CHAPTER 7. SYNTAX AND SEMANTIC FIELD RESEARCH

7.1. Introduction

Syntactic fieldwork, like all the other structural domains we have considered until now is very complex and, again, exceeds the capacities of a single chapter. In this chapter I follow the approach adopted in the previous chapter on morphology and present *types of phenomena* that should be accounted for by any fieldworker. I present these mainly in terms of communicative functions and the formal realizations of these functions in the syntax.

7.1.1. What is syntax

Let's begin with a definition of syntax. I understand syntax roughly as the set of restrictions on how words are put together into larger units, the structure of those units, and the way that these units present and structure information.

What should the linguist give priority to among the potential areas of syntactic investigation that emerge from this definition? Whatever they discuss about the syntax and other aspects of the language under study, there will be large, disparate audiences ready to pore over what has been said, looking for data related to themes they are interested in, which may not overlap with the fieldworker's original interests in the description. Of course, no linguist can satisfy everyone. But there is no getting away from a certain responsibility that field workers incur to make their results useful and clear to as wide an audience as possible. Therefore, the fieldworker, to repeat a recurrent theme of this book, must endeavor to read widely and inform themselves with a systematic reading program.[41]

In the first part of this chapter, I focus primarily on questions of structure, though I attempt to give the reader a taste for all the other approaches just mentioned. This reflects my own training, personal history, and continuing interests, just as any scientist's decisions are affected by their background. In the next section of the chapter I discuss the syntactic hierarchy and various assumptions that are widespread among linguists, some of which need to be rethought, at least by the fieldworker.

The final section of this chapter addresses the role and method of argumentation in linguistic description. All too often in my experience field linguists fail to recognize that even descriptive accounts of linguistic items must argue for all their conclusions. Therefore, I consider the components of good descriptive argumentation.

We begin with a discussion of syntactic field methodology, move to semantic field methodology, and move, finally, to a discussion of argumentation. Before beginning syntax proper, let us consider some additional relevant considerations for syntactic field research.

7.1.2. The fieldworker is not a blank slate

Occasionally one hears fieldworkers advocate – or be advised to produce – so-called theory-neutral descriptions, in order to serve as wide an audience as possible (and so as not to be overly biased at the outset of the fieldwork). Insofar as this means avoiding theory-specific jargon when writing for a general audience, it is reasonable advice. But it is otherwise naive. Like newborns, fieldworkers are not blank slates. Moreover, they should not be. The purpose of training is not to make us 'open-minded'

---

[41] I suggest to my graduate students the program I maintained for a good part of my career, namely, reading 50 pages per day, minimum, seven days a week, ranging across topics e.g. phonology, syntax, semantics, descriptive linguistics, etc.

but to make us keenly observant, able to recognize from among subtle alternatives the best-motivated analysis. To maintain the 'observer's edge', we need to be slightly biased.

The fieldworker cannot avoid learning and interacting with at least some linguistic theory. This is because all of us accumulate a perspective on how language works during our linguistic training and experience, either from our teachers, our reading, or our natural predispositions coupled with our experiences analysing linguistic data. And theories should not be avoided. They provide a framework of expectations (a source of the 'abductive surprise' discussed in ___ below) and a source of ideas (the 'abductive 'B's/hypotheses' also discussed in ___). And theories provide strategies and boundaries for arguing for conclusions. They are essential training for thinking more precisely. The choice of a specific theory is less important than the recognition that some theory must be selected. However, other things being equal, it is better to choose as your theoretical base a theory that has a track-record of crosslinguistic usefulness (which is why I suggest that fieldworkers familiarize themselves with ROLE AND REFERENCE GRAMMAR (VanValin and LaPolla (1997) in addition to any other theory they might choose to work with).

Theories can be misused of course. They are misused when a linguist fails to consider analyses or overlooks facts that have no obvious 'niche' in his or her own theory. They are also abused when the linguist simply *translates* their analysis into theoretical terminology, i.e. when the theory is only used 'for show' and is never causally implicated in the analysis proper, an all too common occurrence. Syntactic fieldwork always confronts the fieldworker with structures unlike any the have confronted previously. Often a structure that is 'weird' and apparently irrelevant for a particular theory's objectives will be ignored. But the same structure could turn out to be very important for linguistics more generally. This is why a fieldworker must not only familiarize him or herself with a particular theory, but also read widely, beyond the boundaries of that theory.

Ultimately, the fieldworker should strive to simply give an honest day's work for an honest day's pay. That means that they do their best to present the language (the aspect(s) of it they are concerned with at any rate) clearly and fully (never intentionally sweeping facts under the rug which appear to bear on analysis at the center of the fieldworker's description). In return they receive career-related emoluments. To do this work, they certain basic concepts about syntactic field research, to which we now turn.

7.1.3. Crucial syntactic concepts for fieldwork

Probably ever syntactic theory recognizes that the syntax of most, if not all, languages will be composed of constituents of different sizes (though see ___ below on recursion for a discussion of ways syntacticians differ). Almost all linguists agree on the following, in one way or another:

(7.1) Syntactic hierarchy
- a. words
- b. phrases
- (c. clauses)
- d. sentences
- (e. paragraphs)
- (f. discourses)
- (g. conversations)

Some theories, not all, recognize clauses, paragraphs, and conversations (as indicated by the parentheses). All linguists recognize that discourses and conversations exist, but not all theories have a place for them (most lack any well-grounded treatment of conversations). But the items in () not in parentheses are universally recognized. So at a minimum the syntactic description of any language should include a discussion of each of these constituent types, how they are put together, how they relate to one another, how they are like one another, and how they differ from one another. In addition, some linguists believe that there is a correct order of analysis, such that one should begin with the largest constituent, e.g. the conversation (but this is theory-dependent), and then work one's way down the hierarchy, one layer or level at a time.

Each constituent is built out of or 'manifested by' constituents at the next level down, though the precise way in which one constituent is built out of another varies. The standard article on the building or analysis of syntactic constituents is Wells (), still worth reading today. The methodological implications of this hierarchy are discussed in __ below.

Note however, that there are huge differences in the cohesiveness of syntactic units as we proceed 'up' the hierarchy. This is important because it ultimately entails different methodologies and forms of analysis and argumentation for different levels. It is also important theoretically, because it could indicate that levels of this hierarchy have different theoretical status (consider, for example, Chomskyan theory's refusal to consider anything above the sentence as syntax proper. This is not an entirely unmotivated position to take, as Everett () argues in depth. But, methodologically at least, discourses and sentences must be studied in relation to one another.) A conversation can vary tremendously in the units that compose each of its exchanges and these could in principle be rearranged in a number of ways without loss of intelligibility. At the other extreme, the word level, constituents are much more rigidly ordered (but see Bickel, et. al. (2005) for a potential counterexample). Moreover, speakers' varying ability in constructing these units becomes at once more obvious and more crucial as the units get larger. For example, it makes little sense to ask 'Who forms words best around here?' But it does make sense, and the fieldworker should ask, 'Who tells stories best around here?' before deciding on a language teacher for (certain kinds of) discourse studies and text collection.

Therefore, each level of the syntactic hierarchy is associated with a specific methodology. I want to begin the discussion of syntax proper by a review of fundamental notions of syntax shared by most linguists.

7.2. Dominance and precedence (or Configurationality and linearity)

Two fundamental ways in which discourses and conversations differ from constituents lower on the hierarchy are the relevant restrictions of configurationality and linearity (or dominance and precedence) at each level. Consider the following:

(7.2) Words
- a. running
- b. * ingrun

Word-internal constituents, in general, perhaps universally, cannot be moved. So placing *–ing* at the beginning of the word, as in (b) produces severe ungrammaticality. Moreover words allow a range of configurational relationships, ranging from binary

branching (as in the English examples above) to the periphrastic constructions discussed in chapter ___. Phrases, on the other hand, have slightly less rigid linearity constraints:

(7.3) Phrases
a. John's book.
b. *book John's
c. book of John
d. *of John book
e. book of John's

Phrase-internal order is also fairly rigid (and (e) shows that *of* and *–s* are not simply 'allomorphs'. Yet, as (a), (c), and (e) show, it does allow some flexibility. Clauses and sentences show even lesser constraints on precedence and, in some theories, greater freedom in configurational relationships. So, for example, in some languages prepositional phrases are only binary branching, with only one level of branching, though clauses in the same language may have various levels of branching, such as when prepositions allow no modifiers, only a head and a complement, whereas clauses are complex. (Yarawara, Dixon (2005) is one such example.)

(7.4) Clauses and sentences[42]

a. John ran to the store.
b. ?Ran John to the store.
c. To the store ran John.
d. To the store John ran.
e. *To the ran John store.
f. *To the ran store John.
g. Yesterday, John ran to the store.
h. John ran to the store yesterday.
i. John ran, yesterday, to the store. etc.

And discourse orderings are freer yet:

(7.5) Discourse
a. John came in. Peter asked him if he wanted a drink. John sat down later.
b. %Peter asked him if he wanted a drink. John sat down later. John came in.
c. %John sat down later. John came in. Peter asked him if he wanted a drink.
d. John was writing. Peter was writing. Susy was writing.
e. Peter was writing. John was writing. Susy was writing.
f. Susy was writing. Peter was writing. John was writing.

---

[42] Morphological constraints can supplant syntactic restrictions. So in Latin (and other languages with rich morphology) word order is relatively free, because morphology is so consistent:

LATIN EXAMPLE

Discourse orderings are interestingly and significantly different from clause orderings because they are constrained by very different principles. For example, (b) and (c) are anomalous (%) because in the way we normally understand the discourse, Peter cannot ask John anything until he comes in. And John can't sit down until he comes in. But now imagine that John is just outside the window near a lawn chair. Peter asks him, through the window if he wants a drink. John sits down after that, thinking he is going to be served outside. He finds out he needs to come in for his drink, so he does. This little scenario renders (b) just fine. And a bit of rethinking makes (c) similarly fine. So the % only indicates that the order is strange in some contexts, but not in all. And yet the ungrammatical examples of words, phrases, and clauses seems unsalvagable by context. They are always bad. Conversational linearity is constrained by yet different principles:

(7.6) Conversation

a. A Hi. Wanna beer?
   B Yeah, sure.

b. % A Yeah sure.
   B Wanna beer.

c. A So, what were you up to yesterday?
   B Oh, about 5'8". Ha ha. Oh, not much really.

d. % A Oh, about 5'8". Ha Ha. Oh, not much really.
   B So, what were you up to yesterday?

