Quantitative analysis of the linguistic individual

Guy Bailey (*Univ. of Texas Rio Grande Valley*)**,** Patricia Cukor-Avila (*Univ. of North Texas*)

Although a number of qualitative studies have provided significant insights into "the linguistic individual," the primary focus of quantitative sociolinguistics has typically been the group. The focus on groups rather than individuals is partly a consequence of methodological convenience (often the number of tokens from individuals is too small for definitive conclusions) and partly a result of theoretical considerations. Labov has argued that the proper object of study in quantitative sociolinguistics is the speech community rather than the individual because "linguistic analysis cannot recognize individual grammars or phonologies. Individual rules or constraints would have no interpretation and contribute nothing to acts of communication. In this sense, the individual does not exist as a linguistic object (2001: 34)." Recently, though, several quantitative studies have explored individual variation, arguing that no account of variation that ignores the individual can be adequate.

This paper reexamines the issue of the quantitative analysis of individual variation by exploring variation among 58 African Americans from Springville, Texas. The paper analyzes four features among these Springville residents: zero 3rd singular, zero copula, invariant habitual *be*, and quotative *be like*. The individual differences in the use of these features are remarkable. For instance, the frequency of zero 3rd singular ranges from 7.14% to 100%, while the frequency of zero copula ranges from 4.08% to 90.91%. Neither standard social categories (e.g., sex and age), nor identity-related factors (e.g., orientation toward Springville), nor interview context account for very much of the variation. The remarkably wide differences among individuals and the failure of explanatory factors to account for these differences would seem to confirm the need to analyze the individual as well as the group.

A closer examination of the data, however, suggests that this is not the case. Much of what appears to be individual variation in the use of these four features is actually just an artifact of the number of tokens per informant. All the outliers in the data reflect informants for whom there are small numbers of tokens, and as the number of tokens per informant increases, the more likely an informant's use of a feature will approximate norms for the group as a whole. Intra-individual variation further shows that small numbers of tokens lead to aberrant data. This study includes multiple interviews with more than half the informants. Interviews with small numbers of tokens tend to vary widely from the individual's norms just as informants for whom the number of tokens is small tend to vary significantly from group norms. Finally, our data suggests three general principles for sociolinguistic analysis. First, quantitative analyses of individuals require substantial amounts of data – far more than most studies have included and more than the analysis of groups. Second, the larger the number of constraints affecting a feature, the more data needed both for groups and individuals. Third, much of what has been called individual variation is almost certainly just an artifact of small amounts of data.