

## Caveat Emptor:

### Using Data from Standard Malay/Indonesian to Make Claims about Universal Grammar

The Malay/Indonesian language figures prominently in a wide range of theoretical studies concerned with various domains of grammar. Notwithstanding a recent increase in the use of data from colloquial varieties, a large majority of such studies are still based on data from the standard varieties of Malay and Indonesian. This paper argues that the use of Standard Malay/Indonesian to make claims about Universal Grammar is methodologically unsound, for the simple reason that Standard Malay/Indonesian is not a natural language in the usual sense of the word.

Part 1 of this paper presents an overview of a sociolinguistically-based typology of Malay/Indonesian varieties. An important distinction is that between traditional and koiné varieties. Traditional varieties include a range of dialects such as Riau Malay, Kelantan Malay and Sambas Malay, while koiné varieties include varieties such as Riau Indonesian, Jakarta Indonesian and Sabah Malay. In addition to traditional and koiné varieties, there are also migrant varieties, such as Peninsula Chinese Malay, as well as diaspora varieties, such as Sri Lankan Malay. Of concern to us here, however, is an additional category, a potpourri of special modalities and registers that do not fall into any of the previous types, and are of a derivative nature, nobody's native language, of restricted distribution, used in a limited range of contexts, and structurally aberrant. Among such special modalities and registers are manual signed systems such as SIBI; the special orthographic conventions of social media such as Bahasa Alay; assorted literary genres such as the Pantun; youth languages such as Bahasa Gaul; gay languages such as Bahasa Binan; ludlings such as Babahabasaba; code-switching systems such as Bahasa Rojak; pidgins such as Bazaar Malay; and many other such systems — none of which are bona fide languages.

Building on this typology, Part 2 of this paper argues that Standard Malay/Indonesian is itself most appropriately viewed as a special register of Malay/Indonesian, rather than a real variety of the language. To make this point, it is shown that Standard Malay/Indonesian shares five crucial features with most or all of the other special modalities and registers of Malay/Indonesian: (1) It is acquired non-natively, in later childhood or adulthood; (2) It is of restricted distribution, mastered to varying degrees by different members of the community; (3) It is used in a limited range of contexts; (4) It is structurally aberrant, as compared to other varieties of Malay/Indonesian, other languages of the region, and languages worldwide; and (5) It is parasitic, derived not just diachronically but also synchronically from other varieties of Indonesian. Thus, Standard Malay/Indonesian is not a real language, but rather a derivative register of Malay/Indonesian, of similar ontological status to other special modalities and registers such as pidgins, youth languages and ludlings. What this means is that linguists cannot use data from Standard Malay/Indonesian to make straightforward inferences with regard to what is possible in human language, any more than they can allow, say, for a language with a phonological rule inserting a [b] into every syllable peak just because there is Malay/Indonesian ludling that is formed in this way.

Following on the characterization of Standard Malay/Indonesian as a special register, Part 3 of this paper examines some of the ways in which Standard Malay/Indonesian differs systematically from other bona fide varieties of Malay/Indonesian. Focus is on the overt marking of grammatical categories, specifically those pertaining to clausal organization and the assignment of thematic roles. Results are presented of an experiment measuring the degree to which the assignment of thematic roles is encoded in the grammar by means of strategies such as word order, NP flagging, and verbal voice marking. The results show that thematic roles are significantly more strongly grammaticalized in Standard Malay/Indonesian than in the colloquial varieties, the latter allowing for a substantially greater degree of underspecification. In conclusion, cross-linguistic results of the same experiment show that the greater degree of grammaticalization of thematic roles Standard Malay/Indonesian as compared to colloquial varieties is not an idiosyncratic property of Malay/Indonesian, but rather a characteristic feature of grammatical variation between standardized and colloquial varieties of macrolanguages, as evident, among others, in Japonic, Sinitic and Romance. Thus, the focus on standard varieties of major world languages engenders a systematic bias in our view of what the grammars of languages are like, resulting in an overemphasis on the grammatical structures associated with such standardized varieties.