Generally, conversations are structured in pairs of utterances, at a minimum, exchanges between different conversational participants. There are ways to think of the %'d examples such that they become more or less acceptable, however, as with the discourse examples. Although discourse does not share the 'interlocutor exchange' structure with conversation, it does share with conversation the fact that context and imagination can render otherwise (context-free) anomalous utterances or exchanges sensible and acceptable.

Therefore, at each level of the syntactic hierarchy, the linearity principles differ. Also, the overall structure, the configurationality of the constituents at different levels of the hierarchy is achieved by different principles (one structures a novel differently from a noun phrase, for example).

No one theory of syntax offers a comprehensive treatment of the syntactic hierarchy (indeed, as indicated by the parentheses in (), not all theories even recognize all the constituents posited). Tagmemics (Pike & Pike (1976)) attempted to do so, but the results were hard to follow at best and unconvincing and confusing to most linguists. Therefore, the fieldworker will have to supplement his or her chosen theory of expertise with additional, eclectic readings on the other constituents, rather than finding all the answers in any one study, book, or theory.

We now want to consider another notion that is basic for many linguists.

7.3. Basic word order

Linguists often talk about the 'basic word order' of a language. My guess is that most linguists that use this phrase know that it is fundamentally imprecise. Each word of the phrase is problematic. So, for example, 'basic' has no scientific significance, i.e. it is not a technical term of linguistics. The term 'word' is also imprecise because what

most linguists mean here is 'immediate constituents of the sentence or clause', rather than words, which are usually themselves not immediate constituents of the sentence or clause. And, finally, 'order' is not very useful either since to many (but by no means all) linguists what is crucial is the hierarchical structure of phrases, not their linearity.

Given this imprecision, it is not surprising that what linguists actually mean by 'basic word order' includes several concepts, not all clearly related. At least the following are included in one way or another:

(7.7) a. Information-theoretic default constituent ordering.
b. Discourse-functional constituent ordering.
c. Initial/underlying constituent configuration – the hypothesized configuration that serves as the basis for non-default orderings.

Consider (a). In most or all languages there is a default information-theoretic ordering of constituents (see, for example, Van Valin (2005) and Lambrecht (1994)) and __ below. English usually reserves the subject position for old information and the direct object position for new information. In fact, many languages (Van Valin and LaPolla (1997--)) place new information immediately to the left or right of the verb and the old or topical information farther out or on the other side of the verb. To get a feel for how information structure works, consider the familiar examples in (), from Van Valin (), in ___ below.

So for a given language, it is important that the fieldworker understand (by examining constituent orderings and intonation), among other things, how topic and focus are distinguished (see ___ below). To some people, this will be the core meaning of 'basic word order'. But there is complementary concept that any fieldworker could profitably consider, namely, discourse function of different constituent orders, not so directly tied to information structure.

Discourse functions of sentences involve things such as the use of passive clauses to mark backgrounded information or information with some other special purpose in structuring the discourse, or the use of one constituent ordering to open a discourse but a different ordering to mark denouement. Or subject-less or object-less sentences to mark the discourse body but use of full NP subjects or objects to indicate the beginning of end of a discourse. If such discourse notions are found to be useful in a given language, they may very well differ practically from the other concepts in (). If such different discourse functions exist, they may be discovered in one of several ways. First, there is no available excellent software for discourse filing (see the websites mentioned in ___). Second, there are some still useful manual methods. One method I used to employ was to color code sentence types throughout texts (e.g. red for active, blue for passive, etc., depending on the types of sentences you are interested in tracking) and then compare discourse genres and discourse functions or positions (e.g. beginning, climax, denouement, end) to see if there were patterns to the colors. This is a psychedelic experience for those of my generation while also being a useful field methodology.

Let's proceed in our discussion of major syntax notions to the concept of dislocation.

7.4. Dislocation

Most formal theories, to one degree or another, concern themselves with the phenomenon of 'dislocation', i.e. the situation where related meanings are expressed by different word orders or where some constituents are not where they might be in a simple declarative. These theories often take one of the orders as underlying or default or basic and derive the other orders from this. So consider examples like those in ()-():

(7.8) a. John saw James.
b. James was seen by John.

(7.9) a. The mail is here.
b. Is the mail here?

(7.10) a. John likes Mary.
b. Who does John like?
c. John likes who(m)?

(7.11) a. John thinks you like Mary.
b. Who does John think you like?
c. Who does John think likes Mary?

(7.12) a. John thinks that you like Mary.
b. Who does John think that you like?
c. *Who does John think that likes Mary?

In most theories, the (7.8a)-(7.12a) examples represent the basic forms and the other forms the derived forms. The question words in () –() are said to be 'dislocated', i.e. not in their 'expected' positions (where they would be in a simple declarative). Harris (1947) proposed that such sentence alternations should be understood in terms of specific discourse functions for each alternate word order (which he analysed as 'transformations', but without a sense that one was derived from the other). Formal theories have more elaborate means of accounting for dislocation. But whether one works in a formal theory or not, the fieldworker must come to grips with Harris's observations, because they are vital to understanding the grammar of the language in question. It is perhaps always the case that different word orders have different discourse functions. And describing those functions is part of an adequate descriptive grammar or understanding of a language's syntax. Whatever theoretical account of such facts the fieldworker wishes to suggest is welcome, though many (myself included) would see this as secondary to finding out the communicative functions of the alternations or so-called dislocations.[43]

Also, it is important to note that dislocation, while a prominent part of the syntax of many languages, is insignificant in many others (except in very abstract, theory-internal ways that are of little use to the average fieldworker, at least in my experience).

[43] Urban () contains a number of suggestive analyses illustrating potential cultural implications of discourse structures, e.g. the use of the passive vs. the active and the types of heroes that emerge, say, in cultures with high percentages of passive sentences for major characters, vs. those in which active sentences express the actions of main characters. .

Therefore, if the fieldworker goes to the field primarily trained in a theory with a strong emphasis on dislocation, to the exclusion of, say, information structure, then, however interesting and important that theory may otherwise be, they are likely to find themselves relatively 'unarmed' in the face of facts which are complicated, but where dislocation is not part of the complexity.

Before proceeding to a consideration of discourse analysis and syntactic elicitation methodology, I want to consider another notion that is widely assumed by many syntacticians to be universal, but which also seems to have exceptions. Such exceptions, if valid, require that the concept not be used as an obligatory mold for structures in the grammar.

7.5. Tree-structure

7.5.1. Endocentricity

Another widespread assumption about syntactic structure, built into the foundational assumptions of many formal theories, is that all syntactic structures are *endocentric.* As Hudson (2002, 5) says: "Almost all modern theories of syntax accept **endocentricity**: every phrase has a single head which determines the characteristics of the entire phrase. " In this section I want to give reasons why the fieldworker should *not* assume that all structures are endocentric. So consider the following sentences from Wari':

(7.13) [$_{S}$ [$_{Nuc(leus)}$ [$_{S}$ [$_{Nuc}$ **Ma' co mao**]**na -in$_{i}$**
that:prox:hearer m/frp/p go(sg) 3s:rp/p -3n
Guajará$_{i}$]]
Guajará (Brazilian city)

na$_{j}$ -nam$_{k}$ 'oro narima$_{k}$' taramaxicon$_{j}$].
3s:rp/p -3pf collective woman chief
'"Who went to Guajará?" (said) the chief to the women.'

(7.14) [$_{S}$ [$_{Nuc}$ [$_{S}$ **Cao' xi' carawa**]] **nana hwijima'**].
eat 1pincl:rf animal 3p:rp/p children
'The children will eat food.' (lit: '"We will eat food," the children (say).')

As Everett (2007) argues, these sentences have a predicate that is not a verb, the embedded quotative clause, and therefore they are exocentric constructions, but fully productive in Wari' grammar. This means that any attempt to analyze Wari' clauses based strictly on endocentricity will fail.

7.5.2. Simple syntax vs. complex syntax

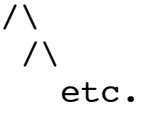
In an extremely interesting and important new approach to syntactic theory, Culicover & Jackendoff (2005:5) state their hypothesis as follows:

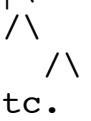
"**Simpler Syntax Hypothesis** (**SSH**)
The most explanatory syntactic theory is one that imputes the minimum structure necessary to mediate between phonology and meaning."

Culicover and Jackendoff go on to demonstrate throughout their book why the structure hypothesized for a given construction should be guided by the SSH rather

than the standard 'theorems' of phrase structure in syntactocentric theories, e.g. Minimalism (Chomsky 1995). Most linguists share the view that figuring out the hierarchical constituent structure of a given language is a very important and essential step to understanding the language. Naturally, for such an important area of linguistics, there are various theories on the constraints on possible tree structures, including the SSH.

Since Kayne ()'s work on restricting tree structures, many generative linguists have become persuaded that trees are binary branching, i.e. that (), but not (), is a possible structure of a human language. The SSH, on the other hand, would allow either, depending on the facts of the language under investigation.

(7.15)


(7.16)


It may in fact be possible to fit any language into either of these structures, more or less convincingly. In fact, it might even be possible (see section ___) to avoid tree structures altogether and use only violable (e.g. Optimality Theoretic, Prince and Smolensky (1993)) constraints on linear precedence in conjunction with semantic-based adjacency to account for the syntax of some languages. We might call the binary branching hypothesis 'complex syntax', the non-binary branching hypothesis in () the 'simple syntax' hypothesis, and the absence of tree structure in a language (whether the entire language or only some structures) the 'no syntax' hypothesis. From my own experience, each one of these views has advantages and disadvantages for the fieldworker (though the inductive simple syntax and no syntax possibilities are spiritual kin, complex syntax has nothing in common with them in its strong deductive approach, forcing structures without considering alternatives).

