Deriving Categorical and Continuous Properties of Javanese Speech Levels

Christopher Davis, University of the Ryukyus

Overview. I propose an analysis of the Javanese speech level system in which (i) a set of binary features on speech-level marked lexical items combine to generate the three traditionally recognized speech levels (*ngoko*, *madya*, and *krama*), and in which (ii) pragmatic principles operating on particular paradigms of lexical alternatives generate continuous gradation of perceived formality within the *madya* level.

Speech Levels and Lexical Classes. Javanese utterances are traditionally sorted into one of three levels, called *ngoko*, *madya*, and *krama*. The *ngoko* speech level is canonically used towards intimates and addressees of similar status, while *krama* speech level is used toward addressees of higher status and low intimacy. The *madya* level serves as a "half-way house" (Wolff & Poedjosoedarmo 1982) between these two endpoints, canonically used in situations where the factors determining the choice of speech level are in conflict.

The speech level of an utterance is signaled by the choice of lexical items belonging to classes which are themselves traditionally labeled in terms of the speech level with which they are used. These lexical items belong to paradigms of suppletive alternants with identical semantic content but differing in terms of the speech levels with which they are compatible. The most basic classification sorts such lexical items into either *krama* or *ngoko* classes. The *krama* speech level is then signaled by the exclusive use of *krama* variants, and the *ngoko* speech level by the exclusive use of *ngoko* variants. The *madya* level, meanwhile, is signaled by a mixture of *krama* and *ngoko* lexical items.

This simple taxonomy is complicated by the existence of what Clynes (1989) calls "style markers". First, there is a small set of lexical alternants that are *only* compatible with the *madya* speech level. Second, among the *krama* and *ngoko* lexical alternants are a subset of items that are *not* compatible with the *madya* level. In other words, among the class of items traditionally labeled as *krama* there is a subset that is compatible *only* with the *krama* speech level, and similarly for those items traditionally labeled *ngoko*. The taxonomy of lexical alternants is thus divided into five classes, which following Clynes 1989 I treat in terms of two binary features, [±K] and [±N].[1] These features indicate the type of context (speech level) that the lexical item is compatible with: *krama* level is signaled by utterances whose lexical items collectively encode the features [+K,−N], *ngoko* level by [−K,+N], and *madya* by [+K,+N]. The five lexical classes are then categorized as follows: [+K,−N] items can be used only with the *krama* speech level, [+K] items with either *krama* or *madya* speech level, [+K,+N] items only with the *madya* speech level, [+N] items with both the *madya* or the *ngoko* speech level, and [−K,+N] items only with the *ngoko* speech level.

Categorical versus Continuously Graded Levels. The feature-based system outlined above generates a categorical distinction between the three speech levels, and hard co-occurrence constraints that block the mixing of lexical items with incompatible feature values (ie, mixing [+K] and [−K] or [+N] and [−N] lexical items). As Clynes convincingly demonstrates, this latter property is empirically valid, since such combinations are in fact judged as ungrammatical by native speakers (examples left out for space reasons). While there is thus strong evidence that Javanese speech levels are categorically divided, and that this division can be modeled by a simple feature system,

[1] Clynes attempts to directly capture the meaning of the associated speech levels by using the features [±STATUS] and [±INTIMATE]. This reduction leads to difficulties that I leave aside for reasons of space.

there is also a widespread (Uhlenbeck 1970, W&P 1982, Errington 1985) intuition that there is a continuous gradation in the *madya* level, with a greater proximity to either *krama* or *ngoko* levels indicated by the proportion of *krama* (our [+K]) and *ngoko* (our [+N]) lexical items used. That is, *within* the categorically determined three-level system, there is a gradient between "more *krama*-like" or "high" *madya*, and "more *ngoko*-like" or "low" *madya* (W&P 1982). The question is how exactly this gradient status is to be calculated.

A simple answer, which implements a suggestion found in Clynes 1989, is this: Assume that speech levels are continuously valued on the interval [0,1], with *ngoko* level valued at 0, *krama* valued at 1, and *madya* ranging over all the values between. Lexical items, in turn, would be numerically valued so that [+N] items are valued 0 and [+K] items are valued 1.[2] We could then calculate the numerical value of the speech level associated with a particular utterance by averaging the values of the lexical alternants used, which would in turn mean that *madya* level utterances would have different intermediate values depending on the proportion of *krama* to *ngoko* items that they employ (cf McCready 2019 for a similar approach to honorifics in Thai and Japanese).

The problem with this solution is that, according to W&P, different *krama* items do not "raise" the level of a given *madya* utterance to the same degree: "These forms are not all equal, and the choice of some of them implies a much higher type of Madyo than the choice of others" (W&P:36). In order to account for this variation in degree, we would need to assign different numerical values to different [+K] lexical items. But then we would lose the ability to model the categorical distinction between *krama* and *madya* speech levels and their associated utterances. Moreover, it would leave us with no explanation for *why* certain [+K] items raise the level of a *madya* utterance to a greater degree than others; since there are over 500 [+K] items in the language, this amounts to a very high amount of arbitrary continuous variation in the lexicon.

I will argue instead that this variation derives from differences in the particular paradigms that individual items are part of. Building primarily on the descriptive data in W&P 1982, I argue that paradigms of lexical alternatives can be sorted into ten subtypes on the basis of what assortment of feature values the competing items in the paradigm have. For example, one simple type of paradigm contains two alternants, whose feature values are [+K] and [+N], respectively. Both alternants are thus compatible with the *madya* speech level. In such a situation, I argue, the choice of the [+K] alternant suggests a "more *krama*-like" speech level than the [+N] alternant does, but *only* because there is competition. This contrasts with a minimally different type of paradigm (List Three in W&P 1982) containing two alternants whose feature specifications I analyze as [+K] and [−K,+N]. In this paradigm, only the [+K] alternant is compatible with the *madya* speech level, and thus no competition arises. The [+K] form is thus compatible with all degrees of the *madya* speech level, despite being featurally identical to the [+K] form in the other type of paradigm. In the full talk I discuss the full range of paradigm types, how competition applies in each one, and show how they correspond to the different "degrees" discussed by W&P.

References. Clynes, A. 1989. Speech Styles in Javanese and Balinese: A Comparative Study. Errington, J. 1985. Language and Social Change in Java. Poedjosoedarmo, S. 1969. Wordlist of Javanese Non-Ngoko Vocabularies. Robson, S. & S. Wibisono. 2002. Javanese English Dictionary. Uhlenbeck, E. M. 1970. The use of respect forms in Javanese. Wolff, J. & S. Poedjosoedarmo.

[2] [+K,+N] items (the *madya* style markers) might be given the intermediate value 0.5.

1982. Communicative Codes in Central Java.