"One of the first tasks which confront the linguist in examining a new language with a view to discovering and describing its structure is the identification of the minimal meaningful units of which the language is composed." Nida (1974 [1946], p6)

CHAPTER 6. MORPHOLOGICAL FIELDWORK

6.1 INTRODUCTION

In my early days of research among the Pirahãs, every single utterance was nearly impossible to understand. Since the Pirahãs do not speak Portuguese, there was no way for me to get a translation of anything they said. And, for reasons I was not to understand for many years, they didn't seem to have simple verbs. Where I expected a simple verb for, say, smelling, I would get long multisyllabic utterances. Each word I was able to isolate and understand was a costly accomplishment. Learning what a particular morpheme meant was cause for celebration.

In polysynthetic languages, e.g. many American Indian languages, the hardest nut to crack is the morphology. And the difficulty of understanding the meaning of individual morphemes is not likely to be significantly easier in any language, even in languages that lack inflectional morphology (e.g. Wari', Everett (2005a, 2005b) or even have no morphology (if there are in fact such languages). If a language has little in the way of morphology proper, this will simply move the problem of understanding the meanings and functions normally conveyed by affixes one level higher, to the word or, more likely, to the phrasal level. In other words, no fieldworker can avoid analyses of words and how they fit into the grammar, a matter that always entails morphological principles.

Therefore, what I want to do in this chapter is to help equip the fieldworker for successful morphological field research. Of course, a single chapter, or two or three, is not really enough space to provide significant detail about all aspects of morphological analysis. There are, after all, entire books dedicated to these matters (e.g. Nida () and Pike (). Nevertheless, this chapter aims to provide at least a 'check list', with examples to remind the fieldworker of the minimum required for data-collection and morphological analysis in the field.

One might think we could begin by simply answering the question 'What is morphology?' and then get on with a methodology for it. Morphology can be usefully defined as the study of words and how they are formed (see Everett (in progress) for various approaches to the study of word formation).

And yet it is not always clear where the dividing line falls between morphology and syntax, a problem that can confound both the fieldworker and the theoretician. Syntax is often defined (at least informally) as the study of how words get put together to form sentences. But in practice, the fieldworker can often only understand word structures if they understand the sentential contexts (structural and semantic) in which the various word-forms may appear and vice-versa. So one needs to study syntax and morphology together, even though one might believe that morphology and syntax are ultimately distinct components of grammar (e.g. that they are organized in different ways and/or have different principles – see Ackema and Neeleman (2004) for useful discussion).

Although the morphology and syntax of a language need to be studied together, there are still principles and methodological procedures unique to word structure. So I focus on those here. But where necessary, syntax is brought into the discussion to

explicate morphological concepts and their implications.[25] I want to offer three caveats on morphological fieldwork before beginning with a consideration of morphological fieldwork proper.

6.2. CAVEATS

6.2.1. PHONOLOGY IS NOT MORPHOLOGY

You will hear a cacaphony of confusion when you first enter an area where you do not speak the language (e.g. as I do now as I write this in Leipzig and *ich spreche nur ein bisschen Deutsch*.) And it can depress you if you think too long about that fact that it is your responsibility to bring order to your confusion, to find euphony in the apparent cacaphony.

In my first linguistics course, Professor Kenneth L. Pike asked the class how we might analyze the form in (1), which he claimed to be common in English (class lectures, University of Oklahoma, 1976):

(6.1) **squeat**

I did not know what he was talking about. What does this mean? Pretending for the moment that we do not speak English, let's approach this problem like fieldworkers. We first need to consider the context in which () is uttered. It turns out that it is usually uttered by a member of a group preparing to dine together. The linguist must observe this context him/herself. They cannot ask speakers directly what an utterance means without supplying them with a context. In fact, native speakers should only be asked to opine about data – and then only when they fully understand the relevant context of the items under discussion in very carefully controlled circumstances. After providing a context, on the other hand, if we then asked speakers of this dialect what *squeat* means, some speakers might understand it and might be coaxed to repeat it more carefully. If they do, they are likely to say something like:

(6.2) Oh, you mean '*let's go eat'*.

With this more careful repetition, it becomes clearer that the phonology of () obscures its morphology. This is an important point. To see this perhaps more clearly, let us go deeper into the phrase. How might a field linguist analyze the **s** at the end of [lɛts], for example? They might be tempted to conclude that this **s** is part of the word [lɛt], basing this morphological decision on the phonetics of the word form.[26] But this

---

[25] Interestingly, in my experience when a language has a complex morphology, that morphology is more difficult to analyze than the syntax of the language, however complex the latter.

[26] This is not a textbook on morphology, but some terminology would be useful here nonetheless:

word form: a phonological 'word', without regard to its morphological composition, a constituent which fits the phonological definition of word in the language. So, for example, **let's** is a word form in English. As is **am**, **is**, **were**, **gone**, **going**, etc.

would be wrong, clearly. Although **s** is indeed part of the *phonological* unit that precedes it, it is best analyzed as an independent word syntactically, a shortened form of the pronoun, [ʊs]. This is what Zwicky (1977, page?) labels a 'simple clitic'. In other words, the phonological data in this (simple) case once again conflicts with the morphosyntactic data. This leads me to offer the following related warning:

(6.3) Be careful not to misanalyze clitics, affixes, and function words because of their phonology.

A subtype of (3) is (4):

(6.4) Cliticization is often confused with affixation.

Consider, for example, how a novice fieldworker might analyze the English examples in (5) and (6):

(6.5) a. John's hat.
b. The man's hat.
c. The man who was running's hat.
d. The man you spoke to's hat.

(6.6) I'll come tomorrow.

In (5) there is an enclitic which attaches to the final word of the noun phrase. This enclitic, as it turns out, is indeed morphologically, syntactically, and phonologically part of the preceding phrase. By analogy, the fieldworker might analyze the ***'ll*** of (6) in the same way, assuming that it too 'goes with' the word on its left. But semantically and syntactically, ***'ll*** forms a unit with the verb that follows it, not with the word that precedes it. That is, its phonological host is unrelated to it syntactically. This phenomenon is often labelled 'wrong-way' cliticization in the literature, a case where the phonology and morphosyntax conflict. An additional example of wrong-way cliticization, further underscoring the importance of keeping phonology and morphology separate in fieldwork, comes from Yagua, a language spoken in Peru, from (Everett 1989, 343, who takes it from Payne & Payne ()):

(6.7) Sa_i -pú̱ú̱chi(*-níí) $Pauro_i$ rooriy(*-níí) viimú -$níí_j$ $Anita_j$.
3sg.cl -carry Paul house into 3sg.cl Anita
'Paul carries Anita into the house'

---

word: an abstract morphosyntactic unit that is not heard, but posited to account for morphosyntactically relevant word behaviors.

lexeme: the minimal unit of the lexicon.

Similar (though not identical) definitions for these are offered in Matthews (1991, 24ff).

Everett (1989) attempts to explain, among other things, the observations of Payne () and Payne and Payne () that Yagua object clitics, like English auxiliaries, undergo 'wrong-way' cliticization. The clitic **–níí** in () must attach to the immediate left of the object, in this case, **Anita**. It cannot attach directly to the object, but it must nevertheless be immediately to its left. So although the clitic and the object do not form a constituent phonologically, they do syntactically (at least according to Everett (1989)).

To sum up these points aphoristically: *What you hear is not what you get.* Or more scientifically, *articulation obscures morphosyntactic boundaries in normal speech.* You must think, analyze, and argue for all your conclusions. Nothing is obvious in linguistic analysis, especially not in morphological analysis. Do not be fooled by confusing phonology with morphology! Let us consider now another caveat, namely, that:

6.2.2. SEMANTICS IS NOT MORPHOLOGY

The fieldworker must also be careful not to assume a direct or transparent relationship between semantics and morphology. A meaning that is expressed as a single morpheme in one language may not be expressed at all in another language or it may be expressed by multiple words or morphemes. Although semantic arguments, like phonological and phonetic arguments, can be useful in identifying and classifying words, word classes, morphemes, etc. in a language, there is no algorithm for translating from one to the other or any necessarily direct relationship.

To understand this a bit better consider the category of tense. Consider the question, "Does English have a future tense?" in light of the examples in ():

(6.8) a. I will go tomorrow.
b. I am going to go tomorrow.

Are either of these the future tense in English? Well, if you mean semantics, then the answer is 'yes', they both express (among other things) a future meaning. But if you mean morphology, then the answer is 'no', neither of these examples contains a morphological future tense. In fact, morphological and semantic categories are often confused in just this way in the professional literature. Everett (1993), for example, argues that Pirahã has no morphological tenses, though it does have 'absolute tenses' (Comrie (1985)) semantically. But I argue there as well that Pirahã has no relative tenses, either morphologically or semantically.

It is common as well for novice field linguists to find it difficult to keep semantics and syntax separated as they prepare a grammar of an unstudied language. For example, linguists I have worked with in following some sort of questionnaire e.g. the Lingua Descriptive Series Questionnaire (LDSQ), often initially answer affirmatively to some of the structural or constructional questions of the LDSQ (of the type 'Does the language have relative clauses?') even when the language in question lacks such structures, just because there is a way to express that meaning in the language under investigation.[27] So, for example, does English have 'evidentials' (see section ___ below) – morphological markers that indicate the source/type/quality of evidence for a given assertion? Morphologically, no, English does not have such markers. But of course an

[27] The LDSQ can be found at: http://lingweb.eva.mpg.de/fieldtools/linguaQ.html

English speaker can always add to a statement, 'I saw that with my own eyes' or 'So they say', using these expressions like evidentials semantically. So English can express evidentiality, even though it has no evidentials. This is a very important point that is not nearly as easy to recognize in fieldwork as it is in the linguistic classroom. In other words, when linguists investigate *structure*, they are not thereby investigating the circumlocutions that might be used to express the meaning or function of the structure. They are concerned with understanding both form and meaning, of course, recognizing however that they should be studied separately (initially at least) in order to better understand each. An excellent discussion of these issues is found in Comrie, et. al. (1993).

A final caveat before beginning our discussion of morphological fieldwork proper is that:

6.2.3. SYNTAX IS NOT MORPHOLOGY

Although, as I said, syntax and morphology need to be analyzed nearly simultaneously due to their mutual influence, it *is* possible to tease them apart. They are not the same thing, at least not operationally in the field (see Ackema and Neeleman (2004) for interesting discussion of how these might be distinguished in the grammar (or not) and the consequences of different decisions). They differ most strongly, in my opinion, in that word formation rules are *paradigmatically constrained* (see Everett (2005) and ___ below).

And yet, morphology and syntax often appear to overlap – which can and will confuse the fieldworker unprepared for this. As Ackema and Neeleman (2004, 48ff) show, morphology and syntax can often 'compete' with one another. So, for example, consider the contrast in (9) and (10) (again, see Everett (in progress) for a survey of approaches to morphological analysis):

(6.9) a. *felicitouser
b. more felicitous

(6.10) a. bigger
b. *more big/more bigger

If an adjective cannot take the comparative suffix, **-er**, then it is made comparative syntactically, or periphrastically. The suffix might be said to *block* the use of the syntactic form. That is, the syntactic or periphrastic form and the affixal form are not in free variation in most dialects (though this is an empirical question that has not been well-investigated to my knowledge, even though its consequences for morphosyntactic theory are potentially important). We will see more evidence for this in section ___ on periphrasis. Such interactions reinforce both suggestions made here, namely, that syntax and morphology must be analyzed together by the fieldworker, yet they are not the same thing. Understanding English comparatives, in other words, requires us to understand the different roles of syntax and morphology in constructing comparatives, as it also requires that we analyze both together to see the full set of patterns emerge from the data. These are the main caveats. Now we are ready to begin our discussion of morphology fieldwork proper. I prefer to discuss morphological field analysis in terms of semantic distinctions and how these are marked, since as a fieldworker, I found this to be the most natural way to think about the issue.

6.3 SEMANTIC AND GRAMMATICAL DISTINCTIONS

6.3.1. NOMINAL CATEGORIES

The two most important questions for morphosyntactic fieldwork are (i) What semantic and grammatical distinctions are marked on words? and (ii) How are these distinctions marked? That is really all there is to morphology. There is a large, but fortunately, not too large, number of answers to each of these questions and many combinations thereof. But we can in fact offer a couple of simplified lists that, I believe, will to better conceptualize the issues.

I am using 'semantic distinctions' to refer to properties of the lexicon (as I see it) that are relevant for the syntax, e.g. aspect, number, class, case, tense, etc. Another way of putting this is as in (11):

(6.11) A morphologically relevant semantic distinction is any meaning contrast of a language x that is reflected in the *words* of x.

The terms we use at this point are not crucial. What we are after here is a counterpart to the philosopher J.L. Austin's (1962) famous *How to do things with words*, namely 'how to do things *to* words'. The most important rule of thumb in morphosyntactic field research is to *be alert* and *be creative* in your thinking, following *Postal's Maxims* (see ____).

We begin with an overview of the major semantic distinctions encountered on nouns, grouped under my headings:

6.3.1.1. REFERENCE: specific vs. referential vs. definite

This is a fundamental distinction that is perhaps a semantic universal, though its morphological reflex (marking on word forms) is less so.

Consider the following sentences:

(6.12) a. John bought a book about Vietnam.
b. John bought the book about Vietnam.

These examples are, respectively, indefinite and definite. In general (but see Lyons (1999)) the definite article indicates shared knowledge between speaker and hearer, or old information from the discourse. The indefinite is often used to introduce new participants into a discourse, or to offer a different type of quantificational reading. But there is more to them than this. In both of these examples the speaker can either have a very specific book in mind or not. For example, both can be continued as in (13) and (14):

(6.13) a. John bought a book about Vietnam, in fact he bought *Fire in the Lake*, by Frances FitzGerald.
a. John bought a book about Vietnam, but I don't know which one.

(6.14) a. John bought the book about Vietnam, or at least that is how he was describing it, I really don't know anything else about it.
b. John bought the book about Vietnam, you know, the one in the window we saw yesterday?

What is going on in these examples? Well, they show that in addition to definiteness, examples can be crossclassified by the notion of *referentiality*, that is, whether they refer to a specific individual, i.e. one known to the speaker. In some languages, this can be marked morphologically or syntactically. So consider the following example from Persian (from the LDSQ):

(6.15) a. *Hasan yek kita:b-ra: xarid*
Hasan bought a specific book
b. Hasan yek kita:b xarid
Hasan bought some book or other

Both of these examples are indefinite. But the special marking, **-ra** on **kita:b** 'book' in (15a) is there to indicate that the example is *referential*, that is, that it refers to a specific individual. Such distinctions, though 'only' semantic in English, can be vital to understanding the morphology of some languages.

Another semantic distinction vital for morphosyntactic analysis is between major participants of the designated verbal event or state:

6.3.1.2. GRAMMATICAL ROLES
GRAMMATICAL RELATIONS

Perlmutter & Postal () launched an extremely interesting and very influential research program on grammatical relations, Relational Grammar, by explicitly adopting the hypothesis that grammatical relations are syntactic primitives which are causally implicated in numerous syntactic operations, especially those based on VOICE (see ___) and INFORMATION STRUCTURE (see __). One of the appeals of Relational Grammar was that its concept of grammatical relations and their relation to syntactic operations was intuitively very appealing. Initially, at least, the theory also seemed to enjoy a great deal of empirical success and conceptual simplification in handling numerous operations that were much more clumsily handled in standard Transformational Grammar, for example. Many linguists still find it convenient to talk about notions like subject, direct object, indirect object, and so on. To see how such notions can be useful, consider that the passive in English and other languages can be generalized to a rule like ():

(6.16) Passive: Promote the direct object to subject position and demote the subject to an oblique relation.