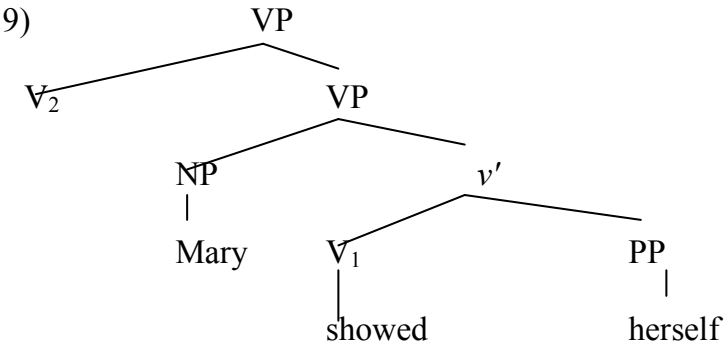
Following the complex syntax hypothesis may be a requirement of the theory you have adopted. But it is no less a procrustean bed for that. If a field linguist refuses to consider non-binary structures, they may indeed produce a description fully compatible with their chosen theory, yet utterly lacking in insight into various aspects of the language and of little long-lasting value (remember the fate of the famous Hidatsa grammar, discussed in ___).

The simpler syntax model may or may not be restrictive enough. My general rule of thumb is to propose the most restrictive trees possible (binary) initially, but always look for (and *accept*) evidence that this analysis is wrong, either relaxing it to (b) or (c) (on the latter also see ___). For example, consider the following discussion of English syntax from Culicover and Jackendoff, based on English examples like those in (7.17) and (7.18):

(7.17) I showed Mary_i herself_i.
(7.18) *I showed herself_i Mary_i.

Following the standard assumptions of complex syntax, i.e. imposing a deductive structure on the data that requires *c-command* to capture binding facts, without concern for the thesis of SSH, Larson (1988) argues that such data require binary trees plus a special 'verb shell', with subsequent movement from V_1 to V_2:

(7.19)



However, Culicover and Jackendoff (2004, 52ff) argue that simple precedence could account for such examples by requiring antecedents in the VP to precede their anaphors (this grossly oversimplifies their account, so the reader is advised to consult Culicover and Jackendoff (2004) for full details of their model).

Let's move now to briefly consider what a language without hierarchical syntax might be like, drawing from some suggestions based on my own work on Pirahã.

7.5.3. No syntax

Almost every language manifests the property of recursion, what some linguists consider to be the foundation of syntax (see Hauser, Chomsky, and Fitch (2004)). Simplifying slightly, recursion has two forms:

Embedding:

(7.20) A → AB

System recursion:

(7.21) a. A → BC
b. B → DE
c. C → AF

Linguists vary in the importance they attach to recursion. So, for example, Hauser, Fitch, and Chomsky (200??) argue that recursion may be the only (or core, depending on interpretation) component of Universal Grammar. Others recognize it, but do not attribute such genetic or species-defining importance to it.

Everett (2005) argues that there is no evidence for recursion in Pirahã. And there is some reason to believe (Hale (), Dixon (), Austin (p.c.), Nordlinger (p.c.) that recursion may be largely lacking from some Australian languages as well. In any case, the fieldworker must exercise care and caution in attributing the property of recursion (or indeed any property) to syntactic structures. This is because recursion is not the only way of combining syntactic units of one level of the hierarchy into another. Another way to put units together is *parataxis* (Bloomfield ()), where two more units are set side by side as it were, with no further structure, in particular without one structure being contained inside another of the same level. So, for example, () is the result of recursion, () is parataxis, and () is arguably parataxis as well:

(7.22) [$_{NP}$ John's [$_{NP}$ father's [$_{N}$ hat]]] is nice.

(7.22) is recursion, because one NP is found inside of another.

(7.23) The lion, the witch, the wardrobe, what have you, are all fictional objects in this story.

The first part of (7.23) is potentially a case of parataxis. The phrases (the lion, the witch, the wardrobe) are linked paratactically, rather than recursively (at least by my analysis). As another example of parataxis, consider Tagalog.

Some linguists, e.g. Schachter and Otanes (1972 - **PAGE**) and others have pointed out that some Tagalog sentence structures are like equations, where the two sides are brought together in a larger unit (the sentence) via parataxis:

(7.24) Predicate/Comment Topic

(7.25) *Titser* *ang babae.*
teacher nom woman
'The woman (is a) teacher.'

(7.26) *Maganda* *ang babae.*
stative-beauty nom woman
'The woman (is) beautiful.'

(7.27) *Umalis* *ang babae*
leave-actor focus, completive nom woman
'The woman left.'

Therefore, we can conclude that, at least for some constructions in some languages, recursion is irrelevant. Further examples can be found in Pirahã, according to Everett (2005). Consider the following:

First Pirahã lacks coordination, disjunction, and embedding. It uses non-recursive 'circumlocutions' to express each one of these structural types.

COORDINATION:

(7.28) *Ti ?isigihii kohoaipi. ?itii?isi pi-ai.*
I eat meat. Fish does also/now.

(7.29) *Ko?oi (hi) kohoaipi. ?aibigai (hi) pi-ai.*
Ko?oi he eats. ?aibigai does also/now.

DISJUNCTION:

(7.30) *Ko?oi (hi) kohoaipi. Hai, ?aibigai (hi) kohoaipi.*
Ko?oi eats. Hmm. ?aibigai eats.

(7.31) *Hi ?isigihii kohoaipi. Hai hi ?itii?isi kohoaipi. Ti kos-aaga.*
He eats meat. Hmm. He eats fish. I am ignorant (do not know).

Notice that in the following quotative, it is the verb of saying that bears what Everett (1986) analyzes as the nominalizer, rather than the otherwise 'embedded verb'. And yet in the subsequent apparently embedded clause 'arrow-making', the 'embedded' verb has the nominalizer. Everett and Van Valin (in progress) analyze this **–sai** as a marker of secondary or topical information, rather than as a nominalizer. The same analysis would hold true for (). The suffix **–sai** is in fact a verbal form and the language has a nominal corresponding to **–sai**, namely **–si** (found on the autodenomination of the language of the Pirahãs, **?apaitíis*i***. The nominal suffix marks nominal secondary or topical information (see Everett (1986)) for details. Many other examples are discussed by Everett (in progress) on Pirahã and language evolution.

EMBEDDING:

(7.32) **ti gái -sai kó'oí hi kaháp -ií**
1 say -nom. name he leave -intention
'I said that kó'oí intends to leave.' (literally 'My saying **kó'oí** intend-leaves')

(7.33) **hi ob -áa'ái kahaí kai -sai**
he see -attractive arrow make -nom.
'He knows how to make arrows well.' (lit: 'He sees attractively arrow-making')

One way to account for the apparent lack of recursion or configurationality of Pirahã is to adopt a no syntax hypothesis and analyze the Pirahã facts in terms of lexical semantics and linear precedence, as in (7.44):

(7.45) Linear Precedence + Lexical Semantics (> = 'immediately precedes'):
Agreement > NUCLEUS
NUCLEUS > Modifier
Topic > Comment
Argument > Agreement

These rules capture most of the facts of Pirahã 'syntax' without recourse to recursion. And I would be surprised if Pirahã were the only language for which the no syntax hypothesis is the simplest account of the data.This brings us to a more recent development of syntactic theory, another place where hierarchical, recursive structures have little role, namely, syntactic 'constructions'.[44]

7.6. Constructions

Words (and for some linguists morphemes as well) are considered by most linguists to exemplify the Saussurian sign, an arbitrary association of sound and meaning. Syntax, as the general reasoning (used to at least) goes is not arbitrary but

[44] We might well ask, of course, why such considerations are relevant for the fieldworker, as opposed to the syntactic theoretician. The reason it is crucial for fieldworkers to be knowledgeable and argue well is that if they simply apply the complex syntax model or some other theory blindly, however 'basic' or 'uncontroversial' they take it to be, their work can fail to teach all of us what it should have taught us about the language in question. The fieldworker should take no structures for granted.

strictly the output of regular constraints or rules. Goldberg (2006, 3) summarizes an alternative position,

> *"Many linguists with varying backgrounds have converged on several key insights that have given rise to a family of ... constructionist approaches. ... constructionist approaches emphasize the role of grammatical constructions: conventionalized pairings of form and function."*

In other words, there are many linguists who are persuaded that some or most of syntax is convention and not merely the output of formal rules or constraints. (But cf. Borer () for an alternative view of constructions.)

Some of the evidence for constructions does indeed seem very strong. So, for example, consider the 'Xer the Yer' or co-variational construction of English:

(7.46) a. The more you think about it, the less you understand.
b. The hurrieder I go, the behinder I get.
c. The bigger they are, the harder they fall.

This co-variational construction, according to Goldberg (2006, 6) is "... interpreted as involving an independent variable (identified by the first phrase) and a dependent variable (identified by the second phrase)." These properties of the construction do not follow merely by reading the semantics directly off of tree structures, as some theories might maintain. Rather, they are 'added' to the sentence by convention, the latter following from the matrix culture (see the papers in Enfield (2004)). In the view of many linguists, all languages have constructions for which part of the meaning is compositional and structure-based but where other parts of the meaning are added by convention. If these linguists are correct, then a fieldworker armed only with tree structures and form-based constraints will fail to account for vital, construction-based components of the grammar.

So how would one go about identifying and describing constructions? I suggest the conceptually simple, though methodologically rigorous, procedure in ():

(7.47) Methodology for identification of constructions
a. Understand the meanings of words and morphemes in the language (morphology fieldwork, see chapter __).
b. Understand the constraints on syntactic object formation in the language (chapter __).
c. Understand the syntax-semantics mapping in the language (section __ below).
d. See if any meaning is leftover that is constant for each instance of the hypothesized construction.

