

5 Escaping Eurocentrism: fieldwork as a process of unlearning

David Gil

1. Learning and unlearning

In her adventures in wonderland, Alice fell into a deep pool of her own tears, and then met a mouse:

“O Mouse, do you know the way out of this pool? I am very tired of swimming about here, O Mouse!” (Alice thought this must be the right way of speaking to a mouse: she had never done such a thing before, but she remembered having seen, in her brother’s Latin Grammar, “A mouse – of a mouse – to a mouse – a mouse – O mouse!”) (Lewis Carroll, *Alice in Wonderland*)

Like other children of her time, Alice had been brought up to believe that not only Latin but also English has six cases: nominative, genitive, dative, accusative, ablative, and vocative.

How this came about is quite obvious. In those days, grammarians worked within traditions that were based on the classical languages of antiquity. So when they first began to examine English, they encountered a language without nominal case marking. Accordingly, they concluded that the Latin cases were there – only invisible.

Today, the discipline of linguistics is more enlightened: we think we know better. But do we really? It is a conspicuous fact about contemporary linguistics that it was developed primarily by speakers of European languages, is practiced mostly in European languages, and even today exhibits a disproportionate concern with the study of European languages. Inevitably, the European history and sociology of the field results in a Eurocentric bias with regard to its content. If in previous centuries it was Latin that was imposed on English and other European languages, today it is English, or Standard Average European, which, via Eurocentric linguistic traditions, is being imposed, often inappropriately, on languages spoken in other parts of the world.

However, in recent years, there is an emerging consensus that this bias must be overcome, through increased efforts directed towards the investigation of languages spoken in other parts of the world. And indeed, as suggested by the various chapters in this volume, one of the best ways to escape

Eurocentrism is to engage in fieldwork, and to immerse oneself in a non-European language.

Among the well-known joys of linguistic fieldwork is the discovery of exotic new linguistic objects, patterns, categories, and structures. As a graduate student in a field methods class at UCLA in the late 70s, I remember trying to investigate the syntax and semantics of quantifier scope in Maricopa, a Yuman language of the southwestern USA. (A comprehensive description of Maricopa grammar can be found in Gordon 1986.) Having painstakingly prepared a set of drawings showing so many men carrying so many suitcases in a variety of complex and confusing combinations, I presented these drawings to the speaker, Polly Heath, and asked her how she might describe the various states of affairs depicted. To my surprise and disappointment, she found the task extremely difficult, and after several attempts at getting reliable data, I gave up in despair. Fortunately, however, this was a classroom situation, and our instructor, Pam Munro, was there to come to the rescue. The problem, she explained, had nothing to do with the intricacies of quantifier scope that I was interested in, but rather with something much more mundane: the verb “carry,” and the suitcases, or, more specifically, the shape in which I had, rather arbitrarily, drawn them. In Maricopa, it appears, there are different verbs “carry” for different shapes of objects carried – and the rectangular suitcases which I had drawn fell smack in the middle between compact objects, for which there is one verb, and elongated objects, for which there is another. Thus, the speaker’s apparent inability to describe the drawings stemmed from her difficulty in choosing the right verb “carry.” Not being familiar with the phenomenon of verbal classification, I was unable to make sense of her confusion, until I was enlightened by the teacher. The results of my work on quantification in Maricopa eventually found their way in to my Ph.D. dissertation (Gil 1982), in which the awkwardly rectangular suitcases were duly replaced by unambiguously elongated sticks. At the time, though, I should have been able to do better by myself. But that was the point of the course – to teach us how to unearth, recognize, and then analyze those exciting and unexpected new linguistic objects, like verbal classification, that are out there among the world’s 5,000 or 6,000 languages.

However, when confronted with a new language, it is sometimes easier to recognize the *presence* of exotic, unexpected, and hitherto-unknown items than to come to grips with the *absence* of familiar, commonplace, and presumed-to-be-universal entities. Our native language imposes a strait-jacket from which it is often difficult to break free, in order to realize that certain grammatical categories, obligatory in our own language, may be absent in the language under investigation. Moreover, if our native language is European, then this straitjacket is likely to be reinforced by the

weight of Eurocentric linguistic traditions, which either implicitly presuppose or else explicitly assert that certain grammatical categories are universal. Thus, fieldwork involves not only the *learning* of new items but also the *unlearning* of old and familiar ones.

The problem is mainly with the “zeroes.” Within most linguistic theories, there are a variety of zero elements: null items, empty positions, or otherwise noumenal entities which, even though they cannot be seen or heard, are still believed to be there. To cite just one example in morphology, in a paradigm such as the Hebrew verbal past tense (using the verb ‘act’), *paʕal-ti* (1:SG), *paʕal-ta* (2:SG:M), *paʕal-t* (2:SG:F), *paʕal* (3:SG:M), *paʕal-a* (3:SG:F), *paʕal-nu* (1:PL), *paʕal-tem* (2:PL:M), *paʕal-ten* (2:PL:F), *paʕal-u* (3:PL), the third person singular form *paʕal*, is commonly considered to contain a zero suffix, *paʕal-Ø*. This is because in all other forms in the paradigm there is an overt suffix marking person, number, and gender. Paradigmatic patterning is one good reason for positing zeroes, and there are other, equally good ones. But there are also some pretty bad reasons. All too often, we posit a zero element in a language just because our European languages, or our Eurocentric theories, lead us to believe that there should have been something there. Becker (1995: 291) notes that “each language, from the point of view of another, appears full of holes.” But “appear” is crucial here: the hole, or zero element, has no existence other than in eye of the beholder.

This issue can be most keenly appreciated in the treatment of Southeast Asian languages, in which many of the staple categories of traditional grammatical theory are apparently lacking – see Ratliff (1991), Riddle and Stahlke (1992), Huang (1994), Bisang (1996), and Gil (in press). For example, theoretically-oriented syntactic descriptions of Southeast Asian languages often posit a category of verbal INFL(ection): see Ernst (1994) for Mandarin, Huang (1991) for Hokkien, Lehman (1998) for Thai, Ramli (1995) for Malay, and many others. Why so? Because English inflects its verbs; therefore this must be part of Universal Grammar; therefore Mandarin, Hokkien, Thai and Malay must do so, too. Even if they seem, on the surface, not to. Moreover, the problem is not limited to the practitioners of specific theoretical frameworks. Many would-be atheoretical fieldworkers – while deriding their theoretically-oriented colleagues for their excessive abstractness and lack of concern with linguistic diversity – fall into the same trap when characterizing the grammars of Southeast Asian languages in terms of categories whose justification stems, once again, from Standard Average European. Some of the many grammatical categories often imported uncritically from traditional grammatical theory are parts of speech such as noun, adjective, and verb; grammatical relations such as subject and direct object; and a host of more specific construction types, including relative clauses, conjunctions, reciprocals, and many more. As a

result, descriptions of Southeast Asian languages often resemble Alice's grammar, with Eurocentric paradigms such as [mouse]_{NOUN}, [mouse]_{ADJECTIVE}, [mouse]_{VERB}, and so forth.

This chapter provides a personal account of my fieldwork experiences with several Southeast Asian languages, focusing on Hokkien, Tagalog, and Malay/Indonesian. It is the story of my ongoing struggles to unlearn the grammatical categories of my native languages (Hebrew and English), and liberate myself from a Eurocentric linguistic education (degrees from Tel Aviv University and UCLA). It is the tale of my efforts to come to grips with the patterns and structures of diverse languages and to describe them as they should be described, on their own terms. And finally, it is an attempt to share some of the excitement of viewing the world from a new perspective, that of a non-European language, which, through familiarity, ceases to be exotic.

2. Macrofunctionality in Hokkien

Different languages carve up reality in different ways. Color terms are a celebrated case: for example, the three Hebrew words *txelet* 'light blue', *kahol* 'dark blue', and *yarok* 'green' cover the same domain as the two English words 'blue' and 'green', which in turn cover the same domain as the single Riau Indonesian word *hijau*. Kinship terms are another well-known case: for example, the two English words 'brother' and 'sister' span the same semantic field as the two Riau Indonesian words *kakak* 'elder sibling' and *adik* 'younger sibling', but each language slices up the pie differently – English by sex, Riau Indonesian by age. Similar cross-linguistic variation can be found in just about every semantic domain. Discovering such variation, and seeing the world through the perspective of a different language, is one of the greatest joys of language learning, and of linguistic fieldwork.

In some cases, a single form in English will correspond to two or more forms in the new language, for example English 'blue' and Hebrew *txelet* and *kahol*. In other cases, a single form in English will overlap with another form in the new language, for example English 'brother' and Riau Indonesian *kakak*. In yet other cases, two or more forms in English will correspond to a single form in the new language, for example English 'blue' and 'green' and Riau Indonesian *hijau*. In cases of the latter kind, the single form in the new language will appear to exhibit the property of *macrofunctionality*, being associated with a function which, from an English perspective, seems to be surprisingly large, encompassing the two or more functions associated with the corresponding English forms.