(6.17) a. John hit Bill.
b. Bill was hit *by John* (John has been 'demoted' to the oblique object of *by*).

Unlike transformations that operate on syntactic categories, the use of grammatical relations rather than categories such as NP, PP, etc. as the basis of grammatical rules offers clearer and more satisfying cross-linguistic accounts of voice alternations. The reader is urged to consult Perlmutter & Postal (), and the large literature on Relational Grammar for further discussion of the usefulness of grammatical relations in understanding syntactic processes.

MACROROLES

An alternative to grammatical relations is proposed in Role and Reference Grammar (Van Valin and La Polla ()), under the label of MACROROLES. Macroroles are different from Grammatical Relations in that, among other things, they are linked simultaneously to the semantic roles specified in a verb's lexical entry and the ways in which the specific nature of these lexical roles are 'neutralized' in the syntax. Macroroles overlap with Grammatical Relations in many simple cases, but they are by no means isomorphic to Grammatical Relations.

For example, ACTOR is a generic term for all semantic roles that are associated with causing or controlling, etc., an action in a specific clause. That is, it is the label for the set of roles whose specific semantics are neutralized in a particular construction. UNDERGOER Is the label for the neutralization of roles associated with arguments that suffer the consequences of the action in some sense. Consider the following examples, from Van Valin (), with the actor vs. undergoer roles labeled:

(6.18) a. The *teacher* is reading the words. Actor of transitive V
b. The *teacher* is singing. Actor of intransitive V
c. The *teacher* fainted. Undergoer of intransitive V

Example (6.18c) is interesting because it shows that subjects of simple clauses in which there is no obvious dislocation can also be Undergoers, i.e. that subjects do not map directly to Actors.

(6.19)
a.*The *teacher* are reading the words. *Undergoer of transitive V
b. The *words* are being read by the teacher. Undergoer of passive V

Example (6.19a) shows that Macroroles do not directly govern agreement in English. In Relational Grammar this is because subjects, not macroroles govern agreement in English. However, Van Valin and La Polla (1997) argue that Grammatical Relations are really not what is needed to account for agreement and other apparently Grammatical Relation-sensitive processes in syntax. Consider other ways in which Macroroles may be manifested in the syntax.

(6.20) a. *Chris* wants to drink a beer. Actor of transitive V
b. *Chris* wants to sing in the park. Actor of intransitive V
c. *Chris* wants to be stronger. Undergoer of intransitive V
d.*$Chris_i$ doesn't want the journalist to interview $\_\_\__i$.*Undergoer of transitive V
e. *Chris* doesn't want to be interviewed by the journalist. Undergoer of passive V

This example is interesting because it shows that Macroroles can only govern verb agreement or appear in a dislocated position if certain conditions on Voice are met. The next examples show that English is quite liberal in the semantic roles that may

head relative clauses. In many languages only the Actor or the Actor plus the Undergoer roles can head relatives (all such cases are discussed in detail in Van Valin (2005)).

(6.21) Mary talked to the man (a) who [A] bought the house down the street.
(b) who [U] the dog bit.
(c) to whom [recipient] Bill showed the house.
Mary looked at the box (d) in which [locative] the jewelry was kept.
(e) out of which [source] the jewelry had been taken.

Verbs have a semantic structure that, among other things, specifies for a given verb how many arguments it needs and what the semantic relationship of those arguments is to the verb.[28] Languages rarely, if ever, mark these various lexical relations directly in the morphology or the syntax (though cf. VVLP's treatment of Acehnese (pp255ff)). More commonly, languages choose to 'neutralize' (VVLP pp251ff) some number of these distinctions by grouping the more finely-grained lexical argument distinctions under larger categories. Tagmemics (Pike & Pike (1976)) and Role and Reference Grammar (VVLP), suggest the labels 'Actor' and 'Undergoer' as the two basic 'macroroles' (VVLP, 139ff), under which the more specific semantic roles are neutralized. These macroroles can have important consequences for a language's mrophology, as seen in examples like (16)-(19), from Kulina, an Arawan language (Wright 1988):

(6.22) **Kodzo ts^{h}ite*i* -na -hari.**
lizard.M shoot***3on3*** -AUX-COMP.M
'S/he shot the lizard.'

(6.23) **Aoi dzoho *i* -na -haro.**
tapir.F carry ***3on3*** -AUX-COMP.F
S/he carried the tapir.'

(6.24) **Kodzo ts^{h}ite∅ -na -hari.**
lizard.M shoot***3*** -AUX-COMP.M
'**He** (*She) shot the lizard.'

(6.25) **Aoi dzoho ∅ -na -hari.**
tapir.F carry ***3*** -AUX-COMP.M
***'He** (*She) carried the tapir.'*

Wright (1988) takes the examples with the ***i-*** '3on3' prefix (third person Actor and third person Undergoer) ((16) and (17)) to be the basic/underlying forms. If one makes this decision, then in the basic form there is agreement with the object. In the 'derived' form, the agreement is with the subject and the object no longer triggers agreement. From this perspective, the ***i-***/∅ alternation's effect on gender agreement looks (somewhat) like an ergative/antipassive alternation (see __ below on ergativity and antipassive). On the other hand, if one takes the ∅-forms to be basic, and the ***i-***

[28] Arguments are expressions required to *saturate* a verb's valence (see ___ below).

forms to be derived, then the structure will look more like a passive (this is the analysis of Campbell and Campbell 1990 for Yamamadi (a language closely related to Kulina), where ***hi-*** is the cognate of ***i-***. It should be noted that that the ***hi-/i-*** prefixes only appear when both subject and object are third person (any number). When either or both is nonthird person, there is no special agreement prefix and only the gender alternation is seen. Cf. (20) and (21):

(6.26) **Kodzo ts^{h}ite*o* -na -haro**.
lizard.M shoot*1* -AUX-COMP.F
'I shot the lizard.'[29]

(6.27) **Kodzo ts^{h}ite*o* -na -hari.**
lizard.M shoot *1*-AUX-COMP.M
'I shot the lizard.'

In these examples the only visible difference is the gender of the verb. In (20) it is governed by the subject and in (21) by the object.

For our purposes, the crucial observation is that Arawan has morphological strategies for distinguishing different pragmatic functions of Actors vs. Undergoers. It is not necessary to label them as either 'passive' or 'antipassive'. One way to think of the alternations in Arawan, as well as related, but nonidentical, alternations in other languages, e.g. passive and antipassive, is that a language can select either the Undergoer or the Actor as the unmarked morphological (and pragmatic as well, in most cases) argument. When the topicality of that unmarked argument is altered (by becoming more or less topical), then the morphology will usually 'kick in' to signal this.[30]

6.3.1.3. POSSESSION: alienable, inalienable, intermediate

Languages often distinguish between *alienable* vs. *inalienable* possession, i.e. nouns that need not be marked as possessed in the morphosyntax and nouns which *must* be so marked (or, alternatively, must have a special form to signal that they are violating the normal expectation that they be possessed). Below I illustrate all of these points from Banawá and Wari.

As in other Arawan languages, there are two main syntactic types of possession in Banawá, alienable and inalienable. The former is subdivided into kinship vs. nonkinship. All nonkinship alienably possessed NPs include the clitic **ka** 'possession' occurring between the possessor and possessum. Kinship terms do not always take the **ka** clitic and in such cases are similar to inalienably possessed forms. They are *un*like

[29] First and second persons and all plural forms are feminine throughout the Arawan languages, regardless of the sex of the real-world referent.

[30] In addition, Actors, Undergoers, and other semantic relations, can be further neutralized by interposing another level of organization on the verb's arguments, namely, grammatical relations, e.g. Subject, Object, etc. But not all theories recognize these grammatical relations as having any usefulness to morphosyntactic analysis. Even theories as otherwise distinct as Role and Reference Grammar and Minimalism reject any role for grammatical relations in their view of grammar.

inalienably possessed forms, however, in that the gender of a possessed kinship term is determined by the kinship term rather than its possessor.

Alienable possession:

(6.28)	**Inaso**	**ka**	**yomai**	**rabikei**	**no,**	**bede**	**ka**	**-wei.**[31]
	Ignacio	ka	dog.M	sick.M		past, run	comit	-cont.M
	'Ignacio's dog got sick and ran away.'							

Inalienable Possession

(6.29)	**kosiba**	**efe**	**wiri**	**to**	**-witi**	**-ei**
	babassu:palm	leaf	strip	away	-end.point	-M
	The babassu palm leaf was stripped off to the end.'					

In example (6.29), the gender agreement on the verb can be controlled by the possessor (**kosiba** 'babassu palm'), but not the possessed item (**efe** 'leaf') and there is no **ka** possession marker allowed. These are two of the criteria for recognizing inalienable possession in Banawá. Banawá also shows a somewhat intermediate category of possession for kinship:

Kinship possession:

(6.30)	**oda**	**ka**	**ami**	**-rawa**	**oda**	**da**	**-daba**		**-bisa**	**towa**	**-maro**
	1P	ka	mother	-pl.F	1P	red	-accomp	-also	be	-past	
	'We stayed with our mothers also.'										

(6.31)	**ami**	**da**	**-daba**	**-bisa**	**towa**	**-maro**
	mother	-pl.F	red -accomp	-also	be	-past
	'You stayed with your mother also.'					

In example (25), the possessor of **ami** 'mother' is omitted, but trivially recoverable because **ami** is the suppletive term for second person's mother. With third person kinship terms (also marked by suppletion in Arawa) following Campbell (), the

---

[31] In personal communication, Alan Vogel suggests that the crucial portions of the examples in ()-() in the text should in fact be analysed as below:

() *bede kawei*
bede ka- na -wei
run COMIT- AUX -now+M

() *wiri towitiei*
wiri to- na -witiei
strip.off away- AUX -out+M

() *towamaro*
to- ha -maro
INCH- AUX -past

Vogel may well be right. But this relatively abstract analysis has not yet been demonstrated to my satisfaction.

possessor of the kinship term can be omitted if it is recoverable from the discourse, unlike normal alienable possession, but like inalienable possession. Thus kinship can occasionally behave intermediate between the two types. Everett (in progress) accounts for this in terms of qualia (see ___ below).

Kinship terms and possession in Wari' are discussed in detail in Everett and Kern (1997:434ff.). There are two formation classes of kinship terms in Wari'. The first includes what Everett and Kern (1997) call ***-xi'*** nouns, where ***-xi'*** is the first-person plural inclusive possession suffix. Only inalienably possessed nouns take possession suffixes and these in turn can be inputs to other word-formation rules, as discussed in Everett and Kern (1997:235ff.). In these forms, the citation or base for the paradigm is first-person plural inclusive. The first-person singular inflection of six of the terms is suppletive; otherwise they all inflect for possession with the ***–xi'*** suffix. The complete list of ***–xi'*** nouns, all in their '1pl inclusive' forms, is given below (first-person singular alienably possessed forms are in parentheses).

Simplex:

(6.32) a. *'aramanaxi'* (*'aramana*) 'sister' (lit., female of the species; male ego)
b. *cainaxi'* (*caina*) 'daughter' (female ego)
c. *cawinaxi'* (*cawiji*) 'son' (female ego)
d. *humajixi'* (*humaju*) 'children' (female ego)
e. *japinaxi'* (*wijapi*) 'wife's mother'
f. *jinaxi'* (*jina*) 'granddaughter' (male ego or female ego's daughter's daughter)
g. *manaxi'* (*mana*) 'wife' ('hole'; *mana* 'my wife/my hole')
h. *namorinaxi'* (*namori*) 'wife's sibling'
i. *tamanaxi'* (*tamana*) 'husband's mother'
j. *taramajixi'* (*taramaju*) 'brother' (male of the species; female ego)
k. *taxixi'* (*taxi*) 'husband'
l. *wijinaxi'* (*wiji*) 'grandchild' (female ego's son's child)
m. *winaxi'* (*wina*) 'grandson' (male ego or female ego's daughter's son)
n. *xerexi'* (*xere*) 'sibling'
o. *xinaxi'* (*wixi*) 'sister's son' (male ego)
p. *xi'* (*na*) 'mother'

The only way to understand how '***–xi'*** nouns' differ from other nouns semantically and morphologically in Wari' is to recognize the (non-universal) semantic distinction of inalienable possession (note that it is *not* necessary to understand the semantic basis per se as to why a given noun is an inalienably possessed noun. This can be helpful, but for diachronic reasons can be fairly asystematic).

Moreover, even languages that do not have inalienable possession morphologically can show traces of it in their syntax, as the following examples from Portuguese show:

(6.33) Joao operou a perna.
(i) John$_i$ operated on the or his$_{i/j}$ leg.
(ii) John's leg was operated on.

In this example, quite common in Brazilian Portuguese, the 'passive' reading in (6.33ii) is available for part-whole relations, the kind that are often the basis for inalienable possession. Consider also:

(6.34) O carro furou o pneu.
'The car's tire was punctured' (literally, 'The car punctured the tire.'

This example, also common in Brazilian Portuguese, is possible because ***pneu*** 'tire' is part of ***carro*** 'car'. Everett () explores these constructions in detail. The relevance of part-whole/inalienable possession relations to the syntax of such constructions is shown by ungrammatical examples like ():

(6.35) *A casa roubou o carro.
'The house was robbed of the car.'

Although the syntax of this example is the same as (6.34) (perhaps coming from an underlying form e.g. ___ ***roubou o carro da casa***, again, see Everett () for details), the example is ungrammatical because () does not involve a part-whole relation (cars are not parts of houses).

Notice too that wherease inalienable possession shows up (mainly) in the morphology of Wari' and Banawá, it appears in the syntax of Brazilian Portuguese. This again shows how semantic distinctions should not be confused with their modes of expression. Form is not function.

From the semantic category of possession, let's move to consider the category of number.

6.3.1.4. NUMBER: singular, plural, intermediate

Apparently all languages of the world, except Pirahã (Corbett (2001)), mark number morphologically. The most common distinction is singular (one entity) vs. plural (more than entity). But there are many languages that make intermediate number distinctions, e.g. 'dual' (two) and 'paucal' (3-5 or so), in addition to plural vs. singular. Therefore, when eliciting number, be sure to test for all these categories, especially if you already know that other languages of the area or the family make such distinctions. You can do this fairly simply by eliciting phrases, e.g. 'One man came'; 'Two men came'; 'Three men came'; etc. plus 'I saw one man'; I saw two men'; 'I saw three men'; up to five at least (be sure to collect numbers in at least the subject and object positions). This will cover most of the possibilities. However, once again, always 'keep your guard up' for distinctions that didn't show up in initial, simple tests, but which might appear in more subtle ways, i.e. elsewhere in the phrase, as you progress in your understanding of the language.

Another important, number-based semantic distinction made in most languages is the distinction between *count* vs. *mass* nouns, i.e. nouns that can be counted and nouns that cannot be, at least grammatically. So consider English examples like 'Some water' vs. *'Two waters' (where 'water' refers strictly to the subtance H_20, and not some standardized forms for containing or channelling it, e.g. cups or rivers) or 'Two arrows' vs. *'Some arrow'. Although linguists almost universally attribute these distinctions to countability, in my view, they may arise from something even more basic than counting, namely, visual perception. For example, the Pirahã language of Brazil (Everett 2005) lacks number, numerals, and counting, yet it apparently has a 'count' vs. 'mass' distinction. This could be evidence for the hypothesis that 'count' vs. 'mass' is a distinction of some 'universal grammar', so that it is found even if a language lacks means of expression for the related cognitive faculty of counting. But the simpler

hypothesis is that this derives from the fact that some entities are easier to perceive as made up of smaller units (i.e. they can be 'individuated'), whereas others are not so obviously decomposable visually. Thus it is possible that the distinction here has nothing to do with universal properties of grammar but, rather, with visual perception. The point is that although the fieldworker should be aware of distinctions drawn in a huge percentage of the world's languages, they must always exercise creativity, thinking 'laterally' to use a term from de Bono (1973).

