se prevajalec ne osredotoča le na jezik, pač

¹ Korejska lastna imena zapisujemo v vrstnem redu priimek-ime, pri čemer pregibamo le zadnjo sestavino. V danem primeru ohranjamo zaporedje ime-priimek, s čimer spoštujemo avtorjeve lastne preference in njegovo pripadnost poldrugi generaciji korejske manjšine v ZDA.

pa tudi na kulturne značilnosti izvirnega in ciljnega jezika ter nanje bralce opozarja s pomočjo jezikovnih prvin. Globlje razumevanje literarnega dela nam lahko pomaga tudi pri boljšem razumevanju literature korejske diaspore in njeni vlogi v literarnem svetu.

2 Roman *Native Speaker*

Chang-rae Lee se je rodil v Seulu leta 1965, pri treh letih pa se je z družino preselil v Ameriko. Najprej se je v Ameriko preselil njegov oče, ki je bil po poklicu psihiater, nato je za njim prišla še preostala družina (Bang, 2009, str. 135). Lee je diplomiral iz angleške književnosti na Univerzi Yale, nato je delal na Wall Streetu. Leta 1993 se je vpisal na Univerzo v Oregonu in tam magistriral iz ustvarjalnih umetnosti. Leta 1995 je izšel njegov prvenec, roman *Native Speaker*.

Sodeč po informacijah v spletnem brskalniku *NAVER Academic*, je v samo 20 letih (med letoma 2000 in 2020) izšlo 132 akademskih člankov na temo Leejeve literature. Poleg tega je bilo o njegovih delih napisanih še 41 magistrskih nalog ali doktorskih disertacij tako s področja angleške literature kot tudi korejske literature. Med članki je več kot 25 takih, ki se ukvarjajo z romanom *Native Speaker*. Nekaj raziskav je tudi o obeh prevodih romana in te v veliki meri obravnavajo tri osnovna področja. Prve proučujejo vprašanje identitete manjšine, druge proučujejo elemente, v povezavi s podobo Koreje, tretje pa odnos med identiteto in jezikom (Park, 2019, str. 4).

Roman *Native Speaker* govori o življenju Američana korejskega rodu, Henryja, ki je vohun. Protagonist se je rodil v korejski družini in si prizadeva postati "popolni" Američan. Verjame, da mu bo to uspelo, ko bo usvojil angleški jezik in ga govoril tako kot ostali Američani. Nemalokrat zaide v konflikt s svojim patriarhalnim očetom, ki je del prve generacije korejskih priseljencev. Tudi v odnosu s svojo ameriško ženo Lelio prihaja do napetosti. V nesreči zakoncema umre sin Mitt, kar je povod za trenja med njima. Henry kot vohun začne delati v volilnem štabu uspešnega politika Johna Kwanga, ki se poteguje za mesto župana New Yorka. Roman skozi Henryjeve odnose z očetom, ženo in Johnom Kwangom oriše protagonistov proces iskanja samega sebe.

Vprašanje identitete je glavna tema romana. Ključ do razumevanja tega vprašanja je jezik, saj glavni element protagonistove identitete ni le angleščina, pač pa tudi korejščina. Vprašanje identitete v Leejevih delih ni zgolj področje, o katerem pogosto pišejo in govorijo raziskovalci ter literarni kritiki, nanj je opozoril tudi avtor sam. Pravi, da je zanj vprašanje identitete zelo pomembno in da o tem vedno razmišlja. Kljub temu da se njegovo življenje ne razlikuje od življenja Američanov, zase pravi, da ni Korejec, ki živi v Ameriki. Čuti močno povezanost z Južno Korejo, kjer se je tudi rodil, in trdi, da bo vedno le Korejec (Kim, 2006, str. 387). A te močne povezanosti ne moremo vedno dojemati kot nekaj pozitivnega. Do nje ga je privedel proces izgube in iskanja identitete ter končne ugotovitve, da je identiteto moč najti le s pomočjo povezanosti obeh kultur (Ji, 2004, str. 306).

Ključni faktor za razumevanje romana *Native Speaker* je torej razumevanje in interpretacija "identitete". Glede na interpretacijo protagonistove (Henryjeve) identitete se razlikuje tudi interpretacija dela, njegova kritika in celo prevod. Še več, glede na identiteto se razlikuje tudi pomen "korejskosti" dela. Pravo razumevanje romana je moč doseči le z razumevanjem jezika v romanu *Native Speaker* in prevoda, ki ima pomembno vlogo pri korejski recepciji romana.

3 Pomen jezika v romanu *Native Speaker*

Razmišljanje glavnega lika v romanu *Native Speaker* je v veliki meri podobno razmišljanju Frantza Fanona, ki je eden vodilnih filozofov postkolonializma. V delu *Črna koža, bele maske* (2015, str. 23-54) je pisal o odnosu med kolonialističnim subjektom in jezikom.

Fanon je v ospredje postavil del črnskega prebivalstva, ki je nekaj časa živel skupaj z belim človekom, in se zato ob vrnitvi v domovino odločil, da ne bo več uporabljal svojega rojstnega jezika. Ti ljudje so se od drugih prebivalcev v domačem kraju razlikovali po rabi jezika, zato so se imeli za drugačne. Ker so se odločili za drug jezik, je bila njihova močna želja po razločevanju med njimi in ostalimi prebivalci v črnski skupnosti še vidnejša. Jezik torej predstavlja kulturno sredstvo, s katerim lahko izražamo samega sebe, z drugačno rabo jezika pa se od drugih tudi razlikujemo. Po Fanonu je v ozadju prepričanje črnega človeka, ki si je želel postati beli človek in je verjel, da mu bo lahko podoben s popolnim obvladovanjem njegovega jezika. To prepričanje je v končni fazi zmotno.

