

Generative Approaches to Syntactic Typology

George Gibbard

Table of Contents

Sec.	Heading	Page
0	Introduction	1
1	Introduction to Theoretical Issues in Syntactic Typology	4
1.1	Basic Word Order Typology	6
1.2	Subject, Object, Accusativity, Ergativity	10
1.3	Review	20
2	Syntactic Typology and Western Malayo-Polynesian Languages	23
3	Is Tagalog ergative? If not, what is it?	26
3.1	Case	27
3.2	Voice	28
3.2.1	Active	29
3.2.2	Transitive Passive	30
3.2.3	Intransitive Passive	33
3.3	Comparison of Tagalog and Acehnese	36
3.4	Review	37
3.5	Voicelessness	39
3.5.1	Existentials	40
3.5.2	"Recent past"	45
3.5.3	The meaning of "voicelessness"	47
3.6	Two Levels of Subject Assignment	50
3.7	Syntactic Status of Two Subjects	53
4	Syntactic Typology Revisited	59
5	Comparison between English and Tagalog	63
	Bibliography	70

Generative Approaches to Syntactic Typology

0. Introduction¹

This thesis aims to be a limited exploration of the field of syntactic typology, examining critically two different proposals that have been made for classifying languages in different subfields of syntax. It is my contention that neither of these widely accepted classifications is really workable, as both attempt to treat problems in one area of syntax independently of that area's interactions with other syntactic rules. I find that the treatment of any given area of syntax must generally be dependent on assumptions about the proper theoretical description of other areas of syntax, and thus that any typological theory with such limited scope is bound to fail.

The study of linguistic typology in general has two distinct but complementary goals. The first of these is to aid in the search for linguistic universals, which must be regarded as the major overarching project of 20th century scientific linguistics. The goal is to be able to classify and explain the diversity of human language in order to show that this diversity actually reflects minor variation in the application of an innate "universal grammar" or "language instinct." A fairly standard attitude is that cross-linguistic variation in the rules of grammar represents the application of different "parameters" to a universal base (the exact form of these "parameters," however, is uncertain, and by this view the most important thing linguistics must try to determine). More recently, optimality theory has argued that grammatical variation is only the result of different rankings of importance assigned to cross-linguistically identical parameters. If either of these is true, linguistic typology would represent a way to account for and classify the different parameters and combinations of parameters that are attested in human language. In fact Bernard Comrie (1989) has written that the study of linguistic universals and of linguistic typology are "just different facets of a single research endeavor." Linguistic typology aims to further search for linguistic

¹ I would like to thank Prof. Fernald whose feedback has been extremely helpful to developing the ideas in this thesis. I am also indebted to Faith Fraser and Kathryn Manz who provided valuable comments on early stages of this work.

universals either by surveying the diversity of human language as broadly as possible in order to be able to say in what areas human languages do not show diversity, or by examining this diversity closely in order to establish "implicational" universals, that is to establish meaningful correlations between the occurrence of specific typological features in different areas of a language's grammar.

The second goal of linguistic typology has been in a sense the opposite of the above: it is the search for meaningful criteria that can be used to quickly and compactly describe features of a language's structure by locating that language within a specified range of diversity exhibited by human language. This is the philosophy behind describing a language as, say, "SVO" or as "ergative." While such descriptions obviously only address a small area of a language's grammar, each is derived from a theory that asserts that each of these is, within a known range of diversity in this aspect of grammar, a meaningful type, i.e. that "SVO" and "ergative" each refers to a single phenomenon that works similarly in all languages matching the typological description. In other words, this kind of use of typological labels is entirely dependent on the correctness of the universal (even if only in one area of the grammar) typological scheme it relies on.

Conversely, any universal typological scheme is dependent on the correct analysis of the languages it classifies. Thus the two sides of linguistic typology are dependent on each other, and no statement either about universal schemes of typological classification, or about the typological characteristics of a given language, can be regarded as definitive while doubts remain in the other area. How, then, can progress be made? Clearly we must proceed tentatively, exploring in detail the complex interactions of different areas of the grammar of each language, and comparing our results to those of similar detailed examinations of other languages. In the meantime, existing typological descriptors, such as "SVO" or "ergative," may be used as a kind of a "shorthand" to introduce features of a language's grammar, as long as their use is recognized as being only this, and it is accepted that they may have no very deep theoretical underpinnings.