The next important semantic distinction languages often make in one way or another on nouns (or their containing phrases; see section ___ below on ways of marking semantic distinctions) is class membership.

6.3.1.5. CLASS MEMBERSHIP: gender and more

The classes into which languages may divide their nouns are likely to have some sort of semantic motivation, at least diachronically. But in general, this motivation is obscured, perhaps completely, by normal historical change of the language.

The simplest class distinction is masculine vs. feminine. Languages that use this simple class system, as well as those that include a third class of 'neuter', are said to have gender systems. This is fine as a label, but the linguist should not forget that this is nothing more than a simple class system and that it is absurd to think that speakers of a language with 'gender' actually see every object in their universe as male or female. A number of objects in each linguistic gender class will indeed have a link to male or female in the real world (e.g. 'man', 'woman', 'girl', etc.) but certainly not all. For example, in Brazilian Portuguese, one says *A cola* 'The (feminine) cola', but *O guarana* 'The (masculine) guarana', where in this case guarana and cola are just two forms of softdrink. No Brazilian I have ever talked to thinks that *cola* is female while *guarana* is male in the real world. More complex classification systems have less obvious semantic unity in each class, by and large, and it is almost always a waste of time (i.e. a violation of the 'law of diminishing returns') to look for semantic motivation for each member of each class. What the field linguist really needs to understand is how these classes, such as they are, are causally implicated in the morphosyntax of their containing clauses and discourses. This is not to say that it will always be pointless to look for the semantic motivation behind class membership, not at all. But it is the case that looking for such motivation is less likely to be useful to initial fieldwork than carefully describing how the classes work structurally.

Class is often marked as a separate morphological category. For example, consider how gender can be marked in the Arawan language, Banawá. Banawá is interesting in that person and number agreement are marked on the rightmost boundary of the verb, i.e. as prefixes or proclitics, whereas gender is marked at the end of the verb, as a suffix. Moreover, gender agreement is not directly governed by either the subject or the object, but by the topic of the sentence which can be subject, object, or the inalienable possessor of the subject or object (see Everett (in progress). Dixon (2004) discusses similar facts in the related language, Jarawara).

We saw some examples of Arawan gender marking earlier from Kulina (()-()). When the Actor is the topic (what Dixon () refers to as the A-construction), the Actor governs gender agreement on the verb:

(6.36) a. kobaya **mere** waka-nei
pig wild:M 3P.OBJ break-AUX:M
'He killed pigs.'

b. **o$_i$ -man$_k$ wete ∅$_j$ -ka -*sei*$_j$**
1S.F -arm.M tie.up 3 -comitative -aspect.M
'He tied up my arms.'

When the Undergoer is the topic (Dixon's () O-construction), it controls gender agreement:

(6.37) **dee** ka -maki yomai owa nabowa bone[32]
2P MOT -follow:F jaguar:M 1S.OBJ kill INTENT:F

awa -tee o -ka abi ati
maybe -RECENT.PST.QUOT+F 1S -POSS father say

-nei
-AUX:M

'You all come here. The jaguar will kill me, my father said.'

But, just as with Kulina, when the Actor or the Undergoer is inalienably possessed, its possessor controls gender agreement:

(6.38) **Mate, mate ime yama hi- kaba metemone**
buttock:F buttock:F flesh:F thing:F 3- eat long.ago:F
mate
buttock:F

hodi wa -metemone, mate tone wa
hole:F stand -long.ago:F buttock:F bone:F stand -

-fara -metemone.
in.open.area -long.ago:F
'The flesh of her upper thigh was being consumed by something, it was open, and her thigh bone was exposed.'

When the Undergoer is treated as topic and where the Undergoer is inalienably possessed, the possessor of the Undergoer, rather than the Undergoer itself, governs gender agreement:

Possessor (of Undergoer) controls gender agreement:
(6.39) **o$_i$ -mano$_j$ wete hi -ka -sa$_i$.**
1S.F -arm.M tie.up 3/3 -comitative -aspect.F
'My arms were tied up by him.'

[32] Vogel (p.c.) also says that **bone** 'intent feminine' should in fact be **bona** 'intent masculine'. But my source is fairly clearly **bone.** I suspect that Vogel is right here, knowing far more than I about these languages, but I leave it this way just the same.

These examples are very useful because they show the interaction of possession types, voice alternations, and gender. Each of these semantic distinctions and the way that each is marked must be carefully analyzed or their interactions will be mysterious.

Everett () argues that the simplest analysis of the facts is that gender agreement is pragmatically governed, whereas person and number agreement are syntactically governed. Pragmatic government of the process requires the whole rather than the part to be the topic. That is, to consier (), you cannot tie up my arms without affecting me. It is 'my' that is the topic, therefore, not 'arms'. The point relevant to this guide is that CLASS is subtle, has effects throughout the clause, and care must be taken in its analysis.

6.3.1.6. PRAGMATICS

Arguably, transitivity alternations in languages have the function, among others, of tracking the pragmatic functions of nominals.

It is useful for languages to mark such alternations. We have already seen examples of this from Salish (() – ()) and Arawan (() – ()). The point of this section is simply to remind the fieldworker to look for a variety of possible correlations between pragmatics and morphology. And the fieldworker must be aware that there are also important morphology-pragmatics connections unrelated to information structure, e.g. politeness markers and honorifics. So consider the examples below from Brazilian Portuguese and Japanese:

(6.40) Tu vai(s) para o cinema hoje, rapaz?
'Are you (most informal) going to the cinema today, fellow?

(6.41) Voce vai para o cinema hoje, rapaz?
'Are you (informal) going to the cinema today, fellow?'

(6.42) O senhor vai para o cinema (, % rapaz)?[33]
'Are you (formal) going to the cinema?'

Generally dialects of Brazilian Portuguese differ as to whether the informal usage is expressed by ***tu*** or ***voce***. However, in some dialects, there is a three-way distinction, with (34) least formal and (36) most formal. Although not all dialects of Brazilian Portuguese that use ***tu*** mark the verb differently (hence the parentheses on the ***s*** in ()), some do. These examples thus show that politeness forms can affect word-formation.

In Japanese, on the other hand, there is a much more elaborate pragmatic-based system, with the forms referred to as honorifics (Potts & Kawahara (2004, 1)):

Japanese Honorifics

(6.43) a. **Sam -ga warat -ta.**
Sam -NOM laugh -PAST
'Sam laughed.'

b. **Sam -ga o -warai -ninat -ta**
Sam -NOM subj.hon -laugh -subj.hon -PAST

[33] % indicates that the addition of this word would be pragmatically strange.

i. 'Sam laughed.'
ii. 'The speaker honors Sam.' [subject honorific]

c. **Sam -ga warai -yagat -ta**
Sam -NOM laugh -antihon -PAST
i. 'Sam laughed.'
ii. 'The speaker views Sam negatively.' [antihonorific]

The authors go on to illustrate how honorific marking can be used to mark speaker attitude not only to other individuals but to the content of propositions as well (p2ff):

(6.44) **Kathyrn -wa Sam -o hai -ken -shi -nakat -ta**
Kathryn -TOP Sam -ACC obj.hon -see -do -not PAST
i. 'Kathryn did not see Sam.'
ii. 'The speaker respects Sam.' [object honorific]

"*We divide our attention between these cases and the honorifics represented in (39) and (40)*[my numbers, DLE]*, which associate semantically with some aspect of the propositional content rather than with the denotation of an argument nominal."*

(6.45) **Mary -ga ringo-o tabe -mashi -ta**
Mary -NOM apple -ACC eat -perf.hon -PAST
i. 'Mary ate the apple.'
ii. 'I am speaking nicely to you.' [performative honorification]

(6.46) **John -wa [Mary -ga nesugoshi -chimat -ta] -koto-o**
John -TOP Mary -NOM oversleep -antihon -PAST –fact

shitteiru.
know
i. 'John knows that Mary overslept.'
ii. 'It sucks that Mary overslept.' [antihonorification]

As they continue:

"*The first is called performative honorification in Harada (1976) (certainly a suggestive name). It goes by the name 'polite speech' as well. We call the second kind antihonorification. Here, the morpheme* ***chimau****, which appears as* ***chimat-****, signals that the speaker has contempt for the proposition expressed by the clause in which it appears."*

The Japanese case thus illustrates that pragmatic markers like this can be 'co-opted' for other grammatical functions. Therefore, such categories are doubly important to recognize, with potential ramifications that go beyond their original functions.

6.3.1.7. CASE

Verbs are said to 'govern' cases, which means that in some languages verbs partially determine the shape of the nouns that are specified in their lexical representation. English has vestiges of this in its pronoun system:

(6.47) a. I saw him.
b. He saw me.
c. *Me saw he.

Of course, anyone who has studied Latin, Greek, Finnish, or any one of a huge number of languages will have learned about case systems much more complex than English. Case is an interesting feature because it is in a sense 'displaced'. It is perhaps best understood as a verbal feature that is borne by nouns. That is, what the case will be on a noun depends on what the verb is. But the verb itself does not bear case. In Minimalism (Chomsky 1995), case is one of the 'uninterpretable' features. But these features in Minimalism seem to be exactly those features that are displaced in this specific sense. To understand case in a language requires a solid understanding of verbal lexical structure.

Case is functionally straightforward. It is used to help interlocutors track actions and their participants. Consider the template for a simple transitive clause (order irrelevant):

(6.48) S V O

Now in a language like English, with rigid word order, the use of additional word markers (case affixes or prepositions) to indicate that one NP is Subject or Actor and the other is Object or Undergoer is unnecessary. A language could distinctively mark these participants but it does not need to. On the other hand, if a language allows its nominals to be ordered more freely than English, a morphological marker on one or both arguments of the verb becomes very useful. Marking just one of them is more efficient (a lesser effort to results ratio) than marking both, although all possible options are found in natural languages. The three possibilities are shown in (43):

(6.49) a. S_{mark} V O
b. S_{mark1} V O_{mark2}
c. S V O_{mark}

The system in (6.49a) is called 'ergative' or 'ergative-absolutive'. The case that is marked on the subject is called the *ergative* case and the unmarked case on the object is called the *absolutive* case. The case marked on the object in (6.49c) is called the *accusative* case and the unmarked case on the subject is referred to as the *nominative* case. The (6.49b) example is rarer, because less efficient, but can fall into either ergative or accusative patterns or yet a separate kind of pattern. Part of knowing how to interpret (6.49b) is learning how the language marks the single argument of an intransitive clause:

(6.50) a. S_{mark} V
b. S V

Marking of the sole argument of an intransitive clause is unnecessary, so (b) is the most common case. When coupled with (c) the null case on the S is called nominative. When coupled with (a) it is called the absolutive. When coupled with (b), if the case of the intransitive S matches the case of the transitive S, we have a nominative-accusative system. If it matches the case of the O, then the system is ergative. If it matches neither, the system is neither. (But again, labels are not all that crucial. Clearly presented, accurate analyses and descriptions are.)Variations are possible, the two most common being split-ergativity and active-marking.[34]

In a split-ergative system, the case marking is nominative-accusative for some purposes and ergative-absolutive for others. From our definition in () below of *prototypical transitivity*, it is not surprising to learn that the dividing line is often punctiliar past and/or definiteness of the arguments. In general, in split-systems, ergative marking is more common with the punctiliar past and nominative-accusative with deviations from prototypical transitivity. But there are other ways to divide the case-marking system. So *beware*. One final observation is that languages do not have to exploit case-marking to show at least traces of ergativity. In Wari' for example (Everett & Kern (1997, 331ff), many verbs take a suppletive form when the Object or the intransitive Subject is plural. Since those two are grouped together by this process, the process is ergative in a sense. Or consider (from Lyons ()) the vestiges of ergativity in English:

(6.51) a. John ate breakfast.
b. John ate.
c. *Breakfast ate.

(6.52) a. John broke the window.
b. The window broke.
c. *John broke.

Verbs like *break* and *eat* in English, as in many other languages appear in transitive and intransitive clauses. However, when they do, the *eat* verbs allow the Actor to appear as their sole argument, never the Undergoer (c), while the *break* verbs allow only the Undergoer as their sole object, never the Actor (c). Therefore, the *break* verbs group together the Subject of the intransitive form with the Object of the transitive form and, as Lyons () so rightly pointed out, they are ergative verbs in this sense.

Another kind of distinction, the last we consider for now, is what some call an *active-stative* case system. Consider the hypothetical examples in (47):

(6.53) a. John$_{mark}$ fell down (on purpose/*accidentally).
b. John fell down (*on purpose/accidentally).

What is going on in such examples? In this type of system, volitional Actors (acting intentionally) are case-marked while non-volitional Actors (things happen to

[34] The best crosslinguistic typological comparison is WALS, Haspelmath, Martin & Dryer, Matthew & Gil, David & Comrie, Bernard (eds.) 2005. The World Atlas of Language Structures. (Book with interactive CD-ROM) Oxford: Oxford University Press. ISBN: 0-19-925591-1.

them without their intention), are not case-marked. Such systems of case-marking are called active-stative.

The key to case analysis is to be alert for ways that nominals are being grouped together. This is rarely arbitrary (modulo vestiges of some historical process now almost gone from the language). Labels are not that important initially. Recognizing patterns is.

Let's move now to another important semantic domain with morphosyntactic consequences for nominals, *qualia*.

6.3.1.8. QUALIA

Morphosyntactic research must also pay attention to a semantic set of categories used in classifying nouns in subtle ways. Pustejovsky (1996, 76ff and 85-105) discusses these in a detailed and helpful manner, labeling them *qualia.* As Pustejovksy (85) explains these, the qualia are the four factors which 'drive our basic understanding of an object or a relation in the world. They furthermore contribute to (or in fact determine) our ability to name an object with a certain predication'.

What this means for the fieldworker is that until they at least roughly understand the qualia of a particular lexical item, they cannot fully understand that item's appearance in sentence structure. This is relevant for morphosyntactic field research because subregularities in word behavior, both in verb and nominal or NP structures (e.g. selection of determiners, numerals, and, occasionally, affixation), can be determined by the qualia of a given item. There are four qualia, according to Pustejovsky (85-86).

(6.54) "1. CONSTITUTIVE: The relation between an object and its constituents, or proper parts.

i. Material
ii. Weight
iii. Parts and component elements

2. FORMAL: That which distinguishes the object within a larger domain.
i. Orientation
ii. Magnitude
iii. Shape
iv. Dimensionality
v. Color
vi. Position

3. TELIC: Purpose and function of the object.
i. Purpose that an agent has in performing an act.
ii. Built-in function or aim which specifies certain activities

4. AGENTIVE: Factors involved in the origin or 'bringing about' of an object.
i. Creator
ii. Artifact
iii. Natural Kind
iv. Causal Chain"

For example, it is possible that qualia can be reflected in noun classification. It is likely that qualia can cause some nouns to behave in otherwise confusing ways, so that they would group with some nouns for some purposes according to qualia, but with others, according to, say, gender/class. Consider, for example, the following examples from Brazilian Portuguese:

(6.55) *João usou uma faca nova no frango*.
'John used a new knife on the (cooked) chicken.'

(6.56) *João usou um terçado novo no leitão*.
'John used a new machete on the (whole cooked) pig.'

