Register specific properties of Jakarta Indonesian SMS: A corpus study

This paper details the distinctive linguistic features used in a corpus of Indonesian SMS that are particular to this, and by extension possibly other forms of informal media.

The literature on the distinctive linguistic properties of communications produced and received via informal media channels is limited in many ways. Most importantly from the point of view of under-resourced studies, the languages of focus have been largely limited to a few, mostly Indo-European languages, and Chinese and Japanese. This language bias may have resulted in correspondingly biased outcomes of the research questions posed. The linguistic communities of Indonesia are enthusiastic users of electronic communications technologies[1]; with its large, mostly multilingual population and geographic expansiveness, Indonesia thus provides a test bed for understanding the sociolinguistic consequences of electronic communications. Moreover, because of its nature as a relative IMA language, that is Isolating, Monocategorial, Associational language (Gil 2005), Indonesian, and even more so its non-standard(ized) varieties, such as Jakarta Indonesia, potentially push the envelope of its grammatical tendencies. That is, SMS language in general tends towards abbreviation, truncation, ellipsis, and other grammatical processes that are already characteristic of IMA languages. While in many ways the corpus under study bears out the prediction that JI, as expressed through SMS, behaves even more as an IMA language than the spoken language, there are also significant ways in which this does not hold.

We detail here both the ways in which JI in SMS is both more IMA-like and less IMA-like.

Another significant gap in the research involves the linguistic properties of SMS in the wider landscape of tweets, blogs, and online newsgroups and bulletin boards (hereafter "general informal media"); all of these media allow for various non-standard language features, including those of spelling and syntax. By their nature, SMS communications are those of people known to each other, and for the most part represent one-to-one communications (a notable register study is Hei, 2014, a comparative analysis of SMS-medium condolence messages among three groups of Malaysians). They also resemble conversations in being produced and received synchronously, unlike emails and blogposts; the turn-taking behavior reflects this near-instantaneous form of interaction. While we have reviewed the literature on general informal media and, even more widely, computer-mediated communication, for hypotheses about SMS register properties, we note that (a) SMS is predicted to differ from other informal media in ways that reflect its inherently interpersonal purpose and (b) register analyses of any informal media (as distinct from studies that correlate particular register properties with, say, gender or education level) are relatively rare in the literature. The current study aims to help fill that gap.

Gil, David. 2005. Isolating-monocategorial-associational language. In Cohen, H. and C. Lefebvre eds. *Handbook of categorization in cognitive science*. 347-379. Oxford: Elsevier.

---

[1] As of 2014, Indonesia counts as the fourth largest population in active users of Facebook (http://addictivelists.com/top-10-countries-with-most-facebook-users-2014/) and the third highest in active users of Twitter, accounting for 6.5% of all active users. (http://www.statista.com/chart/1642/regional-breakdown-of-twitter-users/). Strikingly, some 91% of Indonesians who access the internet, are also active Facebook users. (22% of Indonesians have access to the internet.).

Hei, K. C. 2014. Ethnic Variations in Malaysian SMS condolences. Presented at MICOLLAC 2014. Draft available at
http://www.academia.edu/7780759/Ethnic_Variations_in_Malaysian_SMS_condolences_2014_