In cases of apparent macrofunctionality, a range of analytical strategies present themselves. At one end of the spectrum, one may characterize the

macrofunctional form as having a single, unified meaning, and as vague with respect to the distinctions associated with the different English forms. At the other end, one may describe the form in question as being ambiguous, each of its meanings corresponding to that of one of the English forms. In between these two extremes, one may analyze it as being polysemous, that is to say, as associated with a set of distinct but related meanings: distinct in the sense that they bear different consequences with regard to grammatical structure, related in the sense that they share a single core meaning, or are related through a network of common meanings.

In general, the default hypothesis in all cases of macrofunctionality should be to posit a single, unified meaning. One of the dominant design features governing the structure of language is the principle of one-form-one-meaning, which says that each form in a language has a unique meaning different from that of each other form. An overwhelming body of empirical evidence can be cited in support of this principle – see, for example, Tobin (1990).

Nevertheless, the principle of one-form-one-meaning is far from exceptionless: ambiguity and polysemy are widespread throughout language. Thus, for each case of macrofunctionality that is encountered, it is necessary to invoke a set of objective criteria in order to choose between the various possibilities. One such criterion is the obvious one. If it is possible to define a single common and coherent meaning without recourse to an *ad hoc* listing of sub-meanings, then the form in question is vague or polysemous; if, on the other hand, no such common meaning exists, then it is ambiguous. A second criterion is cross-linguistic replicability. If a particular broad meaning is associated with a single form in a variety of geographically, genetically, and typologically diverse languages, then in each language the form in question is vague or polysemous; if, however, an apparent instance of macrofunctionality occurs in just one language, and does not recur cross-linguistically, then the form in question is ambiguous. (Some of the issues involved in distinguishing vagueness, polysemy, and ambiguity are discussed in Zwicky and Sadock 1975.)

Consider, for example, the Riau Indonesian form *hijau*, corresponding to English ‘blue’ and ‘green’. Clearly, this form has a single unified meaning, which can be easily defined as a continuous area on the color space. Moreover, words with a similar meaning recur in a wide range of languages, for example Wari’ (a Chapakuran language of western Brazil), Setswana (a Bantu language of southern Africa), Welsh, and Japanese. Therefore, the form can safely be characterized as vague or polysemous. In contrast, consider the Riau Indonesian form *tahu*, corresponding to English ‘tofu’ and ‘know’. No amount of mental gymnastics will come up with a single unified meaning here; and in almost all languages these meanings are expressed by different forms. Hence, Riau Indonesian *tahu* can reasonably be characterized as ambiguous.

In practice, however, the treatment of macrofunctionality often suffers from a Eurocentric bias, whereby fieldworkers and other researchers tend to characterize a form as ambiguous simply because it corresponds to two or more forms in English, or is associated with two or more grammatical categories in the investigator's theoretical framework. (Indeed, this bias is even reflected in the prevalent terminology. Rather than *macrofunctionality*, many writers use the terms *polyfunctionality* or *multifunctionality*; however, these latter terms presuppose that the function is constituted from a plurality of more specific functions, thereby implying that the form in question is ambiguous or polysemous, as opposed to being simply vague – as per the default assumption.)

The following is an example of macrofunctionality from Hokkien, a language belonging to the Southern Min group of Sinitic. (The data presented below are from the dialect of Hokkien spoken in Singapore.) In Hokkien Chinese, nominal attribution is expressed via a construction of the form ATTRIBUTE e^{24} NOUN, with the marker e^{24} occurring between the attribute and the head noun. (There are additional ways of expressing nominal attribution, involving other markers, but we will not be concerned with these here.) Some examples of nominal attribution containing the marker e^{24} are given in (1) – (6) below:¹

- | | | | | |
|-----|----------------------------|----------------|----------------------------|----------------------------|
| (1) | $a^{44(>44)}-ber\eta^{24}$ | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ | |
| | Ah Beng | ASSOC | apple | |
| | Ah Beng's apples | | | |
| (2) | $si^{21>53}$ | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ | |
| | four | ASSOC | apple | |
| | four apples | | | |
| (3) | ci^{t4} | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ | |
| | DEM:PROX | ASSOC | apple | |
| | these apples | | | |
| (4) | $a\eta^{24}$ | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ | |
| | red | ASSOC | apple | |
| | red apples | | | |
| (5) | $tou^{21>53}$ | $tie\eta^{53}$ | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ |
| | table | top | ASSOC | apple |
| | apples on the table | | | |
| (6) | $a^{44(>44)}-ber\eta^{24}$ | bue^{53} | $e^{24>22}$ | $p^he\eta^{24>22}-ko^{53}$ |
| | Ah Beng | buy | ASSOC | apple |
| | apples that Ah Beng bought | | | |

As shown above, the attributive e^{24} construction in Hokkien corresponds to (at least) six formally distinct attributive constructions in English: the genitive -'s construction in (1); the numeral construction in (2); the demonstrative construction in (3); the adjectival construction in (4); the

prepositional-phrase construction in (5); and the relative clause construction in (6). So to a Eurocentric eye, it looks as though there are several different e^{24} s in Hokkien, or at least a number of different usages of a single e^{24} form (Bodman 1955). But is this indeed really the case?

Invoking the above two criteria, it is clear that Hokkien e^{24} is not multiply ambiguous. The unified meaning of e^{24} can easily be characterized; in fact it already has been above, namely as a marker of nominal attribution. More specifically, in a construction of the form *ATTRIBUTE* e^{24} *NOUN*, the attribute is interpreted as being associated, in some unspecified way, with the head noun. For example, in (1)–(6) above, the apples are understood as being associated in a contextually appropriate manner with, respectively, a person, a cardinality, an act of deixis, a color, a location, and an activity. Moreover, paradigms replicating (1)–(6) above, in part or in whole, recur cross-linguistically: for example, in many languages of sub-Saharan Africa (Welmers 1973) and in other parts of the world (Aristar 1991).

What this suggests, then, is that Hokkien does not have distinct attributive constructions corresponding to English genitive, numeral, demonstrative, adjectival, prepositional phrase, and relative clause constructions. Rather, it has a single attributive construction, involving the form e^{24} . (Analogous arguments for other East Asian languages have been proposed by Comrie 1996, 1998.)

For the fieldworker, the moral is quite clear: translational equivalence does not entail structural equivalence. Meaning is meaning, form is form – the two should not be confused. For example, just because the Hokkien $a^{44(>44)}-ber^{24}$ bue^{53} $e^{24>22}$ $p^h e r^{24>22}-ko^{53}$ translates into an English relative clause construction ‘apples that Ah Beng bought’, this does not mean that the form in question is a relative clause construction in Hokkien. Thus, when encountering macrofunctionality, the fieldworker must unlearn the specific constructions of his or her native language, and of his or her theoretical framework, in order to be able to describe the constructions of the new language, as they really are.

Some additional, more specific lessons can also be drawn from the above example. Many speakers of Hokkien, when asked about constructions such as (1)–(6), will maintain that there are neither six different e^{24} s, nor a single one; rather, they will insist that there are two e^{24} s – one occurring in (2) and (3), the other occurring in (1), (4), (5) and (6). When asked why, they will offer two related explanations. The first is that the two different e^{24} s correspond to two distinct forms in Mandarin Chinese, *ge* and *de* respectively. Needless to say, this is irrelevant. For a Chinese person, Hokkien may be a “dialect,” or debased form, of Mandarin, the so-called “proper” Chinese; but for the linguist, Hokkien and Mandarin are simply two different languages. To impose Mandarin grammar, with its two distinct attributive con-

structions, on Hokkien, is every bit as unjustified as is imposing English, with its six or more distinct constructions, on Hokkien. The second explanation that is offered is that the two different e^{24} s are written with two different characters. But linguistics is about spoken languages: writing is of a secondary, derivative nature, and should not be invoked in support of one linguistic analysis or another. Indeed, in the case at hand, the writing system is not specifically Hokkien; rather, it is a pan-Chinese system modelled largely after Mandarin. (For more detailed discussion of these and other methodological issues involved in the study of Sinitic languages, see Ansaldo 1999 and Matthews in press.) Cases such as this underscore the need for the field linguist to be on constant guard against admitting the speakers' own extra-linguistic biases as evidence for particular hypotheses. Speakers are just that, speakers; it is the fieldworker's job to engage in linguistic analysis.