Because of what *faca* 'knife' and *terçado* 'machete' mean at the level of qualia, we can use a generic verb like ***usar*** 'to use' and still know that cutting/chopping is being talked about. If a language morphosyntactically distinguishes, say, an aspectual difference for cutting vs. other actions, then () and () could function identically with regard to aspect, as a result of the qualia of the nouns involved, yet fall into separate classes with respect to gender (as in this case, feminine vs. masculine, respectively).[35]

Or consider an analysis of the Banawá possession facts in terms of qualia. As Pustejovksy (1995, 79) observes, "The CONSTITUTIVE (henceforth CONST) quale refers not only to the parts or material of an object, but defines, for an object, what that object is logically part of, if such a relation exists." (p98)

To illustrate let us use Pustejovky's (1995, 99) lexical structure for **hand**:

(6.57)

hand		
ARGSTR =	ARG1 =	**x:limb**
QUALIA =	FORMAL =	**x**
	CONST =	**part_of(x,y:body)**

"That is, the relation in the CONST allows for reference to what something is constituted of as well as what it constitutes, in part; i.e., a hand is part of a body, and a body has a hand..." (Pustejovksy (1995, 99))

My conclusion is that topic-agreement (gender) in Banawá picks out the index of the whole (*body* in (51)) in the qualia structure for its referential index. For inalienable possession this will be the possessor (that is the whole) but for alienable possession it

[35] This has been useful to me in my on-going analysis of Banawá possession. Topic-agreement (gender) picks out the index of the whole in the qualia structure for its referential index. For inalienable possession this will be the possessor (that is the whole) but for alienable possession it will be the possessed item (because that is the whole). Moreover, this allows us to say why kinship is intermediate. It is like inalienable morphologically (the cases without ka(a)) but like alienable syntactically (in terms of agreement). The latter fact is because the possessor of a kinship term is not the whole. The former fact is because it IS inalienable possession without ka(a). In other words, there are two things happening: (i) inalienable vs. alienable possession and (ii) part-whole relations. Each is important in different ways.

will be the possessed item (because that is the whole). Moreover, this allows us to say why kinship is intermediate. It is like inalienable possession morphologically (the cases without **ka**) but like alienable syntactically (in terms of agreement). The latter fact is because the possessor of a kinship term is not the whole. The former fact is because it IS inalienable possession without **ka**. In other words, there are two things happening: (i) inalienable vs. alienable possession and (ii) part-whole relations. Each is important in different ways.

Other semantic distinctions may be less 'deep' or not inherent. One such property is comparison.

6.3.1.9. COMPARISON: more, less, better, best, etc.

Comparison is a ranking of objects relative to one another and is thus not inherent in a single object, by and large, as qualia are. Not all languages have comparison as a morphosyntactic category, but most do. So consider English:

(6.58) a. A good man.
b. A better man.
c. The best man.

Not only is comparison reflected on the modifier of the English NP, choices of comparison (e.g. (c)) can also affect other characteristics of the NP, e.g. marking for definiteness. Comparison, like other semantic categories, may be marked on the N, NP, Verb, or Clause (section ___ below). Be aware that regularities on one word may be caused by or associated with another word. Therefore, it is vital once again that phonology not be confused with morphology (see ___ above). Let's move now to consider a different type of semantic categorization which can affect the morphology of a given language, namely, possible distinctions for different classes and subclasses of words, e.g. common nouns, proper nouns, pronouns, abstract nouns, concrete nouns, quantifiers, number words, etc. All or any of these might play a significant role in the morphosyntax of the language under study. We consider just a few possible subclasses of nouns here.

6.3.1.10. NOUN SUBCLASSES

PROPER NOUNS:

Proper nouns are those which name specific individuals, e.g. *John*, *Fido*, *Bank of America*, *Brazil*, and so on. Morphosyntactically they *usually* do not cooccur with definiteness marking (since they are inherently definite), plurals, or much other morphology usually associated with common nouns.[36] And they behave differently in discourse in some if not all languages. Recognizing them, however, is not always as easy as it sounds. In Pirahã, for example, all proper nouns are morphologically complex or derived from common nouns. Here are a few:

(6.59) a. ʔahoaógií 'big night'

[36] But consider Brazilian Portuguese examples like ():

(i) Vamos na casa *do* João.
'Let's go to *the* John's house.'

b. Kohoibiíhai 'blood eater'
c. Toisií 'eagle'
d. Piihoábigí 'frog' (literally: 'water-side-dirt')
e. Piiopí 'capyvara' (literally: 'enter water')

The first proper noun in (53) is a modified noun, the second a nominalized verb + object, the third comes from a common noun, the fourth from a phrase which is most commonly used as a common noun, and the last from a verb + goal, and is used commonly as the noun for capyvaras. In other words, there is little in Pirahã to distinguish proper nouns from other nouns morphosyntactically, at least in isolation. Only in context can proper nouns be distinguished (mainly by syntactic position and the type of clitic used preverbally), as in (54):

(6.60) a. *Piiopí* *hi* *ʔiaagá*.
name masc. is hungry
'Piiopí is hungry.'

b. *Piiopí* *ʔísi* *ʔiaagá*.
animal
'The capyvara is hungry.'

The proclitic on the verb tells us whether the noun is a proper noun (referring to a person named capyvara), as in (54a) or a common noun (a capyvara) as in (54b).

PRONOUNS, CLITICS, AND AFFIXES

The first caveat offered at the beginning of this chapter is that the fieldworker must be on guard to avoid confusing phonological facts with morphological facts. This is perhaps nowhere truer than in the analysis of clitics, pronouns, and nominal affixes. To get an idea of why this is hard, consider the following examples from Wari' (Everett ()):

(6.61) **Quep** ***na*** ***-in*** **xirim** **te** **pane** **ta**.
do *3s:rp/p* *-3n* house father:1s rem:past emph
'My father made a house long ago.'

(6.62) **Ten** ***ta*** **wao'**.
weave *pass:3s* type of basket
'Baskets are woven.'

(6.63) **Mi'** ***non*** ***-on*** **con** **hwam** **hwijima'** **mon** **tarama'.**
give *3p:rp/p* *-3pm* prep:3sm fish children coll man
'The men gave the children fish.'

The italicized material almost always is found to the immediate right of the verb, as in the examples shown. Should this material be analyzed as a sequence of pronouns, as clitics, or as suffixes on the verb? Such decisions are non-trivial and require consideration of syntactic, morphological, phonological, and, occasionally, semantic evidence. In the case of Wari', we conclude that these are clitics for three main reasons: (i) Vowel Harmony operates within words, but is blocked from spreading from a verb

to an agreement marker, i.e. across a clitic boundary; (ii) stress falls on the last syllable of each word, but clitics are never stressed as part of any adjacent word, receiving either no stress or an independent stress in clitic 'clusters' (as in ()-()); (iii) these constituents may follow embedded sentences, which are not words, and thus can appear without any potential word for them to affix themselves to (Everett (in progress)). Perhaps the best summary of the differences between clitics and affixes is given in a well-known article by Arnold Zwicky and Geoffrey Pullum (19--). Another useful discussion on distinguishing clitics, affixes, and pronouns (or words more generally) comes from the article by Anderson and Zwicky () in the second edition of the Oxford International Encyclopedia of Linguistics:

"Since the unusual properties of clitics ... are bound up with their ambiguous status between affixes and words, we may consider some criteria which distinguish affixes (determined, bound, reduced) from words (undetermined, free, full):

(a) The typical word, but not the typical affix, has an independent accent.

(b) The phonological shape of a word must be listed in the lexicon, but the phonology of an affix is described in general by saying how the shape of some stem is altered (so that affixes can have 'process', as well as affix, realizations) [See Affixation].

(c) Separate, language-specific restrictions can govern the possible phonological shapes of words vs. affixes. In particular, affixes, but not words, are often non-syllabic.

(d) Syntactically, words belong to (lexical) categories, i.e., word classes; but the assignment of affixes to such categories is problematic...

(e) Syntactic rules introduce word classes as co-constituents with other syntactic categories; but an affix is syntactically dependent, described by rules which locate it by reference to syntactic elements (e.g., on Nouns, on the head of VP, in the first word of S, at the right edge of NP).

(f) For each affix, morphological rules specify the class of words with which it can occur, and the properties of the resultant combination; but the syntactic rules distributing words typically make reference to phrasal categories rather than to word classes. From this, it follows that affixes are typically very selective in the word classes with which they occur, but words are unconstrained with respect to the word classes that happen to occur adjacent to them.

(g) Syntactic rules cannot alter morphological structure. In particular, syntactic rules cannot allow a word to interrupt a stem + affix combination; a word attached to such a combination must have edge position.

(h) Syntactic rules which introduce a lexical category are blind to the morphology and phonology of its co-constituents, but rules which introduce an affix may be contingent on such properties of its stem. From this it follows that there can be arbitrary gaps and morphophonological idiosyncrasies, including suppletion [q.v.], in the set of stem + affix combinations, but not in the set of word + word combinations.

(i) Alternative orders of words within a constituent are common, but the ordering of an affix – with respect to its stem, and to other such affixes – is fixed (although the same affixes may combine in different orders to express different meanings: e.g., the passive of a causative is not the same as the causative of a passive, and the affixers involved may reflect this in their ordering)."

The question here, of course, is how do we formalize such observations and turn them into rules of thumb for fieldworkers? First, consider the basic structures distinguishing clitics from affixes proposed in Everett (1997, PAGE) (in these examples ()s = phonological boundaries and []s = morphological boundaries):

(6.64) a. Y is a free word ([X]) ([Y])
b. Y is a clitic ([X] [Y])
c. Y is an affix ([X [Y]])

Affixes are morphologically parts of words. Clitics are phonologically dependent (though their placement may also be syntactically constrained, e.g. Everett (1989) argues for Yagua). Free words are independent in both senses. The problem is how to determine the type of dependency, morphological or phonological. Thus, by the view represented in (58), the terms 'clitic' and 'affix' are nothing more than labels for morphophonological relationships. It is those relationships which are crucial for the field researcher. Zwicky and Anderson's discussion offers very useful considerations for this.

The three questions the fieldworker must answer, put in a very general way, for affixes, clitics, and free words are:

(6.65) a. Does the item behave like a free word?
b. Or does it behave like an affix?
c. Or does it have mixed qualities, e.g. behaving like a word syntactically but like an affix phonologically?

Put in this way, it is clear that correctly understanding a particular item as a clitic, affix, or free word requires knowledge of linguistics, the language, and, especially, how to define a word in the language under study. Let us turn now to consider another subset of nominals/nominal modifiers (depending on the language).

NUMERALS AND QUANTIFIERS

There are also special classes of nouns that can vary a good deal from language to language as to whether they have idiosyncratic or relatively normal noun behavior. These include, but are not limited to, the following: number words, quantifiers, abstract nouns, and color words.

To take number words as the first example, consider the following examples from Russian (Rappaport (2002, 329)):

> *"Numeral phrases in Russian have long presented a puzzle to those attempting a theory of the syntax of this language. The puzzle itself is not difficult to formulate. When the phrase as a whole stands in a position to which the syntax assigns a direct case (nominative or accusative), the numeral acts like a nominal head of the phrase, assigning the genitive case to the quantified noun (and its modifiers) as if it were a complement. When this same phrase stands in a position assigned an oblique case, the quantified noun acts like the head of the phrase, with the numeral falling into line with the other modifiers in agreeing with that quantified noun. For example, contrast (60a,b)*[my numbers, DLE]*:*

(6.66) a. **videt' pjat' krasivyx ptiček.**
to see five$_{ACC=NOM}$ beautiful$_{GEN.PL}$ birds$_{GEN.PL}$
'to see five beautiful birds'

b. **vosxiščat'sja** **pjat'ju** **krasivymi** **ptičkami**.
to-be-enthralled-by five$_{INST}$ beautiful$_{INST.PL}$ birds$_{INST.PL}$
'to be enthralled by five beautiful birds'"

Or consider the English quantifier *all*:

(6.67) a. All the children came today.
b. *The all children came today.
c. The children all came today.
d. *The children came all today.
e. *The children came today all.

The English quantifier, *all*, 'floats', that is, it can appear in different positions in the phrase, unlike other words. In this case, it shows no morphological irregularity, though its syntax is clearly linked to its morphological class. Therefore it is important in classifying words, part of morphological analysis, to determine what special features they have phonologically, morphologically, or syntactically.

6.3.1.11. NOMINAL TENSE

Fieldworkers will expect certain semantic categories to be marked on a closely corresponding set of syntactic categories for the most part. For example, it is natural to expect tense to be marked on verbs because a principal function of tense is to situate events, expressed almost always by verbs, in time. On the other hand, tense is not marked exclusively on verbs crosslinguistically. For example, a word like **ex-wife** in English arguably marks the tense of 'wife' as past by the prefix **ex-**. Nordlinger & Sadler () have studied the distribution and typology of nominal tense extensively. They say that:

"Although the categories of tense, aspect and (to a lesser extent) mood are traditionally considered to be properties of verbs, the morphological expression of tense, aspect and/or mood (henceforth TAM) on nominals is attested across a range of languages. We distinguish two major subcases of nominal tense, that is, of cases in which TAM marking occurs on a nominal or other constituent of NP/DP. In some cases, a dependent nominal or nominal phrase (of whatever grammatical function) bears some TAM marking which serves to temporally, aspectually or modally specify the clausal predicate which is itself distinct from the nominal argument. This is illustrated by examples (62)-(63)[my numbers, DLE]*) in which the case marking of the dependent NPs changes to reflect the tense (future vs. nonfuture) of the clausal predicate:*

(6.68) Ngamari -lu ngunytyi -ka
mother -ERG give -PAST

ngali -nha mangarni -marru -nga -nha
we.DU -ACC bone -having -GEN -ACC

kathi -nha.
meat -ACC.

'Mother gave us the doctor's meat.' (Pitta Pitta (Australia), Blake 1987:60, ex. 4.12)

(6.69) Ngamari -ngu ngunytyi ngali -ku
mother -NOM.FUT give we.DU -ACC.FUT

mangarni -marru -nga -ku
bone -having -GEN -ACC.FUT

kathi -ku.
meat -ACC.FUT.
'Mother will give us the doctor's meat.' (op. cit., ex. 4.14)"

Thus, initial expectations based on training notwithstanding, we must be careful in fieldwork to look for tense on categories other than the verb. In general we must look for the possibility that many of our semantic categories will not match up as expected with our expected morphosyntactic categories.

6.3.2. VERB CATEGORIES

We now turn to consider the hardest nut to crack in morphological fieldwork, the verb. As one example, verbs in Pirahã have sixteen suffix positions (giving 2^{16} possible verb forms, modulo semantic incompatibilities), none of which include tense or number or person. Pirahã affixes encode notions like evidentiality, aspect, directionality, negation, and mood, among others (see Everett (1986, ---)). So although Pirahã's verbal morphology is very complex, one of the most complex systems I have encountered, its morphological categories overlap surpisingly little with those of English.

Nevertheless, many categories are commonly marked in verbal morphology cross-linguistically and I want to survey those very briefly here.

6.3.2.1. TENSE

Tense is a deictic ('pointing') category that locates an action or event with respect to the moment of utterance (absolute tenses) or another tense referenced on the verb (relative tenses). (See Comrie (1985)).) Consider the examples from Portuguese:

(6.70)
a. Ele fala. 'He speaks.'
b. Ele falou. 'He spoke.'
c. Ele falará. 'He will speak.'
d. Ele falara/tinha falado. 'He had spoken.'
e. Ele terá falado. 'He will have spoken.'