Črnskemu kolonialnemu subjektu je tako razmišljanje vcepil belski gospodar. Fanon pravi, da gospodar subjektu obljublja enakovreden položaj, če bo popolnoma obvladal njegov jezik. Gre za popolno iluzijo, ki se ni uresničila. A subjekti, ki so pod njenim vplivom, so si jezik prizadevali obvladati. Jezik kolonialnega gospodarja je torej "bela maska", o kateri govori Fanon. Četudi si črnski kolonialni subjekt nadene "belo masko (jezik svojega gospodarja)", ne more postati belski gospodar. Uničenje iluzije črnskega subjekta, ki ne more doseči položaja, enakovrednega belskemu, le še poglobi njegove stiske povezane z identiteto.

Fanonov argument lahko hitro potrdimo s primerom iz sodobne korejske zgodovine. Med obdobjem japonske kolonialne nadvlade (1910-1945) so bili Korejci prisiljeni prevzeti japonska imena in uporabljati japonski jezik. Rečeno jim je bilo, da bodo na ta način lahko postali subjekti japonskega cesarja. Korejski kolonialni subjekti niso dosegli enakovrednega položaja Japoncem. Nepriznavanje različnih kultur z nadvlado jezika je tipična imperialistična ideja.

V romanu *Native Speaker* se protagonist Henry ujame v past Fanonove "bele maske". Henry verjame, da bo s tekočim angleškim jezikom postal popoln Američan.

Odloči se, da si bo nadel belo masko in postal beli človek. A to ni fantazija, ki jo je ustvaril Henry sam. To je iluzija, ki se je razvila v ameriški družbi. Popolna ameriška angleščina, ki jo uporablja višji sloj belskega prebivalstva je tudi simbol tako imenovanih ameriških intelektualcev in višjega sloja (v skrajnem primeru se posmehujejo drugim različicam, na primer britanski ali avstralski angleščini). Amerika, država, ki so jo ustvarili migranti, je družba, ki trdi, da je globalno stičišče kultur, a je hkrati podrejena ideologiji vodilnega sloja, ki zahteva, da se vse novo asimilira na ameriški način. Da ljudje, ki živijo v Ameriki, lahko postanejo eno, to dosežejo z uporabo ameriške angleščine in preko nje pridejo v stik z različnimi vidiki ameriške kulture. Tako delovanje odseva imperialistične težnje po uveljavitvi standardnega jezika. Cenjena je ameriška angleščina, ki jo govori manjšinski, višji sloj belskega prebivalstva, nepopolna angleščina je obravnavana kot šibkost, drugi jeziki, kot so španski ali azijski, pa niso sprejeti kot ameriški. Brez filtriranja teh imperialističnih idej se zato imigranti, v tem primeru korejska manjšina, soočajo z vprašanjem identitete (Bang, 2009, str. 148).

Za Henryjevo obsedenost z angleškim jezikom je na nek način odgovoren njegov oče s svojimi preteklimi dejanji. Oče si želi, da Henry odraste v popolnega ameriškega prebivalca in si ne beli glave s sinovo pomanjkljivo korejščino. Svojega sina vzgaja na ameriški način in tako je neizogibno, da tudi Henry prevzame njegovo zmotno mišljenje. Henry svojega sina ne uči korejščine. Boji se, da bo s svojo nepopolno angleščino vplival na sina, zato mu ne bere knjig.

Henryjeva obsedenost z angleškim jezikom pa seže še dlje. Jezik igra ključno vlogo pri vzpostavljanju odnosov z drugimi. Bolj kot ideje, ki so izražene z jezikom, mu je pomemben jezik sam. Njegova žena Lelia je po poklicu logopedinja. Zaradi narave svojega poklica jezik obvlada in pomaga popraviti različne jezikovne hibe. Žena je predmet Henryjeve zavisti, a hkrati ga privlači. Henry Lelio označi kot "osebo, ki dobro govori (she could really speak)" (Lee, 1995a, str. 10). Tudi uspešnega korejskega Američana Johna Kwanga prikaže kot osebo, ki "brez strahu govori z močnim in jasnim glasom (with his voice strong and clear, unafraid to speak the language)" (Lee, 1995a, str. 304). Bolj kot na vsebino se Henry osredotoča na zunanje lastnosti jezika. S tem si Henry, ki ima "črno kožo", nadene masko, ki se imenuje jezik, in sodi sebe ter druge. Sam se očitno trudi, da bi nosil popolno "belo masko".

A Henryjevo prepričanje se izkaže za zmotno. Henry zapostavlja očeta, ki si je zanj želel, da bi odrastel v popolnega Američana. Ironično je vzrok za njegovo obnašanje prav jezik (angleščina). Ignorira očeta, ki ni uspel popolnoma obvladati angleškega jezika, oče pa ima enak odnos do žene ter do delavcev, ki so prišli iz Južne Amerike in ne znajo angleško tako dobro kot on. Henryjev sin Mitt je umrl, še preden bi lahko s pomočjo jezika razvil lastno identiteto. Žena Lelia Henryju očita, da je "lažniv govorec jezika (false speaker of language)" (prav tam, str. 1) in odide. Tudi korejski Američan John Kwang, uspešen politik, brez težav obvlada angleščino in sanja o newyorškem

županskem stolčku, na koncu pa, v nasprotju s Henryjevimi pričakovanji, zapusti ameriško družbo.

Glede na Fanonove argumente "popolna angleščina", Henryjeva bela maska, ne more spremeniti barve njegove kože. Obvladanje jezika ne more imeti popolnega vpliva na identiteto. Četudi govori popolno angleščino, ne more postati popolni Američan. Henry, njegov oče, sin in nenazadnje John Kwang, ki ga Henry označi za svojega drugega očeta, so ujeti v prostor, v katerem niso ne "Američani" in ne "Korejci". To stanje ujetosti med dvema kulturnima prostoroma, je definirano kot "hyphenated condition" (Ling, 1999, str. 13).