If a particular phrase, clause, or sentence is a construction in the technical sense, its meaning will not be exhausted by the preliminary steps in (a-c) above, but it will instead be circumscribed by them (i.e. it will not be completely arbitrary, e.g. an idiom like 'kick the bucket', but neither will its meaning be a simple sum of the meanings of its consituents). In other words, the fieldworker is satisfied that they understand (a-c) in the language under study, the part of the word or phrase that is left over, and that this leftover portion follows regularly from the construction in its various manifestations.

7.7. Discourse-down analysis

In this section we consider why it is important to collect the vast majority of one's data in the form of natural texts. It is the best method for collecting *acceptable,* natural examples. We discuss the basic text genres that should be represented in any data sample – especially what I consider to be the four basic genres: *expository* (e.g. teaching someone why certain things in the culture work the way they do), *narrative* (e.g. a historical text), *procedural* (e.g. how to make a particular object or perform a particular ritual), and *hortatory* discourses (e.g. exhortations on how to solve a community problem, sermons, etc.).

Not all cultures have stories about topics we might expect them to have stories about. For example, when I first told people that the Pirahãs had no creation myths or discourses about their history, anthropologists were skeptical.[45] One anthropologist actually went to the Pirahã village with the express purpose of collecting creation myths. His method was to ask in Portuguese and hope for the best, since he immediately recognized the Pirahãs' lack of knowledge of more than very rudimentary Portuguese. "How was your world created?" he asked. He recorded his questions and the answers for me, in order for me to help him translate the material he was collecting. After asking the question in Portuguese, he waited for the Pirahãs to translate the question into Pirahã. "The world is created," replies one of the assembled men in his own language. "Tell me how your god made all this?" the anthropologist presses on. "All things are made," comes the answer. The interview lurches on for a few more minutes, until suddenly, the question and answer session is overtaken by a deluge of excited banter as the assembled Pirahã vie to be heard.

"I've cracked it," said the anthropologist to me as he handed me his tape recording a few weeks later. "Here is the Pirahã creation myth." I must admit to being a bit dubious. In the past three decades, I have spent a total of seven years living with the Pirahã in the Amazon rainforest and am one of just three outsiders, along with my ex-wife and a missionary who spent time with them in the 1960s and 1970s, who is fluent in their language. I have long maintained that they are among the few people on Earth who have not devised a story to explain their existence. Others, including this particular anthropologist, find the idea difficult to accept.

So I listened to the tape. After the short, stilted exchange, some bright spark points out that this guy asking them odd questions doesn't know their language, so he will need to get help from me to translate the tapes. "Hello, Dan!" comes a chorus of Pirahã voices. "How are you?" "When will we see you?" "When you come, bring us some matches." "And bananas." "And whisky." And so on. Nice try, but no creation myth here.

7.7.1. Discourse and its relation to syntax

Understanding what the principal formal genres or 'distinguishing parameters' of discourse are in a given language, how they are structured, and how sentences are

[45] This box on creation myths is take largely from the report on my work by Kate Douglass in *New Scientist*, March 18, 2006, a story I told her during her research.

distributed within them, what kinds of information are talked about, what kinds of information are avoided, and what kinds are implicit, are each fundamentally important to the fieldworker's quest to understand the language they are describing. As Chafe (1994), among many others, demonstrates, there is much to learn about culture, cognition, and grammar in different languages via the study of discourse. With regard to content and cognition, as James (1890, 1:243) [The principles of psychology, 2 volumes, New York: Henry Holt. Reprinted 1950 by Dover Publications, New York.]puts it (also cited by Chafe (1994, 3)):

> *"As we take, in fact a general view of the wonderful stream of our consciousness, what strikes us first is this different pace of its parts. Like a bird's life, it seems to be made of an alternation of flights and perchings. The rhythm of language expresses this, where every thought os expressed in a sentence, and every sentence closed by a period.*"

On the other hand, as the generative tradition has shown over the past several decades, a large portion of syntactic structure does not simply 'emerge' as a by-product of discourse study. Among other things, discourses do not provide systematic ungrammatical examples – and yet these are vital for adequate analyses. The fieldworker should therefore conduct careful investigations of discourse structures, realizing, however, that these investigations will not be sufficient, however necessary, for a full analysis of the syntax of the language.

However, there *are* significant methodological advantages in beginning one's syntactic investigations by working 'downward' from the discourse. We consider a few of these in the remainder of this section.

Before you have learned to speak the language or studied its structure extensively, there is a plethora of opportunities to be misled as to the appropriateness, grammaticality, or translation of different sentences. There is ultimately no way to avoid this (another reason to remember Postal's Maxims). On the other hand, beginning syntactic study with the discourse ensures that all data are embedded in an appropriate context, are (likely) felicitous, and are grammatical (or at least acceptable). Thus, in this sense, the data are more reliable at the initial stages of analysis than the isolated, de-contextualized test sentences that might otherwise be used in elicitation.

Also, by studying sentences in their discourse context, the way the language presents and structures information can be studied more easily, naturally, and effectively in the initial effort to understand how it works. Eventually, other methods will be necessary, but it is difficult to imagine a more effective way of beginning studies of information structure than via the discourse.

Further, intonation can be more studied with more confidence at the outset (see Chapter ___ below for many more detailed suggestions on the study of intonation) because it is naturally contextualized when looked at initially in discourse. Initial hypotheses on intonation can at least assume that the intonation recorded in discourse is appropriate and natural and that it fits the particular information structuring of that part of the discourse.

Another advantage of beginning syntactic analysis with the study of discourse is that the entire array of sentential features, e.g. grammatical voices, constituent orderings, moods, evidentials, pronoun omission, etc. are observed in natural distributions. This is invaluable to undertaking their study. Finding them in appropriate linguistic environments is almost inconceivable without initial study of natural

discourses. Of course, the initial analysis will not exhaust their function and structures, but it will provide a better beginning point.

Functions of different sentence types will also be more readily visible when looked at in natural discourse initially. For example, are certain sentence structural types, certain phrase types, or words more common at the beginning, middle, or end of a particular discourse genre? Even such simple questions could provide important clues and insights into the final analysis of specific structures and constructions. But this simple source of insights depends crucially on studying discourse at the outset of syntactic research. And, as we see in ____, Lowe () argues that understanding discourse is vital to understanding morphology and lexical meanings in many languages.

7.7.2. Discourse genres and analytic methodology

This section makes no attempt to offer anything approaching a full introduction to discourse analysis. The main reason for this lack of ambition is that discourse analysis is an interdisciplinary set of research programs and not a simple step-by-step set of tasks. This interdisciplinarity is natural since discourse is the nexus of culture, philosophy, language, literature, cognition, social relationships and conventions, and grammar, among other things. Arguably, our discourses are essential to identify and distinguish languages, cultures, and psyches of individuals and groups. For that very reason, discourse analysis is a vital part of field research. That is, if I am right to argue that an understanding of a given grammar should go hand in hand with an understanding of the containing culture, then discourse study becomes more important yet, as the crucial link between the two.

7.7.2.1. Discourse genres

Discourse genres or types or varieties, etc. are hard to summarize. What might be to one fieldworker two tokens of the same genre could be to another fieldworker examples of very different types of discourse. Distinguishing discourse genres is largely a matter of what your background, expectations, and objectives are. Nevertheless, there are some distinctions that one should keep in mind, as a set of suggestions.

One type of genre that might be encountered in many societies is what we can label 'stylized or belletristic texts', that is texts that form part of the acknowledged cultural traditions and oral literature of a particular culture. These texts, regardless of their content, are very important because they represent careful reflection on both form and content and can reveal a tremendous amount about the culture and about the linguistics of discourse. However, at the outset of the field research they can be somewhat problematic. The reasons include the facts that they often include stylized or archaic language that is not idiomatic (which will be important to the eventual analysis but confusing at first) and they can be very long. They will be much more involved, harder to figure out in some ways. Moreover, since transcription time is roughly 5-6 times greater than actual recorded material and in fact much more at the beginning, this type of text can compromise your freedom to look at other things in the grammar. And if texts are not processed right away, at least not while in the field, they are simply worthless. Therefore, though no grammar would be complete without a careful discussion and analysis of such material, I recommend that one try to begin field analysis of discourse with simpler texts, e.g. simple reports.

The field linguist should also be sensitive to the possibility that speakers of the language may have their own recognized genres which may or may not correspond to

the field linguist's own categories. Treat these differently if you do encounter such cases. Almost certainly there is something in content or form that will distinguish them from one another and it will be worth your time to attempt to figure this out.

For a linguist, the most important classification of discourses will be in terms of their *structural distinctions*, things such as predominant tenses, aspects, construction types, overall organizational principles, etc. For example, let us consider four types I find useful to distinguish in my own fieldwork, namely, narrative, procedural, expository, and hortatory. To these we can also add religious, brief reports and informal discourse, and, last but by no means least, conversations.

Narratives are the easiest to study. They tend to use mainly indicative or a related mood, e.g. realis, and an array of past tenses. It is interesting to observe how the narratives are organized. For example, are events and participants introduced or presented in chronological order or in some other way, e.g. cultural relevance, logic, etc.? How are main participants distinguished from secondary participants? What kinds of sentences mark the main event line, as you understand the text? Do different kinds of sentences mark events that do not advance the main events of the discourse? How about the way that new participants are introduced? Are all characters of the discourse introduced in the same way? Or does it turn out that characters central to the telling are introduced differently (e.g. more sentences about them, different aspects or moods, more relative clauses, etc.) than peripheral characters? Are mythological characters treated formally different from 'real' characters? And so on. The possibilities are almost literally endless. Part of what distinguishes a mediocre field research from an insightful fieldworker is the effort invested to interrogate structures from different perspectives, formal, functional, cultural, societal, historical, etc.

Procedural discourses, things like how to make a bow and arrow, how to bake a cake, how to catch a fish, etc. involve step-by-step procedures that can illustrate moods and tenses (e.g. imperative and sequential, respectively) that are not found as readily in narratives. Since the subject of most procedurals is the addressee ('You do this, then you do that', etc.) then there is more likely to be subject omission in procedural discourses, due to the facility of subject identification.