The first three tenses are absolute. This is because they are defined with respect to the moment of utterance, i.e. when they are spoken. So each means 'he speaks as I am speaking' or 'he spoke relative to my current speaking' or 'he will speak after I am done talking now', respectively. The last two examples are relative tenses. They refer to events situated in relation to other events. So you might use (64d), for example, in a sentence like (6.71):

(6.71) Ele tinha falado quando eu cheguei na sala.
'He had spoken when I arrived in the room.'

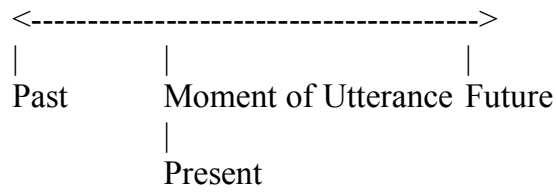
And you might use (64e) in a context like:

(6.72) Ele terá falado quando voce terminar de comer.
'He will have spoken when you finish eating.'

In (6.71) and (6.72) the events in the initial clauses are situated temporally in relation to the events of the second, embedded clauses, rather than directly to the moment of utterance.

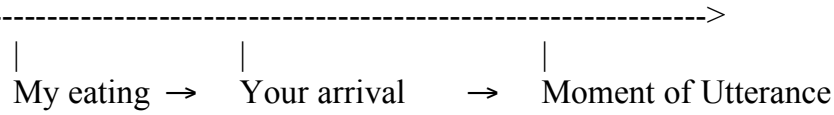
Not all languages necessarily encode tense as a linear category. So for European languages, which do seem to categorize temporality as principally linear, we may represent tense by a line extending metaphorically into the past and into the future, generally representing the future as leading off to the right and the past as to the left.

(6.73) Metaphorical representation of tense as temporal linearity:



By this line, Past, Present and Future are all absolute tenses because they are defined directly with respect to the moment of utterance (simultaneous = present; preceding = past; following = future). Relative tenses are defined on the line with respect to the temporal placement of other events. So, for an example like 'When you arrived I had eaten', our linear representation would look like:

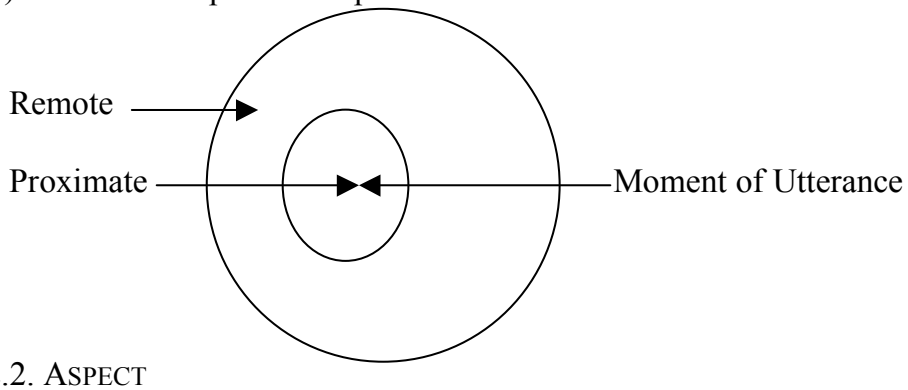
(6.74)



So here the 'eating' event of 'I had eaten' is located before the 'arrival' even of 'when you arrived'. Thus the former is not situated in relation to the Moment of Utterance but in relation to the arrival event.

On the other hand, the linear conception of time is not the only one imagineable nor even the only one implemented in natural languages. For example, in Piraha, according to Everett (1986 – an analysis I would not swear by anymore. I need to do some more thinking about this) there are two tense-like suffixes, **-a** and **–i**, marking 'remote' and 'proximate', respectively. Remote in this analysis refers to events either remote in time from the present (moment of utterance) whether future or past, or out of the speaker's control. Proximate on the other hand refers to the converse, something close to the moment of utterance temporally or within the speaker's control. If this analysis is correct, then rather than a linear conception of time, the best way to express the Piraha conception would be as 'concentric' layers:

(6.75) Possible metaphorical representation of Piraha tense:



6.3.2.2. ASPECT

Understanding a report about an event or state can be made much easier if the speaker gives his or her perspective on how to classify the temporal states being reported. Such temporal classification is known as 'aspect'. So consider the examples below from Brazilian Portuguese:

(6.76) a. Eu estive com ele ontem.
I was with him yesterday.

b. Eu estava com ele ontem.
I was with him yesterday.

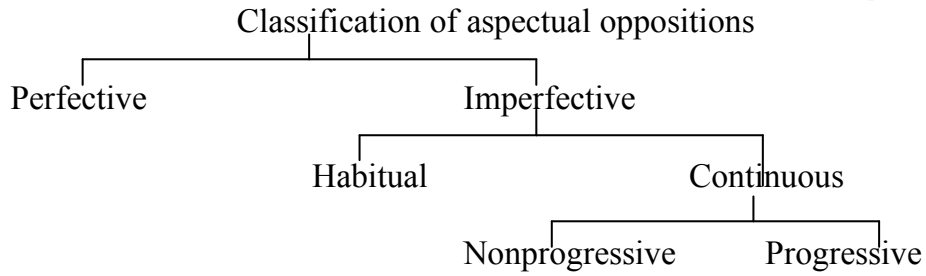
These examples are distinguished not by tense but by aspect, a speaker's perspective on the classification of the 'measuring out' (Tenny (1989)) of an event or the non-deictic temporal properties of an event or state. Both are past tense. The punctiliar aspect in (67a), i.e. a reference to a single being with him at a point in time, is contrasted with the non-punctiliar 'smearing' across time in (67b).

Languages can draw many distinctions based on aspect. A few of these are:

(6.77)
a. Perfective and Imperfective
b. Habitual
c. Telic and Atelic
d. Progressive
e. Punctual and Durative
f. State and Dynamic
g. Iterative
h. Simultaneous
i. Terminative
j. Ingressive
k. Semelfactive

Perfective aspect marks the verb to represent that the speaker is viewing the event as an indivisible whole (Comrie 1976, 16ff). Imperfectivity, on the other hand, is used to make "explicit reference to the internal temporal structure of a situation" (Comrie 1976, 24ff). Comrie's important introductory work goes on to break down a number of aspectual relationships into a single hierarchical relationship, as in ():

(6.78)



This chart represents the broader set of classifications likely to be found in field research. There are other possibilities, however, as languages introduce 'fine-tuning' or specific twists into this classification. Like other linguistic categories, the semantics and formal nature of aspect can only be gotten at more or less reliably by (i) studying the distributions of forms in the grammar and (ii) following a semantic methodology based on conditions of felicity and truth-conditions (see ___ below).

Some other aspectual distinctions the fieldworker may encounter include things like those in the remained of the list in (). Aspectual distinctions are fairly standard across theories, though different theories may emphasize different features of the oppositions. Just to get a better feel for these oppositions, let us consider the one that I consider the most important distinction for the syntax, the telic vs. atelic opposition. This is relevant for the syntax because, as Tenny (1990) shows, it is crucial for understanding the relationship of the verb + undergoer to the actor.

Telic vs. Atelic

Telic refers to an end point, what Tenny () refers to in relation to verbs as the 'measuring out' of the event, in particular, marking when the event has reached its lexically specified ending point or goal. This need not be a morphological category, but is often reflected in the language independent of morphology, e.g. in the syntax of English:

(6.79) a. John ate spaghetti. (Atelic)
b. John ate the spaghetti. (Telic)

In (6.79a) 'spaghetti' is treated like an unspecified mass and so there is no claim that 'John' finished eating the spaghetti, merely that he engaged in the activity of eating it. In (6.79b), on the other hand, the definite article draws a boundary around the spaghetti and indicates that there is an identifiable whole of spaghetti, such that when 'John' eats that he can in fact be said to have completed the goal of his eating. Hence this is a telic clause. Telicity (as Van Valin and La Polla (1997, ---) show) has implications for the syntactic behavior of clauses. But in many languages, e.g. Piraha, telicity is also a morphological category:

(6.80) **Tí soʔóá ʔísigíhií kohoáip -á**.
1sg. already meat eat -telic
'I already ate (the) meat.'

(6.81) **Tí soʔóá ʔísigíhií kohoáip -i**.
1sg. already meat eat -atelic
'I already ate meat.'

Note that in the examples given, the presence of the telic suffix favors an interpretation of the direct object as definite, even though Piraha otherwise lacks a definite/nondefinite distinction morphosyntactically. This is a natural correlate of telicity, however, as we saw in the English examples in ().

There are various other possibilities that involve differential marking (syntactic vs. morphological) vs. semantics of aspect. Although the fieldworker should be familiar with the basic categories mentioned here, they must be careful to arrive at conclusions about 'their' specific language by distributional studies specific to the language in question.

6.4.3. VOICE, VALENCE, AND TRANSITIVTY

Voice is, roughly, is the relationship between the action or state expressed by a verb, and its arguments (nominals to which it assigns a semantic role, e.g. agent, patient, goal, etc.). It is a rather sloppy term that is often used to refer to either Valence (also referred to as Valency) or Transitivity. When applied to Transitivity, different voices increase or decrease the degree of transitivity of a clause, as in the Montana Salish –(**e**)**m** morpheme discussed below. When applied to Valence, these mark changes from avalent to trivalent:

There is a great deal of confusion in the literature, even in books which ought otherwise to clear up this confusion, between VALENCE and TRANSITIVITY. The former refers to the lexically specified number of arguments of a verb. The latter refers to the way that these arguments are 'mapped' to the syntax. Let's consider a simple example from English and then consider some more complex examples from Montana Salish. This issue is not all that complex, but it is very important and often overlooked (though Hopper & Thompson (1980) and VanValin and LaPolla (1997, 147ff) are both noteworthy for their care in distinguishing these concepts. Consider the following examples from English which illustrate the differences between the two:

(6.82) a. The captain sank the ship in order to collect the insurance.
b. The ship was sunk (by the captain) in order to collect the insurance.
c. The ship sank.
d. *The ship sank in order to collect the insurance.

(6.83) a. The captain scuttled the ship in order to collect the insurance.
b. The ship was scuttled (by the captain) in order to collect the insurance.
c. *The ship scuttled.

The a. and b. examples in (6.82) and (6.83) are bivalent. That is there are two arguments lexically required by the verb 'to sink'. In the a. examples and in the parenthetical material in the b. examples, each of the two arguments is overt. But when the parenthetical information in the b. examples is missing, the two arguments are still both there, one explicit and one *implicit* (i.e. understood to be present in the meaning but not heard, like a well-behaved child). This can be seen by the fact that the purpose clause 'in order to collect the insurance' is still grammatical, whether or not the Actor is overt. But purpose clauses can only be used when an agent is either present or understood as part of the meaning of the clause. When the Actor is implicity, it is often nevertheless indicated by voice morphology, e.g. the English passive, which signals the presence of a less topical, potentially implicit Actor (if it appears in the 'by phrase' of English, e.g. 'It was seen *by John*' then it is overt). The c. examples are what some refer

to as the 'middle voice'. With middle voice, the agent is not even implicit. Therefore, with not even an implicit agent allowed with the middle, no purpose clause is permitted in (d).

It is interesting that there is no middle voice for the verb 'to scuttle', (c). The reason is that the middle voice eliminates the agent altogether, i.e. it can be thought of as altering the lexical representation of the verb/deriving from a separate lexical representation without an agent. But 'to scuttle' will lack a middle voice altogether because it has a very specific, non-alterable meaning of 'an agent sinking/destroying' a vessel. Therefore, (c) is ungrammatical. With regard to Valence vs. Transitivity, we can say that the a. examples are bivalent and transitive; the b. examples are bivalent and intransitive; the c. examples are monovalent and instransitive.

Another example might be helpful. In English, as for so many semantic categories, the distinction between causative vs. non-causative is periphrastic or syntactic:

(6.84) a. John cried. (non-causative)
b. Mary made John cry. (periphrastic causative)

A mark on a word that increases the number of arguments required for a lexical item, as in () or periphrastic representation of an added argument, as in English, is said to increase the valency of the lexical item. The causative is one of the most common valency-increasing operations. (Others include the applicative, 'dative shift', or, arguably, simple object-addition, as in English 'Bill ate' vs. 'Bill ate eggs'.) In Piraha, causative is a morphological category:

(6.85) a. Baóhoipaíi ʔi koabáipí
cat it died
'The cat died.'

b. Baóhoipaíi ʔi koabá -a -ip
cat it die -causative -completive

-í
-proximate
'Someone killed the cat.'

There are several types of voice alternation in natural language, including (with my rough descriptions) Antipassive (the detopicalization of an Undergoer); Passive (detopicalization of an Actor), Applicative (increased topicalization of a benefactive or similar prepositional object), Dative Shift (increased topicalization of a GOAL argument. To illustrate some of these, let's consider the extremely interesting case of Montana Salish (from work by Thomason, Thomason, and Everett ()). In this language, as in most Salish languages, there is a morpheme, –(**e**)**m**, that marks a number of apparently unrelated types of meaning. However, when looked at more closely in relation to the Valence vs. Transitivity distinction, a simple analysis emerges. Here are some of the facts.

First, –(**e**)**m** can mark Antipassives, that is a voice alternation whereby the Undergoer of an action is marked (and is in many languages further indicated by

marking the Undergoer with an oblique case) as less topical in the current discourse context. In the (73a) and (74a) examples, we have the antipassive, compared to the normal transitive in the (73b) and (74b) examples. In these examples, the morpheme appears in the position normally reserved for agreement with the Undergoer.

