INDONESIAN JOURNAL OF APPLIED LINGUISTICS
Vol. 10 No. 2, September 2020, pp. 382-396

Available online at:
https://ejournal.upi.edu/index.php/IJAL/article/view/28610

https://doi.org/10.17509/ijal.v10i2.28610

Towards developing colloquial Indonesian language pedagogy: A corpus analysis

Halim Nataprawira and Michael D. Carey*
School of Education, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia

ABSTRACT

This study was motivated by the situation that many students studying Indonesian language have problems to understand and communicate in spoken Indonesian. This is because Indonesian is a diglossic language in which different sets of grammar and vocabulary are used between the high and low diglossic variants, whereas students are usually only taught the high diglossic variant. Only the high diglossic variant of formal Indonesian has an official status, while the low diglossic variant of colloquial Indonesian does not. Sneddon observed that in everyday speech the linguistic features of high and low diglossic variants are merging into a middle variant that Errington called Middle Indonesian. This study examines the extent to which a middle variant of spoken Indonesian has formed by quantifying the amount of high and low linguistic elements that are present in a corpus of everyday spoken Indonesian derived from audio-recordings and written texts containing spoken language. We collected and classified a 14,000+ word corpus of spoken Indonesian. With reference to published descriptions of high (formal) and low (colloquial) diglossia, each colloquial item in the corpus was counted and calculated as a ratio to the total N of the corpus. Colloquial features were found with an average proportion of 0.39 across the corpus, indicating that colloquial Indonesian lexicon and grammar may contribute as much as 39% to everyday spoken Indonesian. This result evidences the need to include this middle variant of spoken Indonesian in the design and resourcing of materials within the Indonesian language curriculum.

Keywords: Colloquial Indonesian; corpus analysis; diglossia; teaching spoken Indonesian,

First Received:	Revised:	Accepted:
18 August 2019	29 July 2020	1 August 2020
Final Proof Received:		**Published:**
5 September 2020		30 September 2020

How to cite (in APA style):
Nataprawira, H. & Carey, M. D. (2020). Towards developing colloquial Indonesian language pedagogy: A corpus analysis. *Indonesian Journal of Applied Linguistics, 10*(2), 382-396. https://doi.org/10.17509/ijal.v10i2.28610

INTRODUCTION

Internationally, many students studying Indonesian as a foreign language have problems to understand and communicate in spoken Indonesian. This may be due to the lack of appropriate learning resources to teach informal spoken Indonesian to foreign learners. Coinciding with this lack of resources, a formal high diglossic variant of standard Indonesian is often misrepresented as the informal everyday spoken language of Indonesia for language teaching purposes. This is because Indonesian is a diglossic language (Errington, 1986; Sneddon, 2003a) in which different sets of grammar and vocabulary are used between the high and low diglossic variants, whereas students are usually only taught the high diglossic variant. Only the high diglossic variant of formal Indonesian (FI) has an official status, while the low diglossic variant of colloquial Indonesian (CI) does not (Smith-Hefner, 2007; Sneddon, 2003b). An understanding of the features of Indonesian diglossia is critical to redress the misrepresentation of the spoken language by Indonesian language teachers and resource developers.

Diglossia is a situation in which a single language community uses two dialects or languages. In addition to the community's vernacular, or everyday language variety (labeled

* Corresponding Author
Email: mcarey@usc.edu.au

"L" or "low" variety), a second, highly codified variety (labeled "H" or "high") is used in certain situations such as literature, formal education, or other specific settings, but not used for ordinary conversation (Errington, 2014; Ferguson, 1959). The reality of the Indonesian linguistic landscape is much more complex than the diglossic paradigm that is addressed in this article when regional languages and dialects are brought into consideration (Tamtomo, 2019). This article primarily addresses the Jakartan-origin middle variant that we hypothesise has become the common contemporary spoken language of Indonesian popular culture.

Research on Indonesian diglossia was pioneered by Errington (1986) and subsequent extensive research was continued by Sneddon (2001). Linguistic descriptions have been undertaken by Nothofer (1995), Sneddon (2001, 2003, 2006), Djenar (2006, 2008), Djenar & Ewing (2015), Tjung et al. (2006), Smith-Hefner (2007) and Kushartanti (2014). Many of these studies concentrated on the social and grammatical functions of selected lexical items. Sneddon (2003b) raised the possibility of a future merging of FI and CI into a middle variant. The gap in the research is that this merger is yet to be empirically investigated with a contemporary sample of spoken Indonesian. It is the objective of this current study, using both qualitative description and quantitative measures, to investigate Sneddon's FI-CI merging postulation. In this paper it is referred to as 'the M (middle) hypothesis' - that a middle variant has become the common spoken Indonesian (SI) language. To affirm the M hypothesis, CI must be an integral feature - alongside FI - in a corpus of informal spoken language.

Indonesian diglossia has arisen from the different Malay dialects that were spoken throughout the Malay Archipelago (Errington, 2014; Ewing, 2016; Gil, 1994; Manns, 2014). Formal Indonesian (FI) is derived from Royal Riau Malay court language which became the basis of Classical Malay literature and was well established as the language of literature by the time of European arrival in the 16th Century (Sneddon, 2003b). There were also several varieties of Market Malays, used by commoners in everyday transactions. Some of these varieties are the antecedents of colloquial Indonesian (CI). The CI variety that is treated in this study is the CI of Jakarta which is strongly influenced by Jakarta's Malay dialect Betawi Malay (Grijns, 1991; Sneddon, 2003a). Betawi Malay itself is a form of Malay that is influenced by Sundanese, Javanese, Balinese, Hokkien Chinese and Dutch, and these language features have in turn been inherited by Jakartan CI.

The emergence of Jakarta as the capital of independent Indonesia led to the formation of a language hybrid that we call spoken Indonesian (SI) in this article, an everyday spoken language that consists of FI and CI. This SI was largely driven by the 'new Jakartans', the post-independent generation of the capital who began fusing CI Betawi linguistic features with FI (Sneddon, 2003b). The Jakartan population, the youth especially, created many new words and phrases, even though the linguistic patterns, grammar, phonology and morphology did not evolve beyond those of Betawi Malay. It has been noted that children in Jakarta and the surroundings grow up speaking a register of Indonesian that leans strongly towards CI (Kushartanti, 2014).

While CI originated in the Jakartan speech community and its surroundings, in time, due to the prominence of Jakarta as the capital city and as an exporter of culture through its command of the media and literature, it spread to other parts of Indonesia (Sneddon, 2006). For example, outside the capital Jakartan CI can be commonly heard in radio broadcasts in regional cities such as Bandung, Denpasar and Padang as young speakers in regional cities use it during inter-ethnic interactions, as an in-group code and to project youth identity (Manns, 2014).