Expository discourses explain things about the world. They may not turn out to be a lot different from simple narratives. But it is useful to begin by assuming that simple story-telling could be different from explanations and to examine these types of texts carefully for distinguishing formal and cultural characteristics.

And finally, hortatory texts are a superb source of cultural and linguistic insights (if the language in fact formally distinguishes them, though it likely will). These will be a rich source of deontic modals, different types and scopes of negation, cultural values, aspects, subject forms, etc.

With this barest list of suggestions on how to approach the general issue of discourse, let us consider in more detail an accompanying methodology.

7.7.2.2. Discourse analytical methodology

First, it is vital to have video or audio recording equipment present at different events associated with different discourses or the whenever there is a likelihood of recording specific texts, e.g. when someone returns from hunting, preparation for ritual, etc. Second, you must be creative in getting what you are after across clearly to your language teacher. Often in my experience, language teachers have felt that it is somehow wrong to give complex answers and prefer to give short, one sentence or one word answers to questions. Unless one can effectively communicate to the language

teacher what one is after, the field session can be very harrowing and unrewarding. And it is often just hard to get speakers to give a discourse in such cases. One suggestion is to build up texts line by line, by going back and forth between speakers to get more and more information on a particular topic and then ask someone to recap, rewarding the person who tells the longest story about it.

Taperecorders have a use beyond recording of sounds. They are essential in discourse analysis as well. For analyzing discourse, I recommend that you use at least two recorders, or a recorder and a computer, etc. That is, that you have one device dedicated to playback and another device dedicated to recording. Simple analog recorders with 'cue/review' buttons are the most convenient, but if you have a computer for playback this can be almost as easy to use. First, record a text with a native speaker. Second, transcribe the text and translate it with a native speaker. Third, go through the text with another native speaker (there are some ethical considerations here: make sure that the first speaker has given you permission to play the tape to others in the village, that the tape contains no sensitive or embarrassing material, and that the genre is one appropriate for a general audience, or carefully select language teachers that can and are willing to help you with sensitive material). This is done by playing back a portion of the tape and asking the new language teacher to repeat what was said, slowly. Record what the new language teacher says, including their comments about the text. This slower repetition will almost certainly introduce changes of pronunciation and lexical choice, but that is good. You want paraphrases and alternative pronunciations. The commentary provided will provide cultural and linguistic insights that are difficult to obtain in any other way. For a particularly interesting or challenging text, I would ask up to three additional speakers for comments, corrections, and translations.

After collecting texts, however obtained, they need to be processed. One set of suggestions that I have found useful is: (i) record the text digitally, with the speaker using a headset (this is discussed in the chapter on phonetics); (ii) transcribe it the first time alone, with no speaker present; (iii) check your transcription with the speaker who gave the text originally and write their corrections and comments in a different colour of ink. This will require also the use of two recorders. One to play back the original recording and one to record the speaker's comments. (Alternatively, one can record directly onto the computer. But I recommend separate digital or analog recorders and then transferring the data to the computer so as to have automatic backups of the data, among other reasons.); (iv) check the second transcription with another speaker and use two recorders, again, one to play the original text and one to record the second speaker's comments, corrections, etc.; (v) do a four-line transcription of the entire corrected text (all of this in the field!), with the following lines: Line 1 = morphophonemic transcription; Line 2 = phonemic/underlying segments transcription – these will differ occasionally; Line 3 = gloss; Line 4 = free translation. These are illustrated in the following example from Wari':

(7.48) **Mon** **womu** **cara** **ne**
ma'-on **womi-u**
that:prox:hearer-3sm cotton-1s that:rec. rec:past
'Where are my clothes that were just here a minute ago?' (lit: '... my recently absent clothes?')

Also, above the top, or morphophonemic, line one could also add an intonation line, e.g. in ToBI format (see chapter ___ for much more detail). This entails a significant amount for each text transcribed, but I will discuss why this is worth it and suggest that not every single text needs this level of detail. (The reader should also refer back to section 4.7.1. for more on metadata.)

Last but not least, all field linguists need to attempt to record conversation. The box here gives some initial bits of a conversation recorded in July 2004, among the Banawa people of Amazonas, Brazil.

Consider the following examples of a conversation transcribed and analyzed by Ms. Julia Reinbold, a research associate on a grant of mine.

Speaker A: Batao
Speaker B: Bido

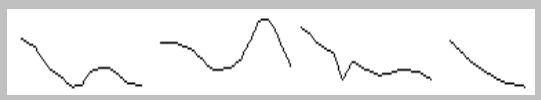
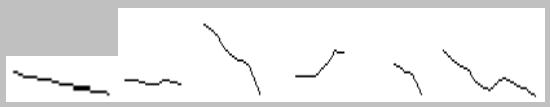

1. A: ta bom. e Bidoo. ebihiyarabone faiyama maniha / da Danieo badi
   e Bidoo. ebi-hiyara-bone faiyama maniha / da Danieo badi
   Bidu ?-converse-INT. this day ? Daniel big
   'Bidu, let us us talk today.'

2. B: m ee



3. A: e bihiyarabonee e faroee atiya / ebihiyarabone Danieo /
   e bi-hiyara-bone e faro ee ati-ya ebi-hiyara-bone Danieo
   converse-INT ? our ? voice-with ? converse-INT Daniel
   'Let us talk in our language.'

4. B: m

5. A: Danieo kamei hidabani fame / eremee e atini me no fa fa i e

   Danieo kamei hi -dabani -fame ere-mee
   Daniel came 3p all together 1PL OBJ-3PL
   atini me nofa fa i
   language PL like this much
   'Daniel and the others came all together. They like our language much'

This type of study is extremely important at various levels of study of the language and culture, revealing not only formal components of conversational organization in the language, but natural intonation, intersentenial anaphora, and so on.

By carefully looking at syntax, intonation, and discourse from natural data, we have an extremely useful tool and data source for subsequent analysis.

Of course, getting two people to sit in front of a researcher and converse naturally is not easy. The Banawa were particularly good at this, however, and although the conversation began slowly, within a few seconds, both interlocutors were conversing naturally, looking at each other and not at the researchers, the recorders, or the video camera. Further, both speakers, Sabatao and Bidu, wore headset microphones and each recorded onto a separate channel of our stereo digital recorder. If you are unlucky and cannot get people to converse naturally in front of you (the Pirahã won't for me, even after all these years), then there are other ways to approach the problem. You can get the people's permission to leave a voice-activated recorder in their house. You can try to record phone conversations (our first Suya text was one side of a phone conversation, recorded more or less by mistake as the language teacher's cellphone rang during a language lab session). These all have obvious disadvantages, but still they can produce useful data for discourse and conversational analysis. Other possibilities are to record radio broadcasts, video discussions of community beliefs, or rituals, etc. When asking questions you hope will produce discursive answers, be sure that the questions are framed broadly enough so that a free-ranging answer is more appropriate than a simple yes or no. Watch how they ask questions on your favorite talk show (modulo the rehearsal that goes into these overproduced spectacles). How does the talk show host get guests to open up and talk freely (at least between commercials)? How does a reporter interviewing a nervous subject? Take notes on technique and apply what you learn to your discourse-collection techniques. It is particularly important that researchers avoid framing questions so as to impose their own categories, values, orderings, abstractions, and length expectations, among others, on the language teacher (i.e. insofar as possible. It is never possible to completely eliminate our own culture, nor do we want to because our grammar-writing goals are a *cultural* value of our own. The very doing of field research and research more generally are themselves cultural goals, so we are obviously as bound by our culture during the data-collecting process as the language teacher is, if not more so.)

Finally, let me conclude with a brief mention of the importance of implicit information – what is *not* said (e.g. sample Pirahã text) in studying natural language discourse. This kind of information can be important in understanding how the culture works. The reason for that is that what is not said often reflects values and knowledge that all hearers and speakers of the culture hold in common. Consider for example, the following hypothetical exchange:

(7.49) Speaker A: Do you want drinks at the reception?
Speaker B: If you want to clean up and referee.

This is a simple example. Members of American (and other Western) cultures will know that this conversation is about alcoholic beverages and that people who drink such beverages can get messy and rowdy. That is what the references to 'clean up' and 'referee' refer to, this implicit cultural information which is simply not present in any theory of the literal semantics of this exchange. This seems obvious to members of the

culture, but it is not obvious to outsiders and this kind of thing will be puzzling for the field researcher until they have mastered the culture of the language under study (which they may never do of course).

7.7.3. Syntactic methods

When a discourse begins, we can be sure that its beginning marks one and probably several layers of morphosyntacic boundaries (e.g. morpheme, word, phrase, clause, sentence, discourse). Likewise, the end of a discourse will be the right boundary of at least one and likely several morphosyntactic constituents (and layers of constituency).

Also, we can identify morphosyntactic boundaries via discourses by identifying points at which speakers backtrack to reinterpret the sentence (e.g. for hard to follow or garden-path sentences, etc.), to correct errors, and so on, even without sophisticated experimental equipment) .

Spontaneous discourses also are marked by hesitations of different kinds by the speaker.These hesitations will often come at morphosyntactic or prosodic boundaries and so may be used as a source of hypotheses for investigating these boundaries.

Each discourse genre can also be expected a priori to be a likely source of certain kinds of grammatical features. So, for example, in a narrative discourse, most sentences are likely to be in past tense and indicative mood in most languages. In hortatory discourses, a large number of sentences are likely to manifest imperative mood and conditional or other special tenses. In procedural discourses we will get sequential relationships between sentences, present tenses, indicative and imperative mood, and so on. In other words, regardless of the theoretical status or relationship between discourses and sentences (and see Everett () for a review of many ways in which they might be profoundly different), the study of discourses can provide extremely important methodological support to understanding sentences. Let's consider some methodological rules of thumb that I have found useful over the years.

As I have said, textual information alone is inadequate to underwrite a grammar of a given language. This is because individual sentences need to be studied and permuted for tests of grammaticality. Nevertheless, sentences are usefully extracted from discourses for further testing by the following procedures.