(6.86) a. **tʃn tʃ'n'eyn -m t o'ol'ó**
1sgSItr trap -ANTIPASSIVE OBLIQUE marten (small weasel)
'I trapped some/a marten'

b. **tʃ'n'ey -s -t -∅ -n łu o'ol'ó**
trap -TR -TRANS -3OTr -1sgSTr 2ndary marten
'I trapped the marten'

(6.87) a. **∅ -n -'úłx^{w} -m t lúk'^{w}**
3SItr -LOC:in -bring -ANTIPASSIV OBLIQUE wood
'He brought in some wood'

b. **n -'úłx^{w} -st -∅ -s łu lúk'^{w}**
LOC:in -bring -TRANS -3OTr -3STr2ndary wood
'He brought in the wood'

This same morpheme can also be used to mark the passive, an unusual fact. To get a passive reading from –(**e**)**m**, simply put –(**e**)**m** in the position on the verb normally reserved for agreement with the Actor:

(6.88) a. **'íłn -∅ -∅ -m**
eat -TRANS -3OTr -BCK.AG
'Someone ate it'

b. **'íłn -∅ -∅ -m łu scnc'á t wílwłt**
eat -TRANS -3OTr -BCK.AG 2ndary frybread OBL. scoundrel
'The fry-bread got eaten by a scoundrel/scoundrels'

Compare:

(6.89) **'íłi -∅ -∅ -s łu scnc'á t wílwłt**
eat -TRANS -3OTr -3STr2ndary frybread OBL scoundrel
'The scoundrel(s) ate the fry-bread'

(6.90) a. **x^{w}úy -st -∅ -m**
go -TRANS -3OTr -BCK.AG
'Someone made him go'

b. **x^{w}úy -st -∅ -m łu Coní t Malí**
go -TRANS -3OTr -BCK.AG 2ndary Johnny OBL Mary
'Johnny got made to go by Mary'

Th e Salish –(**e**)**m** can also be used to mark what Thomason, Thomason, and Everett () refer to as 'Derived Transitives'. The traditional Salishanist understanding is that in Derived Transitives a verb not normally transitive (in fact monovalent) becomes transitive (more accurately bivalent), i.e. the number of arguments increases. Or, as L. Thomason () puts it, more rigorously:

> "*Atelic stems can be turned into telic stems through the addition of -m and a transitive marker. The meaning of such derived stems is lexicalized to a greater or lesser degree; locative prefixes frequently help to direct the newly-telic action towards an object. Two examples of derived transitives are given in [...]:*"

(6.91) a. **tʃn 'áyχ^{W}-t**
1sg SItr tire-STATIVE
'I am tired'

b. **'áyχ^{W} -t -m -n -∅ -∅ -∅**
tire -STATIVE -DER.TR -TR -TRANS -3OTr -1sgSTr
'I tire of him/it, I find him/it tiresome'

c. **'áyχ^{W} -t -m -n -∅ -∅-∅ łu s-q^{W}lllú-s**
tire -STATIVE -DER.TR -TR -TRANS -3 OTr-1sg STr NOM-story-3 Poss
'I tire of his story/stories'

Another use of this morpheme is to mark continuative aspect on transitive verbs, as in (79) and (80), in which case it follows the object marker. This is referred to as the 'Transitive Continuative':

(6.92) **es -ƛ'e' -∅- ém -s łu ululím**
stative -seek -3OTr-CONT -3Poss 2ndary money

t Coní
OBL Johnny
'Johnny is searching for the money'

(6.93) **k^{W}u es -'ác'χ -m -s t Coní**
1sgOTr stative -look.at -CONT -3Poss OBL Johnny
'Johnny is looking at me'

The reader may have also noticed a suffix **–n**(**t**) (e.g. examples (73), (74), and (78) above) which marks the verb as being greater than monovalent. Although it is refered to in the Salish literature (and it is roughly pan-Salishan) as a transitivizer or as 'the transitive apparatus', its real function is to mark valence (Thomason, Thomason, and Everett ()), whereas the **–**(**e**)**m** suffix marks deviations with regard to prototypical transitivity, (). There are two questions the Salish morphologist must answer with respect to these examples. First, why does the language have both the valency marker, -**n**(**t**), and, second, why does –(**e**)**m** have so many functions? If we distinguish, as we

should, the lexical notion of Valence, a nongradient concept, from the syntactic notion of Transitivity, which is gradient, then we predict that, as in Salish, a language can mark variations in each with separate morphology (just as we predict that they can vary independently in the first place). And the different functions of the transitivity morpheme, –(**e**)**m**, can be understood as marking, in the first instance, deviations from the prototypical notion of transitivity in (81) (which is meant not as a theoretical statement or commitment to prototype theory at all, but merely as a very useful rule of thumb):

(6.94) Prototypical Transitivity: The prototypical transitive situation is one in which a definite Actor affects a definite Undergoer in punctiliar past.

In light of (81), consider the different uses of the transitive morpheme just looked at. Why does the same morpheme mark both antipassive and passive, for example? Because it indicates, by appearing in the relevant 'slot' on the verb, that the Actor or Undergoer is indefinite, hence that there is a deviation from Prototypical Transitivity. Why does it mark the continuative in transitive verbs? Same reason – a deviation from (). How can it mark a Derived Transitive? In this instance, it marks a valence change not a transitivity change. It is crucial to note that in the Derived Transitive, –(**e**)**m** appears to the left (i.e. closer to the root) of the –**n**(**t**) morpheme. This relative proximity to the root means that it (following Bybee (1985)) is likely to affect the meaning of the root more severely. In this case that is the right prediction – it does indeed affect the meaning more – it alters the valence (a lexical property of verb meaning), not merely the transitivity (a property of clauses). In these cases it moves the verb closer to Prototypical Transitivity, by signalling that the verb has a sufficient number of arguments to be transitive, thus licensing the occurrence of the bivalence marker/'transitive apparatus'. In other words, unless we carefully distinguish transitivity from valence, we have no account of Salish verbal morphology.

Now why have I given so much attention to this particular analysis of Salish here in a field methods guide? Because it is a very clear illustration of the vital distinction between valence and transitivity that, if it is not controlled for, can mislead or puzzle the field researcher for a long period of time. And because most linguists fail to recognize the distinction.

6.4.4. MOOD (or MODALITY):

Mood is used to indicate the relationship of the reported event or state to reality. There are various kinds. The simplest mood system with opposing moods is two: realis (action or state is perceived as real) vs. irrealis (action or state perceived as potential, hypothetical, future, etc). There are more complex mood systems, e.g. indicative, subjunctive, imperative, hortatory, optative, etc. in Romance languages. Mood is often marked on the verb directly, but not always. Consider English (82) vs. Portuguese (83):

(6.95)
a. If I were you.
b. If I was you.
c. If it had rained, I would not have come.
d. If it rains tomorrow, I will not come.
e. ?It rains tomorrow.

Although the English subjunctive is disappearing, a mood used to indicate hypothetical or conditional statements, a vestige of it is seen in (82a). For dialects that use (82a), this is a subjunctive form (it would be ungrammatical if used as an indicative, as in *'I were too here yesterday'). And (82e) needs the help of *might*, as in 'It might rain tomorrow', with the indicative ending *–s* missing from the verb, to indicate the subjunctive mood.

(6.96) a. Se eu fosse voce.
b. *Se eu sou voce.
c. Se tivesse chovido, nao teria vindo.
d. Se chouver amanha, nao virei.

In Portuguese, unlike English, as (83) shows, mood is an obligatory morphological category.

Mood is also important (as are numerous other verbal categories) for syntactic reasons as well, because it can indicate the degree of formal relatedness between clauses, as the Portuguese examples below show:

(6.97) a. Se fosse você, não viria/vinha/*vim/*verei amanhã.
if went:pastsubj. you not would come/was coming/came/will come tomorrow
'If I were you, I would not come/was not coming/*came/*will come tomorrow.'

b. Se for amanhã, terá/tem/*tinha/*tive que levar um casaco.
if go:fut.subj. will have/have/*was having/had to take a coat.
'If you go tomorrow, you will need to take a coat.'

The correlation between the mood of the protasis (the embedded 'if clause') and the tense and mood of the apodosis (the matrix 'then clause') indicates embedding in Portuguese and can indicate different degrees of clausal interconnection in other languages, so can serve as a useful diagnostic for embedding and other aspects of analyses of interclausal relationships.

6.4.5. DIRECTIONALS

Verbs are further restricted in some languages by affixation of directionals. These affixes may be used *in addition to* a similar range of prepositions or they may obviate the need for most prepositions (so I advise the fieldworker to look with particular care for verbal directionals if there are only one or two prepositions in the language!). Directionals primarily function to fix the event in space. But they are often extended in meaning. Intransitive prepositions in English often function as directionals, as seen in ():

(6.98) a. The plane takes off
b. Look up the tree; look down the street.
c. He threw up; he upchucked (from the stomach to the mouth is seen as an upward motion, but the extension of the directional meaning is also present).

But also (see Lakoff and Johnson for further extensions of verb + particle constructions in English):

(6.99) a. Look him up vs. ?/*look up him.
b. Look Bill up vs. look up Bill.
c. He is moving up in the world.

To my mind the contrast in (a) shows an interaction of constraints on English cliticization (where the prepositions are simple clitics, merely phonologically reduced forms needing a 'host' in Zwicky's (1977) terminology). A prominal clitic in English, e.g. [m] 'him', attaches to what it is an argument of. If it is an argument of the verb, it attaches to the verb, as in (). If it is an argument of a preposition, as in (), then it attaches to the preposition, producing a very different reading. As with other morphological categories, directionals have a fairly straightforward and literal basic function with subtle semantic extensions that can take considerable research time to understand.

So compare () above with (): (see Lakoff and Johnson () for a large number of such extensions) consider:

(6.100) a. Look him up; */?Look up him.
b. Look Bill up; Look up Bill.
c. Take on city hall.
d. He is moving up in the world.

The contrast in (a) shows an interaction between the constraints on English cliticization (simple pronoun clitics, i.e. those which are merely phonologically reduced forms of full pronouns, but otherwise unexceptional syntactically, see Zwicky ()), such that a pronominal clitic must attach to the verb before a prepositional clitic. A pronominal clitic in English, e.g. *him* [ɪm] in (), attaches to what it is an argument of. If it is an argument of the preposition it attaches to the preposition, as in (), but if it is an argument of the verb it must attach to the verb. Thus () is bad but () is good. The reading where it is an argument of the preposition *up*, (), is quite different, of course, than where the pronoun is an argument of the verb.

Examples like (), and many more in Lakoff & Johnson, illustrate how the uses of directionals can be extended to be interpreted in ways other than their etymological source might indicate. As with other morphological categories, directionals have a fairly straightforward and literal basic function with subtle semantic extensions that can take considerable semantic research to understand.

6.4.6. AKTIONSART

Aktionsart, 'kind of action', is a semantic classification of verbs, based on the inherent temporal characterization of a verb's specified action or event.[37] In VanValin and LaPolla (1997, __), as in many other works, the aktionsart of verbs is an important component in the analysis of their morphological (e.g. the range of affixes they can

[37] This section borrows heavily from the excellent introduction to verbal semantics in Van Valin ().

take) and syntactic (e.g. the range of constructions in which they may appear) behavior. This notion is introduced clearly by Van Valin (---):

"Verbs can be subclassified in various ways. Some researchers have found some ways of subclassifying them more useful than others. One system that has been widely adopted is to subclassify verbs based on Aktionsart ('kind of action', from Vendler ()), the inherent temporal characterizatoin of a verb's specified action or event. There are four subtypes of verbs in the traditional Aktionsart classification proposed by Vendler:

(6.101) a. States: be sick, be tall, be dead, love, know, believe
b. Achievements: pop, explode, shatter (the intransitive versions)
c. Accomplishments: melt, freeze, dry (the intransitive versions); learn
d. Activities: march, walk, roll (the intransitive versions); swim, think, snow, write, drink

States depict static situations which are inherently temporally unbounded (atelic), and both achievements and accomplishments express changes of state, which are inherently temporally bounded (telic): achievements are instantaneous, while accomplishments are not. Activities are dynamic, inherently temporally unbounded (atelic), states of affairs. Vendler proposed this taxonomy based solely on the analysis of English verbs, and yet it has proved to be of great cross-linguistic validity.

It would be reasonable to hypothesize that these distinctions are the universal basis of the organization of verbal systems in human language. There is one important non-Vendlerian Aktionsart class, namely semelfactives (Smith 1997). Semelfactives are punctual events which have no result state. Examples are given in

(6.102) a. The light flashed.
b. Chris coughed.
c. The tree branch tapped on the window.
d. Dana glimpsed Kim.

There is a derivational relation between two classes which is very important crosslinguistically, namely that between activities and what are called ACTIVE ACCOMPLISHMENTS, the telic use of activity verbs. This general pattern relates activity verbs of motion (e.g. run), consumption (e.g. eat), and creation (e.g. paint) to the corresponding active accomplishment verbs. This is illustrated in (87) for English.[my numbers, DLE]

(6.103) a. The soldiers marched in the park. Activity
a′. The soldiers marched to the park. Active Accomplishment
b. Dana ate fish. Activity
b′. Dana ate the fish. Active Accomplishment2
c. Leslie painted (for several hours). Activity
c′. Leslie painted Mary's portrait. Active Accomplishment

There are a few verbs in English that are lexical active accomplishments, e.g. devour and go; that is, they do not alternate with an activity counterpart like the verbs in (6.104).

Most fundamental is the distinction between static and non-static verbs, which distinguishes verbs which code a 'happening' from those which code a 'non-happening'. In other words, with reference to some state of affairs, one could ask,

'what happened?' or 'what is happening?'. If, for example, a sentence like Bob just ran out the door could be the answer to this question, then the verb run is [– static]. On the other hand, a sentence like John knows Bill well could not be the answer to this question, because nothing is taking place. Hence know is a [+ static] verb. By this criterion activities, achievements, semelfactives, accomplishments and active accomplishments are [– static]. States, however, are [+ static]."

However, it would be a mistake, a serious one, for a fieldworker to attempt to simply translate the diagnostics for Aktionsart from English. They will likely not be useful in many other languages, at least without considerable adaptation. Finding useful diagnostics for linguistic categories, especially diagnostics that give such (apparently at least) clear-cut distinctions is not easy and in a particular field situation it is possible that nothing corresponding to these will be found. Aktionsart is an important concept to the fieldworker not because it will always be useful in every field situation, but because it *may* be very useful in figuring out how the morphology works in some languages.

But beware, again, of simple solutions. It may be very difficult to come up with satisfying, watertight classifications. Nevertheless, your morphological research will benefit tremendously by investigating at least the possibility of verb or verb combinations classified on the basis of Aktionsart.

6.4.7. FINITENESS

Verbs can be more or less *finite*. Finiteness is the degree to which the verb is referentially anchored. The greater the amount of inflection on the verb, the more precisely is it anchored referentially. Consider the following examples from English:

(6.104) He owns the place.

(6.105) a. He wants to own this place.
b. He thinks he can own this place.

(6.106) Owning is not always a good idea.

In (88) the verb is marked for present tense, indicative mood, third person, singular number, and has an aspectual reading that indicates on-going ownership. This verb thus provides the hearer with a good deal of information to help situate the 'owning' in time, reality, and relationship to identity of owner, among other things. But in (89) the same verb is not marked for these semantic categories. This less finite form can be overtly marked by an item such as syncategorematic *to* (Pullum ()), (89a) or by no marking at all, (89b). But in these examples it is temporally-aspectually unanchored, i.e. it refers to no specific 'owning' of 'this place'. Its gerundive form in (90) likewise marks it as lacking anchored temporal-aspectual, etc. reference. And the gerundive marking further indicates that it is 'less verby' (see __), becoming more like a thing (an abstract thing, i.e. the act of ownership).

Finiteness can manifest itself in different ways in different languages. So consider the examples below from Karitiana (Arikem family) and Pirahã:

(6.107)

ÿn	**nãka**	**paka**	**-j**	**ÿn**	**pykyp**
I	affirmative	clean	-tense	I	clothes

'I will clean the clothes.'

(6.108) **ÿn i paka ÿn pykyp**
I it clean I clothes
'I will not clean the clothes.'

(6.109) **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')

(6.110) a. **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')

b. **kahaí kai -sai hi ob -áa'ái**
arrow make -nom. he see -attractive

c. ***hi kahaí kai -sai ob -áa'ái**
he arrow make -nom. see -attractive

Karitiana is interesting here because a negated verb takes a non-finite verb form. Pirahã is interesting for other reasons. Let us consider first the Karitiana example. Apparently in Karitiana, the negation of a verb removes, in standard cases, its verbal referentiality. It no longer takes finite verbal morphology because a negated action or state simply never occurred and thus can be left unanchored.

The Pirahã case is perhaps even more interesting. Everett (2005) argues that Pirahã lacks recursion, at least in the syntax. Yet, at the same time, some functions of clauses are less finite, corresponding partially, but by no means completely, to what in other languages would be embedded clauses. On the other hand, the reduction of finiteness in this language does not mark embedding. For example, in (), it is (what most linguists would take to be) the *matrix* verb that receives the **–sai** suffix, where as in () it is what might otherwise be taken to be the embedded clause which is so marked. This difference in marking violates normal expectations of finiteness as an embedding marker (see Koptjevskaja-Tamm (1993a and 1993b for detailed discussion of nominalizations and finiteness more generally). Everett and VanValin (in progress) analyse the **–sai** suffix as an information marker, indicating topical or old information.

Nonfiniteness of the verb can mark embedding. One reason that this is common is because, as Cristofaro (2003) argues, subordinate clauses are not assertions. Therefore, it is common for their verbs to be less referential in the sense of this section. However, that is not the only function of nonfiniteness. It can also indicate that the verb has 'moved' in the sense of Ross's (1973) scale towards 'nouniness'. An analysis associating nonfiniteness exclusively with embedding could not capture the generalization that Pirahã otherwise lacks any evidence for embedding.