"SVO" and "ergative" are descriptors from two different widely accepted typological classifications, each treating a subfield of the linguistic subfield of syntax. Yet each of these schemes—basic word order typology following Greenberg (1966), and the

ergativity-accusativity typology first systematically proposed (I believe) by Edward Sapir—has been accepted rather uncritically, in isolation from their place in a wider theory of syntax or of syntactico-semantic interactions. My thesis is that there are problems in treating either Greenbergian word order typologies or ergativity v. accusativity as meaningful typologies. Each of these descriptive typologies rests on assumptions about other areas of the grammar: in order to claim either "SVO" or "ergative" about a language, one must first have resolved all problems of syntactic structure that relate to what syntactic position corresponds to the English "subject." If one actually attempts to resolve these issues, one often encounters problems of types linguists do not normally expect. And finally, if one accepts a theory that successfully makes the "subject" position unproblematic, such a theory renders typologies like "SVO" and "ergative" fairly meaningless, each of these being revealed as just one possible way to slice a really much wider diversity in the respective areas of the grammar these typologies address. As I will make clear in conclusion, what one is left with is a need to analyze in greater detail the differences between complex procedures different languages use to translate from thematic and discourse roles to syntactic positions, avoiding the obscuring simplifications that earlier attempts at syntactic typology necessitate.

1. Introduction to Theoretical Issues in Syntactic Typology

This section aims to provide definitions of various technical terms used in this paper and to emphasize the diversity of coding strategies whereby linguistic structure gives clues to underlying grammatical structure. The experienced reader may want to move on to the next section. There are many different ways surface structure can make clear the grammatical relations and discourse roles of the different participants in a given event. 'Grammatical relations' refers to such roles in traditional European grammar as 'subject,' 'direct object,' 'indirect object.' Such terms clearly often express something about the semantic relationship of a noun phrases (NP) to the verb of the sentence. On the other hand it may be hard define precisely what. While a primitive

linguistic theory may use only one level of terminology to describe grammatical relations and related phenomena, linguists find they need several sets of terms to describe separate levels of relationships the NPs in a sentence bear to one another, the correspondences between these levels being sometimes difficult to trace.

Most broadly, linguists need to recognize three different structural dimensions in which any given utterance exists: these are semantic structure, discourse structure, and finally linguistic structure proper. "Semantic" here is understood as referring only to truth-value semantics; on this level, NPs take one and only one "thematic role" or "theta-role," which is a semi-formalized expression of what kind of role the NP plays logically in the event described by the sentence. An NP having some theta-role in reference to another word is called an argument of that word, and . Theta-roles may be formally written as small capitals, examples being AGENT, EXPERIENCER, INSTRUMENT, RECIPIENT and so forth. This semantic level needs to be distinguished from that of discourse structure: this is the level at which we gauge how central or important to a speaker's conception a given NP is, whether it is well known, somewhat unfamiliar, or an entirely new departure in the discourse at hand. This is the level which is responsible for giving us the distinction in English between definiteness and indefiniteness; however we need to bear in mind that discourse structure is a more complex underlying stratum that these linguistic structures only serve to crudely translate.

Taken together, semantic and discourse structures constitute the pre-linguistic structure of meaning that a speaker intends in generating an utterance. The structure of language itself, then, is a formal mechanism for encoding these two structures, balancing their demands or privileging one or the other to some extent. Linguistic structure relating different lexical items is conventionally thought of as operating on at least two levels, those of morphology and syntax, syntax being defined as the structure relating words and morphology being the internal structure of words; what a word is I can't really say. Syntax, too, is considered to have nested levels of structure, just as morphological structures are nested within syntactic structures; thus a sentence may be thought of as having a structure in some ways paralleling the internal structures of its constituents, such as noun phrase, verb phrases, etc. Finally, word order may or may not be independent of syntax in some cases, while in others it is clearly strictly regulated by syntax.

In all these levels of linguistic structure there are several types of ways in which the semantic and discourse roles of arguments can be encoded. Firstly, at the morphological level, both NPs and verbs may be affixed to provide clues about grammatical relations. A 'case' is a marking of a noun to clarify its role in syntactic structure; English shows case only in a few pronouns, such as 'he' versus 'him.' Marking of a verb which clarifies its argument structure is often called 'voice,' the distinction of active versus passive verbs in English being an example. Syntax also encodes grammatical relations in a variety of ways. In English the most important part of how we absorb the underlying syntactic structure of a sentence lies in word order, but other clues to syntactic structure insofar as it is not revealed by morphological markings may be sought in the 'transformations' that can reorder and combine syntactic groupings. Word order in some languages can be much freer than in English; particularly when rich morphological markings are enough to make syntactic structure readily apparent, word order may serve primarily to encode discourse structure instead of thematic roles. English shows discourse structure partly through such markers as articles, partly through word order, but largely through such marginally formal features of language as stress, timing, and intonation. Such 'prosodic' features are very common markers of discourse structure, but also may show things about formal syntactic structure.

It is clear then that syntactic typology will have a very hard time in classifying all the combinations of these features that a language uses to encode grammatical relations. Fortunately, the field is predicated on the notion that the combinations are not arbitrary, and that common patterns will continually reemerge if enough languages are examined in enough detail. So far, linguists have generally tried to work as much as possible from formal features of linguistic structure, seeking to moderate their task by looking for common trends and types in specific subareas of formal grammar. The following are two proposals that have been made for schemes of syntactic typology.