A second lesson to be drawn from the above example relates to the relevance of phonology to syntactic and semantic argumentation. The reader will have noted the detailed notation of lexical tone in the above examples, and may have puzzled over the frequent use of the symbol $>$, as, for example, in $e^{24>22}$. This symbol marks the occurrence of *tone sandhi*, a phonological process whereby the tone associated with a syllable changes to another one: for example, in the form $e^{24>22}$, the basic, or citation tone 24 changes to the derived, or sandhi tone 22. (For discussion and analysis of tone sandhi in various dialects of Hokkien, see Cheng 1968, 1973, Chen 1987, and Peng 1994, 1997.) The tone changes that take place in Hokkien are given in (7) below:

(7)	basic tone	>	sandhi tone
(a)	24	>	22
(b)	22	>	21
(c)	21	>	53
(d)	53	>	44
(e)	44	>	44
(f)	32	>	4 (before <i>p, t, k</i>) / 53 (before ?)
(g)	4	>	21

As suggested above, Hokkien has an inventory of seven lexical tones, listed in the first column of (7). When sandhi occurs, each tone changes to another one of the same seven tones, as indicated in the second column of (7). (The dialect of Hokkien described here differs from that described by Chen 1987 in the following two respects: (a) the 22 and 21 tones in (7b) and (7c) are phonetically indistinguishable but differ with respect to their sandhi tones, whereas in Chen's dialect they are phonetically distinguishable; and (b) the 44 tone in (7e) remains unchanged, whereas in Chen's dialect its sandhi tone is 22.)

What is interesting from a syntactic point of view is *when* tone sandhi occurs. The distribution of tone sandhi in a given construction can be described in the following way:

- (8) (a) ... [\$\$\$] [\$ \$] [\$] [\$ \$ \$ \$ \$] ...
 (b) ... [S S B] [S B] [B] [S S S S B] ...

A sequence of syllables, denoted '\$,' is parsed into constituents, or *tone groups*, as for example in (8a). Within each tone group, the last syllable retains its basic tone, denoted 'B,' while all non-final tones in the tone group change to their sandhi tones, denoted 'S,' as shown in (8b). Thus, the occurrence of tone sandhi effects a parsing of each and every construction in Hokkien, which correlates in part with the syntactic constituency of the construction.

With this in mind, let us examine the paradigm in (1)–(6). In every one of the examples, the marker e^{24} surfaces in its sandhi form, $e^{24>22}$, showing that it belongs to the same tone group as the following noun: ... $e^{24>22}$ $p^h e \eta^{24>22}$ $-ko^{53}$. Of interest to us here is the syllable preceding the attributive marker $e^{24>22}$, namely, the last syllable of the attributive expression. In example (2), the numeral $si^{21>53}$ 'four' occurs in its sandhi form, showing that the whole construction constitutes a single tone group, [$si^{21>53}$ $e^{24>22}$ $p^h e \eta^{24>22}$ $-ko^{53}$]. In contrast, in the remaining five examples, the last syllable of the attributive expression occurs in its basic tone. This shows that [$e^{24>22}$ $p^h e \eta^{24>22}$ $-ko^{53}$] constitutes a complete tone group, while the attributive expressions preceding it constitute another, separate tone group: [$a^{44(>44)}$ $-be \eta^{24}$] in (1), [$ci \eta^{4}$] in (3), [$a \eta^{24}$] in (4), [$to u^{21>53}$ $tie \eta^{53}$] in (5), and [$a^{44(>44)}$ $-be \eta^{24}$ bue^{53}] in (6). Thus, the tone sandhi facts suggest that the syntactic bond between the attributive marker e^{24} and a preceding numeral, as in (2), is stronger than that between the attributive marker and other kinds of attributive expressions. (Interestingly, in the dialect described by Chen, the last syllable of the attributive expression undergoes sandhi if it is a numeral or a demonstrative; in his dialect, then, sandhi would occur also in (3).) This in turn suggests that e^{24} may exhibit a certain degree of polysemy, entering into two distinct attributive constructions, one with numerals (and for Chen also demonstratives), the other with all other kinds of attributive expressions.

While some fieldworkers like phonology, others are more interested in syntax and semantics, which is only fair. Some syntacticians and semanticians, though, are uncomfortable with phonology; indeed, many seem to feel particularly intimidated by lexical tone. At least in part, this is due to a Eurocentric perspective, and the absence of lexical tone from most or all of the languages of Europe. As a result, data from tonal languages are all too often cited with the tonal information omitted. For example, in volume 7 (1998) of the *Journal of East Asian Linguistics*, six of the ten articles are

concerned with tonal languages; however, only one of these six articles cites data with lexical tones marked. The remaining five articles, containing data from several Chinese languages as well as two tonal African languages, Yoruba and Ewe, fail to mark the tones. But this is just as though a Hokkien linguist chose to cite English data without final consonant clusters simply because he or she could not hear or pronounce them. However, as the above example shows, phonological, and in particular tonological information can be relevant to syntactic and semantic analysis. What is more, when working on a new language, there is no way of knowing in advance which aspects of the phonology will turn out to be relevant to a syntactic/semantic analysis, and to what extent. The conclusion to be drawn is clear: the field-worker cannot study the syntax and semantics of a language without also giving due consideration to its phonology.

The Hokkien attributive marker *e*²⁴ is but one of a lengthy list of cases which I have encountered in the course of my work, in which a form whose range of usages appeared, at first, extraordinarily broad, turned out upon closer reflection to be related, either through polysemy, or as instantiations of a single construction with a unified function. When I began working on my Ph.D. dissertation, it emerged that a distributive numeral, such as Japanese *sankozutu*, could mean either ‘three each’, or ‘in threes’; the resulting dissertation (Gil 1982) accounts for this range of meanings in terms of a single unified semantic relation of distributivity applying over a variety of syntactic domains. Some years later, working on Malayalam, it turned out that one and the same suffix, *-um*, could mean ‘and’, ‘also’, ‘even’ and ‘every’; in a series of articles (Gil 1994a, 1994b, 1995a), these meanings are assigned a unified semantic representation in terms of the notion of conjunctive operator. At present I am studying the range of usages of the Riau Indonesian form *sama*, which, in different contexts, appears to be endowed with a wide range of functions, including ‘and’, ‘with’, ‘same’, reciprocal, agentive, oblique, and object of comparison; in work in progress, I am attempting to come up with a common core meaning underlying all of these usages, something along the lines of ‘together’. Each of these cases, and many other similar ones, brought with it the satisfaction of suddenly realizing that what seems, from a Eurocentric point of view, to be an array of disparate entities is, from the perspective of the language itself, simply one and the same macrofunctional item.

3. Syntactic categories in Tagalog

Tagalog, the major language of the Philippines, is a relatively well known language, with several good reference grammars, such as Blake (1925) and Schachter and Otanes (1972), and lots of discussion in the linguistic

literature, for example Schachter (1976, 1977), Carrier-Duncan (1985), and Kroeger (1993). As a graduate student, I was fascinated by what I had read and heard about the language, and set off to the Philippines to check it out for myself. For a period of several years I elicited data from native speakers while at the same time learning to speak the language simply in order to be able to communicate with people. It took me too long, but finally I realized that the language that I was obtaining through elicitation and then writing articles about was not the language I was learning to speak.

Ask a speaker of Tagalog how to say ‘The chicken is eating’, and you might get a sentence such as the following:

- (9) *Ang manok ay kumakain.*
 TOP chicken INV PROG-ACT.TOP:REAL-eat
 The chicken is eating.

As suggested by the above gloss, the structure of the Tagalog sentence bears a superficial resemblance to its English counterpart. The word for ‘chicken’, *manok*, is preceded by a grammatical formative *ang* which, among other things, marks it as definite; and the word for ‘eat’, with stem *kain*, is marked for voice, tense, and aspect by reduplication, *kakain*, and subsequent infixation of *-um-*, *kumakain*. Finally, *ang manok* is linked to *kumakain* with the grammatical marker *ay*, which occurs in a position reminiscent of the English ‘is.’ Indeed, sentences such as (9) above are still cited in some Tagalog pedagogical grammars as evidence for the claim that basic sentence structure in Tagalog is the same as in English, namely *subject – copula – verb*.

Nevertheless, most modern descriptions of Tagalog recognize the fact that sentences such as (9) occur relatively infrequently, and only in formal registers. When asked to translate the same English sentence, a speaker of Tagalog is actually more likely to provide the following:

- (10) *Kumakain ang manok.*
 PROG-ACT.TOP:REAL-eat TOP chicken
 The chicken is eating.

In (10), *kumakain* ‘is eating’ precedes *ang manok* ‘the chicken’, and the marker *ay* is absent. In fact, most linguistic descriptions of Tagalog consider the construction illustrated in (10) to be the simple or unmarked one, and accordingly characterize Tagalog as exhibiting *predicate–topic*, or *verb–subject*, basic word order. As for the construction exemplified in (9), it is usually taken to be more complex or highly marked, the additional form *ay* being characterized as an explicit marker of *inversion*.

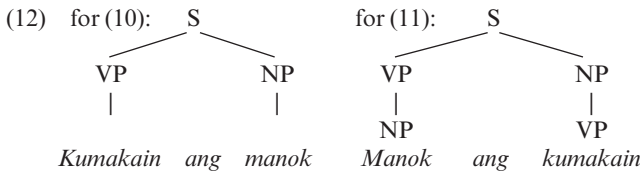
That is what the grammar books say, and that is what I was getting from elicitation from native speakers. But the language I was learning to speak, through simple immersion and the usual processes of second language

acquisition, was turning out to be quite different. Just as frequently as the construction in (10), I was encountering, both in speech and in writing, constructions such as that of the following:

- (11) *Manok ang kumakain.*
 chicken TOP PROG-ACT.TOP:REAL-eat
 The chicken is eating.