First, isolate clauses in the text, using phonology and phonetics to probe for relevant constituent boundaries, as well as any knowledge available of the morphosyntax. Once each sentence of a particular discourse has been identified, then these can be used to generate paradigms, where each position and hypothesized constituent of the sentence can be checked by substitution and distribution tests. So, for example, assume that you have extracted the sentence, *John will probably eat beans on Wednesday.* Then this sentence can be extracted and tested as per (), listing English grammaticality judgments (my own) by way of illustration:

(7.50) a. John will probably eat beans on Wednesday.
b. ?John will on Wednesday eat beans probably.
c. *Beans will probably eat John on Wednesday.
d. The boy will probably eat beans on Wednesday.
e. The girl will probably eat beans on Wed.
f. Bob will probably eat beans tomorrow.
g. Bob will probably smoke beans tomorrow.
h. etc.

As you work through every sentence in every text in this way, adding to this process data observed in perambulatory elicitation, you will quickly come to have a very solid foundation for analysis and 'attacking' the entire grammar of the language.

Filing your data is vital for your own subsequent analysis as well as for linguists and others of the future for whom your data will be important.

Some linguists, usually those of the above 40 category, will prefer to file their data on paper. There are in fact advantages to hard-copy filing. For example the linguist works 'manually' with their data in this type of filing and for some people, this can be an aid to remembering the data. Certainly your field transcriptions, hand-written notes, etc. will be filed in hard copy.

By and large, however, I recommend the use of the very useful software programs available for language data storage. A number of such programs are available through SIL International, for example (http://www.sil.org/computing/catalog/index.asp). Other software producers include the Nijmegen Max Planck Institute for Psycholinguistics (http://www.mpi.nl/DOBES/INFOpages/applicants/dobes-techframe-main.html). The general rule of thumb for filing of any kind is to provide ready and clear access to all categories of grammar and semantics that the fieldworker has determined to be of relevance and significance for understanding the language in question. Thus, filing will not be the same from one language to another, but will require the judgment of the field linguist. Categories such as discourse type, clause valency, subordinate vs. matrix clauses, aspect, tense, mood, case, and so on are all obvious initial divisions in the filing system adopted for any language. The software collected will also provide numerous useful ideas for filing. On the other hand, since judgment is important in establishing, maintaing, subdividing, and applying categories appropriate for each language, the linguist will want the software to exhibit flexibility. Some programs can turn out to be fairly 'wooden' in practice. Therefore, before investing time and money in data filing software, you should inquire from the software producer and several users of that software something of its problems, advantages, and overall rating or utility. On the other hand, the more flexible the software, the less user-friendly and more demanding of the fieldworker's computer knowledge is likely to be.

7.8. Semantic fieldwork

7.8.1. Introduction

Semantics fieldwork is difficult under the best of circumstances. Part of the reason for this, in fact a large part of the reason for this, is that semantics, the study of meaning, must involve knowledge of culture, since in a sense, culture is the source of a great deal of meaning in natural languages. Another part of the reason is that for the large part of how meaning is related to syntax in the language in question, as well as various formal aspects of meaning (as discussed below) are extremely difficult to get at in the best of circumstances. Here methodology is crucial. In this chapter, I provide a brief overview of some of more salient issues in semantic fieldwork, drawing both from my own experience and from the very important article on semantic methodology by Matthewson (2005).

Semantic fieldwork can involve at least the following topics:

Ethnoclassification and Semantic fields – These related areas have to do with the way that cultures organize the world around them into meaningful categories. The world we live in can appear to be self-classifying to the naive observer. There are things like water, forest, fire, sky, ground, animals, people and minerals in the world.

Why would the linguist's task be any more complicated than working through 'Adam's task' in reverse? (Where by 'Adam's task' I mean just the biblical reference to Adam giving names to the things around him as though his pre-existing language matched one-to-one the world around him, without the necessity for an intervening culture to give meaning to that world.) That is, why doesn't the linguist simply collect the names for things in the a priori universe? Let's consider this problem in more detail, via the real world entities that we refer to in English as 'sky' and 'ground'. Isn't it perfectly obvious that these are distinct entities in the world?

In my first day of research among the Pirahã, I learned that they used a single word to refer to the entities labeled by 'sky' and 'ground' in English, their word **bigí**. But it took many more years of research to find out why this was the case. Initially I thought that perhaps I might have gotten the tones wrong or that there were actually two forms that shared a single truncated form. But it wasn't too difficult to determine that in fact it is the same word. So could the two words be homonyms, like 'too' and 'two' in English or 'janus' words (words that are their own antonyms) e.g. English *fast* 'to move quickly' and *fast* 'immobile' ('to stand fast'). These hypotheses seemed self-serving and unrevealing. There was clearly something going on here that I was missing. Ultimately, it emerged that the Pirahãs view the universe differently than we do. To the Pirahãs, the universe is structured in layers and the name for the boundary between any two such layers is **bigí**. So above the 'sky' is one layer of the universe. Below that is another (our biosphere). Below that is another, the boundary corresponding in a very broad way to what English calls 'ground'. Thus there is no simple mapping between English words for nature and the Pirahã **bigí**. Pirahãs and English speakers structure these parts of their universe differently.

The field linguist, if they are to give a comprehensive or accurate account of the semantics of the target language must provide at least an initial list of categories and subcategories of the natural and mythical worlds found in the language under study. But this will clearly involve a study of the culture of the native speakers – at least a synchronic study and often diachronic studies as well. Further, for detailed ethnosemantic studies it is crucial to enlist the support and advice of ethnobiologists, either in the field or in your prefield preparation. A great number of important pieces of cultural knowledge, linguistic structures, and meaning relationships depend on careful ethnosemantics.[46]

Semantics proper

Semantics-syntax interface: How does the syntactic form of a sentence constrain its meaning? Is semantics simply a summation of the meaning of the words of a sentence plus the contribution of a specific word order or hierarchical structure? We have already seen in ___ that grammatical *constructions* are important because, among other reasons, important aspects of their meaning are not derived by strict composition in the sense just given, but also involve what we might refer to as 'annotations', specific bits of noncompositional meaning associated with a particular construction which must be memorized as part of one's knowledge of the language in question. This then raises the question of what other types of meanings and meaning-form relationships might be introduced in the syntax. To get at the meanings that are linked to different syntactic

[46] A very good study of this kind in Brazil is Jensen's () study of bird classification among the Wayampi of Northern Brazil. A huge portion of Brent Berlin's research program has been dedicated to such classification, with important insights for all of us interested in the language-culture-real world connection (see Berlin (), (), ()).

forms and to tease apart different kinds of semantic meaning requires, to put it mildly, careful and painstaking work. In this subsection, we look at some suggestions on how to proceed in this area, borrowing heavily from Matthewson (2004).

Word and sentence meaning: Our basic question in this aspect of meaning is this: what does one know when one knows the meaning of a word or sentence? There are two broad components of the answer to this question of importance to most linguists. First, to understand the meaning of a word or a sentence is to know when it is appropriate to utter it. Saying 'John is a bastard' when meaning to say that 'John's parents were not married at the time of his conception' might be literally true in some archaic form of bastard. But this would not be a felicitous use of the term in general since to most speakers 'bastard' is principally used to comment negatively on the character of the individual in question, rather than the marital status of his parents. The second basic component of our knowledge of the meanings of words and sentences are what the world would have to be like for the utterance of the word or sentence to be true. The first component of meaning mentioned here is referred to as FELICITY CONDITIONS. The second component is known as TRUTH CONDITIONS. On the other hand, as important as these two notions are in developing a theory of meaning or describing meanings in a particular language, there are other fundamental concepts that must be explored: ENTAILMENT; IMPLICATURE; AMBIGUITY; and VAGUENESS. For sentences there is an additional meaning component, discussed in ___, viz., information structure. Because we discuss information structure separately, I want to focus here exclusively on these other concepts just mentioned, all of which are causally implicated in an understanding of sentence or word meaning.

First let us consider the truth conditions of a sentence or word. Arguably the most important component of meaning is found via the truth conditions. To see this, consider the following examples:

(7.51) He ate *all* the fish.
(7.52) He ate *none* of the fish.

How must the world be like to native speakers of English to judge the truth value of each of these examples? Let's first consider how we might go about determining this in a monolingual setting. One idea would be to put some food in front of a helper (not the language teacher), having previously instructed them to eat it all as soon as you give the signal. Then you could repeat the experiment, having them eat *most* but not all of the fish. In the simplest case, the language teacher will accept your description of 'He ate *all* the fish' in the first instance and reject it in the second, where part of the fish was not eaten. A very similar set of scenarios could be created to get at the truth conditions of *none*. On the other hand, it is clearly vital that *much* thought needs to go into testing for truth conditions. The linguist tests for truth conditions using complete, grammatical sentences, to check meanings that are already largely understood. Checking for truth conditions is not a tool for initial field research.

Using a metalanguage, the procedure is easier in some respects. For example, consider how we might probe the truth conditions of 'all' using a metalanguage. We will assume that the same third party eats the fish again. You ask, 'Can I say 'he ate all the fish'? (in the metalanguage) The speaker will presumably then say 'Yes, you can say that he ate all of the fish'. (Again, make sure that all exchanges between linguist and native speaker are complete sentences. It is easy to misinterpret results if partial utterances are given. And partial utterances cannot be appropriately labeled as either

grammatical or ungrammatical.) Further probing will reveal that when the speaker eats only part of the fish, even the major part, that the native speaker will say that 'No, then you cannot say he ate all of the fish, because there is some left.'

To get at felicity conditions, consider the following example from Matthewson (2004, 401):

> "(7.53) Situation: There are two cats in the room, and they are both asleep.
> The cats are awake. FALSE
> The cat is asleep. INFELICITOUS"

How does one distinguish reliably between the falsity of a sentence vs. the infelicity of a sentence? This involves discussion with native speakers, attempts to paraphrase, or, as in ()-(), changing grammatical marking of some sort or another.

