Journal of Language and Literature
Vol. 21 No. 2, October 2021, pp. 403-417
DOI: 10.24071/joll.v21i2.3252
Available at https://e-journal.usd.ac.id/index.php/JOLL/index
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

The Addition of Indonesian Prefixes *meN-* and *di-* to English bases: A Corpus-based Study

Alifa Camilia Fadillah[1], Ika Nurhayani[2], Sri Endah Tabiati[3]
fadillahalifa@gmail.com[1], inurhayani@gmail.com[2], stabiati@gmail.com[3]
Linguistics Department, Universitas Brawijaya, INDONESIA

Abstract

This paper serves as an initial identification of the addition of Indonesian inflectional prefixes meN- and di- to English bases of any word class through a corpus-based study. With the prevalence of English influence in Indonesian native speakers' linguistic repertoire, particularly within the scientific and computational domain, there emerges a tendency to resort to the original terms in English than those of the Indonesian equivalences. This phenomenon, addressed as leksikalisasi timpang *or unequal lexicalization, refers to the use of words in source language to make up for the lack of corresponding lexicalization in target language. This leads to a linguistic innovation to 'localize' English words by adding Indonesian inflectional prefixes such as meN- and di-. Out of 1 million sentence size Web corpus obtained from The Leipzig Corpora Collection, this paper is able to yield approximately 489 (0,21%) combinations of* meN- *+ English bases with 2,813 (0,018%) word tokens and 475 (0,20%) combinations of* di- *+ English bases with 2,377 (0,015%) word tokens. Six allomorphs of* meN- *are also attested, namely* meng-, men-, mem-, me-, menge-, *and* meny-, *with* meng-, men-, *and* mem- *as the most used allomorphs by word frequency and type. This investigation backs up the hypothesis that the process of word assimilation leads to nasal sound changes. This paper also observes that there are 13 most used typographic forms shared between the combinations of meN- and di- + English bases, and 7 other forms on a very low frequency. The words observed in this paper's database are then grouped into three semantic clusters based on their use in context: computer-related (CR), non-computer-related (NCR), and both (NCR/CR), where computer-related words are observed to dominate the database. The findings indicate that this linguistic creativity is the outcome of how familiar Indonesians are with English terms than the official equivalences, especially towards technology and computational vocabulary.*

Keywords: *Indonesian prefix; English base; corpus study; morphology*

Article information

Received: 27 March 2021

Revised: 20 June 2021

Accepted: 29 June 2021

Introduction

As the first foreign language to be officially acknowledged in Indonesia, the influence of English in the evolution of Indonesian vocabulary is inevitable (Lauder, 2008; Percillier, 2016; Sneddon, 2003). Indonesian has borrowed a considerably extensive amount of borrowing words from English occurring in most global domains, such as sport, movies, music, popular culture, business, banking, politics, trading, military, science, medical, and computational (Kachru & Nelson, 2006; Lowenberg, 1991; Sneddon, 2003). To be successfully assimilated, these loanwords have gone through some nativization process such as nasal sound change, semantic shift, word order, not to mention the process of turning phrasal verbs into verbs such as '*mem-back up'* (from 'to back up'), which all contribute to the conformity of the Indonesian grammar (Sneddon, 2003). On the one hand, there are also instances of unassimilated loanwords where they retain their original spellings. In most cases, these words have their own equivalences in Indonesian, such as 'download' (*unduh*), 'upload' (*unggah*), 'print' (*cetak*), 'copy' (*salin*), and 'paste' (*tempel*). Kadarisman (2005) addresses such phenomenon as *leksikalisasi timpang* or unequal lexicalization where signs in source language do not have their suitable equivalence in target language, resulting in the process of adopting the signs as they originally are to keep the information flow smoothly. That is, as Kadarisman (2005) states, English words like 'CPU', 'monitor', 'printer', 'printout', and others are mostly used by Indonesians as they originally are without the need to translate all of them to Indonesian.

Although later on many of these unassimilated loanwords have been translated into their respective equivalences in the 2008 edition of Indonesian official dictionary *Kamus Besar Bahasa Indonesia* (KBBI), the fact that the original words reign longer than the corresponding equivalences, with respect to computer-mediated communication, help make up for the lexical familiarity (Kadarisman, 2005; Manns, 2010). This is also supported by Sneddon's (2003) statement in which common words in English are often much more memorable than the Indonesian version prior to the prestige status pertaining to English mastery and the image of modernity. Henceforth, with the help of mass media and the vast advancement of computational technology in particular, English has become 'naturalized' in a way it is assimilated with Indonesian's linguistic features, affixes, for example (Kadarisman, 2005; Saddhono & Sulaksono, 2018; Smith-Hefner, 2007; Sneddon, 2003). To quote Sneddon (2003, p. 183):"As the word becomes more common, it becomes more assimilated; increasing numbers of people feel comfortable to use it as if it were a native word".

The combination of Indonesian prefixes and English bases that emerges from this phenomenon, such as *men*download "download", *mem*posting "posting", and others, prove that indeed Indonesians have assimilated their linguistic feature with the former to make it sound more "natural" when either spoken or written. It is also found that *N-* which represents nasal sounds also changes to assimilate with the first sound of the English bases, meaning that Indonesians have applied the same phonological rule to English as well (Sneddon, 2003). Interesting enough, this occurrence becomes prominent with the development of the Internet and social media in the 21st century, hence combinations such as *di*-follow "followed", *di*-add "added", *di*-upload "uploaded" which are all references to social media activities are being utilized dominantly, as well as the combination of other English bases in other domains as previously mentioned (Oktavia, 2019; Saddhono & Sulaksono, 2018).

Although Saddhono and Sulaksono (2018) have explicitly mentioned that this unique form exists by providing more data than Sneddon (2003) and Kadarisman (2005) did—in which the examples are reflective rather than factual—the data collected is limited to institutional domain only, where they recorded students' conversations in five universities located in five big cities in Indonesia. In addition, Saddhono and Sulaksono (2018) have yet to provide any morphological analysis to such phenomenon, as they state that such form is a random occurrence—meaning it does not follow any phonological rules of either of Indonesian or English language. On the contrary, Sneddon (2003) has explicitly stated that the

Indonesian phonological rule applied to such combinations, but he only provides a small set of examples that uses the prefix *meN-*, not to mention that the gap between both Sneddon's and Saddhono and Sulaksono's study are quite distant. Taking such consideration in mind, this paper wants to contribute further to investigate the phenomenon of the combination of Indonesian inflectional prefixes and English base words (henceforth English bases) through a corpus-based investigation.