The above example differs from its predecessor in that the two main words, *manok* ‘chicken’ and *kumakain* ‘is eating’ are interchanged. However, the grammatical marker *ang* remains in the same position, and thus, in (11), is marking *kumakain* rather than *manok*.

Turning to the reference grammars, I soon found mention of “predicate nominal constructions,” in which words such as *manok* ‘chicken’ occur in sentence-initial predicate position. But it was harder to find any discussion of constructions such as *ang kumakain*. To the extent that their existence was at all acknowledged, they were characterized as “nominalized predicate” constructions, having undergone a process of “zero-conversion” from VP to NP. A typical analysis of the constructions in (10) and (11) might look as follows:



Occasionally, in line with the above, translations of constructions such as in (11) were offered involving English cleft constructions, for example, ‘The one that is eating is the chicken’.

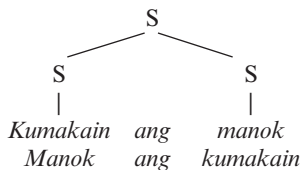
This, however, struck me as unsatisfactory for several reasons. First, I could find no explicit arguments in support of any such rules of conversion, changing a nominal *manok* ‘chicken’ into a predicate, and transforming a verbal *kumakain* ‘is eating’ into an NP. Structurally, (11) appears to be completely parallel to (10). And pragmatically, sentences such as (11) can be used in situations where an English cleft construction such as ‘The one that is eating is the chicken’ seems inappropriate. But secondly, what my ears and eyes were telling me was that constructions such as (11) occur quite naturally and commonly, with perhaps the same frequency as their counterparts in (10). Why, then, I wondered, should they be characterized as more highly marked, involving seemingly unmotivated grammatical processes of *zero-conversion*?

I troubled over these matters for years, but then, finally, the Eureka moment arrived. Here was Alice and the mouse all over again! If *manok*

and *kumakain* can occur in all of the same positions, then what justification is there for assigning them to two different parts of speech, NP and VP? Of course, their translational equivalents in English, ‘chicken’ and ‘is eating’, belong to two different categories, NP and VP, but that is a fact about English, which cannot and should not be carried over to Tagalog. Trying to analyze a Tagalog sentence in terms of English parts of speech is like trying to describe English nominal morphology in terms of the six cases of Latin. In fact, the supposed zero-conversion of *kumakain* from VP to NP in sentence (11) makes no more sense than a would-be zero ablative suffix on an English noun.

Instead, I realized that Tagalog simply does without the traditional parts of speech: it has no distinction between nouns, verbs, and adjectives, nor between lexical categories and their phrasal projections. After some reflection, I came to the conclusion that what it has instead is a single open syntactic category, S, corresponding more or less to the traditional category of sentence. In particular, pairs of sentences such as (10) and (11) share a common syntactic structure, such as that indicated below:

(13) for (10) and (11):



As soon as I had figured this out, the whole language fell into place, metamorphosing almost instantaneously from a strange, exotic, and somewhat bewildering labyrinth into a simple, elegant, and crystal-clear edifice. The absence of parts-of-speech distinctions accounted straightforwardly for a wide range of seemingly unrelated facts, first and foremost among which being the observation that almost all expressions enjoy the same distributional privileges: with but a limited number of exceptions, anything can go anywhere. But this insight only came when I was able to shed the blinkers of Eurocentric grammatical theories, and stop looking everywhere for nouns and verbs, or NPs and VPs; in other words, once I had unlearned the parts of speech of my native language, and of the syntactic theories I had been brought up on.

The possibility that Tagalog may be lacking many parts-of-speech distinctions, foreshadowed by Bloomfield (1917), has recently been raised by Himmelmann (1991) and Shkarban (1992, 1995); however, the above claims remain controversial. But the point here is not to convince the reader that this particular analysis of Tagalog is the best one – the relevant arguments have already been presented elsewhere, see Gil (1993a, 1993b, 1995b) for discussion of Tagalog, and Gil (2000) for more general theoretical con-

siderations. Like any analysis, it may turn out to be quite wrong. Rather, the goal herein is simply to show how the Eurocentric bias in linguistics tends to lead researchers away from proposing certain hypotheses which, regardless of whether they are ultimately right or wrong, are at least plausible, and worth formulating.

But there is a further methodological moral to the story: eliciting data from native speakers is a valuable way of obtaining data, but it can never provide the whole picture of a language. Native speakers can translate sentences, and they can provide judgments of well-formedness, meaning, and appropriateness in context; but they can never produce an exhaustive list of all the interesting constructions in the language. When eliciting data from native speakers, you tend to get what you ask for. As a result, it is very easy to miss out on a whole lot that is there because it never occurs to you to look for it. In the case at hand, if I had stuck to eliciting data from native speakers, I would have encountered many sentences such as (9) and (10), but might never have come across sentences such as (11). It was only my exposure to the language, in speech and in writing, that brought to my attention the existence of constructions such as those in (11), and then convinced me of their prevalence and importance.

It is worth asking why speakers tend to offer certain sentences rather than other, equally grammatical ones. One reason for this is what I like to refer to as the *good informant paradox*: the better s/he is, the worse s/he is. When offering an English sentence to a native speaker, the more skilled informant will provide a translational equivalent that is as close as possible to the English source sentence, not just in meaning but also in structure. Accordingly, the more talented one's informants, the more likely one is to end up with a corpus of sentences which, although grammatical, are actually too much like their English source sentences, thereby providing a distorted picture of the language under investigation, and downplaying the degree to which it differs from English. In the case at hand, Tagalog speakers apparently felt that the English source sentence 'The chicken is eating' corresponds more closely to Tagalog sentences (9) and (10) than it does to sentence (11), and therefore didn't offer sentence (11), even though it is grammatical and an appropriate translation of the original English sentence.

This shows that in general, the elicitation of data from native speakers cannot provide an adequate empirical basis for the description of a language. But neither can any other single method, such as the collection of texts. There is no privileged field method in the study of language: linguists are like the proverbial blind men groping at the elephant, each from his or her own particular angle.

4. Basic sentence structure in Riau Indonesian

Malay/Indonesian is one of the world's major languages, with up to 200 million native speakers. Actually, though, it is not one language, or even two, but a family of languages with about as much internal diversity as the Slavic or Romance language families. In the early 1990s, I got a job in Singapore and began learning the variety of Indonesian spoken on the islands right across from Singapore, in the Indonesian province of Riau. After a short while, it became clear to me that the language that I was learning, Riau Indonesian, was very different from the well-known standardized varieties of Malay/Indonesian, and had not been previously described.

Ask a speaker of Riau Indonesian to translate the English sentence 'The chicken is eating', and the answer might be as follows:

- (14) *Ayam makan*
 chicken eat
 The chicken is eating.

As suggested by the above example, Riau Indonesian is an isolating language with very little morphology: in the above example, each word consists of a single morpheme.

As already pointed out, Southeast Asian languages typically exhibit a large degree of semantic vagueness, with various categories, obligatorily grammaticalized in most European languages, left underspecified. In Riau Indonesian, the characteristic Southeast Asian grammatical indeterminacy is perhaps at its most extreme. In the above example, *ayam* 'chicken' is unmarked for number, allowing either singular or plural interpretations; and in addition it is unmarked for (in)definiteness, permitting either definite or indefinite readings. Similarly, *makan* 'eat' is unmarked for tense and aspect, allowing a variety of interpretations, such as 'is eating', 'ate', 'will eat', and others.

So far, not too surprising, but this is only the tip of the iceberg. Arbitrarily keeping constant the singular definite interpretation of *ayam* and the present progressive interpretation of *makan*, the above construction can still be interpreted in many different ways, some of which are indicated below:

- (15) *Ayam makan*
 chicken eat
- (a) The chicken is being eaten.
 - (b) The chicken is making somebody eat.
 - (c) Somebody is eating for the chicken.
 - (d) Somebody is eating where the chicken is.
 - (e) the chicken that is eating
 - (f) where the chicken is eating

- (g) when the chicken is eating
- (h) how the chicken is eating

Comparing the gloss in (14) with those in (15a)–(15d), we see that in the above construction, *makan* does not assign a particular thematic role to *ayam*: the chicken could be the agent, as in (14); the patient, as in (15a); or any other imaginable role, such as the cause in (15b); the benefactive in (15c); the locative in (15d); and so forth. And comparing the gloss in (14) with those in (15e)–(15h), we find that the construction as a whole may be associated with an interpretation belonging to any ontological category: an activity, as in (14); a thing, as in (15e); a place, as in (15f); a time, as in (15g); a manner, as in (15h); and others.