The taxonomies and coding of Indonesian diglossia

The FI–SI–CI taxonomy in this article corresponds to Sneddon's High, (hypothesized) Middle, and Low varieties. The FI–SI–CI coding we propose is a categorization system that establishes well-defined boundaries of each variant and allows for qualitative and quantitative linguistic analysis. FI, also referred to as standard Indonesian and known in Indonesian as bahasa Baku, is the language of formal spoken and written communication, such as government protocols and news presentations. The everyday spoken language is known by Indonesians as bahasa Sehari-hari. Indonesians certainly recognise the differences between formal and informal forms and switch between the two as the situation demands. However, often in practice there is not always a clear distinction between the use of formal and informal language (Djenar & Ewing, 2015; Sneddon, 2001). Speakers may make their informal speech somewhat more formal by incorporating some features of formal language and thus characteristics of FI are not excluded from informal conversation (Sneddon, 2001). Likewise, the formal language does not always conform to a standard form when used in social discourse. A politician may use less formal language in an unprepared speech to demonstrate his populist intentions when trying to connect to the masses. This linguistic grey zone described above by Sneddon and Djenar is considered in this article as the formal-informal spectrum of SI.

The grammar and identity of FI is well

established and universally accepted. One problem in discussing Indonesian diglossia is the lack of universally agreed terms for the different diglossic language variants and sub-variants. The next section consolidates existing sociolinguistic terminologies into a workable coding system that allows for a systematic analysis of Indonesian diglossia.

Confusion in terminology

Firstly, it is important to clarify terminology used in relation to CI because consensus is lacking across the literature. Sneddon (2001) and Djenar & Ewing (2015) have used the term 'informal Indonesian', and Smith-Hefner (2007) used the term 'spoken informal Indonesian', while Manns (2014) used the term 'Jakartan Indonesian'. Djenar (2006, p. 22) noted that there are many other terms used at different times by different writers in regard to the colloquial variety of Indonesian including bahasa tak baku "non-standard language", bahasa informal "informal language", bahasa gaul "social language", bahasa ABG "teen language", bahasa remaja "youth language", 'informal Jakartan Indonesian' and 'colloquial Jakartan Indonesian' (Kushartanti, 2014; Sneddon, 2006). Our view is that the terms mentioned above are often interchangeable and, in some cases, sub-variants of CI. The most common recent confusion amongst student researchers of Indonesian language is that bahasa gaul (social language) has been mistaken as CI. In this article, we classify bahasa gaul as a sub-variant of CI because bahasa gaul does not have different linguistic features to CI, aside from some extra lexical items created by younger speakers. Smith-Hefner (2007) stated that bahasa gaul functions within the linguistic parameters of CI with additional fad words. Like all living languages, it is constantly changing as new words or expressions become popular and fall out of use. At this point, it is worth clarifying the distinction between CI and SI. CI linguistic features pre-existed in Betawi Malay. SI on the other hand is a modern hybrid that we propose to be a derivative of both CI and FI. SI possesses no linguistic features of its own but is dependent on those of CI and FI. The presence of CI linguistic features in SI defines SI's function as an informal language variant.

This study analyses a corpus of everyday spoken Indonesian language derived from transcribed audio-recordings, such as interviews and films, and written texts containing spoken language, such as novels and short stories. Linguistic features were classified at the lexical and sub-lexical level as CI, FI, or neutral lexemes, and transcribed using the International Phonetic Association's (IPA) set of phonetic symbols. These linguistic features included lexis, phonology, morphology and semantics. The following questions guide this research:

1. In what ways are the linguistic features of CI unique and how can they be identified and described?
2. How prevalent are the linguistic features of CI in a corpus of everyday spoken Indonesian?

METHOD

A corpus-based analytic approach was the chosen research method because corpus-based research assumes the validity of linguistic forms and structures derived from linguistic theory (Biber, 2015). The primary goal of this research approach is to analyse the systematic patterns of variation and use for pre-defined linguistic features. The approach allowed us to ascertain how, and to what extent, pre-defined linguistic features form part of everyday spoken Indonesian. Previous descriptions of CI (Djenar, 2008; Djenar & Ewing, 2015; Kushartanti, 2014; Sneddon, 2006;) were used to classify the features of CI. These non-FI linguistic features were used to inform the qualitative description of CI using the IPA. Each CI item in the corpus was counted and quantitatively measured as a ratio in each data sample and to the total N of the corpus. Lexicon that are 'neutral', namely uninflected base words, are not counted as CI and make up the proportion of the remaining total (neutral + FI). The M hypothesis of Indonesian diglossia is expressed as a null-hypothesis H0: CI/SI = 0 and as an alternative hypothesis H1: CI/SI > 0. The SI in these hypotheses represents the entire N of the corpus of everyday language and the CI/SI ratio is used as a proportional measure to gauge the extent to which CI linguistic features form part of the everyday informal spoken Indonesian.

Data samples

The corpus used in this study is a sample of real-world language data and is therefore assumed to be representative (Chapman & Routledge, 2009; Stubbs, in Davies & Elder, 2008). The corpus was assembled and is available online (Nataprawira, 2017). Samples have been obtained from interview recordings with native Indonesian speakers compiled by Sneddon (2006) as well as samples of spoken texts from media, internet content, billboard advertisements and audio-visual media such as TV shows and films (Table 1).

The data samples were analysed as raw data, meaning that they were not modified from their original form. Audio-visual data samples were obtained from YouTube. The corpora were collected by transcribing parts of dialogues of films, comedies and TV shows. These text samples were selected because they provide a range of discourse registers (field, mode and tenor), including some spontaneous language use (comedies) that represents naturally occurring spoken dialogue.

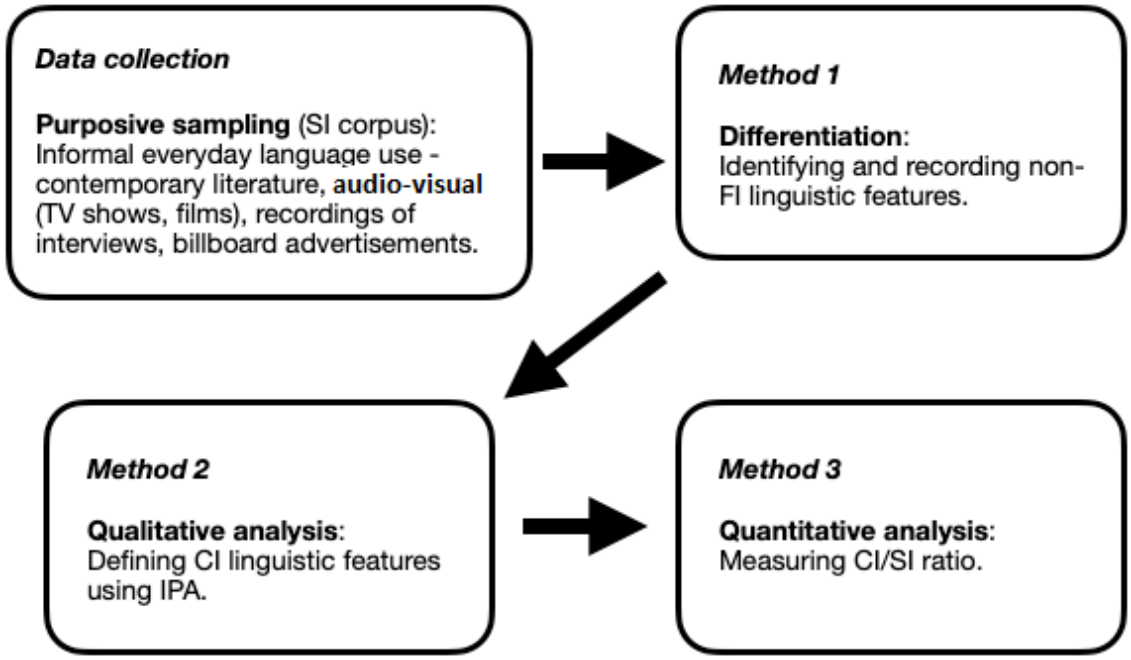
Examples of audio-visual data sources include dialogues from the Opera Van Java comedy show, parts of films such as Buaya Gile and Jakarta Undercover. The billboard data samples were obtained from photographs of billboards. Table 1 shows the number of data samples, the number of lexical items each sample contained and the number of CI lexical items in each sample contained and the number of CI lexical items in each category.