Indonesia recognizes three prefixes which all form transitive and intransitive verbs, *ber-, meN-,* and *di-* (Sneddon, 2010).Prefix *meN-* and *di-* can also be called inflectional voice prefixes as they play role in indicating whether a sentence is in active or passive voice. Regarding *ber-*, which can be attached to either transitive or intransitive verbs, its occurrence is inconstant and oftentimes interchangeable with *meN-* if attached to intransitive verbs (Sneddon, 2010). Therefore, the use of *meN-* and *ber-* especially within the intransitive verbs depends on the person's familiarity towards either prefix, as Sneddon (1996) states "one form is more common than the other".

Unlike *ber-* and *di-*, *N-* in *meN-* alongside *peN-*, and *peN-…-an* represents a sound change depending on the first sound of the base, leading to a phenomenon of nasal allomorphy (Denistia & Baayen, 2019; Sneddon, 2010; Sukarno, 2017). Therefore, to quote Denistia and Bayen (2019, p. 387), *meN-* and *peN-* are examples of "classical phonologically conditioned allomorphy". Sukarno (2017) creates a comprehensive table to illustrate the sound changes of *N-* according to their distributions:

Table 1. Map of *N-* nasal change adopted from Sukarno (2017, p. 48)

Phoneme	Allophones	Distribution/Context
/N/	[m]	If it is followed by a labial sound (p, b, f) and they occur in different morphemes
	[n]	If it is followed by an alveolar stop sound (t, d) and they occur in different morphemes
	[ɳ]	If it is followed by a voiceless stop sound (s) and they occur in different morphemes
	[ŋ]	If it is followed by a velar and a vowel sound (k, g, h) and they occur in different morphemes
	[ø]	Elsewhere

It also appears that when *N-* is attached to successfully-assimilated English loanwords, the allophones also emerge with one condition that differs it from native words: it is likely that the first sound remains instead (Sneddon, 2010). Therefore, in words like *mentargetkan* (to target) and *mengkontrol* (to control), *N-* changes to its respective allophone followed by the first consonant of the word. However, as time goes by and Indonesians have favored the words like they do with the native ones, the original application of *N-* sound change pertains, resulting in the loss of the initial consonants: *mentargetkan* becomes *menargetkan*, and *mengkontrol* becomes *mengontrol* (Sneddon, 2003). Regardless, it has to be kept in mind that such change is rather flexible and often replaceable in accordance with one's familiarity with either form, although oftentimes the mass media is the most significant influence that the latter becomes more common than the former. Even so, the possibility that nasal sound change also applies in unassimilated English loanwords, which are also ubiquitous in Indonesians' linguistic repertoire, has not been further studied in academia, allowing for further observation to shed lights in this particular area.

Prefix *meN-* and its passive counterpart *di-* can also be optionally supplemented with verb-forming suffixes such as *-kan* and *-i* which can affect the semantic roles of a word once either suffix is added (Arka & Yannuar, 2016; Denistia & Baayen, 2019; Sneddon, 2010). The mention of the *suffixes* -kan *and -i* is very important in this study as it was identified in this study's database that there are indeed additional affixes that emphasize the semantic

roles of the combinations of English bases and Indonesian prefixes *meN-* and *di-*. The suffix *-kan* signals causative, instrumental, and benefactive functions, while the suffix *-i* marks locative and repetitive functions. Many verb bases can handle both *-kan* and *-i,* such as *tawar* 'offer' and *masuk* 'put/enter', but they differ in meaning once either suffix is added (Sneddon, 2010), where *-kan* signals object as a patient, and *-i* signals locative and recipient function. In addition, some verbs have no *-kan* or *-i* counterparts, such as *menyewakan* 'to lend' and *menghiasi* 'to decorate' (Sneddon, 2010). But in some cases, very contrary to the former example, there are verb bases that can take both *-kan* and *-i* with the same meaning, and both forms are also commonly acceptable (Sneddon, 2010). It was later that the semantic distribution of suffixes *-kan* and *-i* is formulated with the *distinctive* and *similarity* hypotheses using the hierarchical clustering analysis, resulting in families of derivational roots that group together and those attached to *-kan*/*-i* that are segregated (Rajeg et al., 2019).

This paper aims to be a follow up investigation towards identifying and analyzing the combination of Indonesian inflectional affixes and English bases in the lights of three problems, (i) the addition of inflectional affixes *meN-* and *di-* to English bases and their allomorphs, (ii) the typographic constraints of the combination of Indonesian inflectional affixes and English bases, and (iii) the frequency of the aforementioned combination in two semantic clusters, computer-related and non-computer-related verbs. The investigation of the allomorphy of *meN-* and *di-* when combined with English bases unveils what causes the choice of selected allomorphs with regards to both Indonesian and English morphological system. The typographic constraints refer to how one combination is styled in this paper's database along with the frequency of each typographic forms to compare which forms is used the most. As an example, the English base 'download' has three typographic forms in this paper's database: *mendownload* (N=282), *men-download* (N=63), and *men download* (N=4). Last but not least, since previous studies have claimed that the combination of *men-* and *di-* with English bases emerges with the familiarity around computational terms, the recorded combinations in the database are separated into three semantic clusters: computer-related (CR), non-computer-related (NCR), and both (NCR/CR), since there are combinations where they occur in both semantic clusters following the concordance results.

Therefore, based on these considerations, there are limitations this paper wants to focus on. First, *ber-* is not included with the assumption that *meN-* and *di-* are more prevalent than *ber-* and yields more data when combined with English bases. Second, me*N-* receives sound changes according to the English bases as it does to Indonesian words, allowing for the possibility that it also applies to unassimilated loanwords as it does to successfully assimilated ones. And last but not least, the additional affixes such as *-kan* and *-i* are expected to be found in the data as they emphasize more on the semantic roles of the verb bases.

Methodology

Upon building the database, this paper uses one part of the *Indonesian Leipzig Corpora Collection,* specifically the `ind-com_web_2018_1M`, which is composed of text materials taken from random Web sites (Goldhahn, Dirk; Eckart, Thomas; Quasthoff, 2012). The files can be accessed from https://wortschatz.unileipzig.de/en/download/Indonesian#ind_mixed_2013. This corpora has been widely utilized by fellow researchers on various Indonesian morphology issues, which elicits importance of Leipzig Corpora Collection (henceforth LCC) within the academia (see more Choi, 2019; Denistia, 2019; Denistia & Baayen, 2019; Rajeg et al., 2019, 2020; Rajeg & Rajeg, 2017).