6.4.8. Evidentials

Many languages employ morphological marking to indicate the speaker's epistemological stance with regard to their assertions, as an indication for the hearer(s) of the reliability of or the warrant for the content of the speaker's assertion.

For example, La Polla (2003) discusses the interesting case of the Tibeto-Burman language, Qiang. This language has a suffix, **-k**, which is labelled as 'inferential/mirative', depending partially on the Aktionsart of its verbal host. That is it is 'inferential' with activities, but 'mirative' (i.e. marking unexpected or new information) with resultatives and states. For example:

(6.111) a. the: ʐdʐyta: ɦa qə -k
3sg Chengdu:loc or go inferential
'He went to Chengdu (I infer based on the evidence, e.g. bags are missing and was told earlier about the trip, etc.)

And from Shipibo (Camacho, Elias-Ulloa ()), '...an inferential can be combined with a reportative, yielding a potentially ambiguous sequence: either the speaker reports someone else's inference, or s/he expresses an inference from reported evidence."

(6.112) Ani -ronki i -bira [a] -i jawen jema.
large rep be -inferential -incl poss3 village:absolutive
'It appears the door is open/apparently the door is open.'

As Camacho and Elias-Ulloa go on to observe:

"Several studies have tried to explore whether propositional source-marking obeys some kind of cross-linguistic tendency. Willett (1988, p. 57), for example, ¯nds in a survey of 38 languages, that evidentiality falls within three categories: attested (which can include subdivisions that include visual, auditory or other sensory information), reported, and inferred. Speas (2004, p. 4), on the other hand, proposes four hierarchical categories: personal experience, direct (sensory) evidence, indirect evidence and reported evidence (hearsay).

Aikhenvald & Dixon (2003, p. 3), suggest two broad types of evidential systems: those that state the existence of a source of evidence without specifying it, and those that specify the source. Within the second type, they discuss several subtypes:

(6.113) Two-distinction systems:
1. Eyewitness and noneyewitness.
2. Nonfirsthand and everything else.
3. Reported and everything else.

(6.114) Three-distinction systems:
1. Visual, inferred, reported.
2. Visual, nonvisual sensory, inferred.

(6.115) Four-distinction systems:
1. Visual, nonvisual sensory, inferred, reported.
2. Visual, inferred (2), reported.
3. Nonvisual sensory, inferred (2), reported.

4. Visual, inferred, reported (2).

It seems clear, particularly for three and four-distinction systems, that they are built on Willett's three basic distinctions, with some possible subdivisions within each of those categories. Speas, following Oswalt (1986), suggests an additional category different from visual-sensory: personal experience. Willett, Oswalt and Speas suggest that there is an underlying hierarchy corresponding to the degree in which the source directly involves the speaker's evidence. This scale goes from more direct experience to no experience at all (Speas 2004, pg.258):

(6.116) personal experience >> direct (e.g. sensory) evidence >> indirect evidence >> hearsay."

Evidentials are interesting for what they have to teach us about the interplay of semantic scope between different types of affixes, about what they reveal of different epistemologies, about the nature of morphosyntactic structure, and about the cultures that produce them.

6.4.9. DISLOCATION[38]

Modern syntactic theories focus (to an inordinate degree in my opinion) on just one of the many items that we must 'memorize' about a language, namely, whether or not it allows for constituents to be displaced or dislocated from their purportedly 'underlying' or (perhaps more accurately) their unmarked positions. The functions of dislocation in a language (aside from the extremely theory-internal arguments of some syntactic theories) include marking the *information structure* of the sentence (see Lambrecht () and section ___ of chapter __ below) and scope alternations. But of course dislocation is simply one among many different strategies for marking different information structures and scope, inter alia. Otherways of marking these include prosody, morphological marking, separate construction types, or a combination of some subset of these options, among other strategies. Still, it is true that many languages choose dislocation as perhaps the principal means of pragmatic highlighting, scope alternations, or information structure changing. Occasionally, dislocated elements may have consequences for the morphology of a language, which is why dislocation is important to us in this section. So consider question-related dislocation and morphological marking in Wari:

(6.117) **Ma'** **co** **tomi'na**?[39]
that:prox:hearer m/f:rp/p speak 3s:rp/p
'Who is speaking?'

(6.118) a. **Ma'** **co** **tomi'ca**?

[38] See also the section on dislocation in chapter 8 below.

[39] Example (4) is interesting because it illustrates that questioning the subject of the sentence requires tense in second position, to the immediate right of the question word, and also immediately to the right of the verb. WH-questions of subjects require that tense be expressed twice in the sentence. This, as (5) – (7) show, is not true of any other questioned constituent.

that:prox:hearer m/f:rp/p speak 3sm
'Of whom is he speaking?'

b. **Ma' carawa ca pa' caca mon**
that:prox:hearer animal nrp/p kill 3pm coll

tarama'?
man
'What thing/animal did the men kill?'

(6.119) **Ma' ca para 'aca ca pije ma'**?
that:prox:hearer nrp/p why cry 3sm child that:prox:hearer
'Why is that child crying?'

(6.120) **'om ca mao ca**.
not:exist nrp/p go(sg) 3sm
'He did not go.'

Dislocation is morphologically relevant in Wari' because the nature of the grammatical dislocated (e.g. subject vs. direct object vs. adjunct) results in different kinds of agreement patterns. First, any dislocation causes tense to appear before the verb. This is due to a simple (and crosslinguistically very common) constraint that requires tense to appear following the first constituent of the clause. But notice that in (), when the subject appears in sentence-initial position, tense appears *both* in second position and immediately following the verb (it's 'normal' position). Only one occurrence of tense, however – in second position of the clause, is found when objects and adjuncts are 'dislocated' to the beginning of the clause (() and ()). Whatever the reason for this asymmetry, it is evidence that dislocation can have morphological and not merely syntactic effects.

Another morphological configuration that some languages allow is Incorporation.

6.4.10. INCORPORATION

The first clear reference to Incorporation (though not by that name) I am aware of is Montoya (1585-1652), Jesuit missionary and linguist who worked among the Guarani of Paraguay. Like Dislocation, Incorporation is a common device for marking changes in information structure. It can also be used for other functions, e.g. additional specification of the verb. As Mithun (1986, --) says:

> *"Noun incorporation is perhaps the most nearly syntactic of all morphological processses. Examination of the phenomenon across a large number of geographically and genetically diverse languages indicates that, where syntax and morphology diverge, incorporation is a solidly morphological device that derives lexical items, not sentences. It is used for four different but related purposes; these fall into an implicational hierarchy which in turn suggests a path along which incorporation develops historically. Differences in its productivity from language to language show that this development may be arrested at any point – resulting either in the eventual disappearance of the process, or in its resurgence as a productive system of affixation."*

An example from Yucatec Mayan

(6.121) a. **tinč'akah če'**
t- in- č'ak- Ø- ah če'
COMP- I- chop- it ASP

b. **č'akče'nahen**
č'ak- če'- n- ah- en
tree chop- tree- APASS- ASP- I.ABS
"I chopped a tree."

Noun-incorporation has various manifestations and a range of functions. In the examples of Piraha encliticization repeated below, there is no real incorporation, but the phonological divisions between object and verb have been reduced in a way that doesn't happen with subjects:

(6.122) tíì ʔísìtòí ʔóògàbàgàí → tíì ʔísìtǒògàbàgàí
1psg egg want
'I want an egg.'

(6.123) kàhàí ʔáàgàhá → kàhàǎàgàhá
arrow is
'It is an arrow.'

There is something 'iconic' in this in that some structures show less phonology to represent a tigher semantic relationship. What that semantic 'tightness' actually means in any given language has to be determined by the linguist by careful analysis and argumentation.

6.4.11. AGREEMENT

Agreement is a way of keeping track of verb participants by marking them morphologically, either on the verb (as in English, *John run***s**) or elsewhere in the clause, with Wari's second-position verb agreement markers:

Most languages will have agreement in one form or another. As Everett (1996) observes, it is imperative to analyze carefully the relationship between pronouns, pronominal clitics, and agreement affixes. Everett (1996) considers a variety of languages in which these categories interact to produce a type of 'agreement mosaic'. This is also discussed in the section on clitics above and illustrated in many examples throughout this book.

Let us turn now to strategies commonly employed to mark the distinctions discussed above.

6.5. MORPHOLOGICAL REFLEXES OF GRAMMATICAL AND SEMANTIC DISTINCTIONS

6.5.1. CONCATENATION

Easily the most common morphological marking strategy is the addition of extra material to the root (the base of a word from which all morphology has been stripped), what morphologists know as *concatentation*. Material may be added to the root at the right or left margins (prefixation and suffixation, respectively), at both simultaneously (circumfixation), or within the root (infixation). Of these various strategies, suffixation

is the most common, followed by prefixation, then perhaps infixation, then circumfixation. These are illustrated in (6.124) – (6.126):

Suffixation: adding to the end of a particular morphological constituent (word, root, or stem)

(6.124) John run***s***.

Prefixation: adding to the beginning of a particular morphological constituent (word, root, or stem)

(6.125) John ***en***livened the party.

Infixation
(6.126) Tagalog Infixation (McCarthy & Prince 1993)

a. /um + ʔalis/	→ /ʔum-alis/	*'leave'*
b. /um + tawag/	→ /t-um-awag/	*'call' pf., actor trigger*
c. /um + gradwet/	→ /gr-um-adwet/	*'graduate'*

Circumfixation is the simultaneous addition of material preceding and following the root, stem or word to mark a single distinction. Circumfixation is perhaps best known from the German past participle (*ge- -t* for regular verbs). The verb *spielen*, for example, has the participle *gespielt*.

6.5.2. SUPPLETION

Another common strategy for marking distinctions like those discussed above morphologically is suppletion. Suppletion is the label given to alternations where there is no rule-based phonological relationship between the various allomorphs. It is quite common in the world's languages. So consider the examples below from English and Portuguese:

English *to be*
(6.127) I go, you go, he goes, they go, we go, I went, you went, etc.

Portuguese *ir*
(6.128) Eu fui; voce foi; eles foram; Eu vou; voce vai; nos vamos, eles vao.

It is common to find suppletion in more frequent verbs, e.g. 'to be' or 'to go'. This is likely because it is easier to remember the forms when they are heard more often.

Another not uncommon type of suppletion is found in Wari', where the suppletion follows an ergative pattern, the form of the verb governed by the object of the transitive or the subject of the intransitive. Not all verbs participate in this process (and it would be an interesting study to see if these verbs follow any particular pattern). But those that do behave as in (), from Everett & Kern (1997, 331):

(6.129)	a.	Toc	nana	com.
		drink:sg	3p:rp/p	water

'They drank water.'

b. Tototoc nana com.
drink:pl 3p:rp/p water
'They drank and drank water.'

This example is interesting, because although 'water' has no plural in Wari', the suppletive plural of the verb can also indicate a plurality of events. This scope variation of suppletion, i.e. marking either object plurality or event plurality, is common. Abdias, Walker, and Everett (1984) discuss a similar case for the Mayan language, Huastec. But there are many other examples where the object itself governs plurality:

c. Cao' 'ac cacama na.
eat:sg travel 3pf 3s:rp/p
'Then they ate.'

d. Mon cacacao' 'ac cacama -in capam' con
slowly eat:pl travel 3pf -3n cornbread prep:3sm

jowin ma' na.
monkey:species that:prox:hearer 3s:rp/p
'Then they ate cornbread and *jowin* monkey for a long time.'

The Wari' case is very interesting because the suppletion shows an ergative pattern (see __ above). For a very useful and detailed study of morphological suppletion, the reader should consult Veselinova (2004).

Let's consider one more, often overlooked strategy for morphological marking, *periphrasis*.

6.5.3. PERIPHRASIS[40]

Consider the contrast between aspectual marking in English vs. Portuguese:

(6.130) a. Eu estava aqui ontem.
b. Eu estive aqui ontem.

(6.131) Eu tomava cafe, sim.

(6.132) a. I was here yesterday for a while.
b. I was here yesterday.
c. I used to eat breakfast.
d. I was eating breakfast.

Portuguese marks punctiliar, continuative, and habitual aspects morphologically. English marks the latter two periphrastically. The fieldworker can easily mistake periphrasis, a morphological strategy, for syntax. But periphrasis, as Börjars, Vincent,

[40] This section assumes the paradigm to be a theoretically primitive of linguistic theory. For eliciting lists of related lexical items commonly known as paradigms in a non-theoretical sense, the reader is referred to __ of chapter __ on monolingual fieldwork.

and Chapman (1997) show, is an important component of natural langauge morphology. To see this more clearly, consider the case of periphrastic pronouns in Wari', discussed in Everett ():

TABLE ONE
Paradigm of spatial demonstrative pronouns

	Proximate to Speaker	Proximate to Hearer	Distal
Masculine singular	*co cwa'*	*co 'ma'*	*co cwain*
Feminine singular	*cam cwa'*	*cam 'ma'*	*cam cwain*
Neuter	*'i ca'*	*'i 'ma'*	*'i cain*
Plural	*caram cwa'*	*caram 'ma'*	*caram cwain*

TABLE TWO
Paradigm of temporal demonstrative pronouns

	Heard/not seen	Recently absent	Long absent
Masculine singular	*co paca'*	*co pacara ne*	*co pacara pane*
Feminine singular	*cam paca'*	*cam pacara ne*	*cam pacara pane*
Neuter	*'i cara*	*'i cara ne*	*'i cara pane*
Plural	*caram paca'*	*caram pacara ne*	*caram pacara pane*

The constructions in these tables are not simply words. Phonologically, words in Wari' manifest two important characteristics that these periphrastic forms fail to show. First, Wari' words disallow internal consonant clusters where one of the consonants is a glottal stop. Thus, the initial glottal stop on *'ma'* [ʔmaʔ] 'that proximate to hearer' in Table 1, as in *co 'ma'* 'that masculine one proximate to hearer' and *'i 'ma'* 'that neuter one proximate to hearer', should not occur, since word-medial glottal-consonant clusters do not otherwise occur. By this otherwise inviolable criterion, in the cases of *co 'ma'* and *'i 'ma'*, *co* and *'i* must be interpreted as morphologically independent and not, say, as prefixes. We interpret them, as well as *cam* and *caram*,as clitics (as discussed in Everett and Kern (1997, 413ff)). Second, in Wari' words prefixes always undergo Vowel Harmony with their hosts. The kinship terms discussed in the penultimate section of this paper, which are analyzed as words, rather than periphrastic forms, do indeed show Vowel Harmony between the *co* and the root. The absence of Vowel Harmony in the forms in Tables 1 and 2 is therefore additional strong evidence that these are not words (or, at the very least, that they violate expectations on word phonology).

Among the many arguments Everett (2005) gives that these pronouns are periphrastic morphology, rather than syntax, is the following. Periphrastic pronouns in Wari' mark all three Wari' genders. But neuter gender marking is accomplished by

using the morpheme ***cara***, just in case it it is in a periphrastic construction. ***Cara*** 'heard' otherwise has no inherent gender association (Everett (2005))). This means that ***cara***'s interpretation in the tables above is a function of the *paradigm* (an abstract constraint of morphology but not syntax), rather than the syntax and is thus non-compositional and morphological. Therefore, Everett (2005) concludes that periphrasis is an important morphological strategy in Wari', a way of co-opting phrases to function as 'hybrid' words.