As our research design used descriptive statistics, a measure of statistical power for the number of word tokens collected in the corpus was not required. Instead, we selected word tokens from a range of text types and spoken registers (14711 words across 48 data samples) to obtain a valid representation of SI language (Table 1).

Table 1
SI Corpora Data Samples

Data sample categories	***N***	**SI**	**CI**
Recorded interviews and conversations	6	6408	2130
Contemporary literature	16	4603	1298
Audio-visual media	14	3626	1745
Billboard advertisement	12	74	27
	∑***n*** = 48	∑**SI** = 14711	∑**CI** = 5200

Figure 1
The Mixed-Method Design of This Research



To address the research questions, a mixed-method design consisting of qualitative and quantitative analysis was chosen. The qualitative component defines the CI linguistic features in the SI corpora (research question 1), which in turn are quantitatively measured to obtain an indication of the level of CI frequency and prevalence in SI (research question 2).

Method 1 - Differentiation: Identifying and collecting non-FI linguistic features.
The differentiation method used to investigate if CI was present in the SI corpus involved the identification of linguistic features that were not FI. In this process, lexical items were first classified as FI or non-FI through a broad analysis of the phonological, morphological and semantic features of lexical items in the corpus. The description of FI in this study followed Sneddon (1996, 2000), Quinn (2001) and Djenar (2003).

Method 2 - Qualitative analysis: Defining CI linguistic features using IPA.
Using the findings from Method 1, the CI linguistic features were categorized more discretely using the IPA. We referred to previous use of IPA in classifying the features of CI employed by Grijns (1981) in his study of variations in Betawi Malay.

Data analysis
Three methods of data analysis were used after collecting the raw corpus data (Figure 1).

The morphological analysis follows the common system used to describe affixation in Indonesian such as that employed by Boellstorff (2002). Using various existing descriptions of CI that have been provided by previous researchers, we devised guidelines to identify CI linguistic features. The guidelines included several indicators. Examples of these indicators are provided in the section - Qualitative results: CI in SI corpus:

1. Syntactical ellipsis is a common feature in daily speech (Sneddon, 2006).
2. Morphological variations that are different from FI (Fan, 1990; Kushartanti, 2014).
3. The phonological divergences from FI (Kushartanti, 2014).
4. Elisions and allomorphy (Kushartanti, 2014; Sneddon, 2006).
5. Alternative lexical items not present in FI (Djenar & Ewing, 2015; Sneddon, 2006).
6. Variation in semantic properties that fall outside of FI grammar (Djenar, 2008; Sneddon, 2006).

Method 3 - Quantitative analysis: Measuring the CI/SI ratio.

The aim of this research was to establish quantitatively the number of CI items in the SI corpus. Descriptive statistics were applied to test the null hypothesis that the CI/SI ratio in the corpus is equal to zero; H0: CI/SI= 0 and the alternative hypothesis that the CI/SI ratio in the corpus is greater than zero; H1: CI/SI > 0.

FINDINGS AND DISCUSSION

Qualitative results: CI in SI corpus

The first method of data analysis indicated that there was a substantial amount of non-FI linguistic features in the SI corpus. These linguistic features have sub-components which consist of: non-FI lexicon, non-FI morphological features, non-FI null parameter / ellipsis, non-FI elisions, non-FI phonological realizations and non-FI semantic properties. The presence of CI and FI in the SI corpora supports Sneddon's (2006) assertion of the existence of a middle variant in spoken Indonesian. Concurrently, the notion that a pure form of FI is used as an informal spoken language can be rejected. CI can be positively verified to be an integral part of the everyday language. The second method was then applied which involved a discrete classification of non-FI items using the IPA. The result is a detailed description of CI that demarcates the diglossic boundary between CI and FI. CI consists of CI lexicon, CI morphological features, null parameter / ellipsis, elisions, CI phonological realisations and CI semantic properties. The examples below provide a summary of CI that was identified in contrast to the FI form; for a complete analysis, see Nataprawira (2017):

1. *Word class ellipsis/null elements Ø in the syntax of daily speech.* Three notable common null elements in informal Indonesian syntax are

   a. *The personal pronoun ellipsis in structures such as:*

_	**FI**	
CI		
Syntax	Ø *Mau* Ø *ke mana*?	*Anda mau pergi ke mana?*
Gloss	Ø-pro aux-mau Ø- verb prep-ke wh-mana?	
English translation	Want to where?	"Where are you going?"

_	**FI**	
CI		
Syntax	Ø *Gak mau*.	*Saya tidak mau*.
Gloss	Ø-pro neg-Gak aux-mau	
English translation	I not want	"I don't want to"

   b. *The* adalah *copula ellipsis in nominative structures such as:*

_	**FI**	
CI		
Syntax	*Bapak Ø kepala desanya di sini*.	*Bapak adalah kepala desanya di sini*.
Gloss	pro-Bapak Ø-cop NP-kepala desaDET-nya prep-di NP-sini	
English translation	Mister head village-the in here.	"He is the village head here."

   c. *The common null element parameter of the predicate pergi in phrases such as:*

_	**FI**	
CI		
Syntax	Ø *Lagi* Ø *ke mana*?	*Anda sedang pergi ke mana*?
Gloss	Ø-pro aux-lagi Ø-verb prep-ke wh-mana?	
English translation	-ing to where?	"Where are you going?"