7.8.2. Diagnostics

There are a number of useful diagnostics for the study of meaning in well-known languages. One such test is called the 'wait a minute' test, first introduced by Kai von Fintel.

This test (Kai von Fintel, cited in Matthewson (2004, 34ff)) is useful for distinguishing infelicity from falsity, a crucial distinction for the semantic fieldworker. But like so many other diagnostics, it often does not travel well from one language to another. Consider Matthewson's (2004, 34ff) discussion of this in the box.

"The issue of distinguishing falsity from infelicity is particularly tricky with respect to St'át'imcets and Straits at least, and possibly other Salish languages, because one easy language-internal clue to infelicity due to presupposition failure simply does not work in these languages. The test is the so-called wait-a-minute test, invented by Kai von Fintel (personal communication). Since presuppositions are propositions that discourse participants are assumed to know already at the time of utterance, a presupposition which is not known can be challenged with "wait a minute!" (or another similar expression). On the other hand, an assertion which is news to the hearer cannot be challenged by "wait a minute!" English examples are given
in (52) and (53).

(52) Presupposition of stop:
A: Felicia has stopped smoking
B1: Wait a minute! I didn't even know she smoked!
(presupposition unknown)
B2: #Wait a minute! I didn't even know she stopped!
(assertion unknown)

(53) Existence presupposition of the:
A: Barnaby won the semantics prize.
B1: Wait a minute! What semantics prize?
(presupposition unknown)
B2: #Wait a minute! I didn't know he won it!
(assertion unknown)

Unfortunately for researchers working on Salish, for some reason this test does not work. Even items that are difficult to conceive of as lacking presuppositions (such as 'stop') do not give rise to a "surprise" response with any consultant tested so far. Even when consultants offer or accept overt denials or questionings of the failed presupposition, as in (54), this does not distinguish presuppositions from assertions (as the English "wait a minute" test crucially does). Exactly parallel denials and questions are judged appropriate for assertions which are not already known, as shown in (55).

(54) A: plan cukw kw-a-s mán:x-ˊm s-Bob
already finish det-impf-nom smoke-mid nom-Bob
'Bob stopped smoking'.
B1: ?az L:u? kw-ˊn-s-wá zwát-ˊn kw-s
neg just det-1sg.poss-nom-impf know-dir det-nom
tu? mán:x-ˊm s-Bob
past smoke-mid nom-Bob
'I didn't know Bob smoked'. (presupposition unknown)
B2: wa? ha tu? mán:x-ˊm kw-s Bob?
impf ynq past smoke-mid det-nom Bob
'Did Bob smoke?' (presupposition unknown)

(55) A: plan cukw kw-a-s mán:x-ˊm s-Bob
already finish det-impf-nom smoke-mid nom-Bob
'Bob stopped smoking'.
B1: ?ay L:u? kw-ˊn-s-wá zwát-ˊn kw-s
neg just det-1sg.poss-nom.impf know-dir det-nom
cukw-s
finish-3sg.poss
'I didn't know he stopped'. (assertion unknown)
B2: wa? ha tu? cukw
impf ynq past finish
'Did he already stop?' (assertion unknown)

Why the "wait a minute" test should be inapplicable in some languages is at present not known. For fieldworkers investigating a language that does allow "wait a minute" responses, the test is a very useful tool for establishing infelicity."

Again, there is no magic set of crosslinguistic tests that will enable you to make all the distinctions you want to make clearly. But there are numerous suggestions that, when coupled with your own creativity and thought can provide useful inways into the semantics of the language in question.

Let's move to a final area of study, one of the links between semantics and syntax in fact, *information structure*.

7.9. Information structure

Information structure is to me the heart of syntactic analysis. It is the way that language formally distinguishes between relationships among types of information. To use common, though not completely accurate terminology, it is where we distinguish new, contrastive, topical, and related types of propositional relations and content in the language.

In general, as students of language have recognized for centuries, the sentence contains a basic pragmatic division roughly understood as the contrast between old vs. new information. In general prosody (see ___) is used to provide clues as to the location of focus in a sentence. Thus, since the predicate generally carries new information it will likewise tend to bear the primary sentence stress. This can be seen in the examples below, from Lambrecht (1994, __), in which predicate forms are illustrated from different languages:

(7.54) Predicate focus
Q: What happened to your car?
A: a. My car/It broke DOWN. English
b. (La mia macchina) si è ROTTA.Italian
c. Auto se POKVARIO/POKVARIO se. Croatian
d. (Ma voiture) elle est en PANNE.French
e. (Kuruma wa) KOSYOO-si-ta. Japanese

(7.55)

Sentence:	*My car broke DOWN.*
Presupposition:	'Speaker's car is available as a topic for comment x'
Assertion:	'x = broke down'
Focus:	'broke down'
Focus domain:	verb plus remaining post-verbal core constituents

The ingredients of information in Lambrecht's theory are the *presupposition*, i.e. what the speaker assumes that the hearer knows; the *assertion*, the telling of what is new; *focus*, the term of the assertion that is what the assertion is about; and the *focus domain*, that structural part of the sentence that contains the focus. In Role and Reference Grammar, this focus domain is known as the *actual focus domain* and a further distinction is made of the *potential focus domain*, i.e. where the language allows actual focus domains to occur.

To get a feel for focus domains in English and other languages, for example, consider the following, extracted from Van Valin's work on these matters in Role and Reference Grammar (and sent to me for use here by Van Valin in email, March 2006):

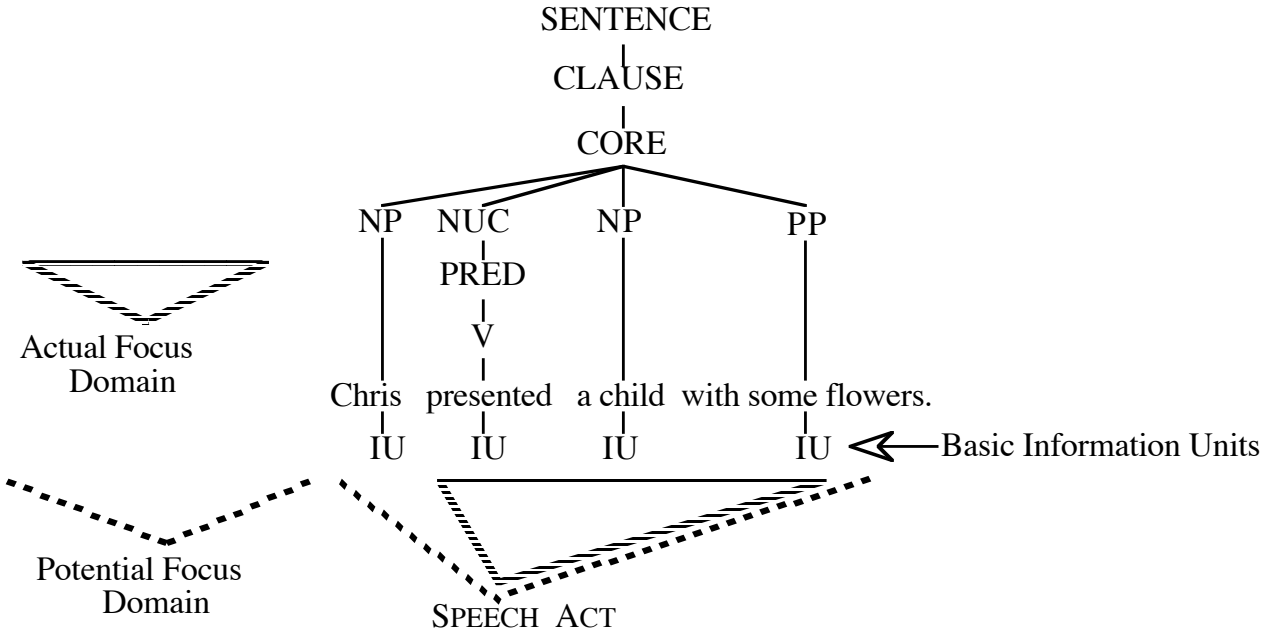


Figure : Predicate-focus construction

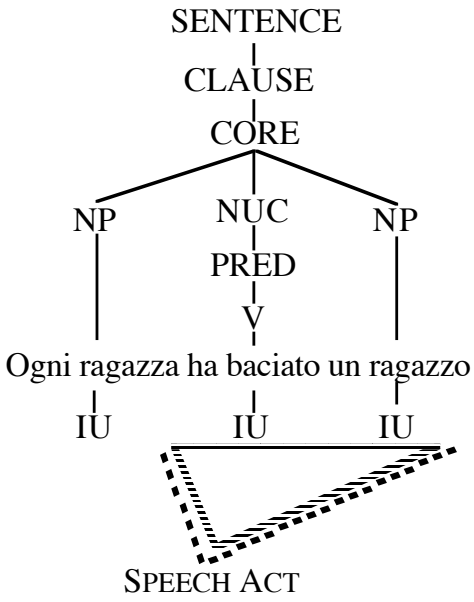


Figure : Predicate focus in Italian

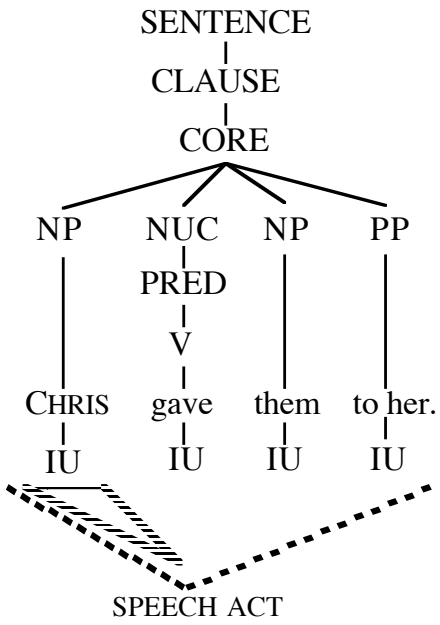


Figure : (Marked) Narrow focus in English

Now let us consider other types of focus. If the entire proposition is the assertion, then the focus is sentential. For example:

(7.56) Sentence focus
Q: What happened?