Jezik določa človekov obstoj. Ne izraža le njegove identitete, je tudi nujno sredstvo medosebne komunikacije. Jezikovne prepreke lahko vodijo do nerazumevanja in občutka odtujenosti, s čimer lahko oseba zapade v krizo identitete. Brez ustrezne komunikacije pride tudi do prekinjenega odnosa z drugimi ljudmi (Bang, 2009, str. 146). Uspešno sporazumevanje je torej pogoj za ohranjanje medčloveških odnosov in navsezadnje zavedanje samega sebe. Henry si je nadel "belo masko", opazoval očeta, otroka, ženo in druge okoli sebe ter tako gojil odnose. Skrival se je za masko in ni govoril resnice. Na površju je morda dajal vtis, da mu komunikacija ne predstavlja težav, a kot mu je očitala žena Lelia, Henry s svojim skrivanjem ni postal nič drugega kot "false speaker of language".

Jezik v romanu *Native Speaker* ni le maska, predstavlja tudi ogledalo. Skozi to ogledalo lahko vidimo sebe in sogovorca. Odsev, ki ga vidimo v ogledalu, je lahko popačen, a je najboljši približek resničnosti. Henry je skozi ogledalo, ki se mu reče jezik, videl svoj odsev in ljudi okoli sebe. Na koncu je spoznal, da ima vsak svojo vrednost in da je vsak drugačen. Ta "drugačnost" je ključna beseda, ki v ameriški družbi razlaga postkolonialno večkulturnost. Konec romana odseva na novo odkrito Henryjevo identiteto. Skupaj z ženo Lelio otroke manjšin uči angleščino. A ne uči jih tako, da bi z "belo masko" lahko zakrili "črno kožo", pač pa, da bi lahko na pravi način našli same sebe. Osebe skozi ogledalo (jezik) gledajo in se v ogledalu tudi vidijo. Med učenjem ščitita identiteto manjšin in otrokom hkrati pomagata pri privajanju na novo (ameriško) družbo. Gresta še korak dlje: Lelia spodbuja otroke k uporabi njihovega pravega imena. S temi dejanji želita sporočiti, da se človeka ne da določiti le z enim jezikom ter da ni le enega jezika, ki bi prevladal na svetu.

4 Korejski elementi in korejski prevod romana *Native Speaker*

Zaradi uspeha romana *Native speaker* v Ameriki, so se zanj začeli zanimati tudi v Južni Koreji, kar je pripeljalo do kar dveh prevodov tega romana v korejščino. Prvič ga je prevedel literarni kritik Hyeon Jun-man leta 1995, torej istega leta, kot je bil izdan izvornik v Ameriki. Ta korejski prevod ima naslov "*Neitibeu seupikeo* (네이티브 스피커 Naravni govorec)", izšel pa je v dveh delih (izvornik je bil izdan v eni knjigi) pri založbi

Miraesa. Drugi korejski prevod je izšel leta 2003, prevedel ga je profesor prevajanja in tolmačenja Chong Young-mok. Svojemu prevodu je dal naslov: “*Yeongwonhan ibangin* (영원한 이방인 Večni tujec)”.

Ob prevajanju literature je treba dobro premisliti o dveh faktorjih: natančnosti in berljivosti. Natančnost se navezuje na izhodiščni (v primeru Leejevih del je to angleščina), berljivost pa na ciljni jezik (korejščina). V primeru “natančnosti” prevod poskuša ostati zvest izhodiščnemu besedilu, kar imenujemo potujčevanje. Pri neposrednem prevajanju je običajno potrebna dodatna razlaga v opombah. Pri literarnih delih dodatne opombe in razlage ovirajo tekoče branje, predvsem v primeru korejščine in angleščine, ki imata popolnoma drugačno strukturo jezika. Besedilo zato postane težko berljivo ali pa je prevod neprimeren. V primeru, da prevajalec poudarja berljivost, tak način prevajanja imenujemo podomačevanje, in v ospredje postavlja bralce ciljnega jezika. Pri tem je najpomembnejša razumljivost v ciljnem jeziku, s tem pa se ni mogoče izogniti možnosti, da se spremeni, popači ali celo izgubi besedišče, ki sicer nosi ključni pomen.

V izogib pomanjkljivostim, ki jih imata ta dva načina, je v zadnjih časih vse glasnejša ideja o tem, da prevajanje literarnih del ni le jezikovno prevajanje, pač pa “kulturno prevajanje”, kar izhaja iz mišljenja, da izbira le enega načina prevajanja ni ustrezna. Pri potujčevanju kulturnih elementov ne moremo prenesti v drug jezik le s spremembo besed, enako pa velja tudi za podomačevanje. Elementov izvirne kulture se ne da razložiti tako, da se jih umetno nadomesti z elementi ciljne. Sodeč po razpravah o “kulturnem prevajanju” se je potrebno izogniti popolni modifikaciji ali radikalnim spremembam jezika. Kultura izvirnega jezika in elementi ciljne kulture pa morajo biti vidni v prevedenem besedilu. Tudi če prevod deluje nenaravno, s tem ni nič narobe, če bralce spodbudi k razmišljanju o razlikah med dvema jezicoma (kulturama), ter se zaradi tega sprašujejo o razlogu, zakaj se je prevajalec odločil za določen prevod (Chon, 2017, str. 127-129). Proces prevajanja je torej razumevanje procesa branja. Pravo “kulturno prevajanje” ne pomeni vztrajanja pri potujčevanju ali podomačevanju pač pa pomeni izbiro metode, ki najbolj ustreza kontekstu.