2. *Morphological features.* Some scholars regard these following phonemic forms as allomorphy of the active me- prefix, but they could possibly also be independent morphemes inherited from Sundanese, Javanese and Balinese.

a. 'm' (/m/) – X

_CI	**FI**	
Lexical item	*Make*	məmakai
Syntax gloss	m-(p)-ake	
English gloss/translation	to use; to wear	"to use; to wear"

Note that the base word pakai this example also undergoes a phonological shift to [pake].

b. 'n' (/n/) – X

_CI	**FI**	
Lexical item	*Nangkep*	mənangkap
CI phonology/morphology gloss	n-(t)-angkəp	
English gloss/translation	Catch	"to catch"

Note that a phonological change also takes place in the base word tangkap ⇨ tangkəp.

c. *'ng' (/ŋ/) X & 'nge' (/ŋə/) – X*

The example ngopi also demonstrates the predication of a NOUN X that does not occur in FI:

_CI	**FI**	
Lexical item	*Ngopi*	minum kopi
CI phonology/morphology gloss	ŋ-(k)-opi	
English gloss/translation	drink coffee	"to drink coffee"

_CI	**FI**	
Lexical item	*Ngirim*	məngirim
CI phonology/morphology gloss	ŋ-(k)-irim	
English gloss/translation	Send	"to send"

_CI	**FI**	
Lexical item	*Ngecek*	məməriksa
CI phonology/morphology gloss	ŋə-cek	
English gloss/translation	check	"to check"

d. X'-in' (/-in/)

This morph replaces both FI's predicate suffixes '-kan' and '-i'. It encompasses all the grammatical functions that these FI suffixes impart (accusative, dative-benefactive, accusative-causative):

_CI	**FI/pragmatics**	
Lexical item	*bikinin*	(mem)buatkan [+benefactive]
CI phonology/morphology gloss	bikin-in	
English gloss/translation	make	"to make something for somebody"

_CI	**FI/pragmatics**	
Lexical item	*benerin*	(mem)bənarkan [+causative]
CI phonology/morphology gloss	bən-ə-r-in	
English gloss/translation	fix	"to fix/correct something"

e. '-ny' (/ɲ/)-X

Like the /ŋ/ phoneme, /ɲ/ is also an allomorphic active prefix of me- (or a proper morph) that operates on base words with first letters 'c' and 's'. Some examples include:

_CI	**FI**	
Lexical item	*nyuci*	məncuci
CI phonology/morphology gloss	ɲ -(c)-uci	
English gloss/translation	wash	"to wash"

_CI	**FI**	
Lexical item	*nyebar*	mənyəbar
CI phonology/morphology gloss	ɲ-(s)-əbar	
English gloss/translation	spread	"to spread"

f. 'ng' (/ŋ/) - X'-in' (/–in/) & 'nge' (/ŋə/) – X'-in' (/–in/)

This is the active form of 1.3b. It is the CI variation of FI's me- X –kan and me- X –i. The example ngapain is a predication of WH- lexical item apa and has two semantic values:

_CI	**FI/pragmatics**

Lexical item	*ngapaiin*	sedang apa; untuk apa [+interrogative]
CI phonology/morphology gloss	ŋ-wh-apa-in	
English gloss/translation	what-ing?"; "What for?"	"what are you doing?"; "What for?"

_CI	**FI**	
Lexical item	*ngebeliin*	məmbəlikan [+benefactive]
CI phonology/morphology gloss	ŋə-bəli-in	
English gloss/translation	buy	"to buy something for somebody"

g. '-ny' (/ɲ/)-X'–in' (/–in/)

This is the active form of 1.3c. It is the CI variation of FI's me-X–kan and me-X–I for base words with first letters 'c' and 's'. Some examples are: nyədiain, nyariin,

_CI	**FI**	
Lexical item	*nyediain*	mənyədiakan [+benefactive]
CI phonology/morphology gloss	ɲ-(s)- ədia-in	
English gloss/translation	prepare	"to prepare something for somebody"

h. 'ke-' (/kə-/) X'-an' /-an/

These are the alternative CI [+excessive] adverbial marker to FI's adverb *terlalu*. Examples include:

_CI	**FI**	
Lexical item	*kegedean*	tərlalu bəsar
CI phonology/morphology gloss	kə -gəde-an	
English gloss/translation	too large	"too large"

i. X'-an' /-an/

This affixation is a CI alternative to the FI adverb lebih [+comparative]:

_CI	**FI**	
Lexical item	*bagusan*	ləbih bagus
CI phonology/morphology gloss	bagus-an	
English gloss/translation	nicer; better	"nicer, better"

3. Elisions, allomorphy and phonological variations different to FI:

a. Elision of first letters 's' and 'h' in some common words

_CI	**FI**	
Lexical item	*ama*	Sama
CI phonology/morphology gloss	(s)-ama	
English gloss/translation	With	"with"

	CI	**FI**
Lexical item	*abis*	Habis
CI phonology/morphology gloss	(h)-abis	
English gloss/translation	Finish	"finished"

b. Elision of prefix me- (or /m/-X allomorphy) in active verbs with first letter 'p'

_CI	**FI**	
Lexical item	*make*	məmakai
CI phonology/morphology gloss	(p)-m-ak-e	
English gloss/translation	Use	"to use"

c. Phonetic realisation [e], [ə] or [ɛ] - in place of the second syllable 'a' vowel in the /a/ phoneme

_CI	**FI**	
Lexical item	*item*	hitam
CI phonology/morphology gloss	(h)-it-ə-m	
English gloss/translation	black	"black"

d. Phonetic realisation [e], [ə] or [ɛ] - in place of the second syllable 'a' vowel in place of /ai/ diphthong:

_CI	**FI**	
Lexical item	*Make*	məmakai
CI phonology/morphology gloss	(m)-ak-e	
English gloss/translation	Use	"to use"

e. The [o] phone substitute for 'u' vowel:

_CI	**FI**	
Lexical item	*Sorga*	Surga
CI phonology/morphology gloss	s-o-rga	
English gloss/translation	Heaven	"heaven"

f. The [o] phone substitute for /au/ diphthong:

_CI	**FI**	
Lexical item	*ijo*	hijau
CI phonology/morphology gloss	(h)-ij-o	
English gloss/translation	green	"green"

4. An existing array of alternative lexical features different to FI, which is often preferred in speech rather than the FI variants (see Table 2).

Table 2
Lexical Features Different to FI

CI	FI	Gloss
Enggak/gak	tidak	"no, do not"
Cuma	hanya	"only"
pake VP segala?	kenapa harus VP?	"why VP"
Mendingan	lebih baik	"it is better to…"
Pengen	ingin, mau	"to want"

5. The frequent use of discourse particles that are absent in FI as can be seen in Table 3.

Table 3
FI Absent Discourse Particles

CI	Pragmatics
Kok	[+interrogative]
Deh	[+agreement]
Sih	[+affirmative]
Dong	[+ request +affirmative]
Loh	[+interrogative]
Mah	[+declarative]
Nah	[+affirmative]

6. The common use of tag questions constructions:

_CI	**FI**	
Syntax	*Bagus nggak?*	*Bagus atau tidak?*
CI phonology/morphology gloss	adj-bagus neg/tag-nggak?	
English gloss/translation	Good not?	"Is it good?"

_CI	**FI**	
Syntax	*Lucu kan?*	*Lucu benar?*
CI syntax gloss	adj-lucu tag-kan?	
English gloss/translation	Funny right?	"Funny wasn't it?"

7. Variation in semantic properties of Indonesian lexica which are not traditionally recognised in prescriptive FI grammar. Some examples are provided in Table 4.

