Moreover, while I have stressed the importance of negative data, elicitation and grammaticality judgements, it is possible to do a great deal of work and get a great deal of data from positive materials alone.

7.7.3 Working in the field language

During your field trip it may become desirable or necessary for you to work in your field language. I have not included information on monolingual elicitation here – see Everett (2001) for more information. You may wish to work with monolingual as well as bilingual consultants, though. It's possible to do this even without fluency in language, and I recommend it highly as a way of helping you to increase your fluency quickly.

Your early aim in working this way should be simply to get people talking. You'll be able to go through what they said and transcribe with bilingual consultants later on. If you have some ability to respond in the language, you will probably get fairly simple structures, as people will use foreigner talk to you.

Some techniques you can use include asking very general open-ended questions as a way of encouraging discussion. You can also present stimulus materials. Asking for vernacular definitions is also very easy to do with monolingual consultants.

7.7.4 Further reading

- **How to set out an argument**: Harris (2000) is an excellent example; see also Green and Morgan (1996).
- **Auxiliaries:** Anderson (2005).
- **Semantics**: Evans and Sasse (2005), Matthewson (2004, 2005).
- **Quantifiers:** Enç (1991).
- **Experimental design**: Crain and Thornton (1998: chs 16, 17, 24).
- **Various morphosyntactic categories**: Aikhenvald (2004), Chafe and Nichols (1986), Comrie (1976, 1985), Dixon and Aikhenvald (2000), Dixon (1994), Lyons (1999), Palmer (1986).
- **Syntax**: Bouquiaux and Thomas (1992), Levinson and Wilkins (2006), Payne (1997), Shopen (1985), Thomas (1975).
- **Word order and (non)configurationality:** Austin and Bresnan (1996), Jelinek (1984), Kiss (1998), Mithun (1992, 2003).
- **Parts of speech:** Schachter (1985).

8
Lexical and Semantic Data

In the previous chapters we have discussed sentences, morphology and grammar. However, we should also talk about the documentation of the lexicon of a language. Lexical documentation can be something done in conjunction with other work on the language, however dictionary making is also an extensive enterprise in its own right. You could just wait for lexical items to appear in the course of your other work (e.g., in text collection); however, the returns on this method diminish rapidly as you gain familiarity with the language (i.e., the number of new words in texts drop off rapidly).

8.1 Getting vocabulary

One way to collect words is just to ask for them. We discussed basic lexical elicitation in §3.1.3. Those methods can be extended to more detailed and less common vocabulary within various semantic fields. Collecting antonyms, converses and hyponyms at the same time as a lexical item can be useful. For example, when you ask for 'little', ask for its opposite at the same time. (Don't just ask for the English opposite, in this case 'big', ask for the 'opposite'.[1]) You may also want to branch out early and ask for 'very little', 'littler' and 'littlest'.

I've always found it easier to do vocabulary elicitation in small groups (three or four consultants). Consultants prompt each other and the arguments about definitions are usually interesting and good sources of conversational data. In a field-methods class this probably will not be feasible.

transcribing and make a note of any new words. A technique that I have found useful is to ask your consultants to pretend that they are speaking to a Martian who crashed their spaceship near the community and who wants to find out what different things are called in the language. This Martian doesn't know anything, so needs to know what absolutely everything is called. Phrasing the task like this provides a bit of light relief from your endless questioning.

8.2 Lexicon compilation

Even if compiling a dictionary is not your main intention, keeping a lexicon database file is very useful. The list of words will help you study phonology, including variation. A wordlist is an excellent source for interlinear glossing, and it is also a useful item to return to the community at the end of your field trip if a dictionary has not yet been compiled. You can compile a wordlist using just about any format – card files, Word document or Excel spreadsheet. But a database is easier than a flat file because you can store information hierarchically and it's easier to code and search.

Here is a list of information that it is useful to have in a dictionary or wordlist. However, even a headword and gloss will be useful. Further information about dictionary compilation is given in §14.8.