A:	a. My CAR broke down.	English
	b. Mi si è rotta la MACCHINA.	Italian
	c. Pokvario mi se AUTO.	Croatian
	d. J'ai ma VOITURE qui est en PANNE.	French
	e. KURUMA ga KOSYOO-si-ta.	Japanese

(7.57) Sentence:	*My CAR broke down.*
Presupposition:	none
Assertion:	'Speaker's car broke down'
Focus:	'Speaker's car broke down'
Focus domain:	Clause

Focus can also take as its scope only a single phrasal constituent:

(7.58) Narrow focus (what Lambrecht calls 'argument' focus)
Q: I heard your motorcycle broke down.

A:	a. My CAR broke down.	English
	b. Si è rotta la mia MACCHINA.	Italian (Lit: 'broke down my car'/ it's
	È la mia MACCHINA che si è rotta.	my car which broke down')
	c. AUTO mi se pokvario.	Croatian
	Pokvario mi se AUTO.	
	d. C'est ma VOITURE qui est en panne.	French ('it is my car which broke down')
	e. KURUMA GA kosyoo-si-ta.	Japanese

(7.59) Sentence: *My CAR broke down.*
Presupposition: 'speaker's x broke down'
Assertion: x = 'car'
Focus: 'car'

English is relatively remarkable in the degree of flexibility it allows for focus domains:

(7.60) a. Leslie sent the book to DANA yesterday.
b. Leslie sent the book to Dana YESTERDAY.
c. Leslie sent THE BOOK to Dana yesterday.
d. Leslie SENT the book to Dana yesterday.
e. LESLIE sent the book to Dana yesterday

As suggested by this example, languages differ (VanValin and LaPolla (1997, ---)) according to the relative rigidity of their focus structures. Some languages allow focus only in narrowly circumscribed syntactic positions, e.g. French and Italian (immediately postverbal position is reserved for focus), whereas other languages allow more flexibility in their potential focus domains, as illustrated in Table __ below:

	Rigid Focus Structure	Flexible Focus Structure
Rigid Syntax	French, Toba Batak	English, Toura
Flexible Syntax	Sesotho, Italian	Russian, Croatian

Table : Typology of the interplay of focus structure and syntax (Van Valin 1999)

Moreover, as Van Valin (1999) has also show, there is an interesting interplay of syntax and morphology with focus structure, as illustrated in (). As () shows, the more marked a particular morphosyntactic object is in a particular information theoretic role, the more phonology it has.

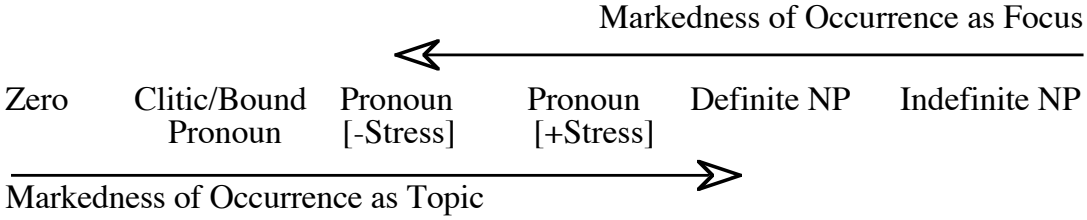


Figure 1: Coding of referents in terms of possible functions

7.10. Argumentation

All syntactic analysis is observation, guessing, belief-formation, and argumentation. Argumentation is understood here as providing warrants for one's own beliefs while removing warrants from plausible opposing beliefs.[47] Unfortunately, too

[47] A 'warrant' can be understood informally here as the linguistically ostensible reason for your belief.

many grammars include only [observation + beliefs] or [observation + beliefs + warrants for those beliefs]. And yet, leaving out the final stage, demonstrating a lack of warrant for plausible alternative beliefs, renders a reference grammar or article much less useful to many readers. It is easy to see why. The audience of a reference grammar or professional article will be linguists and specialists from related disciplines. Good readers must be persuaded, not merely informed. They will be constantly thinking of alternative accounts of the facts. And those accounts will have warrants that are only ostensible by your account. You need to say why, in anticipation of these being raised.

Let me give a very simple example. Let's say that you analyze the following sequences as trochaic stress, oriented left-to-right within the word, i.e. 'stress every even-numbered syllable from left-to-right in the word' (where boldface indicates stress).

(7.62)

a.	ba	**ba**	ba	**ba**	ba				
	1	2	3	4	5				
b.	ba	**ba**	ba						
	1	2	3						
c.	ba	**ba**	ba	**ba**	ba	**ba**	ba	**ba**	ba
	1	2	3	4	5	6	7	8	9

Your analysis works, quite obviously. But are there other analyses (structural beliefs) you might form about these sequences? One possibility is that the system is instead an iambic system oriented right-to-left, i.e. exactly the mirror-image of the proposed trochaic analysis. Just as obviously, this new proposal works as well for the facts as the trochaic analysis (stress every even-numbered syllable left-to-right). Therefore, parallel to your belief about how stress works in the language, there is a belief that it works in just the opposite fashion. The small amount of data in () warrants both sets of beliefs. So how to improve the warrant for your belief while removing the warrant for the opposing belief? Well, expand the empirical set under consideration. What does each analysis predict for words with even numbers of syllables (those above have only odd numbers of syllables)? Consider then the further hypothetical examples in ():

(7.63)

a.	**ba**	ba	**ba**	ba		
b.	**ba**	ba	**ba**	ba	**ba**	ba
c.	**ba**	ba				

If, ceteris paribus, words with even numbers of syllables are indeed stressed as shown in (), then the alternative iambic hypothesis fails, its belief is no longer warranted.

Now we need to ask how one comes to have ideas on how to select between alternative hypotheses. Basically, the ability to argue effectively derives from knowing your own analysis well.

Knowing my own analysis well enough to argue for its superiority to plausible alternatives was hard for me in the early days of my career. And the effect was that all of my initial submissions to journals were rejected. It was embarrassing to have referees point out alternative analyses that worked as well or better than my own. I realized that they were able to propose these counteranalyses because they understood my own claims better than I did myself. So I eventually learned to work more slowly

and to think and reflect about my own analysis so that I would know its implications at least as well as the referees. This enabled me to discuss the predictions of my analysis and why these were superior to the hypothetical alternative analyses.

To know one's analysis requires that one know well the linguistic subfield in which one is working, including the particular theory or framework one has selected to analyze one's data and present one's research report. This means that the fieldworker must read deeply and widely in linguistic theory, familiarizing themselves with the predictions, constraints, and understanding of the relevant aspects of grammar urged by a particular theory.

A good linguistic argument is based first of all on a good hypothesis and solid knowledge of that hypothesis's predictions and implications. Again, good hypotheses come from knowledge of linguistic theory, knowledge of the language, knowledge of other, especially related, languages, as well as intelligence, aptitude, and luck.

One can approach hypothesis-formation in the field inductively, deductively, or abductively, as in ()-():

(7.64) Inductive methodology

a. A fact is found. Analyze this fact. What is your understanding of this fact in your theoretical framework and in your larger understanding of the language (etc.)?

b. Your understanding leads you to believe that future facts encountered in the language will be constrained in specific ways.

c. You systematize facts according to what you think accounts for them best.

(7.65) Deductive methodology (best exemplar is the syllogism)

a. A theory is known. The theory constrains the notion of 'possible fact', along a certain range.

b. Your understanding of the theory leads you to believe that facts encountered will have a certain shape and to so classify the facts you find.

c. You systematize facts according to your understanding of the theory.

In fieldwork, as in all scientific research, both inductive and deductive hypothesis-formation should be used. Using these as a beginning, it is inevitable that some facts subsequently encountered will not fit the understanding mentioned in () and (). When new facts do not fit our expectations, we are surprised. A great deal of the success of field research is the result of being surprised, i.e. coming up with problems to resolve. But, again, coming up with these problems to resolve comes almost exclusively from the reasoning types in () and () in conjunction with the careful study and knowledge on which they are based.

So when one comes up with a surprising fact, how is it to be dealt with, that is, how is it to be integrated in one's emerging understanding of the language? The American philosopher, Charles Sanders Peirce (1839-1914) suggested a method of reasoning which has become extremely influential in accounts of how scientists reason about the world. Peirce () labled this new type of reasoning *abduction* or *retroduction*. I summarize this in ():

(7.66) Abduction

a. You find a surprising fact, A, where by 'surprising' we mean that it fails to conform to or contradicts your predictions.

b. But if another belief, B, a specific one you have imagined, were held, the A would no longer be surprising.

c. The ability of B to remove the surprise of A is evidence, C, for B.

Let's consider an example. Consider that you have discovered several clauses where the subject NP precedes the verb which in turn is followed immediately by the object NP. So based on both your knowledge of theory (deduction) and your experience with these examples (induction) you propose ():

(7.67) Information-theoretic based constituent order hypothesis: old information precedes new information (where subjects carry old information and objects new information).

Next, assume that you find a sentence in which the subject follows the object, e.g. the simple form VOS. You are convinced that in this and other VOS examples, the subject still realizes old or topical information and the object realizes newer or focused information. Therefore, insofar as you believe the hypothesis in (), you are surprised. However, you then have another idea, the hypothesis in ():

(7.68) Information-theoretic based constituent order hypothesis: new or focused information immediately follows the verb.

Hypothesis () works for both sets of examples. If it were true. then VOS examples would no longer be surprising (because () says nothing about the positioning of the subject NP, only the object NP. Therefore, C, you have reason to believe (), the abductive B.

In a sense, abduction is simply a formalization of guessing. But it is useful because it gives us a clear, explicit visual representation of the nature of our task in hypothesis formation. Peirce (1877) is a principal source on the concept of abduction. Every fieldworker would profit by familiarizing themselves with reasoning models and the notion of warranted belief.

Wood, Linda and Rolf O. Kroger. 2000. *Doing discourse analysis: methods for studying action in talk and text*, London, Sage Publications.