When learning Riau Indonesian, it took me some time to become aware of the extent to which underspecification is permitted. Again, it was only through exposure to naturally occurring speech that I was able to appreciate the widespread nature of the phenomenon. However, even after I had realized how pervasive this indeterminacy actually was, it still took me a long time to figure out how to deal with it.

My initial reaction was to attempt to provide a different analysis for each interpretation – or, as I would now say, for each of what seemed to me, mistakenly, at the time, to be distinct interpretations. For example, interpreted as (15a), *Ayam makan* might be analyzed as a “zero-marked passive.” Or, interpreted as (15e), *Ayam makan* might be assigned the structure of a “zero-marked relative clause.” But something was clearly wrong with this proliferation of zero markings. To begin with, it made the language look more abstract and complex, when my gut feeling was telling me that if anything it was more concrete and simple. A more specific objection was that each and every sentence in the language was turning out to be multiply ambiguous. But were these really ambiguities, or was this an artefact of imposing Eurocentric categories on a language that didn’t really have them? It was time to listen more closely to the speakers themselves, and how they use the language.

Poets, diplomats, and a few other people construct ambiguous sentences deliberately, for their own specific purposes; but most ordinary people do not. Take a garden-variety ambiguous sentence such as *The chicken is ready to eat*. This sentence can be understood as either ‘The chicken is ready to eat the food’ or ‘The chicken is ready to be eaten’. However, in any given utterance of the sentence, the speaker will have only one of these two interpretations in mind: it is virtually impossible to imagine a situation in which a speaker would utter the above sentence without caring which of the two interpretations it is assigned. This is a good reason to characterize the sentence as ambiguous with respect to the thematic role of *chicken*. Now consider the fact that, under the latter interpretation, the chicken could be

fried, boiled, stewed, fricasseed, and so on. Although in many situations the speaker might know how the chicken is prepared, in many other situations he or she may not. But in those situations, the speaker simply would not care, and the sentence could still be appropriately uttered. Clearly, in such cases, we would not want to characterize *chicken* as ambiguous with respect to the mode of preparation. Rather, in such instances, as in many other similar ones, the expression may be characterized as vague with respect to the feature in question.

Now let us examine some specimens of spontaneous speech in Riau Indonesian. (In the following examples, the context associated with each utterance is indicated in square brackets.)

- (16) *Kalau ada penjahat mau dia dikejar.*
 TOP exist PERS-bad want 3 PAT-chase
 [a monkey's owner, about his monkey]
 (a) If there's a bad guy, he'll chase him. *singular*
 (b) If there are bad guys, he'll chase them. *plural*
- (17) *Aku Cina tak makan la.*
 I:SG China NEG eat CONTR
 [going out to eat, approaching a Chinese looking place]
 (a) I'm not eating Chinese food. *patient*
 (b) I'm not going to eat in a Chinese place. *locative*
- (18) *Ini bisa juga.*
 DEM:PROX can CONJ.OP
 [playing laptop game, speaker discovers that another key also works]
 (a) This one can too. *theme*
 (b) With this one you can too. *instrument*
 (c) This one makes you able to do it too. *cause*
- (19) *Cantik gol.*
 beautiful goal
 [watching replay of football goal on TV]
 (a) That goal was beautiful. *property*
 (b) That was a beautiful goal. *activity*
- (20) *Ini bagus bajunya.*
 DEM:PROX good shirt-ASSOC
 [putting on a newly bought shirt, admiring it in mirror]
 (a) This shirt is good. *property*
 (b) This is a good shirt. *thing*

Sentence (16) provides a relatively straightforward example of vagueness with respect to number: in the given context, the speaker has no reason to distinguish between singular and plural interpretations of the expression *penjahat* 'bad guy'. However, the remaining sentences provide more far-reaching and surprising instances of vagueness.

Sentences (17) and (18) illustrate vagueness with respect to thematic roles. In (17), *Cina* ‘China’ may be construed as referring either to the food, as in (17a), in which case it would be the patient of *makan* ‘eat’; or to the restaurant, as in (17b), in which case it would be the locative. And in (18), *Ini* ‘this one’ may be understood as the theme of *bisa* ‘can’, as in (18a); the instrument, as in (18b); or the cause, as in (18c). However, in the given contexts, the various construals end up meaning the same thing, and it is hard to imagine that the speaker could have been intending to convey one interpretation to the exclusion of the other or others. Rather, in the contexts at hand, it is a safe bet that the speakers had in mind a single undifferentiated reading encompassing the given glosses.

Sentences (19) and (20) exemplify vagueness with respect to ontological categories. In (19), *Cantik gol* may be interpreted either as in (19a), denoting a property, being beautiful, predicated of an activity, the goal; or as in (19b), denoting an activity, the goal, with an attributed property, being beautiful. Similarly, in (20), *bagus bajunya* may be understood either as in (20a), denoting a property, being good, predicated of a thing, the shirt; or as in (20b), denoting a thing, the shirt, with an attributed property, being good. Once again, in the contexts at hand, the different readings end up meaning the same thing, and it is clear that the speakers were intending a single underspecified interpretation unmarked with respect to ontological categories and whether the property is predicated or attributed.

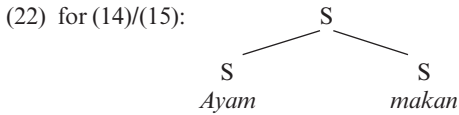
Thus, the above examples show that thematic roles and ontological categories are not obligatorily marked in the grammar of Riau Indonesian. This suggests that basic sentences such as (14)/(15) should be considered not as multiply ambiguous, but rather as vague with respect to thematic roles and ontological categories. More specifically, a sentence such as (14)/(15) may be associated with a single undifferentiated meaning, indicated in the gloss below:

- (21) *Ayam makan*
 chicken eat
 Entity associated with chicken and with eating

The above gloss may sound awkward in English, but the idea behind it is straightforward. Combine any two expressions in Riau Indonesian, and the meaning of the resulting collocation is, quite simply, anything that is associated, in some way or another, with the meaning of the two constituent expressions. Each of the glosses in (14) and (15) can be construed as a particular case of the gloss in (21) above, obtained by the imposition of further semantic constraints.

Syntactically, too, basic sentences such as (14)/(15) instantiate a single general construction type, rather than a variety of distinct, zero-marked

constructions such as zero passive, zero relative clause, and so forth. The simple structure of such sentences is indicated below:



As suggested in (22) above, Riau Indonesian, like Tagalog, would appear to have just a single major syntactic category S: this reflects the fact that in Riau Indonesian, as in Tagalog, almost all expressions enjoy identical distributional privileges – just about anything can go anywhere.

Again, as with Tagalog in the previous section, the specifics of the above analysis may turn out to be controversial. (So far, there has been little discussion of Riau Indonesian in the linguistic literature: a preliminary analysis along the above lines, in Gil 1994c, is treated favorably in Kibrik 1997.) But the details of the analysis are not what this is about. Whatever the optimal treatment of Riau Indonesian may turn out to be, it can only be arrived at by unlearning Eurocentric grammatical traditions, and coming to grips with the structure of Riau Indonesian on its own terms.

At graduate school we are taught that, although languages often seem, at first blush, to be very different from each other, closer scrutiny will reveal these differences to be a superficial veneer just barely cloaking those deeper and more fundamental similarities. The goal of linguistic analysis, so we are told, is to demystify the apparently exotic features of different languages, and to seek out the commonalities which will form the basis for a general theory of language. The exciting and rewarding moments in linguistic analysis, so it is suggested, are those in which a strange and bewildering phenomenon in an exotic language is suddenly revealed, with the correct analytical tools, to be just like a well-known phenomenon in a well-studied language. Indeed, some of my fieldwork experiences have been of the above kind. However, my encounters with Riau Indonesian have led me in the exact opposite direction.

Before I started working on Malay/Indonesian, I had a preconceived notion of the language as being rather boring, with few of the interesting morphological and syntactic features characteristic of, say, related Philippine languages such as Tagalog. In fact, Malay/Indonesian seemed to me to bear a superficial resemblance to many of the well-known languages of Europe. The word order appeared similar, the amount of morphology was just right, and there didn't seem to be any of the complex morphological and syntactic patterns associated with Tagalog. However, it did not take long for me to realize that Malay/Indonesian was a wolf in sheep's clothing.

The more I worked on Riau Indonesian, the more exotic it became, the

more different from Standard Average European. Time after time I found myself puzzling over a construction, only to realize that the best analysis entailed dispensing with the traditional categories of Eurocentric grammar. Again and again I would ask “is such-and-such a this or a that?,” only to apprehend, sometimes after months or years, that I had been asking the wrong question, because the distinction between this supposed category and that simply wasn’t relevant to the grammar of Riau Indonesian. After a certain amount of time working on Riau Indonesian, I came to the conclusion that it does not have the familiar grammatical relations of subject and object. Considerably later, I realized that Riau Indonesian, like Tagalog, does not distinguish between major syntactic categories. It is only recently that I have begun to accept that even the basic notions of reference and predication may be foreign to the grammar of Riau Indonesian. At every stage, I found myself casting off the notions of traditional Eurocentric grammar, in order to gain a more perspicuous insight into the language.