Table 4
Indonesian semantic properties of Indonesian lexica

Lexical item	CI	FI
Jalan	[+V] (" to go")	[+N] ("street"); [+V] ("to walk)
Buat	[+prep] (" for")	[+V] ("to make")
Biar	[+CP] ("so that")	[+V] ("to let be")
Mau	[+aux +tense] ("will")	[+aux +modal] ("to want")
Suka	[+aux +tense] ("used to do")	[+aux +modal] ("to like")
Pada	[+pronominal plural marker]	[+prep] ("on, at")

Quantitative results: CI/SI

The third quantitative method of analysis involved counting every lexical item with CI markings in each of the data sample in the corpus and

statistically analysing these in terms of the CI/SI corpora ratio. SPSS produced an overall mean CI/SI ratio of 0.39. The overall mean result of CI/SI ratio at 0.39 means that H0: CI/SI = 0 can be rejected and that H1: CI/SI > 0 can be accepted.

Figure 2 illustrates the spread of each of the data samples as a CI/SI ratio. This is presented for the reader to provide a visual representation of the ratio for all corpora in their data set categories (AV: Audio- Visual; BB: Billboard; LIT: Literature; RI: Recordings of Interviews). Figure 2 shows that most data samples contained CI below 0.39, while most data samples containing CI above 0.39 ratios are only found in the AV and BB categories.

Interestingly, while the overall CI ratio in the RI category in this study is below 0.39, the RI data samples compiled by Sneddon (2006), show a much higher individual CI word count usage in comparison to the FI equivalent as shown in Table 5.

Correspondence analysis and the formal-informal spectrum

The distribution of the mean ratios for each data set (Table 6) shows that most of the data sets fall within the 0.2 – 0.7 range with the 0.3-0.49 dimension holding the most entries. There were only three data sets that fell within the <0.2 and >0.7 dimensions.

Figure 2
All Data Sets

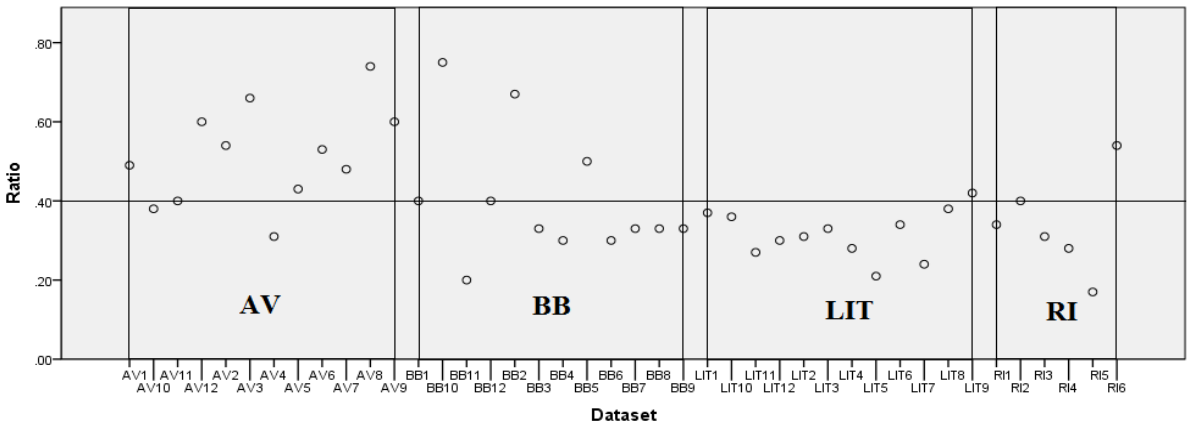


Table 5
Sneddon's Individual CI Word Count in the RI Category

CI lexical item	Percentage vis-à-vis FI equivalent	FI equivalent
Aja	98.8 %	*saja*
Udah	96 %	*sudah*
-in suffix	70.4 %	*-kan/-i* suffix
sama/ama	84.6 %	*oleh*
lagi (aux)	98.9 %	*sedang*
bakal	46.1 %	*akan*
nggak/kagak/ndak	97.9 %	*tidak*
gua/gue	91.8 %	*saya/aku*
cuma(n)	95.9 %	*hanya*
banget/amat	95.3 %	*sangat/sekali*
Entar	62.4 %	*nanti*
Gimana	94.9 %	*bagaimana*
Kayak	84 %	*seperti*
pengen/kepengen/pingin/kepingin	97.9 %	*ingin*
(ng)omong	93.9 %	*bicara/berbicara*
Gede	88.6 %	*besar*

Table 6
Correspondence Analysis of All Data Sets

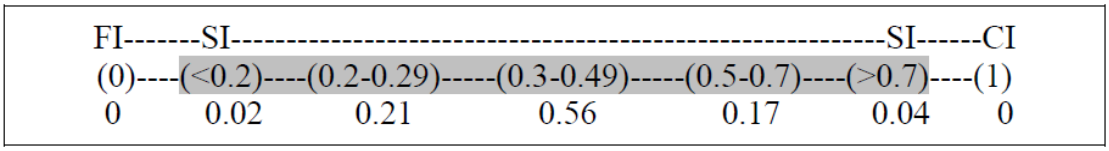
	Dimensions = CI/SI Ratio				
	<0.2	0.2-0.29	0.3-0.49	0.5-0.7	>0.7
BB n = 12		1	8	2	1
LIT n = 16		7	9		
AV n = 14		1	7	5	1
RI n = 6	1	1	3	1	
% of corpus	0.02	0.21	0.56	0.17	0.04

(BB: Billboard; LIT: Literature; AV: Audio-visual; RI: Recordings of Interviews)

The next analysis compares the dimensions of Table 5 with the formal-informal spectrum of the SI continuum (Figure 3). The dimensions of the correspondence analysis are translated as intervallic variables in the formal-informal spectrum to show the spread of the data samples. The left-most 0 on the spectrum represents zero presence of CI while the right-most 1 on the spectrum represents usage containing exclusively CI. The bottom indicator marks the percentage the dimensions occupy as datasets from the corpus. Figure 3 represents this study's quantitative findings located along the informal language continuum of SI (Djenar & Ewing, 2015; Sneddon, 2006).

Figure 3 shows that none of the corpora fell at the extreme end of the intervallic scale (0; 1), indicating that neither FI nor CI in their pure forms are used as an everyday language. The shaded range covering dimensions <0.2 - >0.7 is where the corpora data samples have spread with one RI data sample falling in the <0.2 dimension and one AV and one BB falling in the >0.7 dimension (Table 2). Datasets in the dimension 0.3-0.49 CI/SI ratio occupy the largest share (0.56) of the corpus (Figure 3) suggesting that a formal- informal spectrum with a 0.3-0.49 CI/SI ratio is the most commonly encountered form of SI.

Figure 3
The Spread of Data in the SI Formal-Informal Spectrum



There are plausible reasons why three of the data sets fell outside the <0.2 and >0.7 range. The two data sets below <0.2 involved 1) an interview with an academic, and 2) an after-school-lesson advertisement. In the introduction of this transcript, Sneddon (2006) noted that the interview with the academic was 'somewhat formal and courteous'. Prior to that he has stated that it is usual amongst educated people, even when conversing in informal settings, that speech consisting of CI elements is likely to occur in only short segments and that FI will always dominate the register.'