Kar se tiče romana *Native speaker*, je trenutno v redni prodaji mogoče dobiti le drugi prevod, prvi pa ni več ponatisnjen. Oba prevoda sta bila sprejeta precej različno. Številni raziskovalci, kot so Koh (2002, str. 619), Lee (2008, str. 27), Lee (2011, str. 134) in drugi, so razpravljali o problematiki Hyeon Jun-manovega prevoda. Še celo Chang-rae Lee je izrazil precejšnje nezadovoljstvo in povedal, da je seznanjen s problematiko prvega prevoda (Im, 2002, str. 291). Njegova največja pomanjkljivost so napačno prevedeni ali celo izpuščeni deli besedila. Zaradi tega se v prevodu izgubi nit izvirnega besedila, kar ima posledično tudi velik vpliv na berljivost. Napačno prevedeni deli besedila so posledica napačne interpretacije odnosa med literarnimi liki. Bralci tako ne morejo pravilno razumeti njihovega zapletenega odnosa, ki je sicer zelo pomembna prvina romana (Lee, 2011, str. 134-135). V nasprotju z negativnimi odzivi na pomanjkljiv

prvi prevod, pa je bil Chong Young-mokov prevod deležen pozitivnih kritik. Odobrava ga tudi avtor romana (Lee, 2008, str. 27). Glede na to, da se prva verzija prevoda ne prodaja več, ga korejski bralci ne morejo več brati, zato se večina južnokorejskih raziskovalcev pri svojem proučevanju osredotoča na originalno besedilo in drugi prevod romana. Tudi ta članek temelji na drugem Chong Young-mokovem prevodu.

4.1 Naslov romana

Najprej je potrebno obravnavati naslov prevoda. Prvi prevajalec, Hyeon Jun-man, se je odločil obdržati angleški naslov *Neitibeu seupikeo* (네이티브 스피커). Drugi prevajalec, Chong Young-mok, pa se je odločil za lastno interpretacijo in prevodu dal naslov "Večni tujec". Naslov je tesno povezan z vsebino literarnega dela, zato nosi pomembno sporočilo, poleg tega pa predstavlja tudi protagonista Henryja. Primer prvega prevoda ne odseva prevajalčeve interpretacije. Prav tako korejski bralci iz naslova ne morejo ugibati o vsebini dela. Po drugi strani se druga izbira naslova, "Večni tujec", sicer oddalji od originalnega naslova, a bralca aktivno spodbudi k poglobljenemu razmišljanju o vsebini, prav tako se lahko med branjem bralec ponovno sprašuje o pomenu naslova dela. To je dober primer "kulturnega prevajanja".

4.2 Lik očeta

V današnjem času je korejska družba še vedno izrazito patriarhalna in v ospredje postavlja moške. S tega vidika je očitno, da skozi lik očeta v romanu lahko razumemo elemente korejskosti. Del, ki zahteva podrobnejšo obravnavo, je Henryjev sledeči opis njegovega očeta:

- (1) My father, a Confucian of high order, would commend me for finally honoring that which is wholly evident. For him all of life was a rigid matter of family (Lee, 1995a, str. 6-7).

높은 수준의 유생인 나의 아버지라면 내가 마침내 아주 자명한 것을 존중하게 되었다는 점을 들어 나를 칭찬할 것이다. 아버지에게는 인생의 모든 것이 반드시 가족 문제였다 (Lee, 2003, str. 22).

Nop-eun su-jun-ui yu-saeng-in na-ui a-beo-ji-la-myeon nae-ga ma-chim-nae a-ju ja-myeong-han geos-eul jon-jung-ha-ge doe-eoss-da-neun jeom-eul deul-eo na-leul ching-chan-hal geos-i-da. A-beo-ji-e-ge-neun in-saeng-ui mo-deun geos-i ban-deu-si ga-jog mun-je-yeoss-da.

'Moj oče, konfucijanec visokega statusa, bi me pohvalil, ker sem končno začel spoštovati tisto, kar je samoumevno. Za mojega očeta je bilo vse v življenju nujno družinska zadeva.'

Pomembni izrazi so "my (moj)", "a Confucian of high order (konfucijanec visokega statusa)" in "family (družina)", zato se kakovost prevoda ne bo pokazala v prevajanju dolge povedi, pač pa v izbiri preprostih besed. S temi tremi ključnimi izrazi bo moral prevajalec namreč izraziti Henryjevo mišljenje o očetu.

Chong Young-mok (v Lee, 2003, str. 22) je "my father" prevedel kot "*na-ui a-beo-ji* (나의 아버지 moj oče)". V korejščini se ponavadi pri naslavljanju očeta ne uporabi besede "*na-ui* (나의 moj)", pač pa "*u-li* (우리 naš)". Vseeno pa se je Chong Young-mok v svojem prevodu odločil za "*na-ui a-beo-ji* (나의 아버지 moj oče)". S tem je prikazal, da se je Henry v svojem razumevanju očeta že oddaljil od korejske skupinske mentalitete "*u-li-ju-ui* (우리주의 We-ism)". To pravilno prikazuje tudi razlike v mentalitetah prve in poldruge generacije korejske manjšine.

Chong Young-mok je "a Confucian of high order" je prevedel kot "*nop-eun su-jun-ui yu-hak-ja* (높은 수준의 유학자 konfucijanec visokega statusa)". Ključna beseda tukaj je "*yu-hak-ja* (유학자 konfucijanec)". V romanu Henryjev oče ni konfucijanec. Je migrant prve generacije, ki je sicer diplomiral na priznani univerzi v Seulu, v Ameriki pa ima v lasti majhno trgovino s sadjem. Kljub temu Chong Young-mok izraza "Confucian" ni prevedel kot "nekdo, ki verjame v ideje kitajskega filozofa Konfucija". Namesto tega je uporabil izraz "konfucijanec visokega statusa". S tem je označil očeta za filozofa ali učenjaka. Iz Chong Young-mokovega prevoda je vidno, da Henry nima slabega mnenja o očetovem konfucijanskem mišljenju. Oznaka očeta kot konfucijanca nima sarkastičnega prizvoka ne v izvornem besedilu ne v prevodu. S tem zgolj pove, da je očetovo prepričanje enako prepričanju konfucijanca visokega statusa. V tem prevodu je moč razbrati mentaliteto Vzhoda, kjer je oče na višjem položaju. Hkrati pa izraža tudi odobravanje globine te filozofije.