In choosing ind-com_web_2018_1M to build the database, several considerations are taken into account to make sure that this paper is equipped with a large set of data to analyze. First, compared to other types of corpus (mixed, news, newscrawl, web-public, and web), ind-com_web_2018_1M consists of various Web sites which elicits a range of topics and themes. Second, the corpus is the largest and the newest among others: approximately 237,677 word types and 1,5420,886 word tokens within 1 million

sentences. The data compiled by Saddhono and Sulaksono (2018) is restrained by several limitations, such as the domain and the location of data source. Through the use of Web-based corpus, one can obtain an extensive data organized from diverse sources of various domains as well. In terms of analyzing the corpus, Denistia and Baayen (2019) use MorphInd, a morphological analyzer specifically made for the Indonesian language, but because this study focuses on the combination of Indonesian and English, the corpus was then analyzed using AntConc, a free software corpus analysis toolkit for concordance and text analysis. The use of AntConc is very essential in this study as its Concordancer tool helps to display several data examples that a researcher wants to look for as it is quick and efficient to use (Anthony, 2005). For the best result, version 3.5.8 is used to match the Macintosh OS X 10.14.6, but they also have versions for Windows and Linux, as well as their older versions for another operating system.

Since the corpus is quite large to process at once, several search patterns are customized and utilized in order to obtain all of the combinations existed in the corpus. First, in finding and selecting the construction of *meN-* + English bases and *di-* +English bases, the Word List feature was started to display all the words in the corpus and then they were sorted by their alphabetical order. The results were then cloned for easier navigation and transferred to Microsoft Word once the desired words appeared. At this point both the combination and Indonesian words were still mixed up, hence the transfer to Word for manual separation. However, there are several typographic styles in the database that the combinations are bound to, especially with the construction *meN-* + English bases since *meN-* allows for nasal sound changes. Therefore, word clusters beginning with *men-, meng-, me-, mem-,* and *menge-* with the exclusion of *meny-* were activated in the Clusters feature to allow for more results. The position of *meN-* allomorphs should be on the left and the English bases on the right.

A problem arose with the findings of *di-* + English bases. While there were no significant problems in applying the first pattern, searching for word clusters beginning with *di-* reveals 257,668 word frequency, leading to some technical issues concerning both the software (AntConc) and the hardware (personal computers) when the search was run. To tackle this problem, results from the first data run using Word List were duplicated and converted into a plain text file. The text file then was loaded in the Clusters feature's advanced search where the custom list was used as the search terms. This way one could obtain more typographic styles of a certain *di-* combination, but unfortunately disregarded the appearances of other English bases clustered with *di-*.

Alternatively, one can always use customized Regular Expressions (Regex) by entering the right formula to find the combination of *men-* and *di-* + English bases. For future references, here are some of the possible Regex which can be used in the Concordance feature:

- \bme[a-z]+?[a-z]\b for *meN-* + English bases
- \bdi[a-z]+?[a-z]\b for *di-* +English bases.

While the following formulas can be used in the Clusters feature without turning on the Regex option:

- meN(and its allomorphs)#
- di#

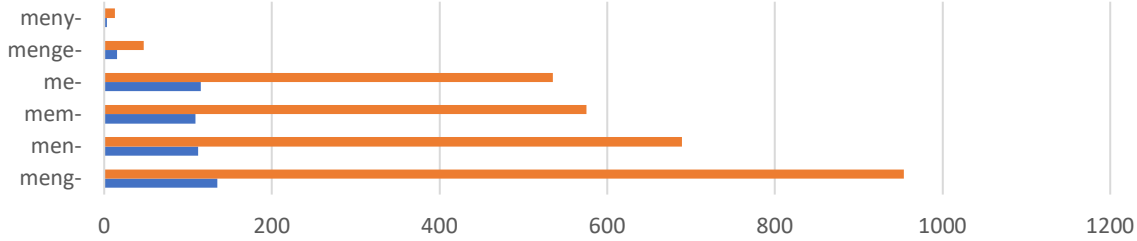
These alternatives still have to undergo manual separation to find the right combination. Cloning the results after the search stops is highly recommended to document the database in the desired format, along with the ranks, frequencies, and word ranges.

Results and Discussion

Allomorphs of meN- and di-

Out of the ind-com_web_2018_1M corpus, this paper is able to yield approximately 489 (0,21%) combinations of *meN-* + English bases with 2,813 (0,018%) word tokens and 475 (0,20%) combinations of *di-* + English bases with 2,377 (0,015%) word tokens. It is also observed that within the database there are six allomoprhs of *meN-*, namely *men-, meng-, mem-, meny-, menge-,* and *me-* (see Figure 1). Since *di-* does not have allomorphs, this paper focuses only on the allomorphs of *meN-*.

Figure 1. The distribution of *meN-* allomorphs in the database



	meng-	men-	mem-	me-	menge-	meny-
Tokens	954	689	575	535	47	13
Types	135	112	109	115	15	3

Figure 1 shows that *meng-* yields the most combinations with 135 types and 954 tokens, followed with *men-* in second place with 112 types and 689 tokens, and *mem-* in third place with 575 types and 109 tokens. Prefix *me-* which does not undergo nasal changes comes in fourth with 115 types and 535 tokens. Lastly, *menge-* comes in fifth place with 15 types and 47 tokens, while meny- is the prefix yielding the least combinations with only 3 types and 13 tokens. However, it is also observed that the combinations attached tothe allomorphs shows inconsistencies with the Indonesian nasal change rules. Inconsistencies occur in allomorph *meng-, men-,* and *mem-,* the top three prefixes with the most word combinations, while the latter halves *me-, menge-, and meny-* are undetected of such variability.

As an allomorph with the most word types and tokens, there are approximately 17 combinations of *meng-* and English bases which detected to deviate from the nasal change rules in Indonesian (see Table 2). Originally, *meng-* occurs before words with vowel and velar initials **/k/** and **/g/,** such as '*meng-coding*' (N=1), '*meng-echo*' (N=1), '*mengimport*' (N=18), '*mengapply*' (N=1), and '*meng-upload*' (N=44).. However, words with initials **/f/**, **/l/**, **/p/**, **/s/**, **/ʧ/**, **/z/**, **/ʤ/**, and **/v/** also appear to be combined with allomorph *meng-*, such as the word '*meng-file*' (N=2), '*meng-folder*' (N=1), '*meng-plugin*' (N=1), '*meng-charge*' (N=1), '*mengsave*' (N=1), '*meng- photo*' (N=1), '*menggenerate*' and '*meng- video*' (N=1).

