Vocabulary-Management Profiles as Unlabelled Tree Diagrams of Discourse

GILBERT YOUMANS

Most empirical studies of literature calculate statistics for entire texts or groups of texts. Statistics such as these reveal little about the constituent structure of discourse. Vocabulary-Management Profiles (VMPs), by contrast, provide direct visual analogues for constituent structure: major valleys on the VMP typically occur near discourse boundaries—near divisions between paragraphs, episodes, and the like—and major peaks normally occur near the middle of paragraphs, episodes, and so on. Furthermore, the deepest valleys on VMPs tend to correspond with the highest-level divisions within texts (such as the boundaries between the numbered sections of Faulkner’s story “A Rose for Emily”), whereas shallower valleys correspond to lower-level divisions (such as the boundaries between paragraphs or between topics within a paragraph).

The simplest binary distinction between words in a text is the contrast between new and repeated vocabulary, and this is the basis for constructing VMPs. First an interval, say thirty-five (or fifty-five words), is selected. Next, a computer counts the number of new words (types) introduced in the first interval (1-35), then in interval (2-36), and so on throughout the text. Finally, the number of new words in each interval is plotted at the midpoint of that interval. The result is a moving average, as it were, of the number of new vocabulary words introduced over thirty-five word intervals—the VMP for the text. When authors introduce a new topic the incidence of new vocabulary tends to increase, whereas when authors elaborate on an old topic, the incidence of new vocabulary tends to decrease. VMPs, then, tend to correlate with the “informativity” of a discourse.

Because VMPs indicate both grouping and relative prominence relations, they are formally equivalent to unlabelled tree diagrams of the type used by Lerdahl and Jackendoff in their analyses of musical structure. (L and J refer to their diagrams as time-span reduction trees.) The crucial question is, how closely do VMPs correlate with the cognitive structure—with the structure that an experienced reader would assign to the discourse?

The evidence so far is that VMPs correlate surprisingly well with authors’ segmentation of their own texts (into paragraphs and the like) and that they also succeed in highlighting quantitative verbal emphasis. Like unlabelled trees, however, VMPs cannot detect narrative transformations (for example, narrative flashbacks), nor can they detect anaphora (as in interlaced narration or allusions to earlier episodes). Finally, VMPs are blind to rhetorical understatement. For example, the most shocking (and therefore the most salient) event in “A Rose for Emily” is recounted in a single sentence at the end of the story: after a Miss Emily’s funeral, a strand of her gray hair is discovered on the pillow beside the skeleton of her long-dead lover. In this case, quantitative verbal emphasis would be not only unnecessary, but anticlimactic (given the strong cultural taboo against necrophilia). In such contexts, the discrepancy between the VMP and cognitive structure reflects a kind of rhetorical irony (deliberate understatement).

References