As was the case previously, with Tagalog, many of my insights into Riau Indonesian derived from the use of naturalistic data, such as that in (16) – (20) above. But in addition, much of my understanding of the language came from being able, after a time, to speak it – not natively, of course, but well enough to pass as a native for a brief period of time, in a dark spot, or over the telephone. It is a common layman’s misconception that in order to be able to do research on a language, a linguist has to be able to speak it. But there is a grain of truth to the claim after all. Although most adults cannot acquire native competence in a new language, they can, with effort, achieve various degrees of near-native proficiency. After spending several years among speakers of Riau Indonesian, I found that I was able to introspect and come up with surprisingly subtle judgments with respect to grammaticality, semantic interpretation, and pragmatic appropriateness. Of course, it would be totally illegitimate to use such non-native intuitions as primary linguistic data. Nevertheless, such non-native intuitions may still serve a valuable function. Specifically, they may suggest various hypotheses which the linguist may then test against reliable sources of data. And in addition, they may provide the linguist with a kind of intuitive backdrop, or reality check, for hypotheses already formulated; in other words, a better “feel” for the language.

On a recent trip, I found myself on a boat, engaged in small talk with a fellow passenger, who, like many Indonesians, wanted to improve his English. Speaking to me in Indonesian, he asked how I would translate various sentences into English. I don’t remember what the exact sentences were, but one of them could easily have been ‘The chicken is eating’. My interlocutor was an inquisitive guy, and he soon noticed that the English

sentences I was giving him contained more words than the Indonesian sentences he was starting with. So what is the meaning of ‘the’?, he asked. But this was not the time and place for a lecture on definiteness. And what is the meaning of ‘is’?, he went on. Then I began to wonder: what *is* the meaning of ‘is’? Or rather, why on earth does English have a ‘the’, and an ‘is’, and all those other little words that Indonesian does so well without? As a linguist, I could provide all kinds of technical answers, but for an ordinary speaker of Indonesian, such answers would not be very satisfactory. Suddenly I realized that, after so many years of immersion in Indonesian, I had managed to unlearn my native language, and was now able to look at English through Indonesian eyes – non-native, admittedly, but still Indonesian. And seen through Indonesian eyes, English was a very exotic language indeed.

That was the moment when I really believed, deep down inside me, that my view of Riau Indonesian as sketched above might actually be right. And that was when I knew that I had finally turned the tables on Eurocentrism, and was able to deal with Riau Indonesian on its own terms.

5. Eurocentrism and language engineering

The preceding sections of this chapter were about Eurocentrism, its effects on the field of linguistics, and what the worker in the field can do to transcend it. But the effects are not limited just to linguistics: in some cases, Eurocentrism also has a profound influence on the actual languages that are the objects of the fieldworker’s attention.

No language exists in isolation. All languages are in constant contact with other languages, and over time, languages in contact begin to resemble one another, lexically and structurally. Most of these changes take place *naturally*, that is to say without any conscious efforts on the part of speakers to borrow from one language to another. Occasionally, however, linguistic changes occur *unnaturally*, through deliberate acts of language planning, engineering, and prescriptivism. And such situations often involve the imposition of European structures and categories on non-European languages.

A vivid example of this is provided by the history of Malay/Indonesian. When the Portuguese and the Dutch came to Southeast Asia, in the sixteenth and seventeenth centuries, they found varieties of Malay being used as a lingua franca over wide areas of the Indonesian archipelago. Recognizing the great value of such a common language, they set about to standardize it, to serve their own goals of proselytizing, trade, and colonial administration. As the colonial era drew to an end in the mid-twentieth century, the two major newly-independent countries, Malaysia and

Indonesia, resumed the process of standardization with renewed vigor, through the establishment of official language academies, the *Dewan Bahasa dan Pustaka* (Institute of Language and Literature) in Malaysia, and the *Pusat Pembinaan dan Pengembangan Bahasa* (Center for Language Development and Cultivation) in Indonesia. (A lively account of the history of Malay is provided in Collins 1996. Discussion of some of the issues involved in language planning in Southeast Asia can be found in Abdullah 1994 and Heryanto 1995.)

During the colonial era, Portuguese, Dutch, and then British prescriptivists often distorted the language in order to force it into a more familiar European mould. After Malaysia and Indonesia became independent, one might have expected this particular motivation for linguistic change to have become defunct; indeed, one of the goals of the language academies is the introduction of indigenous lexical items to replace foreign loan words. However, such linguistic purism is more or less limited to the lexicon – in the domain of grammar, the academies are busy making their language look more and more like English. In both Malaysia and Indonesia, there is a misguided belief that in order for a language to be able to fulfill the functions of a national language, it must have a well-developed system of grammar. Unfortunately, the only type of grammar that the language planners are usually familiar with is the Eurocentric grammar of European languages. Thus, Standard Malay/Indonesian has had a variety of linguistic features artificially grafted onto it that are reminiscent of European languages, including nominal number marking, verbal active and passive prefixes, and others.

In general, language engineering has more of an effect on the acrolectal, or formal registers, than it does on the basilectal, or colloquial varieties. Politicians, newscasters, schoolteachers, and other professional people are obliged to speak *properly* when engaged in their official capacities, but ordinary people in everyday circumstances couldn't care less what the prescriptivists think. In Malaysia and Indonesia, the situation is even more extreme than it is in many other countries: the standardized language and the colloquial varieties have drifted so far apart that they are of very low mutual intelligibility. Whereas Standard Malay and Indonesian are not much more different from each other than Standard British and American English, the colloquial varieties of Malay/Indonesian are as diverse as the different varieties of English, Dutch, and German. In Malaysia and Indonesia, educated people are basically diglossic, while uneducated people may understand very little of the standardized language, even if they are monolingual native speakers of some colloquial variety of Malay/Indonesian.

For a couple of years, I taught linguistics at the Universiti Kebangsaan Malaysia in Kuala Lumpur. My goal was to get the students to work on

their own basilectal varieties of Malay. Whenever I had a new group of students, I would begin the class by asking them to translate a sentence from English to Malay. Imagine you're just coming out of a sports stadium, I would say. You turn to your friend and exclaim: 'That match was great!' How would you say that in Malay? Invariably, the first answer the students would offer would be something like the following:

- (23) *Permainan itu sangat menarik.*
 NOM-play DEM:DIST very interesting
 That match was great.

All in all, the above sentence looks pretty much like its English source. The first word, *permainan* 'match', is a nominalized form of *main* 'play'; together, *permainan* and the following demonstrative *itu* form what appears to be a topic NP, which in turn seems to be followed by the comment *sangat menarik* 'very interesting'. So far so good; but after a little more discussion, I would then ask my students: So what would happen if you really said that coming out of a football match? People would think there is something wrong with you, is the answer I would get. The above sentence is in Standard Malay, and nobody really speaks like that. So how do people really speak, I would ask. Inevitably, the revised answer that I would then get would look something like this:

- (24) *Best la dia main tadi.*
 good CONTR 3 play PST:PROX
 That match was great.

And everybody would laugh, because it sounds so inappropriate to hear colloquial Malay spoken in a formal, university setting.

A brief inspection of (24) will reveal that it is totally different from (23) not just in choice of words but also in syntactic structure. To begin, whereas in (23) the comment *sangat menarik* follows the topic *permainan itu*, in (24) the comment *best la* precedes the topic *dia main tadi*. This alternative word order is one of the expressive devices available in the spoken language, in this case to add vividness to the utterance. Of greater interest, however, is the internal structure of the topic expression, *dia main tadi*. In Standard Malay, like in English, activities are prototypically expressed with VPs, while things are prototypically expressed with NPs. However, if you want to talk about an activity, such as 'play', and then, for example, to predicate something of it, such as being 'great', you have to convert the VP into an NP. This can be done in two ways, either by choice of a different lexical item, such as the English 'match', or by use of a special nominalized form, such as the English 'playing', or the Standard Malay *permainan*. In such cases, then, the prototypical association of semantic and syntactic categories is disrupted, with an activity being expressed, non-prototypically, by

means of an NP, ‘that match’ or *permainan itu*. It is precisely because of constructions such as these that descriptions of English and Standard Malay require recourse not only to semantic categories such as activity and thing, but also to syntactic categories such as VP and NP. But now let us look at the topic expression *dia main tadi* in (24). Although functioning as a topic, it shows no signs of having undergone any process of syntactic conversion, or nominalization; in fact, if it stood by itself, *Dia main tadi* would constitute a complete and well-formed sentence, meaning ‘They played’. In Kuala Lumpur Malay, as in Tagalog and Riau Indonesian, there seems to be no reason to distinguish between NPs and VPs, or for that matter between any other major syntactic categories. Almost any word, or larger expression, can go anywhere; in particular, a complete clause denoting an activity, such as *dia main tadi*, can find itself, unchanged, functioning as the topic of a bigger sentence, as is the case in (24).

