2.2 Formulating Your Research Question: Some Theoretical Considerations

The first step towards constructing your stimulus set is to identify your specific research question. It could be a modest endeavor: perhaps you are investigating the number system in your field language and you have a singular/dual/paucal/plural distinction. You wish to know the semantics of the paucal term. Is it, in fact, a quadral (meaning precisely four)? Does it exclude two? What is the maximum quantity it includes? Where does the boundary between the paucal and plural lie? In order to tease apart these questions you could construct a stimulus set of pictures, depicting differing amounts of various objects and ask participants to describe them. You could then establish the boundaries of the categories, and establish whether there was agreement across speakers.

The research question could be grander in scale. Many classical investigations of how semantic domains are categorized by lexical resources—in studies of colour, emotion, and folk biology, for example—attempt to
p. 58 quantify the extent to which ↳ there are universally shared or culturally relative categories. If you were to investigate number from a cross-linguistic perspective, then a stimulus set might be built differently than if you were only to investigate the boundaries of a 'paucal' category within one language. For cross-linguistic comparison, you would want to first consult existing resources on what the possible distinctions across multiple languages from different linguistic stock might be, and then be sure to build those into your stimulus set (more on this later). In any case, the first step is to articulate the scope and aim of the investigation.

2.2.1 Do I start with form or function?

There are alternate emphases in formulating a research question. Investigations differ in whether their primary focus is on formal class (e.g. adpositions) or conceptual domain (e.g. space). One approach is to begin with a form-class, or rather a specific form, and track the possible functions it can have. An alternative approach, common to most anthropological and psychological investigations, begins with a conceptual domain and then tracks how that is 'carved up' in different languages. Both approaches can lead to valuable insights about semantic categorization. Nevertheless, it is important to be aware of the underlying assumptions and problems of both the 'bottom-up' approach of beginning with form and the 'top-down' approach of beginning with function.

2.2.2 Beginning 'bottom-up'

The advantage of beginning with form is that the phenomenon of interest is relatively easily delineated and identified—at least within a single language, or related languages. To take an example from Haspelmath (1997), one could conduct a study on the semantics of the Instrumental case in Slavic languages, where the modern day reflexes of the Proto-Slavic forms *-*mĭ*,*-*mi*, and so on, were studied. Or we could take a lexical domain and compare cognate terms, for example, the English *break*, German *brechen*, Dutch *breken*, Swedish *bräcka*, etc., all of which spring from the single Proto-Indo-European root **bhreg-* (Majid, Gulberg, et al. 2007).

The disadvantages of this approach become clear as soon as we try to scale up to include other languages. What should we count as equivalent forms (Croft 2001; Haspelmath 2007)? Obviously, as we move away from Slavic languages, we are not going to find suffixes that have the precise instantiation of *-*mĭ*, *-*mi*, etc. We could, in response, expand the scope and criteria of our investigation. We could simply study Instrumental case across languages. The problem with this is that the definition used to identify 'case' differs from researcher to researcher (see Haspelmath 2007). Even if we could agree on a way to identify the