- the **headword** sorted by semantic field or alphabetical order – many database programs allow for variable sorting
- **parts of speech** (be very careful about creating parts of speech labels on the basis of the gloss of the word. This is very misleading.)
- phonological **irregularities**; **pronunciation** if unpredictable from the standard orthography
- a single-word **gloss** for interlinearization
- a more detailed **definition**
- morphological **paradigmatic** information, such as gender, class or conjugation.
- any notes on comments your consultant made about the semantics of an item
- **encyclopaedic** information, e.g., information on an item's usage or ethnographic information about the cultural importance of the item. This could be accompanied by a picture
- synonyms and antonyms, hyponyms or other information about how the word relates to other items in the lexicon

- the source of the word (e.g., if a borrowing; etymology, if known)[2]
- **semantic field(s)** of the item
- any **usage** information – e.g., if it is slang or taboo
- a **reversal** field (so that you can compile an English–Language finderlist from your data)
- **sound clip(s)**, and example sentences
- derived words
- the source of the information (e.g., who told you the word)
- questions for further research.

8.3 Specific domains for lexical elicitation

Here is some discussion of some particular lexical domains for elicitation, based on the list of semantic fields available on the web site. Note that semantic domains are highly specific to region, especially for flora and fauna, so an out-of-area list may need adaptation.

8.3.1 Body parts and products

External body part terminology can be elicited by pointing to the particular body part. For internal organs and body products, you may need to use terms in the contact language or anatomical diagrams. In some cultures it is not appropriate to show pictures of naked bodies (or anatomical diagrams), so find out in advance if this is all right. It may also be frowned on to show pictures of naked members of the opposite sex.

Try to get baby terms and slang terms for body parts as well as the regular terms. Possessive marking can be elicited at the same time as you do body part work. If the language has inalienable possession, this is a place you are highly likely to find it.

It might be easier in some cases to get detailed anatomical information about animals rather than humans. For example, hunter-gatherer groups often have very detailed knowledge of (and terminology for) the anatomy of the animals they hunt, but developed human information might be less detailed.

8.3.2 Artefacts and everyday items

The best way to elicit information about artefacts is in procedural texts (which may be videoed). Get your consultant to explain the use of the item, how it is made, and who uses it. You may need to consult a specialist within the community. For example, most English speakers don't

guild. Your average English speaker is highly unlikely to be able to give you an accurate definition of a warp thread.

Ask about everyday items as well as the exotic. Don't just concentrate on traditional items. If you are trying to learn to speak the language, you will also need to know the words for everyday items.

8.3.3 Flora and fauna

Books with pictures or illustrations of local flora and fauna are useful elicitation prompts. People might not always recognize the item from the picture, though, or they might be misled by the colours in the picture. Collaboration with trained botanists, and so on, will be necessary to make sure identifications are correct, but you can do some on your own. Glosses of the type 'kind of tree' are better than no word at all, but more information is better still!

Flora and fauna can be a great source of simple language elicitation. I have used both to elicit vernacular definitions (see Casagrande and Hale 1964). The set of questions can be used to turn just about any single word into a little story. They are particularly useful for flora and fauna but can be adapted to most items:

- Tell me all about X:
- What does it look like?
- Where is it found?
- What does it eat (and what eats it)?
- How is it different from things that look like it?
- What's it used for?
- If it's poisonous, what happens to you if you eat it?

The advantage of this type of data collection is that it involves preserving cultural knowledge at the same time as getting general language data. The texts make nice little story books too. Each book, or page, can be illustrated by photos (or children's drawings), or they can be made into short video clips.

You might need to gloss species through a third language. The species might not be named in the contact language. English is pretty poor in names for species outside Europe (and even then many people don't know those words). Not every species will necessarily have a name, and often multiple species will have the same name. Don't expect the names to match up to English names or to Linnaean taxonomy. Multiple species might have the same name, or the same species might be called by different names at different stages of its life cycle (e.g., English *goose*, *gander* and *gosling*). For further information, see Cotton (1996) and Berlin (1992), amongst others.