The contrast between (23) and (24) thus highlights the effects of Eurocentrism on the standardized varieties of Malay/Indonesian. Although Kuala Lumpur Malay (24) contains an English loan word *best* (with a somewhat modified meaning ‘good’), the structure of the sentence is radically different from that of its English source. In contrast, the structure of Standard Malay (23) is actually quite similar to that of its English counterpart. This is no coincidence. Rather, it is a product of prescriptivism, and the conscious attempt to force Malay into the grammatical mold provided by European languages, and the theories constructed in order to account for them. Regrettably, the effect of such language engineering is to suppress the spirit of the language, and, by making it look more like English, to diminish the overall amount of linguistic diversity in the languages of the world. Fortunately, however, such prescriptivism generally fails to make significant inroads into the more basilectal varieties, which therefore remain the harborers of the language’s true genius.

6. No description without theory, no theory without description

For me personally, fieldwork is an endeavor of ongoing joy. I love working with people, and I revel in the data that they provide: the tone sandhi, the voice affixes, the sociolinguistic variation, and so forth. Yet at the same time, I delight in the search for the more highly abstract patterns and structures which lead towards a deeper understanding of the nature of human language.

For many linguists, however, these represent two distinct activities in irreconcilable opposition. Two buzzwords, *theory* and *description*, dominate the debate, as rallying points around which the combatants gather to cry out their slogans. In one camp are the self-professed theoreticians, who

declare that the only worthy activity is that of theory construction: for many of them, *description* is a condescending word hurled at those poor, uninteresting souls who have failed to see the true light. In the other camp are the self-styled descriptivists, who accuse the theoreticians of engaging in fruitless, frivolous activities, wasting time and taxpayers' money in arm-chair speculation while languages are dying all around us undescribed. In reality, however, both sides are equally misguided.

While the need to document endangered languages is undoubtedly the most urgent task facing linguistics today, it is an illusion to believe that one can conveniently separate description from theory, and – in the context of the documentation of endangered languages, at least – engage in the former without having to bother with the latter. Since the bare facts about any language are infinite in number, a finite description of the facts has no choice but to posit categories and formulate generalizations governing these categories, which *is* theory. Conversely, any theory that is empirically grounded accounts for a certain range of facts, while leaving others unaccounted for. Which facts get to be dealt with is at least in part a matter of taste and inclination, with respect to which practitioners may legitimately differ. However, for those facts that fall within the scope of the theory, the account in question *is* description. The truth of the matter is that there can be no description without theory, just as there can be no theory without description. (This point is argued forcefully in Dixon 1997.)

For example, many a linguistic description contains a statement to the effect that the language in question has subject-verb word order; typically such a claim is backed up by examples of basic sentences such as 'chicken eat', 'boy run', and so on. However, as suggested in the preceding sections, even such commonplace categories as subject and verb are theoretical constructs, which may or may not be the ones most appropriate for the data under consideration. Indeed, the word order of basic intransitive sentences is potentially amenable to a variety of alternative accounts, making reference to different kinds of categories: for example, NP precedes VP, actor precedes verb, topic precedes verb, participant precedes monovalent activity, less complex constituent precedes more complex constituent, shorter constituent precedes longer constituent, and so forth. And this is anything but terminological hair-splitting. Rather, such alternative accounts bear empirical consequences: each one makes different predictions with regard to the word order of various other sentences. And if the language under consideration is endangered, one can only hope that these predications will be tested before the language is extinct.

The choice that faces the fieldworker is not between description and theory, but rather between two different modes of descriptive/theoretical activity. At one extreme is an approach which may be characterized as top-

down, or *templatic*. This is a method that has been productively institutionalized by the annual field expeditions organized by Moscow State University, as described in Kibrik (1988). Well-known exemplars include the grammatical descriptions produced in the UK and the US by the Croom Helm Descriptive Grammars and Routledge Descriptive Grammars, and in the former USSR by the many publications of Izdatel'stvo Nauka, such as those in the series titled *Jazyki Narodov SSSR* and *Jazyki Narodov Azii i Afriki*. In this system, the author produces a linguistic description in accordance with a pre-prepared and standardized checklist, which spells out the topics to be covered and the order in which they are to be dealt with. This way of doing things is intended to guarantee a relatively complete coverage of the major features of the target language, and to ensure that descriptions of different languages by different scholars be readily comparable. Such descriptions are generally extremely user-friendly. For example, when I was working on my Ph.D. dissertation, a typological study of distributive numerals (Gil 1982), I knew that I could pick up any grammar book by Nauk, zip through the table of contents to a sub-sub-section titled *čislitel'nye* ('numerals'), and home straight in on a brief, usually one-paragraph description of *razdelitel'nye čislitel'nye* ('distributive numerals') in the language in question. However, such convenience comes at a price. Although guaranteed to be free of arcane theoretical terminology, a description of this kind is anything but atheoretical. On the contrary, by its very nature, a templatic description involves the imposition of a universal scheme upon a particular language – and such a scheme necessarily invokes a host of theoretical assumptions concerning the relevant units of linguistic description. And of course, such a universal template is inevitably rooted in Eurocentric grammatical traditions, and may thus be ill-equipped to handle the diversity exhibited by languages spoken in other parts of the world.

Escaping Eurocentrism leads towards an approach that is diametrically opposed to the templatic – one which might appropriately be characterized as bottom-up, or *free-wheeling*. In accordance with this system, the data themselves are taken as the starting point, and the description of the data is what then provides the motivation for the postulation of appropriate categories and structures. Doing linguistics this way involves turning an attentive ear to the language under investigation and listening to what it is trying to say, even if this entails unlearning various aspects of one's native language, and of one's linguistic education. The bottom-up approach frees the fieldworker from having to handle a particular theoretical framework, and squeeze-fit the language into a set of predetermined and possibly irrelevant grammatical categories. However, it presents a greater challenge: that of organizing the data from scratch, identifying the interesting regularities,

and accounting for these regularities by means of various theoretical constructs. Perhaps the most well-known proponents of such an approach were the American Structuralists, as represented by Bloomfield (1917, 1933).

Ultimately, however, the construction of a true bottom-up linguistic description is a chimera. The fieldworker can switch a tape recorder on and point it at a speaker, but in order to do anything with the data, even to transcribe it, there is no alternative but to invoke at least some a priori categories: segments, syllables, words, utterances. Subsequent more detailed analysis may reveal such categories to be inappropriate, but you have to start with something. In practice, then, working with languages involves progressing simultaneously in both directions, top-down and bottom-up, with each of these approaches informing the other at all times. In this chapter, I have attempted to swing the pendulum, as it were, away from what seems to me to be an excessive top-down orientation with its concomitant Eurocentrism, and towards a more bottom-up mode of analysis. However, in any given situation, it is up to the fieldworker to find the right balance between these two idealized approaches to the study of language.

Fieldwork is thus an ongoing dialogue of opposites: the deductive and the inductive; the abstract and the concrete; the general and the particular. Like, for example, when some ideas about ontological categories are confronted with an exclamation uttered by a television football spectator on an Indonesian island, as in (19). For me, perhaps the greatest satisfaction in fieldwork comes from tying all of these opposites together to form a single holistic activity.

ACKNOWLEDGMENTS

Since this chapter reflects two decades of academic activity, it is impossible to acknowledge the many people who have contributed, in one way or another, to its coming into being: I can only mention a few who come to mind now. Ed Keenan was the one who introduced me to the study of linguistic diversity. Sharon Armon-Lotem was a sounding board for my ideas about syntactic categories. Peter Cole, Gabriella Hermon and Uri Tadmor have been around for the last few years, for ongoing discussions about doing Malay/Indonesian linguistics. Michael Israel helped me to a better understanding of the issues relating to polysemy and ambiguity. Andrew Carstairs-McCarthy stepped in at the right moment with some insightful e-mail messages about the non-indispensability of reference and predication. The Hokkien judgments cited in section 2 were provided by Clara Lee Pei San and Peh Cheng Hwee. The Riau Indonesian data cited in section 4 were spontaneously uttered by Ahmadsayuti and Arip. Finally, it was Kairil who introduced me to Riau Indonesian, and gave me the Indonesian

eyes with which to view English. My years of fieldwork described in this chapter were made possible by the institutional support of Tel Aviv University, Haifa University, National University of Singapore, Universiti Kebangsaan Malaysia, and the Max Planck Institute for Evolutionary Anthropology. In addition, the work on Malay/Indonesian described herein was supported, in part by National Science Foundation, Grant numbers SBR-9121167, SBR-9729519, and INT-9423291.