The more formal register in these data samples was likely to result from the education field and high-status tenor between the speakers, which in this case demonstrates the function of FI as a language of education and formality. This serves to remind us that foreign Indonesian language learners still need to be taught about the sociolinguistic implications for their choice of register and their need to be conscious of using FI in appropriate settings.

The audio-visual data set above >0.7 is a comedy scene from a film starring the late Betawi actor Benjamin. The heavily CI-influenced informal register reflects his Betawi cultural background. These data sets are provided in the Appendix as examples to demonstrate how CI and FI were coded in the corpus data sets. To see how all the data sets were coded see 1st Author (2017).

Kohler and Mahnken (2010) have noted how the complexity of Indonesian language variants has been simplified in textbooks and consequently the spoken language is under-represented. This has resulted in learners of Indonesian language being ill-equipped to communicate in informal settings. Many informal dialogues in Indonesian language textbooks, which are usually designed or generated by the writer(s), are presented in FI. This contrasts with the results of this study which found that FI in its pure form is not used as an informal spoken language. The common practice of misrepresenting Indonesian as exclusively FI (Djenar, 2006) is partly due to a lack of understanding of the diglossic situation and because of the traditional educators' perception that the CI language is not appropriate to be taught because it is not 'good and proper' (baik dan benar) (Sneddon, 2006).

CONCLUSION

The main finding from this study is that linguistic features from informal spoken Indonesian CI are prevalent in everyday speech. Corpus data support Sneddon's observations that standard Indonesian FI has merged with CI to form an informal spoken Middle variant SI. This research shows that there are no set quantitative boundaries as to what defines the parameters of SI. This finding suggests that CI lexicon and grammar may contribute as much as 39% to everyday spoken Indonesian (SI).

The intention of this study was primarily to investigate the validity of existing observations and assertions by other scholars of the existence of SI, a middle variant, using qualitative and quantitative methods against corpora of informal language. Questions of SI use in relation to demographics are outside the scope of this article but provide opportunities for further research. The findings of this study may inform further research on SI such as geographic and demographic variations of SI, as well as diachronic CI studies, and the impact that modernity and world languages (notably English) have on SI.

This study and other similar studies on

Indonesian linguistics and sociolinguistics form part of a shifting paradigm in the understanding of the spoken Indonesian language and subsequently changes in the teaching and learning of Indonesian language. A practical outcome of this research is the development of an SI language description which may inform the inclusion of CI in Indonesian language teaching materials to benefit students studying Indonesian as a foreign language.

Research suggests that utilising authentic texts in second language acquisition aides in developing native speaker competency (Gilmore, 2007). Many language learning texts that are created by publishers often do not reflect real-life language usage. Explicit teaching and learning of CI can provide explanations of the hitherto insufficiently understood CI lexis, speech acts, semantics and pragmatics, and allow for Indonesian language teachers to understand and utilise more authentic sources (e.g., contemporary real-life materials from TV, internet and films) as teaching resources.

The findings of this study lay the linguistic foundation for the development of a colloquial spoken Indonesian pedagogy. It is outside the scope of this article to detail this colloquial spoken Indonesian pedagogy here, but the reader can find such detail in the unpublished Doctoral thesis on which this paper is based (Nataprawira, 2017). For future publications on this subject, the authors intend to provide pedagogic models on how to teach and learn colloquial spoken Indonesian. Language aspects to include are authentic texts featuring common native speaker speech acts and explicit analysis of spoken lexis, collocation and intonation, and their semantic and pragmatic implications.

REFERENCES

Biber, D. (2015). Corpus-based and Corpus-driven analyses of language variation and use. In B. Heine & H. Narrog, (Eds.), *The Oxford handbook of linguistic analysis* (2nd ed.). Oxford University Press.

Boellstorff, T. (2002). Ethnolocality. *The Asia Pacific Journal of Anthropology, 3*(1), 24-48. https://doi.org/10.1080/1444221021000170619 6

Chapman, S., & Routledge, C. (Eds.) (2009). *Key ideas in linguistics and the philosophy of language.* Edinburgh University Press.

Davies, A., & Elder, C. (2008). *The handbook of applied linguistics*. Blackwell Publishing.

Djenar, D. N. (2003). *A student's guide to Indonesian grammar*. Oxford University Press.

Djenar, D. N. (2006). Patterns and variation of address terms in colloquial Indonesian. *Australian Review of Applied Linguistics, 29*(2), 16. https://doi.org/10.1075/aral.29.2.07dje

Djenar, D. N. (2008). On the development of a colloquial writing style: Examining the language of Indonesian teen literature. *Bijdragen tot de Taal-, Land- en Volkenkunde, 164*(2/3), 238. https://doi.org/10.1163/22134379-90003658

Djenar, D. N., & Ewing, M. C. (2015). Language varieties and youthful involvement in Indonesian fiction. *Language and Literature, 24*(2), 108-128. https://doi.org/10.1177/0963947015573387

Errington, J. (1986). Continuity and change in Indonesian language development. *The Journal of Asian Studies, 45*(2), 329-353. https://doi.org/10.2307/2055846

Errington, J. (2014). In search of Middle Indonesian: Linguistic dynamics in a provincial town. In G.V. Klinken, & , W. Berenschot. (Eds.), *In Search of Middle Indonesia* (pp. 199-219). Brill.

Ewing, M. C. (2016). Localising person reference among Indonesian Youth. In Z. Goebel, D. Cole, & H. Manns (Eds.), *Margins, hubs, and peripheries in a decentralizing Indonesia* (pp. 26-41). Tilburg University.

Fan, L. Y. (1990). *Speak standard Indonesian: A beginner's guide*. Times Books International.

Ferguson, C. A. (1959). Diglossia. *Word, 15*, 325-340. https://doi.org/10.1080/00437956.1959.116597 02

Gil, D. (1994). The structure of Riau Indonesian. *Nordic Journal of Linguistics, 17*(2), 179-200. https://doi.org/10.1016/b0-08-044854-2/04594-6

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97-118. https://doi.org/10.1017/s0261444807004144

Grijns, C. D. (1981). Jakarta speech and Takdir Alijahbana's plea for the simple Indonesian word-form. In N. Phillips & K. Anwar (Eds.), *Papers on Indonesian Languages and Literature* (pp. 1-34). School of Oriental and African Studies.

Grijns, C. D. (1991). *Kajian bahasa Melayu – Betawi*. PT Pustaka Utama Grafiti.

Kohler, M., & Mahnken, P. (2010). *The current state of Indonesian language education in Australian schools*. Education Services Australia.

Kushartanti, B. (2014). *The acquisition of stylistic variation by Jakarta Indonesian children*. LOT Trans 10.