Table 2. Anomalies of allomorph *meng-*

word	frequency	base	initial
meng-file	2	file	f
meng-folder	1	folder	f
meng-charge	1	charge	ʧ
meng-like	1	like	l
meng-plugin	1	plugin	p
meng-private-kan	1	private	p
meng-scan	1	scan	s
meng-share	2	share	ʃ
meng-tap	1	tap	ʧ
meng-zoom	1	zoom	z
menggenerate	5	generate	ʤ
mengsave	1	save	s
mengstream	1	stream	s
mengsubmit	1	submit	s
mengtap	1	tap	ʧ
meng- photo	1	photo	f
meng- video	1	video	v

Although not as many as allomorph *meng-* , 5 words are detected to exhibit variability outside the Indonesian nasal change rules (see Table 3). Allomorph *mem-* is usually attached to word initials **/b, p, f/** , such as '*mem-browse*' (N=1), '*mempublish*' (N=7), '*memformat*' (N= 44), and '*memfilter*' (N=36). Variability occurs when initials **/k/, /d/, /t/, /w/,** and **/r/** appear after *mem-*, such as '*memchached*' (N=4), '*memdisk*' (N=1), '*memtest*' (N=3), '*mem-wallpapering*' (N=1), and '*memrise*' (N=1). Another noteworthy occurrence pertains the fact that Indonesian acknowledges the loss of initials /p, b/ when attached with *mem-* since their initial sounds are in the same natural class [LABIAL]. Although the database reveal that English bases with initial /p/ and /b/ retain their original form, there are 3 exceptional cases where initial /p/ is lost in '*memosting*' (N=11),

'*memostingnya*' (N=2), and '*memrogram*' (N=2). However, these forms are less frequent compared to '*memposting*' (N=212), '*mempostingnya'* (N=7), and '*memprogram*' (N=23) respectively.

Table 3. Anomalies of allomorph *mem-*

word	frequency	base	initial
memcached	4	cache	k
memdisk	1	disk	d
memtest	3	test	t
mem-		wallpaper	
wallpapering	1	-ing	w
memrise	1	rise	r

On the one hand, regardless of it being in the third position according to word tokens and types, *men-* is detected to yield the most cases of variability with 42 words (see Table 4). According to the Indonesian rule, *men-* occurs when the initial is within the class of alveolar sounds, among them are **/d**, **t**, **ʧ**, **ʃ**, and **z/** (Sneddon, 2010), such as '*men-judge*' (n=6), '*mendeposit*' (N=5), '*men-tattoo*' (N=1), and '*mendownload*' (N= 282). The variability occurs when *mem-* is attached to initials other than that of alveolar sounds, such as **/e/, /k/, /f/, /s/, /ʌ/, /p/,** and **/ɪ/**, as in '*menencourage*' (N=1), '*men-capture*' (N=1), '*mensubmit*' (N=11), and '*menupload*' (N=1). Althoguh there is a particular case where *men-* can be exceptionally attached if the initial is of consonant clusters, as in the words *'mencropping*' (N=1), '*men-scan*' (N=5), '*menframing*' (N=1) and '*menstarter'* (N=2). Nevertheless, since consonant cluster is regarded to be a foreign characteristic and not Indonesian's (Sneddon, 2010), these words are still considered variabilities until a future investigation proves otherwise.

Table 4. Samples of anomalies of allomorph *men-*

word	frequency	base	initial
men encourage	1	encourage	e
men-supply	2	supply	s
mencounter	1	encounter	e
mencover	1	cover	k
menencourage	1	encourage	e
menframing	1	framing	f
mensandwich	1	sandwich	s
menscroll	1	scroll	s
menstarter	2	starter	s
menstimuli	1	stimuli	s
menbackup	1	backup	b
mencapture	1	capture	k
mencopy	2	copy	k
mencropping	1	cropping	k
menfilter	2	filter	f
menscan	2	scan	s

Allomorph *me-* also appears to have variabilities exhibited through 7 cases (see Table 5). Allomorph *me-* is usually followed by environments other than the aforementioned, such as bases with initials **/l/,/r/,/m/, /n/, /w/,** and **/y/,** as in the words '*memanage*' (N=17), '*memonitor*' (N=141), '*me-retweet*' (N=12), and '*melaunching*' (N=12). In contrast, the variabilities are composed of initials outside this environment such as *'mecopy*' (N=1), '*me-blog*' (N=1), '*me-file*' (N=1), '*me-start*' (N=1), '*me-swipe*' (N=1), and *'me-ignore*' (N=1), although compared to other allomorphs, these variabilities' frequency is very low, with only one case per word combination.

Table 5. Samples of anomalies of allomorph *me-*

word	frequency	base	initial
mecopy	1	copy	k
me-blog	1	blog	b
me-file	1	file	f
me-start	1	start	s
me-swipe	1	swipe	s
me-posting	1	posting	p
me-ignore	1	ignore	ɪ

This leaves allomorphs *meng-* and *meny-* as the only allomorphs without variabilities, meaning that the bases attached to them are in accordance with the Indonesian nasal change rules. Allomorph *meng-* occurs when the bases are of one-syllable (usually foreign) words such as '*mengecheck'* (N=7), *'mengetweet*' (N=1), '*mengeshare'* (N=4), and '*mengetest'* (N=4). On the one hand, allomorph *meny-* is followed by initial **/s/** sound where the **/s/** is lost, as in the words '*menyensori*' (N=1), *menyensor* (N=11), and '*menyervice*' (N=1). These instances is the closest to the their Indonesian counterparts which are are all listed as successfully-assimilated loanwords in the Indonesian dictionary (KBBI), '*sensor*' and *'servis*' respectively.

While *di-* does not have allomorphs, both *di-* and *meN-* are observed to share several similar cases, such as the attachment of gerund/progressive verbs and additional

suffixes. Words such as '*membranding*' (N=3), '*mensharing*' (N=1), '*meng-hosting*' (N=4), '*memposting*' (N=212), '*difinishing*' (N=), '*dilaunching*' (N=19), '*didropping*'(N=1), and '*didubbing*' (N=1). Some gerund/progressive bases also occur in both prefixes, as in 'posting', 'launching', 'monitoring', and 'branding', and more. A snippet of these similarities is featured in Table 6, although the words displayed are of the highest frequency among several typographic forms of each word.