Chong Young-mok je bil pozoren tudi pri prevodu izraza "family (družina)". Ta se lahko v korejščino prevede kot "*ga-jog* (가족)" ali "*ga-mun* (가문)", v tem primeru pa se je prevajalec odločil za prvi izraz, *ga-jog*. *Ga-jog* označuje ožjo družino, je najmanjša možna družinska enota, povezana s krvnim sorodstvom, torej starši in njihovi otroci. Nasprotno *ga-mun* označuje širšo družino, je največja možna enota, ki povezuje krvno sorodstvo in vključuje vse z enakim priimkom. V konfucijanski kulturi je navadno pomembnejši *ga-mun* (in z njim povezano dostojanstvo). Prevajalec Chong Young-mok se je odločil očeta, konfucijanca, prikazati kot osebo, ki daje večji pomen ožji družinski enoti, *ga-jog*. To je tudi značilno za prvo generacijo korejskih migrantov. Ti sicer niso pozabili na konfucijanske ideje, ki so jih pridobili v Koreji, a je za njih pomembnejša manjša družinska enota, *ga-jog*, ki vključuje posameznika, partnerja in njihove otroke. S prevodom je tako prikazan kulturni proces, med katerim je Henryjev oče od tam (iz Južne Koreje) prišel sem (v Ameriko). Nujno je, da so bralci pozorni na podrobnosti tega prevoda.

Ti trije izrazi ne izražajo le identitete Henryjevega očeta, pač pa hkrati prikazujejo, v kakšni luči ga vidi njegov sin. Tega prevajalec sam sicer ni eksplicitno izrazil, a to

podrobnost lahko opazijo raziskovalci in pozorni bralci. Dober prevod torej pripomore k razumevanju odnosov med literarnimi liki.

Tudi sledeči izsek prikazuje očetovo identiteto:

- (2) Once when he was having some money problems with a store, he started berating her with some awful streams of nonsensical street talk, shouting “my hot mama shit ass tight cock sucka”, and “slat-eye spic-and-span motha-fucka” (Lee, 1995a, str. 63).

한번은 가게에서 어떤 돈 문제가 생겼을 때, 아버지는 말도 안 되는 거리의 욕설을 끔찍하게 내뿜으며 어머니를 호되게 나무라기 시작했다. 아버지는 “마이 핫 마마 섯 애스 타이트 콕 서카”니 “슬랜트-아이 스피크-앤-스팬 마다-퍼카”니 하는 소리를 질러댔는데, 틀림없이 가게에 온 손님들한테서 주워들었을 것이다 (Lee, 2003, str. 105).

Han-beon-eun ga-ge-e-seo eo-tteon don mun-je-ga saeng-gyeoss-eul ttae, a-beo-ji-neun mal-do an doe-neun geo-li-ui yog-seol-eul kkeum-jjig-ha-ge nae-baet-eu-myeo eo-meo-ni-leul ho-doe-ge na-mu-la-gi si-jag-haess-da. A-beo-ji-neun “ma-i has ma-ma swis ae-seu ta-i-teu kag seo-ka”ni “seul-laen-teu – a-i seu-pig – aen – seu-paen ma-da – peo-ka”ni ha-neun so-li-leul jil-leo-daess-neun-de, teul-lim-eobs-i ga-ge-e on son-nim-deul-han-te-seo ju-wo-deul-eoss-eul geos-i-da.

‘Ko je imel nekoč zaradi trgovine finančne težave, je oče začel grajati mamo in vanjo izpljunil nesmiselne ulične psovke. Vpil je ‘My hot mama shit ass tight cock sucka’ in ‘slat-eye-spic-and-span motha-fucka,’ besede, ki jih je moral pobrati od svojih strank v trgovini.’

V odlomku Henryjevega očeta razjezi problem povezan s trgovino, ki jo ima v lasti, in zato svoji ženi nameni nekaj kletvic. Kot je razvidno, Chong Young-mok kletvic v korejščino ni prevedel, le zapisal jih je s korejsko pisavo. V opombah sicer opozori na njihov rasistični element. To je primer potujčevanja, z njim pa bralec dobi občutek, da je besedilo čudno, neznano. V bralcu vzbudi začudenje nad očetovim načinom govorjenja. Če bi bil ta del preveden v korejščino, je velika verjetnost, da korejski bralci ne bi dobili občutka neznanega. Tak prevod je na nek način primer “kulturnega prevajanja” in ima najmanj tri različne učinke. Prvi je občutek neznanega. Oče je v resnici kričal v angleščini. Pred tem je Henry o očetu dejal, da začne govoriti angleško, kadar se želi zlagati ali pa kadar želi kaj prikriti. Če bi bil ta del, v katerem oče preklinja v polomljeni angleščini, preveden v korejščino, bi ob vprašanju, kaj skuša s tem dejanjem zakriti, občutek neznanega izginil. Drugič, prevod prikaže odnose med osebami. V Henryjevi družini obstaja (hierarhična) jezikovna piramida – položaj člana je odvisen od znanja angleščine. Vrh zaseda Henry, za njim je oče in najnižji položaj zaseda mati. Oče na tak način preklinja izključno v pogovoru z materjo. S sinom, ki nosi svojo “belo masko” popolne angleščine, se ne more kosati v znanju jezika in ne more govoriti