8.3.4 Place names

One area of the lexicon that doesn't often receive much attention is place names. In the Western lexicographical tradition, place names are left out of dictionaries, confined to specialized dictionaries or omitted altogether from linguistic research. Place names should be included in a documentation. They are useful for historical linguists, they often have different syntax from other items, and they are a wonderful cultural resource and a prompt for storytelling and deixis elicitation.

To do site mapping, you'll need a GPS and the best maps of the area you can find. Getting detailed maps can be difficult; often 1:100,000 scale maps can be the best you can find, and they are too small for accurate site mapping (although they can be enlarged on a colour photocopier and you can add detail by hand). Aerial photographs or satellite images may give you more information and may be more useful if good maps of the area are not available.

If your consultant isn't very mobile (or if you are working a long way from their traditional lands), it is possible to elicit place names by asking your consultant to give directions, telling you how to get from point A to point B and what you'll pass on the way. Or you can ask them where their favourite places are for different activities. Unless your consultant is absolutely amazing, you'll get omissions and mistakes (to see how hard this is, try it yourself – try to list all the streets you pass between your home and your university).

It's much better to take a small group of people to the area and get the data *in situ*. Try not to be the driver though; you'll need to write (recording is almost impossible unless you're in a quiet car on sealed roads and even then if you're driving you'll have enough to think about).

8.3.5 Kinship terminology

It's useful to master kinship terminology in your field language. You can observe what people call each other, how kinship affects interaction, and can use them as a prompt for kin-related language (e.g., wedding ceremonies).

Eliciting kin terms can be quite confusing. It's best to use real family situations – your consultants will probably be able to do complex kin calculations in their heads, but you'll get confused without practice.[3] Try to establish the range of each term. Who can be called *son* or *daughter*? Who can be called 'granny'? Elicit information from different people

and compare the results. The scope of kin terms can be very difficult to define, for example, when eliciting from English, as the English kinship system is impoverished compared to many standards. Draw family trees. Ask different members of the same family. Reckoning may be different for men and women (e.g., for languages which classify children according to the sex of the parent, not the sex of the child; cf. Bardi *aala* 'man's child' and *bo* 'woman's child', not 'son' versus 'daughter').

8.3.6 Other domains

There are many other domains for lexical exploration. A few others that are useful or often overlooked are:

- Occupations (and social structures more generally)
- Value judgements – how to talk about evaluation
- Sounds, textures and the like
- Religious or other ceremonial terminology
- Musical terms; other specialized knowledge
- Abstract concepts (ethnophilosophy, moral reasoning)
- Mental vocabulary (thinking, forgetting, etc.)
- Mathematical concepts

8.4 Frequent lexicographic pitfalls

As mentioned in §6.4.1, it is very difficult to know whether the inferences that you have drawn about the meaning of a word are the same as your consultants' inferences. Discovering what a word or sentence means is not a trivial task! Here are some commonly encountered problems specific to lexicographic work.

8.4.1 Preliminaries

Multilingual consultants may give you a word in a different language. This doesn't mean that they don't keep the languages separate when they are speaking naturally, but it is quite difficult to do so when asked point-blank for vocabulary. Factual errors also occur sometimes – that is, your consultant may simply give you the wrong word. You've probably been in the situation where someone has asked what you call something, and you might not be able to remember (although probably came to you later on). If this happens in fieldwork and your consultant can't remember the item, don't worry, go on to the next item. They will probably remember later on and tell you. (See also the comments on translation in §6.4.1.)

8.4.2 Polysemy and homophony

Don't ever assume you've got a complete description of a word's meanings. For example, if someone tells you that Eastern Armenian *tʰert* means 'sheet of paper', don't assume that this is the only meaning of the word (in this case, the word also means 'newspaper', but this was not the context in which the word was originally elicited). Sometimes a consultant will volunteer multiple meanings. It can be worth asking if a word has any other meanings, but the answer to this question is not necessarily reliable.