Table 6. A snippet of similar gerund/progressive verbs attached to *meN-* and *di-*

bases	word	prefix	frequency
posting	memposting	mem-	212
	diposting	di-	174
launching	melaunching	me-	12
	dilaunching	di-	19
monitoring	memonitoring	me-	6
	dimonitoring	di-	2
branding	mem-branding	mem-	2
	dibranding	di-	1

The database also observes that there are bases which are only attached to either *meN-* or *di-*, such as *'mem-explosive-kan*' (N=1) and *'di-finishing*' (N=3) (see Table 7). The base *'finishing*' itself has 3 typographic forms: '*difinishing*' (N=1), '*di-finishing'* (N=3), and '*di finishing*'. However, it does not mean that '*explosive*' or '*finishing*' cannot, theoretically, occur after prefix *meN-* or *di-*, since *meN-* can also be attached to denominative verbs as well, although empirically there is no combination of '*diexplosive*' or '*menfinishing*' in the database itself.

Table 7. A snippet of different bases attached to *meN-* and *di-*

prefix	word	base	frequency
mem-	mem-explosive-kan	explosive	1
mem-	memphoto	photo	1
mem-	mem-wallpapering	wallpapering	1
mem-	memveto	veto	2
meng-	mengglobal	global	10
di-	dikick	kick	1
	di dubbing	dubbing	2
	diremap	remap	1
	diblow-up	blow up	2

Suffixes are also reported to occur in the combination of *meN-* and *di-* + English bases. Interestingly, suffix *-kan* occurs 21 times in both *meN-* and *di-* combinations, as in the words '*mem-balance-kan*' (N=1), '*mentradingkan*' (N=1), *'diemailkan*' (N=1), and *'diprintkan*' (N=1). Suffix *-nya,* on the other hand, appears 35 times in *meN-* + English base combinations as in the words '*meng-unpin-nya'* (N=1) and *'memasternya*' (N=2) as well as 6 times in *di-* + English base combinations as in the words '*di share-nya*' (N=1) and *'direviewnya*' (N=1). Last but not least, suffix *-i* occurs 3 times in both combinations as in the words '*melabelinya*' (N=24), '*melabeli*' (N=3), and *'dipostingi*' (N=1).

Several bases also appear to feature in both *meN-* and *di-*. A snippet of what bases attached to both prefixes is featured in the following Table 8. It has to be noted that this snippet features words with the most frequency among others with the same base but different typographic form. For instance, 'download' has 4 forms when attached to *men-* and 3 forms when attached to *di-*, each with their own frequency. Issues pertaining typographic forms is further explained in the following subchapter.

Table 8. A snippet of similar bases attached to *meN-* and *di-*

base	prefix		frequency	
	meN-	*di-*	*meN-*	*di-*
download	mendownload	didownload	282	93
upload	mengupload	diupload	116	68
upgrade	mengupgrade	diupgrade	29	18
import	mengimport	diimport	18	7
support	mensupport	disupport	21	23
update	mengupdate	diupdate	136	98
follow	mem-follow	difollow	11	4

Typographic forms of *meN-* and *di-* + English bases

A technical issue in gathering data for building the database this paper uses concerns the fact that a single English base, when attached to either *meN-* or *di-*, has more than 1 way of writing. It is estimated that there are 13 forms of writing either combination which covers a single base, additional suffix, and if the base is composed of a phrasal verb. These forms are also observed to be shared between both prefix. Among these forms, ipW is considered to be most used which makes up 192 bases attached to *meN-* and 201 words attached to *di-*. The following list consists of the 13 forms, while Table 9 provides closure into a snippet of how these forms affect the frequency of the same bases:

- ipW
- ip-W
- ip W
- ip-WS
- ip W-P(Adv)
- ip-W-S
- ipWS
- ip-W P(Adv)
- ip-W-P(Adv)
- ipW P(Adv)
- ip-WP(Adv)
- ipW-P(Adv)
- ipWP(Adv)

Table 9. A snippet of the typographic forms of *meN-* and *di-* + English bases

prefix	base	word	frequency	pattern
meN-	download	mendownload	282	ipW
		men-download	63	ip-W
		men download	4	ip W
		men-downloadnya	2	ip-WS
	bully	mem-bully	1	ip-W
		membully	11	ipW
		mem-bully-nya	1	ip-W-S
	upload	mengupload	116	ipW
		meng-upload	44	ip-W
		meng-uploadnya	2	ip-WS
		meng upload	1	ip W
		menguploadnya	8	ipWS
	back up	mem-back up	3	ip-W P(Adv)
		mem-back-up	4	ip-W-P(Adv)
		memback up	2	ipW P(Adv)
		mem-backup	11	ip-WP(Adv)
		memback-up	2	ipW-P(Adv)
		membackup	11	ipWP(Adv)
di-	download	didownload	93	ipW
		di-download	28	ip-W
		di download	55	ip W
	bully	dibully	26	ipW
		dibullynya	1	ipWS
		di bully	5	ip W
		di-bully	11	ip-W
	upload	di upload	27	ip W
		di-upload	22	ip-W
		diupload	68	ipW
	back up	di-backup	6	ip-WP(Adv)
		dibackup	9	ipWP(Adv)
		di backup	7	ip WP(Adv)

ip=inflectional prefixes (*meN-* and *di-*); W=words; S=suffix; P=preposition; Adv=Adverb

It also appears that there are 7 other forms in a lesser frequency and occur in a fewer types compared to the aforementioned forms. Some of the forms are not shared with either prefix, but there is only one that occurs in both *meN-* and *di-*(ip – W). The following is

the list of the forms along with the examples and their frequency:

- ip'W S: *meng'homeschool kan* (N=1)
- ip"W": *meng"outsource"* (N=1), *mem"boom"* (N=1)
- ip-*W: *meng-*global** (N=1), *meng-*update** (N=1)
- ip- W: *meng- file* (N=2), *meng- folder* (N=1)
- ip – W: *me – recovery* (N=1), *di – recall* (N=2), *di – quote* (N=1)
- ip-'W': *mem-'bully'* (N=1)
- ip 'W': *di 'suspend'* (N=1), *di 'booking'* (N=1)

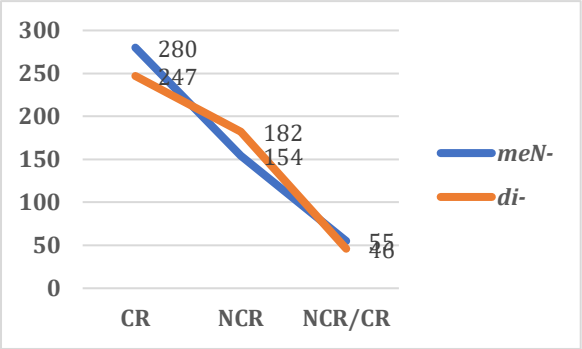
The observation of typographic forms concerns with whether one can ignore these forms when grouping each combination based on their shared English bases, since AntConc does not treat them as the same. In many cases, when *meN-* and *di-* are separated from the bases, allomorphs of *meN-* and *di-* are considered a single word in the Word List feature. On the one hand, when entered in the Clusters feature, they appear to be clustered with other words, regardless of the punctuation marks and white spaces which more often than not occupied 85% of the observed forms. This constraint also leads to how one can decide whether combinations of the same specific base but written in three or four different forms are grouped semantically, since each form yields different results when applied to various contexts. This issue is further addressed in the following subchapter.