Manns, H. (2014). Youth radio and colloquial Indonesian in urban Java. *Indonesia and the Malay World, 42*(122), 43-61. https://doi.org/10.1080/13639811.2014.876156

Nataprawira, H. (2017). *Recognising the sociolinguistic reality of spoken Indonesian: A corpus and usage analysis of a middle*

diglossic variant. [PhD thesis, University of the Sunshine Coast]. USC Australia Research Bank. https://research.usc.edu.au/discovery/fulldisplay?context=L&vid=61USC_INST:ResearchRepository&search_scope=ResearchETD&tab=Research&docid=alma99450901402621

Nothofer, B. (1995). The history of Jakarta Malay. *Oceanic Linguistics, 34*(1), 86-97. https://doi.org/10.2307/3623113

Quinn, G. (2001). *The learner's dictionary of today's Indonesian*. Allen & Unwin.

Smith-Hefner, N. (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. N. (1996). *Indonesian reference grammar*. NSW, Allen & Unwin.

Sneddon, J. N. (2000). *Understanding Indonesian grammar.* Allen & Unwin.

Sneddon, J. N. (2001). Teaching informal Indonesian: Some factors for consideration. *Australian Review of Applied Linguistics, 24*(2), 81-95. https://doi.org/10.1075/aral.24.2.06sne

Sneddon, J. N. (2002). *Variation in informal Jakartan Indonesian: A quantitative study*. Paper presented at the Ninth International Conference on Austronesian Linguistics, Canberra, Australia.

Sneddon, J. N. (2003a). Diglossia in Indonesian. *Bijdragen tot de Taal-, Land- en Volkenkunde, 159*(4), 519-549. https://doi.org/10.1163/22134379-90003741

Sneddon, J. N. (2003b). *The Indonesian language: Its history and role in modern society*. UNSW Press.

Sneddon, J. N. (2006). *Colloquial Jakartan Indonesian*. Pacific Linguistics.

Tamtomo, K. (2019). The creation of monolanguaging space in a krámá Javanese language performance. *Language in Society, 48*(1), 95-124. https://doi.org/10.1017/s0047404518001124

Tjung, Y., Cole, P., & Hermon, G. (2006). Is there pasif semu in Jakarta Indonesian? *Oceanic Linguistics, 45*(1), 64-90. https://doi.org/10.1353/ol.2006.0009

APPENDIX

Data sample BB10

The following are the data samples outside the <0.2 and >0.7 correspondence analysis range. The first two samples have a CI content <0.2 and the last one >0.7. These samples demonstrate how field and tenor can affect language use in the formal-informal spectrum.

BB10 is an advertisement for after school lesson preparing students for the national and general exams. The word count only included the sentence *Dapetin Suksesmu Di Sini* ("Find Your Success Here") - an advertising slogan appealing directly to the target audience. The pragmatic function of this sentence might explain the use of the CI item *Dapetin*, employing language of familiarity to attract student customers. The general information about the course on the banner - all in FI - is not counted, as it is not representative of direct speech or dialogue.

Depatin SuskesMu Di Sini. Dapetin – CI morphology of –in suffix

Total word count: 5. CI words: 1

Data sample RI15

RI15 is the transcript of a recorded interview between the interviewee R, a 47-year-old academic and the 23-year-old interviewer Yuli. The interview took place in R's office. The tenor, a senior academic conversing with the younger interviewer in a somewhat formal setting - R's office, would have informed the choice of more formal language, despite it being a non-formal interview. CI items (in italics) still peppered the conversation used by both speakers. Source: Sneddon, J. N. (2006).

Two speakers:

A: R, 47, female, member of academic staff, Atma Jaya University
B: Yuli, 23, female, interviewer and recorder

The interview was in R's office on 10 January 2001. The opening is somewhat formal and courteous. The interviewee speaks rather slowly and quite fluently. Her story is at times somewhat discursive and not always chronological.

B: Selamat pagi. Ah sekarang saya ada di ruangannya Ibu RJ, kepala PBB yang baru. Aa *Slamat* pagi, Bu R.
A: Selamat pagi Yulianti.
B: Apa kabar Bu?
A: Eh, baik-baik *aja tuh. Gimana*?
B: Ah gini Bu. Saya *mo interview* Ibu *ni*. Bisa *nggak* Ibu cerita kira-kira dari kehidupan Ibu dari kecil *sampe* sekarang?
A: Am *gini* Yulianti. Saya itu *kan* lahir *taun* lima puluh tiga, ya. Lima pulu tiga itu, *skarang* sudah umur empat *pulu tuju* tahun ya? *Udah*, udah tua, *uda* nenek-nenek.
Lalu, saya *mulai* di- saya dilahirkan dari sua- satu keluarga yang sangat besar dengan orang tua yang punya anak dua belas anak. Lalu ayah saya itu seorang miskin ya, dalam arti, aa saya datang dari keluarga miskin. Ayah ibu saya itu, Ibu saya tukang ju- tukang kue. Malu *kan*? Hanya…
B: *Nggak pa-pa*.
A: Tukang kue keliling, *gitu* ya. Tukang kue keliling dan ayah saya itu juga aa mungkin *kalo* sekarang itu tukang loak, ya? Bilangnya ya? Yang di pinggir jalan itu ya. Lalu dia punya anak dua belas. Lalu a… setiap anak itu diajar untuk mandiri. Untuk sendiri-sendiri pokoknya cari makan, *gitu yah*. Supaya *survive*. Tapi ada

