(Automatic) Noun class assignment in Swahili

John Olstad (University of Wisconsin, Milwaukee)

This paper reports the results of a preliminary study that evaluates the systematicity of the Swahili noun class system using a probabilistic classifier model based on the semantic feature analysis proposed by Contini-Morava (1994). The findings show that this probabilistic model, a Naïve Bayes classifier, can do the following:

(1) Evaluate the systematicity of noun distribution over noun class.

(2) Detect prototypical exemplars of each noun class.

(3) Predict which classes are apt to be semantically productive.

Noun class assignment in Swahili is reflected in a complex agreement system. The following example shows the class 7 controller noun *kapu* 'basket', target adjective *kubwa* 'large', target numeral *moja* 'one' and target verb *languka* 'fell', all of which agree with the controller in noun class. The controller and targets take the class 7 prefix *ki-*.

ki-kapu	**ki**-kubwa	**ki**-moja	**ki**-languka
CL7-basket	CL7-large	CL7-one	CL7-fell

'One large basket fell.'

(Corbett 1991:43)

The context for which the present study contributes is a series of attempts to describe what determines the distribution of nouns into noun classes in Swahili and other Bantu languages. The problem is that Swahili has eleven classes, whose members are largely not predictable by generalizations. Earlier treatments have attempted to define noun class semantics by a set of common properties that all members must share (cf. Denny and Creider 1976, Zawawi 1979), which leads to a generalization with a large number of exceptions. Since they exhibit no further semantic coherence, exceptions are thought to be assigned arbitrarily (Corbett 1991:48). Contini-Morava (1994) argues that systematicity can be shown if we "broaden our conception of coherence" to include multiple criteria and the possibility of assigning members to classes that are more prototypical than others. The current study seeks to test this claim quantitatively, using a probabilistic classifier model.

The model is considered successful on (1) above because it is able to predict class based on the set of features significantly better than with frequency counts alone. It is considered successful on (2) above because it is able to output correct predictions at probabilities of .99 (prototypical) all the way down to .2 (non-prototypical) in a few cases. Finally, it is considered successful on (3) above because it predicts class 1/2, the 'human, animate' class, to be semantically productive. This is in line with the child language acquisition data presented in Demuth (2000) regarding several Bantu languages and the generally accepted idea that semantic productivity is limited to the animate class (Corbett, 1991:49).

The results are consistent with what has been shown about human categorization in studies like Rosch (1973), whose prototype theory predicts the successes of a model based on graded classes. Noun class assignment is able to "proceed well with high-but-not-perfect predictability" (Givón, 1986:98), which, according to Givón, is both akin and advantageous to biological organisms.

The predictions of the model lend support to the claim that the noun class system of Swahili is a hybrid of learned (memorized) class assignments and graded predictability based on semantic knowledge of prototypical exemplars.

Aikhenvald, A.Y. 2000. *Classifiers: A typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Ashton, E.O. 1944. *Swahili grammar*. London: Longman.

Contini-Morava, E. 1994. *Noun Classification in Swahili*. Publications of the Institute for Advanced Technology in the Humanities, University of Virginia. Research Reports, Second Series.

Corbett, Greville. 1991. *Gender*. Cambridge University Press.

Craig, Colette (ed.) 1986. *Noun classes and categorization*. Amsterdam: John Benjamins.

Denny, J. Peter and Chet Creider. 1976. "The semantics of noun classes in Proto-Bantu". *Studies in African Linguistics* 7.1:1-30. Reprinted in Craig (ed.), pp. 217-239.

Givón, Talmy. 1986. "Prototypes: Between Plato and Wittgenstein". In C. Craig (ed.), pp. 77-102.

Johnson, Frederick. 1939. *A standard Swahili-English dictionary*. Oxford University Press.

Rosch, E.H. 1973. "Natural categories", Cognitive Psychology 4, 328-350.

Zawawi, Sharifa. 1979. *Loan words and their effect on the classification of Swahili nominals*. Leiden: E.J. Brill.