Semantic clusters of meN- and di- + English bases

Although it is previously mentioned that the English bases have to belong to the category of verbs and/or other word classes with the possibility of being a verb by looking at the context, it is unfortunate that the classification of bases is not approached properly except for the fact that such limitation is significant in building the database since the prefixes attached are of inflectional voice ones.

However, the database provides other information regarding each word's affinity towards a certain semantic cluster: computer-related (CR), non-computer-related (NCR), or both (NCR/CR). By looking at the context, or the concordance result of each combination, this paper is able to group words of the same semantic clusters (see Figure 2).

Figure 2. Raw calculations of semantic clusters based on word types of *meN-* and *di-*



On the one side, as the results of the various typographic forms of each base when combined with either prefix and the attached additional suffixes, some combinations overlap in terms of their semantic clusters (see Table 10). In context, bases such as 'back up', 'launching', and 'update' exist in either cluster, while 'download' and 'install' are grouped in the computer-related one, considering their various typographic forms. Examples for the aforementioned three bases within each cluster are featured in sentences (1) to (7).

(1) *Kami sarankan untuk juga* ***mem-back-up*** *di CD*
'We suggest to also ***back it up*** in a CD.' (ind-com_web_2018_1M:417119)

(2) *Kolaborasi Petronas-Yamaha akan* ***memback up*** *tim Marc VDS yang tengah bermasalah di internal*

(3) 'The Petronas-Yamaha collaboration will **back up** the Marc VDS team which is now in the middle of internal dispute'
<ind-com_web_2018_1M:481455>

(4) *Para pendukung Jokowi itu bahkan sempat* ***melaunching*** *gerakan Antipolitisasi Masjid di Jalan MH Thamrin, saat CFD sedang berlangsung*
'Even the Jokowi supporters had the time to **launch** the Mosque Antipoliticization movement in Jalan MH Thamrin in the middle of CFD (car free day)'
(ind-com_web_2018_1M:654564)

(5) *Jatonic akan* ***melaunching*** *website baru bulan depan*

'Jatonic will **launch** its new website next month'
(ind-com_web_2018_1M:369033)

(6) *Kami juga akan selalu **mengupdate** apabila ada perubahan atau penambahan data harga*
'We will also frequently **update** whenever there are changes or additional data in price'
(ind-com_web_2018_1M:413542)

(7) *Jadi, pilihan terbaik adalah **mengupdate** konten lama dan membuat beberapa link menuju halaman-halaman terbaru*
'That's why the best choice is to **update** the older contents and make some links towards the newer pages'
(ind-com_web_2018_1M:356429)

Table 10. Samples of semantic clusters anomaly within *meN-* + English bases combination

word	frequency	base	affix	semantic cluster
mem-back up	3	back up	*mem-*	CR
mem-back-up	4		*mem-*	CR
memback up	2		*mem-*	NCR
memback-up	2		*mem-*	NCR/CR
me-launching	1	launching	*me-*	NCR
melaunching	12		*me-*	NCR/CR
mendownload	282	download	*men-*	CR
men-download	63		*men-*	CR
men download	4		*men-*	CR
men-downloadnya	2		*men-.-nya*	CR
meng update	4	update	*meng-*	CR
mengupdatenya	2		*meng-.-nya*	CR
mengupdate	136		*meng-*	NCR/CR
menginstall	135	install	*meng-*	CR
menginstallnya	16		*meng-.-nya*	CR
meng-install	19		*meng-*	CR

The same issue goes to the combination of *di-* + English bases (see Table 11). Bases such as 'share', 'blacklist', 'block', 'upgrade', and 'update' occur in either CR or NCR/CR cluster, while 'install' and 'like' in only CR cluster as well as 'blender' and 'translate' in only NCR cluster, with their various typographic forms in consideration. Examples of combinations which occur in either cluster are featured in sentence (8) to (13).

(8) *"Kita mau semua ide yang dibicarakan dan **dishare** dijalankan di daerah masing-masing."*
'" We want all ideas that were discussed and ***shared*** are run in each area respectively."'
(ind-com_web_2018_1M:477395)

(9) *Jadi bahan juga untuk **di share** via blog.*
'(it) also becomes a material to be **shared** via blog.'
(ind-com_web_2018_1M:350495)

(10) *"Dengan jumlah pengguna .... dalam menggunakan KRL,* jadi harus ***diupgrade,***" lanjutnya.
'With the number of users … in using KRL, so (it) should be ***upgraded.***" He continued.'
(ind-com_web_2018_1M: 213517)

(11) *Jadi, RAM, SSD, dan kartu grafis adalah hal yang paling penting untuk **di-upgrade.***
'So, RAM, SSD, and graphic card are the most important things to be ***upgraded**.'* (ind-com_web_2018_1M: 356549)

(12) *Kak, **diupdate** lagi dong pembahasannya dgn soal terbaru, thanks….*
'Kak, please ***update*** the discussion with the newest question again, thanks.' (ind-com_web_2018_1M: 399752)

(13) *Pastikan ClearOS Anda sudah **di update**, jika belum update terlebih dahulu.*
'Make sure your ClearOS has already been ***updated***, if not please update it first.' (ind-com_web_2018_1M:659456)

Table 11. Samples of semantic clusters anomaly within *di-* + English bases combination

word	frequency	base	affix	semantic cluster
dishare	29	share	*di*	NCR/CR
di-share	17			NCR/CR
di share	41			CR
di share-kan	1			CR
di share-nya	1			CR
diblacklist	1	blacklist		NCR
di blacklist	3			CR
diblock	6	block		NCR/CR
di block	4			CR
di upgrade	13	upgrade		CR
di-upgrade	13			CR
diupgrade	18			NCR/CR
di update	33	update		CR
di-update	17			CR
diupdate	98			NCR/CR

The fact that the database still provides raw calculations of each form's frequencies and cluster tendencies means that typographic forms do influence a word's affinity towards a specific semantic cluster. With different typographic forms, observing a certain English base in different contexts can be a little difficult to do since a certain form leads to certain contexts. The words belonging to NCR/CR cluster of both *meN-* and *di-* combinations needs to be sorted out to specify each base's affinity. For instance, three forms of 'update' as in '*diupdate*', '*di update*', and '*di-update*' occur in both CR and NCR/CR clusters. To add, four forms of 'back up' as in '*mem-back up*', '*mem-back-up*', '*memback up*', and *'memback-up*' even occur in all clusters. While it can be inferred that 'update' and 'back up' can be used in either computer- or non-computer-related contexts, quantitative investigation is still needed to clarify this matter.

