Bad Data of the Third Kind

Paul Kiparsky
Stanford University

The distinction between theories and data is not absolute but relative to a specific level of analysis. For example, a grammar of a language is a theory, but the data it is based on are inevitably shaped by prior implicit or explicit theoretical assumptions. And to a typologist, syntactician, or historical linguist the grammar may in turn may be data for their higher-level theories, which may themselves then serve as data for psychological or biological theorizing.

It follows that bad theories at a higher level are bad data. They can obviously be incomplete, or, worse, false, to varying degrees. But third and rather insidious kind of badness arises from the division of labor between “descriptive” and “theoretical” work, typically when a theorist misunderstands the principles that govern the grammatical analysis and the level of abstraction of the representations that it posits. This happens in all branches of linguistics, but I will consider cases where theorists and typologists have mistaken abstract morphophonological representations for phonemic representations, or phonemic inventories as phonetic inventories, and the results have then unfortunately found their way into the theoretical literature.

Evans & Levinson (2009) claim that linguistic universals are “myths”, on the grounds that they are based on incomplete data, and that “an initially plausible pattern turns out not to be universal after all, once the range of induction is sufficiently extended.” The phonological side of their claim, so far not addressed in the critical responses, involves two cases. First, E&L reject Jakobson’s generalization that all languages have CV syllables on the basis of Breen & Pensalfini’s 1999 [B&P] “clear demonstration that Arrernte organizes its syllables around a VC(C) structure and does not permit consonantal onsets.” What B&P’s claim is that Arrernte syllables lack onsets in the underlying representations of the abstract morphophonological analysis they propose. About 25% of Arrernte words as actually pronounced begin with a consonant. B&P’s analysis posits that they have an underlying initial /e-/2/, which is then deleted. Their claim that all Arrernte syllables are closed is likewise about underlying representations; it is not true of words as pronounced. The same is true of the phonemic level of Arrernte. Under any assumptions it has CV syllables (Henderson & Dobson 1994: 23). Since the universality of CV syllables is a claim about phonemic or phonetic syllables, it is perfectly consistent with B&P’s analysis. I furthermore show that B&P’s arguments for their morphophonemic analysis are in any case flawed, and provide substantial positive evidence that Arrernte words conforms to Jakobson’s generalization at all levels of analysis.

Evans & Levinson (2009: 438) further claim that it is “contested” whether all spoken languages have vowel phonemes at all, citing Kabardian. The Kabardian vowel system is a good illustration of my point since the descriptive generalizations of Kabardian phonology are uncontroversial, but different phonological principles dictate wildly different vowel systems for it. If we require linearity (a phoneme cannot correspond to a sequence of sounds) Kabardian has seven vowel phonemes, as posited in much of the older literature and in the UPSID database of phonological typology. If
we don’t require linearity, but do require bi-uniqueness, it has three vowel phonemes; most phonologists adopt this analysis and propose a three-vowel system /a/, /u/, /a:/ (Choi 1991, Matasović 2006, Wood 1994, Gordon & Applebaum 2006, Applebaum & Gordon 2013). If we give up bi-uniqueness, it has two contrastive vowels (Halle 1970, Colarusso 1982). A reduction of the vowel inventory to zero requires notational artifices at odds with all phonological theories and has not found any followers since it was proposed by Kuipers 1960.

Kabardian’s vertical vowel system has also been cited by Dispersion theorists as confirmation for the theory’s prediction that in a system without a backness contrast all vowels will be central. But the representation of the Kabardian phonemes as central vowels is just an arbitrary convention, since all vowels of the language are affected by the rounding, backness, and height features of the neighboring consonants, producing a ten-vowel phonetic repertoire that fills the entire vowel space. In contexts where the vowels are not affected by consonants, it turns out that they are actually front, e.g. /ʔa/ [ʔæ] ‘hand’ (Colarusso 1982: 96, 1992:30). In this case, then, the graphic representations of the Kabardian vowel phonemes with central vowels have been taken as proxies for phonetic transcriptions. The actual phonetics turns out to be not an asset but a liability of Dispersion Theory.