Methodologically, therefore, it is always useful to try to discover unifying principles of morphological groupings, assuming the paradigm as at least a methodological convenience if not a theoretical primitive. This way it will be easier to spot the use of syntax for morphological means as well as 'mixed categories'.

6.5.6. Intercalation

Intercalation is most often referred to in the literature as 'non-concatenative morphology'. However, frm what we have seen in this section, there are various kinds non-concatenative strategies. So I prefer to refer to this type of morphological marking as *Intercalation*. This is illustrated in (110) from Arabic:

(6.133) *TH: Gloss: Morphological analysis:*

a. **qātal** 'he killed' 3 p. masc. sg. perfect
b. **qittēl** 'he murdered' 3 p. masc. sg. perfect intensive
c. **q**o**tēl** 'killing' active participle (qal)
d. ji**qt**o**l** 'he will kill' 3 p. masc. sg. imperfect/future
e. ni**qt**a**l** 'he killed himself' 3 p. masc. perfect reflexive
f. **q**$_e$ **tālām** 'he killed them' 3 p. masc. sg. perfect + pers. pron.

One way to conceie of this is that the morphology of Arabic accesses structures for vowels separately from structures for consonants and then, following precise constraints (e.g. those in McCarthy (1979)) intercalates them.

6.5.7. Prosody

The marking of morphological distinctions can also be accomplished by prosodic features such as tone or intonation, length, and stress. For example, consider the way that Pirahã indicates an inherent vs. an accidental property of an individual via the tone on the verb:

(6.134) Hi ʔii ʔaagá.
'He hungry is.'

(6.135) Hi hiaitíihí ʔáagá.
'He is a Pirahã.'

Or consider the distinction between some nouns and verbs in English:

(6.136) per*mit* (verb) vs. *per*mit (noun); con*tract* (verb) vs. *con*tract (noun), etc.

6.5.8. Reduplication – kamaiura

Another common morphological marking device is reduplication, where some

part of the root or stem is duplicated in order to mark a semantic distinction. One well-known example comes from Kamaiurá (Everett & Seki 1986, McCarthy & Prince 1993):

(6.137)

a. *ohuka*	→	ohuka-*huka* 'he laughed/he kept laughing'
b. *apot*	→	apo-*apo*-t 'I jump/I jump repeatedly' (*apot-*apot*)
c. *o-mo–tumuõ*	→	omotumu–*tumu*–õ 'he shook it/repeatedly'
d. *je–umirik*	→	jeumiri–*miri*–k 'I tie up/repeatedly'
e. *o–etun*	→	oetu–*etu*–n 'he smells/keeps on smelling'
f. *o–ekvj*	→	oekv–*ekv*–j 'he pulls/repeatedly'

According to Everett & Seki (1986), these are examples of suffixation. According to McCarthy & Prince (1993) they are best analyzed as infixation. Whatever the analysis, it is important for the fieldworker to be familiar with this type of marking. A more detailed study of reduplication (but to my mind one with the defect of predicting the absence of the Tupi pattern just investigated) is Inkelas and Zoll ().

6.6. WHERE TO MARK DISTINCTIONS

Now let us consider where the morphological strategies discussed above may appear. For many features they can differentially appear, crosslinguistically, on either the head of the phrase or a dependent (daughter) of the head. This distinction is very important, as Nichols (), the first one to make the distinction in these terms, points out:

"Morphological marking of grammatical relations may appear on either the head or the dependent member of the constituent (or on both, or on neither). Grammatical relations – and whole languages – may be classified according to their propensity for using one of these types of marking. Implicational relations among various marking patterns can be stated: languages display a tendency to use one type consistently throughout their grammar. The difference in patterns provides a typological metric and a functional explanation for certain word-order preferences. For historical linguistics, it provides a diagnostically conservative feature and a clue to genetic relatedness. Although the head-marked pattern is cross-linguistically favored, grammatical theory is strongly biased toward the dependent-marked patterns that happen to dominate in Indo-European."

Consider the two English examples in (115) – (117). The first one marks the relevant morphological distinction on the head of the phrase (the verb in this case), while the other marks it on a dependent of the head (the embedded NP in the possessor position in (116) and the case on the pronoun in (117)):

(6.138) a. The man run*s*.
b. The men run.

(6.139) John*'s* book is wet.

(6.140) a. *I* saw *him*.
b. *He* saw *me*.

Likewise in Portuguese, the gender of the noun is marked on its dependents:

(6.141) a. O homen bonit*o*.
b. A mulher bonit*a*.

Or, as in Wari clausal morphology, the marking may appear on neither the head nor one of its dependents, but in the clausal 'second position', also known as Wackernagel's Position, from the work of the first person to draw attention to this position for clitic placement, Wackernagel (1892).

6.7. Derivation vs. inflection

A venerable distinction in morphological theory is the subclassification of morphological processes which change the category of a root to form a stem as *derivational* and morphology added to the stem, and usually agreeing with a semantic distinction on another word in the phrase, as *inflectional*. Not all morphologists attribute primitive theoretical status to this distinction. Here I want to briefly consider some of the classical reasons for distinguishing these two subclasses of morphology – not because I necessarily endorse the labels, but because recognizing the concepts behind the labels is useful for any fieldworker. Consider the examples in ()-() from English:

(6.142) a. farm
b. farms
c. farmer
d. farmers

(6.143) a. live
b. liven (as in *Liven up the party!*)
c. enliven (as in *Enliven (*up) the party!)*

The suffix *-s* is an inflectional morpheme. It is used to produce 'fit' between a lexeme and its syntactic context (e.g. its phrase or clause), as well as to signal that a meaning option (plurality) has been selected from the relevant inventory of semantic distinctions (number in this case).

Morphology that changes the meaning of a *root* to form a *stem* (an internally-complex morphological input to inflection) often is less productive, more idiosyncratic in meaning and morphology, more constant in its phonological form, among other things, than morphology that does not affect core lexical meaning and that applies to either roots (from which all morphology is absent) or stems. Whatever you call it, this distinction crops up in language after language and it is to the fieldworker's advantage to be aware of it, look for it, and carefully subclassify this and any other distinct mrophological processes that appear to be operative in a given language.

Up to this point, we have considered distinctions often made in natural languages and how and where these may be marked. I now want to offer practical suggestions for morphological fieldwork, i.e. how to collect, analyze, and process morphological data.

6.8. Methodology for morphological analysis

6.8.1. INTRODUCTION

Returning now to elicitation of morphosyntactic data, the researcher must first answer the obvious question of what it means to 'describe', 'document', or 'analyze' *the* morphology of a language. Simply put you have two goals in morphosyntactic description: to identify the primary semantic distinctions in word meanings and to determine how these distinctions are signalled. The task is therefore easy to conceptualize.

Of course, the actual doing of it is much harder. Let's say you hear the two forms in (6.144), for example:

(6.144) a. pandapa 'I ate'
b. pandapã 'You ate'

On the surface, it would look like 2nd person is marked by nasalization of the final vowel, /a/. But here are some obvious questions: (i) are you *sure* there is no nasalization on the final vowel of the first word? (ii) did you get this pair repeated from multiple speakers? If so, did they all use nasalization in (b) and none in (a)? (iii) can (b) also be uttered without nasalization on the final vowel without changing the meaning? These questions may seem obvious now, but they are not quite so obvious in a field situation. The field researcher will soon realize that there is an enormous difference between analyzing printed data and analyzing spoken data. Having 'good ears' is crucial to good morphosyntactic analysis. Acoustic analysis will help, of course, but without good ears, you will be reduced to hours and hours of unnecessary spectrographic labors.

Linguists disagree on how to collect data for morphosyntactic and semantic analysis in the field. Some advocate a text-only approach to data collection. Some claim that specific sentence elicitation can play a useful role in field data collection.

The problem to me is similar to the problem faced by any second language learner. Do you learn a language from books and classes, or from natural conversations in a community of native speakers? The choice is a false one, at least to me. One must converse and develop a *need* to use the second language. But all other sources are also of great potential usefulness and will almost always complement the information otherwise learned, helping in fact to learn constructions and expressions that might have otherwise gone unperceived, at least for a considerable period of time.

Likewise in morphosyntactic research, we need data from a variety of sources. Some linguists are concerned about a 'data problem, arguing that elicited data (one sentence at a time, for example) should be avoided in favor of conversational or text-based data in which sentences appear in natural contexts. Only in such 'natural contexts' we are told, are data reliable. However, I disagree. I believe that elicited data, natural texts, conversations, corpora, and any other source of data can and should be used in constructing a theory of a particular grammar. The 'data problem' is to me a judgment problem. Gathering data requires experience, intuition, a lot of linguistic knowledge, security in what you are trying to do, and the ability to distinguish contrived data from data that could actually emerge in a natural way from the mouths of speakers. (We don't know if they have it in their grammar until we have heard it from their mouth.)

6.8.2. Record, transcribe, annotate (the context and what you think it means)

As you perambulate about the speech community where you are working you will hear words, both words you ask for and words that are offered to you, or just words and

expressions in the environment. You probably will not know at the beginning of your work whether these are words or sentences or phrases (consider the *gavagai* problem discussed in section ___). And you will almost certainly not have a secure understanding of their meaning. Many things you encounter, e.g. rare animal or rituals, or descriptions of culturally sensitive activities you might not be welcome to see again (e.g. something you just innocently stumbled across). So always be prepared to transcribe and, ideally, record in audio and video what you hear. This is where a small, analog recorder, with easy rewind and record functions is very useful. However, there are often occasions that it would be useful to record an image along with the word. My personal digital camera (not a camcorder) allows me to take a picture and a brief recording associated with that picture. This is useful, for example, if I am collecting the names for flora and fauna. I am not a zoologist. Thus whenever I see a new tree or species of rodent, I usually do not know what it is called in English, much less in Latin (i.e. its scientific name). Some fieldworkers do know this. But in my case, when I get back to the village, I pull out one of the flora and fauna books I have taken (see ___ below) and look for what I have just photographed. Failing to find it, I can take the photo with me to a specialist upon returning to my home institution.

6.8.3. Test with native speakers:

You haven't learned a morpheme until you have it down articulatorily and acoustically. You need to test your pronunciation of it (or the words in which it occurs) by practicing with native speakers. As you practice, you need to also listen to make sure you understand how to use the morpheme appropriately in context and any extensions that it has. This is partially pragmatic knowledge of the morpheme and partially semantic. Semantic methodology is presented in __ below.

6.8.4. Get translations and paraphrases from a variety of speakers.

Never rely on the pronunciation, usage, or judgements of a single speaker when analyzing a morpheme. You need to test your hypotheses and judgements with a variety of speakers, at least 3-6 men and women before feeling that you have reached a definitive conclusion.

6.8.6. Text tracking

One of the best examples I am aware of in the literature that traces a linguist's analytical history in relation to a set of morphemes is found in Lowe (1990). Lowe (1990, 544ff) states his problem as follows: "In Nambiquara, four morphemes, namely *–jau*^{3}*-*, *jut*^{3}*-*, *-kxai*^{3}*-*, *-kxe*^{3}*-* figure centrally in expressing the interclausal relationship of cause/reason. Each one of these morphemes occurs in two-clause constructions which can be faithfully translated as expressing a cause or reason relationship. There is the added complication, however that each one of these morphemes is used in other contexts where a causal relationship does NOT [emphasis Lowe's, DLE] seem to be present." Lowe proceeds to discuss the difficulties of analysing these morphemes, showing clearly that looking at their uses exclusively in isolated sentences will lead largely to confusion rather than conclusion. He then shows how an examination of their functions in a wider discursive context leads to a fairly straightfoward analysis. Before reviewing his findings, very useful from a methodological standpoint, I quote his foundational assumption: "The analysis presented in this paper must start with the assumption that each of the morphemes above has a unitary meaning, even though the same morpheme may appear to have different meanings in different contexts. Such an

assumption is justified provided it leads to consistent and insightful descriptions of the function of each morpheme. The danger of assigning disparate glosses for the same surface morpheme in different contexts is that each one of these disparate glosses will be assigned from an English speaker's point of view; *until one struggles to find what, if anything, they have in common, the foreign viewpoint will be all that one has*.' [emphasis mine, DLE].

Lowe goes on to demonstrate, convincingly, that one cannot understand either the notion of causality in Nambiquara, nor the ways that it is marked linguistically, without understanding both the cultural meaning of causality in Nambiquara and the various discursive uses of the different morphemes. In the course of his study, he also strongly supports his assumption of 'unitary meaning' as a vital methodological tool.

Another, similar, example comes from my own study of the intersentential connective, **hoagá** in Pirahã. Consider the following examples and the contexts in which I initially recorded them.

(6.145) **Hi toioʔaagá, *hoagá* hi opaohoaibaáí.**
he old is *hoagá* he works a lot.
'He is old .?.. he works a lot.'

(6.146) a. **Hi gogiísoi ʔigí -ai?**
he what-big:thing with is
b. **Ti *hoagá* pi-obaibaai.**

(6.147) **Ti gi *hoagá* poogaihiai bagaboi.**
'I you *hoagá* give banana.'

6.8.7. Bi-directional translate texts with language teachers

Another potentially useful method of verifying morpheme meanings is to translate with native speakers, who are also fluent in the language of the translation, the words containing the morpheme into the trade, national, or other relevant surrounding language. Then with other speakers, you should translate back into the target language. If you and they have fully understand the morpheme, the translations should produce similar results. If you are wrong, however, you should get unexpected variations.

6.9. Perspectives on morphological alternations
Homonymy, polysemy, and Allmorphy

The fieldworker will find it useful, but not always essential to distinguish polysemy (a single word with multiple meanings) from homonymy (multiple words with a single phonological form). For example, consider the different uses of the clitic *se* in Portuguese. Are these multiple morphemes with one form or multiple meanings of one morpheme? (Each 'se' is proceeded by a label for that particular function.)

Reflexive
(6.148) Sérgio se matou.
'Sergio killed himself.'

Argumental Impersonal
(6.149) Se recebe socos facilmente.

'One receives punches easily.'

Passive

(6.150) Maçãs ***se*** vendem por aqui.
'Apples are sold around here.'

Ergative

(6.151) A janela quebrou ***se*** (*a proposito).
'The window broke.' (*on purpose)

Inherent

(6.152) Tres meninos se desmaiaram. (Brazilian Portuguese)

Each of these uses of *se* from Portuguese (and the phenomenon repeats itself throughout Romance) is different in function. So is there one *se* or five? Just about every hypothesis imaginable has been proposed in the literature at one time or another (see Everett (1996, ---) for a summary).

The answer to this particular problem is not relevant to our present concerns (though I believe that this is only a case of apparent polysemy and that all these functions are related in a simple way). However, the general problem is relevant. What evidence could a fieldworker bring to bear on such a question if encountered in the field (and the odds are high that something like this will confront most fieldworkers). Historical evidence might help, but it is problematic for several reasons. First, it is often unavailable to the fieldworker. Second, even if it could be proven that *se* entered Latin as a single morpheme, that would not entail that it has been preserved as such in any modern Romance language. Things change! We saw a similar problem with the Salish facts in () – () above.

The ultimate answer to this question of homonymy vs. polysemy is found in the linguist's analysis and argumentation. The most satisfying, clearest, and simplest analysis will likely be the one linguists adopt as 'the answer'.

6.10. Cleaning up your messes –follow-up visits

Field researchers must always allow time following the evaluation of data they have collected to find 'holes' in the patterns of their data, for follow-up visits. Follow-up field trips should be scheduled into a project in response to the possibility of discovery, incomplete corpora, change in focus, etc. This is a way to build in error-correction into your project, using these follow up visits to fill in missing bits of the data and to 'clean up your messes', that is, verify conflicting transcriptions, recordings, textual organizations or understandings, etc.