The combination of English base and Indonesian inflectional prefixes *meN-* and *di-* has been acknowledged by several scholars due to its prevalence in either spoken or written discourse, but little to no advances have been made to analyze its morphological transformation further. Saddhono and Sulaksono (2018) notice that this phenomenon does exist, along with other possible combinations including Indonesian suffixes and Colloquial Jakartan Indonesian (CJI) variants, but contend that the structure does not follow either Indonesian or English rules. Sedeng and Indrawati (2019) also acknowledge this peculiarity in their data, but because the focus of their research is on the linguistic level of Indonesian English forms, they do not further address the combination of English base and Indonesian affixes, and instead take their stance by stating that Indonesian people tend to frequently mix English words than phrases or clauses in their conversation. So far, only a study conducted by Oktavia (2019) which explicitly asserts the addition of Indonesian linguistic feature to English words, however, in conformity with other studies mentioned above, there is no further investigation leading to description of the combination itself. With the intention of bridging the gap that the previous studies have left opened, by looking at bigger data and observing the patterns, this study confirms that the combination does lean towards Indonesian *N-* nasal changes as shown in **Figure 1** where allomorphs *meng-, men-, and mem-* appear to be the top three most used allomorphs in the database, which can attest Saddhono and Sulaksono's (2018) claim about the rules itself.

It is also attested that there are 13 most used types of typographic forms shared between the combinations of *meN-* and *di-* + English bases, and 7 other forms attested in either prefix, with the exclusion of one form where it is attested in both prefix but in low frequency and only comprises of 3 words. While it has been mentioned that typographic forms concern how one can semantically determine a certain word, writing foreign, unassimilated words differently has been a shared practice by the Indonesians, especially in printed publications to indicate that the words are not or has not been assimilated to Indonesian (Sneddon, 2003). Sneddon (2003)

mentions that one way to distinguish foreign loanwords in written publications is to italicize them as the original spelling retains. However, the database shows that words in italics is not the only method to do so, but by the aforementioned 20 forms which also includes punctuation and white space plays to draw attention to the foreign words. Although it can also be implied that AntConc simply cannot include italics since the corpus needs to be in a plain text file, resulting in the database inability to record words or characters written in italics.

Furthermore, while it has not been empirically attested in this paper that there is a tendency to use the English equivalences of computer-related bases and not the Indonesian's, the combinations of *meN-* and *di-* + English bases yields another finding pertaining their semantic clusters, in which computer-related cluster dominates both combinations in the database than non-computer related one (see Figure 2). However, since this paper provides only raw calculation as typographic forms affect the decision whether a word can be included in CR, NCR, or NCR/CR clusters, future investigation is needed to clarify each word's affinity towards computer-related or non-computer-related cluster. Expanding the clusters is also a suggested approach since the NCR/CR group needs to be clarified as well, as words like '*diupdate*', '*diupgrade*', *'membackup*', and *'memonitor*' are observed to occur in either CR, NCR, or NCR/CR cluster.

Nevertheless, the database records a total of 527 computer-related words of both *meN-* and *di-* combinations, which also means 3,084 occurrences in the corpus. Not to mention the remaining 101 words which are still in the NCR/CR cluster and in need of future re-investigation. Regardless, this finding has strengthened the hypothesis that Indonesians has familiarized themselves with the English terms, although claiming that the English words are often used than their Indonesian equivalences is considered a premature declaration without adequate quantitative data to back it up.

Although such phenomenon first was generally recognized through spoken discourse, it has affected written discourse as well, especially within the Internet and social media domains. Studies conducted by Saddhono and Sulaksono (2018) as well as Sedeng and Irawati (2019) obtained the data through spoken social intercourse, but Oktavia (2019), however, collected their examples from Facebook, Instagram, and WhatsApp. Though different, the method utilized by this paper and hers demonstrate that the combination exists even in written discourse, especially within the perimeter of computer-mediated communication (CMC). The frequent use of English terminology in these domains, combined with source language's features, allows this prevailing linguistic creativity to be recognized as one of the potential innovations of Indonesian English variant, in which, to quote Lauder (2008, p. 18), "might fit within an EIL (English as an International Language) framework" if taken further.

Nevertheless, there are some issues that need to be addressed, in which this study fails to observe further. First of all, regardless of how typographic forms are used to distinguish English and Indonesian words, they are also essential in determining whether a base of various forms can be considered a single word type. Once this matter is settled, the NCR/CR cluster can be re-investigated to calculate each word's affinity towards computer- and non-computer-related semantic clusters. It is suggested that future research can address this problem and expand the semantic clusters themselves into more specific and distinct clusters than what this paper has done. Second of all, the fact that this study only focuses on the Standard Indonesian (SI) prefixes allows for detailed observation on the Colloquial Jakartan Indonesian (CJI) variety as well, such as *N-* and *Ø-*, and possibly a comparative study between the two prefix varieties to address the productivity of English bases when attached to either variety (Arka & Yannuar, 2016; Inderasari & Oktavia, 2019; Smith-Hefner, 2007). Because the phenomenon is deemed to flourish through the influence of the Internet and social media (Oktavia, 2019; Qory'ah et al., 2019), it is expected that there will be more data to yield from the use of these colloquial variants as they are often omnipresent as the language of the Internet.

Conclusion

This paper reveals how *meN-* and *di-* are attached to English bases as well as the allomorphs of both prefixes, the addition of Indonesian suffixes to the bases, typographic forms encircling *meN-* and *di-* combinations, as well as the semantic clusters of the observed combinations within the database. The investigation concludes that *meN-* allomorphs exist, along with several anomalies concerning the initial bases when attached to a certain allomorph, and unique cases pertaining gerund/progressive verbs, additional suffixes, and similar bases occur in both prefixes. The decision of whether one allomorph is for a specific word depends on the base itself—and the analysis needs to include both the English sound system and the Indonesian *N-* sound change to consider.