z njim na tak način. Oče si svojo "belo masko" nadene le v odnosu z materjo. Če bi bil torej ta del preveden v korejščino, ne bi bilo mogoče prikazati le jezikovne piramide med tremi osebami, pač pa bi bil oče prikazan kot nasilnež, ki se verbalno znaša nad svojo ženo. Z uporabo kletvic, ki jih mati ne more razumeti, oče izrazi svojo jezo, a hkrati tudi spoštovanje do svoje žene. Ona namreč resnično ne razume njegovih kletvic. Tretjič, ta del prikaže očetovo naravnost – tudi on nosi "belo masko". Kletvice so na nek način namenjene tudi Henryju. S tem mu želi pokazati, da se tudi sam trudi postati Američan, in ga prosi, naj razume, kako težak je zanj ta proces. Korejski prevod torej ne razume le ozadja očetovega izbruha, pač pa tudi dobro prikazuje odnose med osebami in zgoraj omenjen občutek tujega.

Dialog znotraj literarnega dela je eden izmed najzahtevnejših delov prevajanja. Glede na osebnost literarnih likov in situacijo, v kateri se dialog odvija, se razlikujeta ton pogovora in stopnja vpljudnosti. Chong Young-mokov prevod se trudi kar najbolje ohraniti občutja, ki jih ima bralec ob prebiranju originala. Z odločitvijo, da ohrani očetove kletvice, ne prikaže le njegove osebnosti, pač pa tudi situacijo korejskih izseljencev v ZDA, zato je to dober primer prevoda in njegove vloge.

4.3 Lik žene

Lelia, Američanka, je žena protagonista Henryja, zato je eden izmed glavnih literarnih likov romana in ima ključno vlogo pri Henryjevem procesu iskanja identitete.

V naslednjem odseku protagonist omeni svojo ženo. Zopet se pojavi dilema, kako bomo prevedli osebni zaimек "my".

(3) I(Henry) should have warned my American wife (Lee, 1995a, str. 8).

나의 미국인 아내에게 미리 알렸어야 하는 건데 (Lee, 2003, str. 24).

Na-ui mi-gug-in a-nae-e-ge mi-li al-lyeoss-eo-ya ha-neun geon-de.

'Svoji ameriški ženi bi moral povedati vnaprej.'

Če bi želeli, da bi ta stavek, ki je del Henryjevega monologa, v korejščini zvenel naravno, bi ga prevedli: "*a-nae-e-ge mi-li mal-haess-eo-ya haess-neun-de* (아내에게 미리 말했어야 했는데 ženi bi moral povedati vnaprej)". Če bi Henry sogovorniku predstavil svojo ženo, bi običajno v korejščini uporabil množinski zaimек, torej "*u-li a-nae* (우리 아내 naša žena)", a v primeru Henryjevega monologa je prevod zaimeka odveč, zato bi ga bilo bolj naravno izpustiti. Chong Young-mok je ta del prevedel "*na-ui mi-gug-in a-nae* (나의 미국인 아내 svoji ameriški ženi)", kar v korejščini zveni zelo nenaravno. Vendar ne bi mogli reči, da gre za primer potujčevanja, saj ni značilno za ameriško kulturo, da bi se ženino državljanstvo namerno poudarilo. To je Henryjev namen. Namerno je poudaril, da je "njegova (moja)" žena "Američanka". Izraza "moja" in "ameriška" prikažeta Henryjevo identiteto in hkrati odnos, ki ga imata z ženo. Čeprav

izbira prevoda v korejščini zveni nenavadno, pa je očitno, da je prevajalec dobro razumel izbiro Henryjevih besed. Z “moja žena” nakaže, da sta si blizu, z “ameriška (torej ne korejska)” pa, da nimata istega ozadja in zato ne moreta biti na popolnoma enakem nivoju. To je bistvo njunega odnosa, zato je prevajalec z izbiro besed na to opozoril in s tem vzbudil zanimanje pri bralcih, da se sprašujejo o namenu tega prevoda in globlje razumejo literarno delo.

Še en del, ki prikazuje njun odnos in ga je potrebno omeniti:

(4) “(Lelia) Sorry, but I was a low-down bitch.”

“(Henry) No you weren’t.”

“(Lelia) Damn it. Henry!”

She got up. “Shit. I’m sorry. Do you want a drink?” (Lee, 1995a, str. 116-7)

“미안해. 당시에 나는 비열한 년이었어.”

“아냐, 그렇지 않았어.”

“젠장, 헨리!”

털리아는 일어섰다. “제기랄. 미안해. 한 잔 할래?” (Lee, 2003, str. 182)

“*Mi-an-hae. Dang-si-e na-neun bi-yeol-han nyeon-i-eoss-eo.*”

“*A-nya, geu-leoh-ji anh-ass-eo.*”

“*Jen-jang, hen-li!*”

Lil-li-a-neun il-eo-seoss-da. “Je-gi-lal. Mi-an-hae. Han jan hal-lae?”

“Oprosti. Do tebe sem se obnašala kot zlobna prasica.”

“Ah, ne, nisi se.”

“Pizda, Henry!”

Vstala je. “Šit. Oprosti. Bi kaj spil?”

