2.5 Coding the data

The aim of coding your data is to be able to identify the referential range of the expressions that were used within the task. To do this, you must first transcribe the full responses of consultants, assuming you have recorded responses. You can then identify the constructional resources employed, how the semantic information you are interested in is distributed over the clause, etc. This can be a period to consider whether additional forms or constructions should also be part of the analysis. It is also a useful phase for weeding out responses which were not 'on target', by which I mean that the participant clearly had a different construal in mind than what the stimulus item was designed to elicit. For example, if when working with spatial ↳ stimuli the participant reverses the figure and ground relations ('the table is under the cup' rather than 'the cup is on the table'), then this is not a target response. Or if, in describing a colour chip in a colour-naming task, the participant describes the shape, this is not on target. These can be interesting responses in themselves but do not speak to the intended depicted event or object.

p. 69

In the process of collecting the data, consultants may have volunteered multiple responses, perhaps a first 'spontaneous' response, and then secondary responses. You may also have additional judgements about which other forms could be used, and whether some forms are not applicable. In the process of coding your data, it is important and worthwhile to tag your responses for whether or not they are the first response; whether they are free descriptions or reflective comments; and so on. In the analysis stage, you may want to consider separately each type of response.

2.6 Analysing the data

In order to explore the extensional semantics of a form (word, morpheme, construction), you will need to identify all of the stimuli that were described by that particular form. One simple way to do this is to code your data in a spreadsheet. You can create a column which has an ID for each of your stimuli, and alongside this insert your linguistic responses in the next column. A separate column for each speaker is helpful. You can then use the 'sort' function which will enable you to identify all the stimuli that were grouped together with the same term.

Standardized coding and eyeballing your data will get you quite far in figuring out the sorts of situations that call for descriptions of particular types but there are tools to go beyond this. There are a number of multivariate statistical techniques, such as multidimensional scaling, correspondence analysis, and cluster analysis, to name a few, which can be implemented in ordinary statistical packages, such as SPSS, SAS/STAT, or the open source freeware package R. These techniques are becoming increasingly important in the field of cross-linguistic semantic studies (e.g. Croft and Poole 2008; Majid et al. 2008). A comprehensive overview of these techniques is beyond the scope of this chapter, but for a gentle introduction see Grimm and Yarnold (1995).