

## **Noun Classes Grow on Trees: Noun Classification in the North-East Caucasus**

At first glance, noun classification in Tsez appears to be extremely complicated. While the classification of animate nouns is clearly semantically driven, the assignment of inanimate nouns, which appear in three of Tsez's four noun classes, is much less straightforward, particularly because members of several semantic groups (e.g. clothing and body parts) are distributed among classes II, III and IV.

To shed light on the operation of the Tsez noun classification system, we have analyzed a lexicon of approximately 3,500 Tsez words drawn from Khalilov 1999, as supplemented and confirmed by a native speaker consultant to ensure data accuracy and consistency of class assignment information. Using the Quinlan C4.5 decision tree algorithm, we tested a variety of formal features (including initial and final segments, initial and final syllables, and number of syllables) and a number of possible semantic features to identify which features are most predictive of class assignments. Based on these results, we have produced an analysis of the Tsez noun classification system that appeals to a small number of straightforward formal and semantic features of the type accessible to children acquiring the language.

Our analysis sheds light on the operation of the Tsez system and demonstrates that even apparently complicated noun classification systems, such as that of Tsez, can be explained without appealing to complicated semantics and concepts of semantic relatedness. It also suggests a promising path for the analysis of the classification systems of the other Nakh-Daghestanian languages, which we are currently pursuing.