Chong Young-mok je njun dialog v korejščino prevedel tako, da se Lelia in Henry med seboj tikata in uporabljata neformalno obliko govora. S tem dajeta občutek, da sta med seboj enakovredna. Prevajalec je Leliino preklinjanje v korejščino prevedel neposredno: “*bitch*” kot “*nyeon* (년)”, “*damn*” kot “*jen-jang* (젠장)”, in “*shit*” kot “*je-gil-al* (제기랄)” (Lee, 2003, str. 182). Čeprav govorita v neformalni obliki, preklinja le Lelia. Zato lahko rečemo, da ima v pogovoru vseeno večjo moč Lelia. V korejski družbi običajno oseba na nižjem nivoju ne sme biti nespoštljiva v pogovoru z osebo na višjem nivoju, pred njo ne sme preklinjati niti v primeru, ko so kletvice namenjene samemu sebi. Chong Young-mok dobro razume to korejsko družbeno pravilo in ga tudi primerno prikaže v svojem prevodu. Čeprav se zdi, da sta Lelia, ki je dejansko belopolta, in Henry, ki nosi belo masko, v enakovrednem položaju, pa to ne drži. Zaradi kletvic, ki jih uporablja Lelia, se ta drobna podrobnost v prevodu ni izgubila.

Analizirajmo ta odsek in ga primerjajmo s Hyeon Jun-manovim prevodom.

(5) “미안해요. 하지만 난 정말 나쁜 여자예요.”

“아냐, 당신은 그렇지 않아.”

“그만 좀 해요, 헨리.”

그녀는 벌떡 일어섰다. “미안해요. 뭘 좀 마시겠어?” (Lee, 1995b, str. 179-80)

“*Mi-an-hae-yo. Ha-ji-man nan jeong-mal na-ppeun yeo-ja-ye-yo.*”

“*A-nya, dang-sin-eun geu-leoh-ji anh-a.*”

“*Geu-man jom hae-yo, hen-li.*”

Geu-nyeo-neun beol-tteog il-eo-seoss-da. “Mi-an-hae-yo. Mwol jom ma-si-gess-eo?”

“Oprostite mi. A jaz sem res slaba ženska.”

“Ah, ne, ti že nisi taka.”

“Prenehajte, prosim, Henry.”

Henadoma je vstala. “Oprostite mi. Bi kaj spili?”

Pri Hyeon Jun-manovem prevodu Lelia uporablja formalno, vljudno obliko govora, Henry pa neformalno, zato Lelia skladno z družbeno normo ne more uporabljati kletvic, kadar govori s Henryjem. Prevajalec se je kletvice odločil prevesti: “*bitch*” kot “*yeo-ja* (여자 ženska)”, “*damn*” kot “*geu-man jom hae-yo* (그만 좀 해요 prenehajte, prosim)”, in “*shit*” kot “*mi-an-hae-yo* (미안해요 oprostite mi)”. Očitno je v tem primeru moški, Henry, na višjem položaju od ženske, Lelie. Obravnavani dialog je očiten dokaz, da se glede na prevod popolnoma razlikuje prikaz odnosa med literarnima osebam.

4.4 Lik Johna Kwanga

John Kwang je oseba, ki ima, podobno kot Henryjeva oče in žena, velik vpliv na Henryja, saj je primer uspešnega Korejca. Bil je deležen elitne izobrazbe, je uspešen politik v Ameriki in se celo spogleduje s pozicijo newyorškega župana. Na Henryja, ki identiteto oseb sodi na podlagi njihove izgovorjave angleščine, John Kwang naredi poseben vtis. Henry občuduje njegovo tekoče znanje angleščine. Prav tako ga spoštuje kot nekoga z njemu enako “črno kožo”, ki mu je uspelo postati del ameriške prevladujoče družbe (Chon, 2017, str. 138). A prav ta uspešni Kwang na koncu romana zapusti ameriško družbo. Henry se skozi proces njegovega odhoda zazre vase in prične iskati lastno identiteto. Tako je v resnici Kwang nanj najbolj vplival in je na nek način njegov (drugi) oče.

Pomemben del prevoda je njegov priimek. “Kwang” bi se po korejsko zapisalo kot “*gwang* (광)”, a v resnici tak priimek v Koreji ne obstaja. Chon YoungEui 전영의 (2017, 138) je dejal, da je mogoče, da se je pisatelj, ki ni navajen na korejsko kulturo, zanj odločil pomotoma, saj izrazi “*gang* (강)”, “*gwang* (광)”, “*kkwang* (깡)” in “*kwang* (淸)” zvenijo zelo podobno. A trditi, da je to napaka avtorja, ni nič drugega kot le ugibanje. V Leejevem romanu je s pomočjo Leliine razlage možno sklepati, da avtor jasno razume

korejske, kitajske in japonske priimke (Lee 1995a, 10). Težko je razumeti vzrok, zakaj se je avtor odločil za zapis "Kwang", interpretacija prevajalca Chong Young-moka v korejskem prevodu pa je sledeča: priimek "Kwang" se je odločil prevesti kot "*Gang* (강姜)", ki je eden izmed najpogostejših korejskih priimkov in ima tudi več pomenov. Prvič, splošno je razširjeno prepričanje, da imajo osebe s priimkom Gang trmast značaj. To mišljenje je tako razširjeno, da je zapisano celo v zgodovinskih knjigah (Sahoegwahagwon, 1975, str. 406). Temu ustreza tudi prikaz literarnega lika Kwanga, ki je tip človeka s trmasto osebnostjo. Še ena lastnost, ki se navezuje na izgovorjavo besede "*gang*", je izraz "močan (강強 gang)". Ena izmed pglavitnih značilnosti Kwanga je, da je "močan". S to odločitvijo je prevajalec dobro prikazal značaj tega pomembnega literarnega lika.

5 Zaključek

Literatura diaspore vključuje prepletanje dveh različnih jezikov in kultur, v primeru Leejevega romana *Native Speaker* sta to ameriška in korejska kultura. Prav razumevanje teh dveh kultur je ključ do analize korejskega prevoda. Dober prevod pomaga pri razumevanju in proučevanju korejskosti, različnih interpretacijah in razumevanju literarnega dela. To pa nam služi tudi kot pomoč pri razumevanju literature diaspore korejskega naroda.

