Speech-gesture production in virtual or robotic agents is usually engineered, using fixed repositories and models that select, combine and adjust predefined behaviors. This gives control over the kind and quality of the producible behavior, but it is inherently limited and does not help to elucidate the nature of the underlying mechanisms in humans. I will present work on a computational production model that (1) is integrated in that it encompasses multimodal conceptualization, composition of verbal and gestural forms, and their realization as overt behavior; (2) is flexible in that it creates speech-gesture behavior on-the-spot based on communicative or cognitive constraints; (3) is cognitively and empirically grounded in that it rests upon empirical findings as well as cognitive modeling techniques. I will discuss how our model adopts and refines suggestions from theoretical accounts, and I will show how it reproduces human-like speech and gesture behavior.