satu hal yang saling men- yang *sampe skarang* saya masih inget bahwa orangtua saya mengatakan bahwa kepandaian itu tidak akan hilang. Jadi dia katanya ee sekolah, begitu. Apapun harus sekolah, begitu. Sehingga ee kami mendapat contoh dari yang paling besar, jadi anak yang paling besar, *skarang* dia adalah ginekolog ee spesialis kebidanan, dan dia sukses sekali ya. Ee dia senior begitu ya? Bekas kepala rumah sakit Cirome, Cirebon, dan sebagainya. Dia tantara ya. *Karna* memang di tantara itu *kan dikasi* makan ya, *dikasi* uang lauk pauk dan sebagainya. Jadi dia kuliah di UI, itu menjadi panutan kita semua. Yang paling besar ee jadi panutan. Dia kuliah di UI dan kami tinggal di Bogor. Dan dia harus *naek* kereta api untuk ke UI, san setelah ee *sampe*, *sampe* dia lulus itu kami masih miskin, *nggak* punya apa-apa. Dia paling-paling naik sepeda, *gitu* ya. Lalu ee *kalo* saya *liat* fotonya *tuh* saya sedih *beneran*, *karna* dia begitu kurusnya, kecilnya begitu ya, tapi dia *pengen* selesei. Begitu dia selesei dia masuk ke ee dinas militer ya. Dinas militer, waktu itu dia ditempatkan di Kalimantan, *kalo nggak* salah. Di Kalimantan itu dengan penuh penderitaan dia lalui, dan dia kembali ke Jakarta. Ah saya *masi* kecil. Saya anak kesembilan. Anak kesembilan dari dua belas *besodara*. Jadi *waktu* dia kembali itu, adik-adiknya ikut dengan dia, *walopun* dia masih minim sekali. Dia baru lulus, baru *selesei*, datang ke Jakarta, keadaan masih *nggak* punya tapi kita ikut, *nebeng*, gitu yah? Dibagi-bagi, *ade-ade*nya *tu* dibagi. Ada yang ikut sana, ada yang ikut sini, *gitu*. Saya *tu* termasuk ikut dia. Ee dia *tuh* tantara. Jadi waktu, saya *inget skali*, *waktu* saya sudah mahasiswa, aa nanti kita *flashback* ke *blakang* ya? Waktu saya mahasiswa, itu ada peristiwa Malari, jadi dia punya… apa? Dia ada mobil combi *gitu*, jelek sekali ya, masuk di Kramat, oo dilempari batu oleh siapa *nggak ngerti*. Dan dia begitu sedihnya *karna* dia pecah itu kacanya *gitu*. Dia *nggak* punya apa-apa *gitu* ya. *Na* itu aa kakak saya nomor satu. Tapi itu jadi panutan saya, terutama saya, karna saya tinggal dengan kehidupan keras ya, dalam arti dia punya anak empat yang empat-empatnya sukses, yang paling kecil dia di Amerika *skarang*. Aa.. apa? Kehidupannya *tuh* kehidupan, kehidupan miskin gitu. Kehidupan *nggak* punya. Jadi kami *kalo* punya uang *tuh* hanya bisa aa bisa untuk minum susu segelas barangkali. Itu minum susu segelas juga sulit nyarinya ya? Dengan makan yang tidak seperti sekarang ya. Jadi kami itu makan semua dibagi ya. Jadi piring-piring *tu* dibagi oleh ibu saya. Piring-piring-piring isinya *tu* ada kentang dua biji, dua biji, dua biji *gitu*, *nggak* boleh *nambah gitu lo*. *Nggak bole nambah* sama sekali. Dan kakak saya paling besar ini, yang, yang di fakultas kedokteran, *kan* dia masih kuliah itu, *ngambil* jeroan ah apa *tuh*, yang dibuang di, di kali, di got *gitu*, dibuang, diambil sama dia disikat gitu lo. Disikat untuk dimasak gitu ya? Itu dikasih ke ibu saya. Ibu saya masak lalu dibagi-bagi ke adik-adiknya *gitu*. Ke anak-anaknya. Aa kehidupan kami *bener-bener* sangat, sangat ee miskin *gitu*, *nggak* punya *pa-pa*. Lalu aa itu, itu pada diri saya juga ada sifat untuk *bageimana* supaya bisa. Tapi satu hal yang pasti itu bahwa kami dididik untuk belajar. Sekolah *gitu*, *karna* sekolah *tu* nomor satu. *Nggak* boleh *nggak*, ya. *Walopun* dengan mengemis, minta-minta untuk ee masuk sekolah *gitu*, tidak aa anak *kan* harus bayar. Dan… lalu *stela* itu *udah*, saya tinggal sama kakak saya itu, ee *nggak* ee agak lama sedikit. Setelah itu dia mulai karirnya maju dan sebagainya, mulai kita dibantu ee uang kuliah, uang sekolah, *gitu* ya. Uang kuliah saya *dapet* uang *skola*, kuliah, sehingga waktu saya *dapet* beasiswa Supersemar dari Pak Harto itu, saya *tu* ee uang, uang sekolah saya *tetep* dibayarin *karna* uang Supersemar itu adalah uang saya *gitu lo*. Jadi katanya, 'Itu *kan* jerih payah kamu. IP kamu *kan*, apa? 'nilai kamu *kan* tinggi, jadi kamu *dapet*, itu hak kamu,' *gitu*. Tetep *aja* saya dikasi *tuh*. Saya *inget* saya *dapet* lima belas ribu dari Pak Harto, dari Supersemar. Jadi mulai pertama kali saya *kulia tuh* saya ditawari *karna* aa *stela* semester satu *tu* nilai saya cukup baik yah, bagus-bagus, lalu saya ditawari saya dapet Supersemar.

Total word count: 816. CI words: 126

Data sample AV8

AV8, the data set with CI content above >0.7, is the transcript of a comedy scene from a film starring the late legendary Betawi actor Benjamin. In this comedy scene, Benjamin and his friend are crossing a river and are arguing over the dirty water and how his friend who is being dragged on a sled because he is ill, will have to get partly submerged while Benjamin is riding a horse. Benjamin's Betawi cultural background reflects the heavily CI influenced informal register. CI items are in *italics.*

Ben, pelan-pelan *dong you jalanye*, *aye* sedang *meriang nih*, Ben... *brengsek lu ah*
ah...diem-diem aja lu di situ, *molor aja terus, lu taunya sampe*, *ah...pake meriang* segala, *udah* tau orang mau *ngungsi*, mau *ngikut*, *jage diri lu baek-baek*, *gue nga* buang *aja udah* bagus *lu…ah..let's go* aduh Ben...Ben...tobat *ah...aye* bisa mati di jalanan *nih*...aduh...
slowly...slowly tiger...slowly
Ben..Ben..*plosotan* Ben...pelan pelan Ben...aduh aduh aduh...bisa *nyangkut nih aye*....Ben...mau dibawa kemana *sih* Ben...Ben...mau ke mana?
Sorry dongo...memang nasib lu...c'mon tiger...mudah-mudahan nga dicaplok buaya lu
pake lewat sungai *lagi*...aduh...dingin...*aye* sedang *meriang nih*...Ben...*kira-kira dong*...kau kira aku ini ikan kapus...Ben...*apaan tuh*?... *pada ngambang nih gituan*...Ben... *lekasan dong*
shut up! Merendem aja situ terus...ama gituan aja takut...bencet aja... ah... masa nga ancur...eh ngorok aja situ terus

Ben, pelan-pelan *dong you jalanye*, *aye* sedang *meriang nih*, Ben... *brengsek lu ah*
ah...diem-diem aja lu di situ, molor aja terus, lu taunya sampe, ah...pake meriang segala, udah tau orang mau ngungsi, mau ngikut, jage diri lu baek-baek, gue nga buang aja udah bagus lu...ah..let's go
aduh Ben...Ben...tobat ah...*aye* bisa mati di jalanan *nih*...aduh... slowly...slowly tiger...slowly
Ben..Ben..*plosotan* Ben...pelan pelan Ben...aduh aduh aduh...bisa *nyangkut nih aye*....Ben...mau dibawa kemana *sih* Ben...Ben...mau ke mana?
Sorry dongo...memang nasib lu...c'mon tiger...mudah-mudahan nga dicaplok buaya lu
pake lewat sungai *lagi*...aduh...dingin...*aye* sedang *meriang nih*...Ben...*kira-kira* dong...kau kira aku ini ikan kapus...Ben...*apaan tuh*?... *pada ngambang nih gituan*...Ben... *lekasan dong*
shut up! Merendem aja situ terus...ama gituan aja takut...bencet aja... ah... masa nga ancur...eh ngorok aja situ terus

Total word count: 236. CI words (in italics): 175