Glavna tema opisanega romana je identiteta, "potovanje v iskanju samega sebe", s katerim se spopada protagonist romana Henry, ki je del korejske manjšine v Ameriki, a si želi postati popoln Američan. Henry vidi svet v ogledalu z imenom "jezik". Trudi se, da bi popolnoma obvladal angleščino ter s tem postal popoln Američan, a zanj jezik, ki bi moral biti njegova maska ali osebna izkaznica, postane le ogledalo. Ljudi okrog sebe sodi glede na stopnjo znanja jezika. Zato je jezik tega literarnega dela najbolj pomembno ogledalo, ki prikazuje odnose med literarnimi liki. Henry v svojem ogledalu spoznava ljudi in svet, ter išče samega sebe. Končna identiteta, ki jo najde, ni niti ameriška niti korejska. Najde zgolj samega sebe.

Za razumevanje romana je nujno razumeti pomen besede "jezik", saj literarno delo prikazuje proces iskanja lastne identitete skozi jezik. Za korejskega bralca ima zato prevod romana ključno vlogo, saj mora bralca opozoriti na vse podrobnosti. Prevajalca Chong Young-moka jezikovne ovire, ki jih je moral prebroditi pri prenašanju romana s te (angleške) strani v tisto (korejsko), niso ustavile. Prevajalec s prevodom zbuja pozornost bralcev ter jih spodbuja k razmišljanju. *Native Speaker* tako postane roman, ki se bere skupaj s prevodom in bralce spodbuja k razmišljanju.

V pričujočem članku je prikazana vloga prostega prevoda naslova, ki se navezuje na temo romana, raba zaimkov, ki prikazujejo korejsko miselnost "*We-ism*" in patriarhalnost, sprememba odnosa med zakoncema in nenazadnje globlji pomen korejskih imen (priimkov). Na podlagi primerov in analize prevodov raziskava predlaga

nov pristop pri interpretaciji literarnih del korejske diaspore, kar bi pripomoglo tudi k bolj poglobljenem razumevanju literature diaspore korejskega naroda. Prav tako članek nagovarja k razmisleku in večji pozornosti pri obravnavanju tovrstne literature v Južni Koreji.

Prevedla: Rozalija Stegnar

Acknowledgments

This research was supported by the generous grant from Academy of Korean Studies, as part of AKS Programs for Korean Studies Abroad (AKS-2020-R90).

Reference

- Bang, J. 방정민 (2009). Ichangnae soseol, neitibeu seupikeo yeongu 이창래 소설, 네이티브 스피커 연구. *Bigyomunhak 비교문학*, 48, 135-156.
- Chon, Y. 전영의 (2017). Ichangraeui yeongwonhan ibangine natanan honjongjeok yongmanggwa eoneogwollyeon 이창래의 영원한 이방인에 나타난 혼종적 욕망과 언어권력. *Hyeondaesoseoryeongu 현대소설연구*, 67, 121-147.
- Fanon, F. (2015). *Peau Noire, Masques Blancs*. Montreal: Kiyikaat Editions.
- Im, H. 임혜기 (2002). *Salamdeuleun jakku geugoseul balabonda 사람들은 자꾸 그곳을 바라본다*. Seoul: Wolganjoseonsa 월간조선사
- Ji, B. 지봉근 (2004). Ichangnaeui telgwoneomintelme natanan hangukgye miguginui jeongcheseong: munhwajeok chaiwa japjongseong 이창래의 원어민에 나타난 한국계 미국인의 정체성: 문화적 차이와 잡종성. *Bigyomunhak 비교문학*, 33, 303-322.
- Kim, M. 김미영 (2006). Honseongjeog sahoee-ui seosajeog daeeung - ichanglae soseol-ui teugseong yeongu 혼성적 사회에의 서사적 대응 - 이창래 소설의 특성 연구. *Gugeogugmunhag 국어국문학*, 143, 379-406.
- Koh, B. E. 고부웅 (2002). Ichangnaeui woneomin: bieoinneun gipyoui jeongcheseong 이창래의 원어민: 비어있는 기표의 정체성. *Yeongeoyeongmunhak 영어영문학*, 48(3), 619-638.
- Lee, C. (1995a). *Native Speaker*. New York: Riverhead Books.
- Lee, C. (1995b). *Neitibeu seupikeo 네이티브 스피커* (Hyeon, Jun-man 현준만, Trans.). Seoul: Miraesa.
- Lee, C. (2003). *Yeongwonhan ibangin 영원한 이방인* (Chong, Young-mok 정영목, Trans.). Seoul: NamuwaSup.
- Lee, H. 이형진 (2011). Ichangnaeui Native Speakerui hangeulbeonyeokbon bigyoyeongu 이창래의 Native Speaker 의 한글번역본 비교연구 (A Comparative Study of the Korean Translations of Native Speaker). *Segyemunhakbigyoyeongu 세계문학비교연구*, 36, 131-165.

- Lee, J. 이진실 (2008). Seosulirone uihan ichanraeui Native speaker beonyeokbon bigyoyeongu 서술이론에 의한 이창래의 Native speaker 번역본 비교연구. (Comparative Analysis of the translations of Chang-rae Lee's Native Speaker based on Narrative Theory). Master's thesis. Seoul: Sookmyung Women's University.
- Ling, A. (1999). *Writer in the Hyphenated Condition: Diana Chang*. Philadelphia: Chelsea House.
- Park, E. 박은정 (2019). Ichanglaeui neitibeu seupikeoe natanan eon-eowa jeongcheseong-ui munje 이창래의 네이티브 스피커에 나타난 언어와 정체성의 문제 (Language and Identity in Native Speaker). Master's thesis. Jeju: Jeju National University.
- Sahoegwahagwon 사회과학원 (1975). *Rijo sillok: sukjong 리조 실록: 숙종*. Pyongyang: Sahoegwahag chulpansa 사회 과학 출판사.