While various typographic forms indicate that the words are not of Indonesian origin but the English', it has also affected how one determines a certain word to be grouped in either a computer- or non-computer-related cluster. However, the hypothesis that Indonesians have incorporated English bases in their linguistic repertoire remains as it is revealed in the database that Indonesians assimilate their linguistic features with other languages, especially English, especially when it comes to global terms that most of the time, Indonesian does not have the correspondence lexicalization to match with. Though the phenomenon itself is not novel, little to no researchers have put their interest in explaining the combination of English bases and Indonesian prefixes and the changes of nasal sound *N-* as the results of the combination. It is expected that this research can shed light on new linguistic phenomenon in Indonesia and encourage other linguists to be attracted to the development of the Indonesian language as it inevitably keeps changing over time.

References

Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. *IEEE International Professional Communication Conference*, 729–737. https://doi.org/10.1109/IPCC.2005.1494244

Arka, I. W., & Yannuar, N. (2016). On the morphosyntax and pragmatics of -in in Colloquial Jakartan Indonesian. *Indonesia and the Malay World*, *44*(130), 342–364. https://doi.org/10.1080/13639811.2016.1215129

Choi, H. Y. J. (2019). *A corpus based analysis of -kan and -i in Indonesian* [Nanyang Technological University]. https://doi.org/10.32657/10356/136955

Denistia, K. (2019). Revisiting the Indonesian prefixes peN-, pe2- and per-. *Linguistik Indonesia*, *36*(2), 144–160. https://doi.org/10.26499/li.v36i2.80

Denistia, K., & Baayen, R. H. (2019). The Indonesian prefixes PE- and PEN-: A study in productivity and allomorphy. *Morphology*, *29*(3), 385–407. https://doi.org/10.1007/s11525-019-09340-7

Goldhahn, Dirk; Eckart, Thomas; Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. *Natural Language Processing Group*, 759–765.

Inderasari, E., & Oktavia, W. (2019). Indoglish Phenomenon: The Power Of Media And Business Languages In The Digitalization Era Fenomena Indoglish: Kekuatan Media Dan Bahasa Bisnis Di Era Digitalisasi. *Jurnal Kata : Penelitian Tentang Ilmu Bahasa Dan Sastra*, *3*(2), 194–206. https://doi.org/10.22216/jk.v3i2.4503

Kachru, Y., & Nelson, C. L. (2006). *World Englishes in Asian Contexts*. Hong Kong University Press. https://www.jstor.org/stable/j.ctt2jbztz

Kadarisman, A. E. (2005). Relativitas Bahasa dan Relativitas Budaya. *Linguistik Indonesia*, *23*(2), 152–170.

Lauder, A. (2008). the Status and Function of English in Indonesia: a Review of Key Factors. *Makara Human Behavior Studies in Asia*, *12*(1), 9.

https://doi.org/10.7454/mssh.v12i1.128

Lowenberg, P. (1991). English as an additional language in Indonesia. *World Englishes*, *10*(2), 127–138. https://doi.org/10.1111/j.1467-971X.1991.tb00146.x

Manns, H. (2010). Indonesian Slang in Internet Chatting. In M. A. S. Babatunde, A. Odebunmi, A. Adetunji (Ed.), *Studies in Slang and Slogans* (pp. 71–99). Lincom.

Oktavia, W. (2019). Eskalasi Bahasa Indoglish dalam Ruang Publik Media Sosial. *Diglosia*, *2*(2), 83–92.

Percillier, M. (2016). *World Englishes and Second Language Acquisition* (Vol. G58). John Benjamins Publishing Company. https://doi.org/10.1075/veaw.g58

Qory'ah, A. N., Savira, A, T, D., & Inderasari, E. (2019). Variasi Bahasa Indoglish dan Idiolek Publik Figur di Instagram. *Transformatika: Jurnal Bahasa, Sastra, Dan Pengajarannya*, *3*(2), 136–149. https://doi.org/10.31002/transformatika.v

Rajeg, G. P. W., Denistia, K., & Musgrave, S. (2019). Vector Space Models and the usage patterns of Indonesian denominal verbs: A case study of verbs with meN-, meN-/-kan, and meN-/-i affixes. *NUSA: Linguistic Studies of Languages in and around Indonesia*, *67*, 35–76. https://doi.org/doi/10.15026/94452

Rajeg, G. P. W., & Rajeg, I. M. (2017). Mempertemukan morfologi dan linguistik korpus: Kajian konstruksi pembentukan kata kerja [per-+Ajektiva] dalam Bahasa Indonesia. In I. N. Sudipa & M. S. Satyawati (Eds.), *Rona Bahasa: Buku persembahan kepada Prof. Dr. Aron Meko Mbete memasuki masa purnatugas* (pp. 288–327). Swasta Nulus. https://doi.org/https://doi.org/10.4225/03/5a0627de02453

Rajeg, G. P. W., Rajeg, I. M., & Arka, I. W. (2020). Contrasting the semantics of Indonesian -kan & -i verb pairs: A usage-based, constructional approach. *Seminar Nasional Bahasa Ibu*. https://doi.org/https://doi.org/10.6084/m9.figshare.11758218

Saddhono, K., & Sulaksono, D. (2018). Indoglish as adaptation of english to Indonesian: Change of society in big cities of Indonesia. *IOP Conference Series: Earth and Environmental Science*, *126*(1), 0–8. https://doi.org/10.1088/1755-1315/126/1/012092

Sedeng, I. N., & Indrawati, N. L. K. M. (2019). The Use of Indoglish in Faculty of Arts, Udayana University-Bali. *Journal of A Sustainable Global South*, *3*(1), 18. https://doi.org/10.24843/jsgs.2019.v03.i01.p04

Smith-Hefner, N. J. (2007). Youth Language, Gaul Sociability, and the New Indonesian Middle Class. *Journal of Linguistic Anthropology*, *17*(2), 184–203. https://doi.org/10.1525/jlin.2007.17.2.184

Sneddon, J. (2003). *The Indonesian Language: Its History and Role in Modern Society*. University of New South Wales Press.

Sneddon, J. (2010). *Indonesian Reference Grammar* (Second Edi). Allen and Unwin.

Sukarno. (2017). The Behaviours of the General Nasal /N/ in Indonesian Active Prefixed Verbs. *International Journal of Language and Linguistics*, *4*(2), 48–52.