Sometimes consultants will say that two homonyms sound different. If you ask if two words are the 'same' (with the intention of discovering if they are homonyms or a minimal pair), the answer tends to be 'no' if the words have very different meanings, whether or not they are homonyms. A less ambiguous question to ask is 'are these the same word, or are they different words which happen to sound the same?' – that will elicit the consultant's feelings about polysemy versus homonymy, and will reveal if the words are not in fact homonymous and you have made a transcription error.

8.4.3 Descriptions instead of definitions

Consultants will often give a description of when a word might be used, which is not the same as the meaning of the word itself. Consider the following Bardi word:

(20) *manbin*
'soft rain'
'dry season rain'

In Aklif (1999), *manbin* is given as 'soft rain'. In fact, a more accurate meaning is 'dry season rain' (i.e., rain that falls in the dry season); such rain is usually light (in contrast to the heavy wet season storms), so the definition is not incorrect, but it is not the core meaning of the word; after all, light rain in the wet season is not *manbin*. A further example from Yan-nhaŋu is given below:

(21) *balgurryu*
'waistband'
'from string'

Here the *real* meaning of this word is 'from string'. It is the word *balgurr* 'string' in the instrumental case. A waistband is an example of

something that is made from string. In other cases, the word may be both a descriptive item and have a more abstract meaning. For example, in Yan-nhaŋu *maŋutji-bu* means both 'something associated with eyes' and, specifically, 'glasses'.

Be on the lookout for times when your consultant says 'it's like when ...'. That is a good indication that the information you are getting is a description rather than a definition. (Of course, descriptions can be useful information too – and if they are given in the target language, they are an excellent sources of similes, short sentences or examples with the semblative case (if there is one) .)

8.4.4 Discourse context

We tend of think that elicitation is 'discourse free', in that there is not the discourse context that is created in narratives or conversational data. However, elicitation can still create its own discourse contexts, and this can interfere with your data. For example, if you use the same participants for each sentence, this greatly increases the likelihood that one or other of them will be identified as a topic (and therefore marked for topichood, by word order, intonation or morphology). One way to try to prevent this is not to have the same participants in two sentences in a row. This increases the artificiality of the elicitation, but it avoids potentially confounding variables like topic tracking.

8.5 Further reading

- **Kinship**: Holy (1996), McKinney (2000: ch. 14).
- **Lexicography**: Mosel (2004), Nichols and Sprouse (2003).
- **Lexical semantics and fieldwork**: Evans and Sasse (2005), Hellwig (2006).
- **Other types of documentation**: Austin (2005), Barwick (2005), Barz and Cooley (1997), Post (2004), Toelken (1996).
- **Ethnobotany**: Berlin (1992), Cotton (1996).
- **Indigenous knowledge systems**: Bicker, Sillitoe and Pottier (2004), Sillitoe, Dixon and Barr (2005).

9
Discourse, Pragmatics and Narrative Data

Field linguists are often told to rely on naturalistic data as much as elicited data. By this we usually think of recorded narratives. However, a comprehensive description of a language should be built not just on elicited data, but also other types of spontaneously produced speech. Narratives and discourse data are covered in this chapter.

9.1 Working with texts

Elicitation will allow you to make a lot of progress, but it will bias your data towards the constructions you chose to ask about and ones that are easy to translate. Also you can't do anything with frequency working from elicited data because there frequency is determined entirely by what you ask. Therefore, you should also use spontaneously generated data.[1]

9.1.1 Text genres and register

It is usually easy to get people to tell stories, but there are many other types of discourse and other genres and registers in languages too. If the language you are working on is regularly used, you should have no difficulty obtaining samples of different registers, genres, and types of interaction. You might be able to record a religious service (Christian sermons are often wonderful data sources for similes, imagery and prohibitions). Is there community radio or TV? You might be able to get tapes and transcripts.

Don't neglect written sources if the language you are working on is regularly written. Sometimes newspapers will have sections in regional languages (e.g., *The Namibian* has a section in Oshiwambo each week).