NOTE ON TRANSCRIPTION

1 In the Hokkien examples, the superscript numerals, as in e^{24} , mark tones, using a scale from 1 to 5 where 1 is the lowest tone and 5 is the highest. In the interlinear glosses, the following abbreviations are used: ACT 'actor'; ASSOC 'associative'; CONJ.OP 'conjunctive operator'; CONTR 'contrastive'; DEM 'demonstrative'; DIST 'distal'; INV 'inversion'; NEG 'negative'; NOM 'nominalizer'; PAT 'patient'; PERS 'personal'; PROG 'progressive'; PROX 'proximal'; PST 'past'; REAL 'realis'; SG 'singular'; TOP 'topic'; 1 'first person'; 3 'third person'.

REFERENCES

- Abdullah Hasan (ed.). 1994. *Language Planning in Southeast Asia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Ansald, Umberto. 1999. Comparative Constructions in Sinitic, Areal Typology and Patterns of Grammaticalization. Ph.D. dissertation, Universitet Stockholms.
- Aristar, Anthony R. 1991. On diachronic sources and synchronic pattern: an investigation into the origin of linguistic universals. *Language* 67:1–33.
- Becker, Alton L. 1995. *Beyond Translation: Essays Towards a Modern Philology*. Ann Arbor: University of Michigan Press.
- Bisang, Walter. 1996. Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in east and mainland Southeast Asian languages. *Studies in Language* 20:519–97.
- Blake, Frank R. 1925. *A Grammar of the Tagalog Language*. New Haven: American Oriental Society.
- Bloomfield, Leonard. 1917. *Tagalog Texts with Grammatical Analysis*. (University of Illinois Studies in Language and Literature, vol. 3, no. 24) Chicago: University of Illinois.
1933. *Language*. New York: Holt, Rinehart and Winston.
- Bodman, Nicholas C. 1955. *Spoken Amoy Hokkien*. Kuala Lumpur: Government Federation of Malaya.
- Carrier-Duncan, Jill. 1985. Linking of thematic roles in derivational word formation. *Linguistic Inquiry* 16:1–34.
- Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4:109–49.
- Cheng, Robert. 1968. Tone sandhi in Taiwanese. *Linguistics* 41:19–42.
1973. Some notes on tone sandhi in Taiwanese. *Linguistics* 100:5–25.
- Collins, James T. 1996. *Malay: World Language of the Ages*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

- Comrie, Bernard. 1996. The unity of noun-modifying clauses in Asian languages. In *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Language and Linguistics*, vol. 4, pp. 1077–88. Salaya, Thailand: Mahidol University, Institute of Language and Culture for Rural Development.
1998. Rethinking the typology of relative clauses. *Language Design* 1:59–86.
- Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Ernst, Thomas. 1994. Functional categories and the Chinese INFL. *Linguistics* 32:191–212.
- Gil, David. 1982. Distributive Numerals. Ph.D. dissertation, University of California at Los Angeles.
- 1993a. Syntactic categories in Tagalog. In *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics (1992)*, ed. S. Luksaneeyanawin, vol. 3, pp. 1136–50. Bangkok: Chulalongkorn University.
- 1993b. Tagalog semantics. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 19:390–403.
- 1994a. Conjunctive operators in South-Asian languages. In *Papers from the Fifteenth South Asian Language Analysis Roundtable Conference*, ed. Alice Davison and F. M. Smith, pp. 82–105. Iowa City: South Asian Studies Program, University of Iowa.
- 1994b. Conjunctive operators: a unified semantic analysis. In *Focus and Natural Language Processing*, vol. 2: *Semantics. Proceedings of a Conference in Celebration of the 10th Anniversary of the Journal of Semantics*, ed. P. Bosch and R. van der Sandt, pp. 311–22. (Working Papers of the Institute for Logic and Linguistics, 7)
- 1994c. The structure of Riau Indonesian. *Nordic Journal of Linguistics* 17:179–200.
- 1995a. Conjunctive operators: a cross-linguistic study. In *IATL: Proceedings of the Tenth Annual Conference of The Israel Association for Theoretical Linguistics*, ed. Alice Davison and F. M. Smith, pp. 72–90.
- 1995b. Parts of speech in Tagalog. In *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*, ed. Mark Alves, pp. 67–90. Tempe: Arizona State University.
1998. Patterns of macrofunctionality in Singlish noun phrases: a questionnaire survey. In *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society*, ed. Shobhana L. Chelliah and Willem J. de Reuse, pp. 147–82. Tempe: Arizona State University.
1999. Riau Indonesian as a pivotless language. In *Tipologija i Teorija Jazyka, Ot Opisanija k Objasneniju, K 60-Letiju Aleksandra Evgen'evicha Kibrika (Typology and Linguistic Theory, From Description to Explanation, For the 60th Birthday of Aleksandr E. Kibrik)*, ed. E. V. Raxilina and J. G. Testelec, pp. 187–211. Moscow: Jazyki Russkoj Kul'tury.
2000. Syntactic categories, cross-linguistic variation and universal grammar. In *Anthology of Word Classes: Empirical Approaches to Language Typology*, ed. Bernard Comrie and Petra Vogel, pp. 173–216. Berlin: Mouton de Gruyter.
- In press. English goes Asian: number and (in)definiteness in the Singlish noun-phrase. In *Noun-Phrases in European Languages: Empirical Approaches to Language Typology*, ed. F. Plank. Berlin: Mouton de Gruyter.

- Gordon, Lynn. 1986. *Maricopa Morphology and Syntax*. (University of California Publications in Linguistics, 108) Berkeley and Los Angeles: University of California Press.
- Heryanto, Ariel. 1995. *Language of Development and Development of Language: The Case of Indonesia*. (Pacific Linguistics, Series D-86) Canberra: Australian National University.
- Himmelman, Nikolaus. 1991. *The Philippine Challenge to Universal Grammar*. (Arbeitspapier 15, Neue Folge) Cologne: Institut für Sprachwissenschaft, Universität zu Köln.
- Huang, James C.-T. 1991. Modularity and Chinese A-not-A questions. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, ed. C. Georgopolous and R. Ishihara, pp. 305–32. Dordrecht: Kluwer.
- Huang, Yan. 1994. *The Syntax and Pragmatics of Anaphora: A Study with Special Reference to Chinese*. Cambridge: Cambridge University Press.
- Kibrik, Aleksandr E. 1988. Čto takoe ‘lingvističeskie ekspedicii’? *Vestnik* 12:94–102.
1997. Beyond subject and object: toward a comprehensive relational typology. *Linguistic Typology* 1:279–346.
- Kroeger, Paul. 1993. *Phrase Structure and Grammatical Relations in Tagalog*. Stanford: CSLI Publications.
- Lehman, F. K. 1998. On the use of dah in Lai Chin questions and the operator syntax of functors. In *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society*, ed. Shobhana L. Chelliah and Willem J. de Reuse, pp. 211–31. Tempe: Arizona State University.
- Matthews, Stephen J. In press. Y. R. Chao and universal Chinese grammar. In *History of Linguistics 1966*, ed. D. Cram, A. Linn, and E. Nowak. Amsterdam: John Benjamins.
- Peng, Shu-hui. 1994. Effects of prosodic position and tonal context on Taiwanese tones. In *Ohio State University Working Papers in Linguistics*, 44, pp. 166–90.
1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25:371–400.
- Ramli Haji Salleh. 1995. *Sintaksis Bahasa Melayu, Penerapan Teori Kuasaan dan Tambatan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Ratliff, Martha. 1991. Cov: the underspecified noun, and syntactic flexibility in Hmong. *Journal of the American Oriental Society* 111:694–703.
- Riddle, Elizabeth, and Herbert Stahlke. 1992. Linguistic typology and Sinospheric languages. In *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*, ed. Martha Ratliff and Eric Schiller, pp. 351–66. Tempe: Arizona State University.
1995. Parts of Speech in Tagalog. In *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*, ed. Mark Alves, pp. 67–90. Tempe: Arizona State University.
- Schachter, Paul. 1976. The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In *Subject and Topic*, ed. Charles N. Li, pp. 491–518. New York: Academic Press.
1977. Reference-related and role-related properties of subjects. In *Syntax and Semantics 8: Grammatical Relations*, ed. Peter Cole and Jerrold M. Sadock, pp. 279–306. New York: Academic Press.
- Schachter, Paul, and Fe T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley and Los Angeles: University of California Press.

- Shkarban, Lina I. 1992. Syntactic aspect of part-of-speech typology. In *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics (1992)*, ed. S. Luksaneeyanawin, vol. 1, pp. 261–75. Bangkok: Chulalongkorn University.
1995. *Grammatičeski Stroj Tagal'skogo Jazyka*. Moscow: Vostočnaja Literatura.
- Tobin, Yishai. 1990. *Semiotics and Linguistics*. London: Longman.
- Welmers, William E. 1973. *African Language Structures*. Berkeley and Los Angeles: University of California Press.
- Zwicky, Arnold, and Jerrold M. Sadock. 1975. Ambiguity tests and how to fail them. In *Syntax and Semantics 4*, ed. John. P. Kimball, pp. 1–36. New York: Academic Press.