

The Emergence of Indonesian Electronic Communication as a Distinct Linguistic Variety

Thomas J Conners
University of Maryland

With the spread of smartphones, the proliferation of social media apps, and availability of high-speed internet, human interactions and communication are increasingly taking place in the digital realm. Electronically mediated communication, or ECM, accounts for an increasing amount of many people’s communications. Indonesia represents an excellent test case; although internet penetration stands at just over 50% in the country, it has the highest rate of increase of internet penetration on the globe (at nearly 51% annually compared with a global average of 8% Nguyen 2017). It further ranks among the top users of a wide variety of different social media platforms.

Recent research on the language of social media, or ECM more broadly, has focused on various aspects of natural language processing, dialect and register identification, and automated extraction of data analytics such as emotion and intensity. Comparatively little work has been done to develop a framework for analyzing and characterizing ECM types as distinct linguistic varieties. Part of the challenge is how to account for the different modes of input (written/typed), and the lack of traditional modes of expression (spoken). Some recent work has explored how the language used on different platforms confirms to a greater or lesser degree to the characteristics of traditional written or spoken language (e.g. SMS as ‘fingered speech’ McWhorter 2013); but no systematic framework has been proposed that allows for a proper description of the unique properties that characterize ECMs, and make them distinct from other ECMs (e.g. the language of Twitter, email, and Instagram as three distinct varieties), and make them distinct from non-ECM (e.g. Sindoni 2013).

This paper presents the beginnings of a framework for exploring systematic differences in the linguistic features that can successfully be used to distinguish and compare varieties of ECMs. The first application of this framework is to a contrastive study of the register-specific features of a corpus of Indonesian twitter and Indonesian SMS. Two broad categories of features are explored, those that hold at the lexical level (phonological changes, abbreviations, expansions, etc., as shown in the table below), and those that hold at the clause level (clausal complexity, length, null anaphora). These features are necessary to characterize an ECM, but not sufficient. Beyond dialectal features, a notion of register, in the traditional sense, is also required. Intended audience, particular platform, and communicative goal are further required to more robustly characterize Indonesian ECMs as distinct varieties.

English gloss	Standard Indonesian	Jl SMS Spelling	Jl Twitter Spelling	Phenomenon
‘thank you’	<i>terima kasih</i>	makaci		s → c
‘okay’	<i>sipsip</i>	cipcip		s → c
‘sad’	<i>sedih</i>		cudih	s → c
particle	<i>deh</i>	dech		_# h → ch
particle	<i>sih</i>		sich	_# h → ch
relative marker	<i>yang</i>	yank		_#g → k
‘older brother’	<i>kakang</i>		kank	_#g → k
‘this’	<i>ini</i>	nich		∅ → ch
‘thanks’	<i>thanks</i>	tq		k → q

International Workshop “Varieties of Malayic Languages”
December 19-21, 2017

1 st person	<i>aku</i>	aq	aq	ku → q
1 st person	<i>aku</i>	qu		k → q
1 st person	<i>gue</i>	gw	gw	ue → w
1 st person	<i>gua</i>	w	w	ua → w
‘wide noodle’	<i>kwetiau</i>	kwto		au → o
‘or’	<i>atau</i>		ato	au → o

Brugman, Claudia and Thomas J Connors. In Press. Comparative study of register specific properties of Indonesian SMS and Twitter: implications for NLP. In *Digital Scholarship in the Humanities*.

McWhorter, J. 2013. Texting is Killing Language. JK! TED talk.
http://www.ted.com/talks/john_mcwhorter_txtng_is_killing_language_jk. Accessed 02/09/2017.

Nguyen, Han. 2017. Indonesia is fastest growing country for internet use.
<https://www.rappler.com/world/regions/asia-pacific/indonesia/bahasa/englishedition/159593-internet-social-media-statistics>. Accessed 11/23/2017.

Sindoni, M. G. 2013. *Spoken and Written Discourse in Online Interactions*. New York and London: Routledge.