1.1 Word Order Typology

Following Greenberg (1966), it is popular to describe languages in terms of dominant word order types. This is in certain cases considered a valuable index that will correlate highly with other ordering phenomena—specifically, if a language is

considered underlyingly V-initial this correlates highly with Head-before-modifier placement in other areas of the syntax, while V-final (especially "SOV") languages tend to have a preponderance of Modifier-before-head ordering. However, while languages show a greater-than-chance statistical tendency to fall into a few Greenbergian typologies, they also very commonly violate or mix these. And while basic underlying initial or final position of the verb is a strong predictor of other ordering phenomena, the very common SVO order is not, as it seems to co-occur with any number of mixtures of canonically "V-initial" and "V-final" features.

According to modern linguistic theory, Greenbergian statements of word order facts such as "SVO" and "SOV" do not have any status in determining syntactic structures. Rather they are side effects of more complex systems of phrase structure and other syntactic rules. The limited predictive power of Greenbergian typology may be considered a natural outcome of the principle that a syntax is simpler if it can use the same principles recursively over various areas of the grammar, which is captured by the cross-categorial generalizations of X-bar theory. A V-final language will likely have modifier-before-head ordering in the noun phrase, as well as having postpositions rather than prepositions, for the reason that adpositions are the head of their phrase and the verb is likewise the head of the verb phrase and of the entire sentence. V-final ordering is the outcome, not a cause, of modifier-before-head tendencies in phrase structure rules.

If this principle were the only one governing word order, we would expect only two possible orderings for human languages, which are the canonical V-initial and V-final types, but this is of course not the case. Crucially, it is now recognized that "dominant word order" at the sentence level is not necessarily the same as "deep structure" ordering, but may be derived from it by "raising" and other transformational rules. The ordering of constituents smaller than the sentence will correlate much more highly with the ordering of "deep structure" produced by phrase structure rules prior to transformational rules. "Raising" is especially used to account for SVO ordering, whence the lack of correlation between SVO ordering and other ordering phenomena: surface SVO often does not reflect deep structure ordering.

As an example, German may be considered at surface dominantly SVO or V2 but underlyingly V-final. This is made likely by the appearance of recessive V-final order in

a number of different sentence types: relative clauses are V-final, and while a main verb without auxiliaries occurs in second position, if there is an auxiliary the auxiliary takes the second position and the main verb is final. If there is an auxiliary in a relative clause, the auxiliary is last and the main verb next-to-last. It is possible to account for these facts by supposing that one verbal element is raised in non-relative clauses from its deep-structure final head position to a higher position in the syntactic tree, resulting in a position preceding the main part of the sentence. The surface V2 ordering should be analyzed as resulting from the raising of one argument of the sentence, the choice of which depends on topicality, to a still higher position preceding the verb. The occurrence of V-initial ordering in yes-or-no questions may be considered a result of the application of the first raising but not of the second (we might presume that in these questions there is a null question complementizer filling the first position).

If we accept the preceding, it is clear that we should not expect to be able to look to the "dominant" SVO/V2 ordering of German to predict ordering in other constituents. Indeed, as mentioned earlier, SVO ordering is not considered a predictor of other ordering principles. Rather, if typological regularities in ordering (insofar as they exist) are the result of recursive application of phrase structure principles, we should look to the "recessive" but apparently DS ordering SOV and expect to find modifier-before-head ordering to be dominant. In fact this is true for the most part (Hawkins 1983): While German has a few postpositions, prepositions are clearly the norm, and in the NP, head nouns are preceded not followed by modifying adjectives, possessives, determiners and numerals. The exceptions to dominant modifier-before-head ordering are genitives and relative clauses modifying nouns. Both modifier-before-head and head-before-modifier orders are possible for both of these, but postnominal genitives are "more frequent and grammatically more productive," and pre-nominal relatives are "nonbasic" and restricted to modifying underlying subjects (pre-nominal relatives take "participle" form, modifying like adjectives but retaining verbal argument structure) (Hawkins 1983, p.227, 229, 249, 333). As predicted, modifier-before-head ordering seems to be very strong in German; the exceptions are probably due to the tendency for genitives and relatives to be much longer than other modifiers and the tendency (usually subordinate to other principles) for languages to place longer

constituents after shorter ones, thereby making syntactic structure perceptible earlier in the utterance(Siewierska 1988, p.29-46).

As the exceptions found in the above discussion should indicate, the predicting power of dominant modifier ordering with respect to heads should not be overestimated. Natural languages do not necessarily adopt the simplest possible syntax. Different syntactic levels and categories also have their own constraints to satisfy and must often be treated by different syntactic laws: A verb phrase and a noun phrase are built of different kinds and numbers of elements and may be governed by different ordering principles; The GB category "Spec" has been used to lump together a number of different kinds of "special" grammatical modifiers in various kinds of phrases, but these have very different functions; An adposition takes a single argument, but verbs must be able to take many kinds of arguments and treat different kinds of arguments differently.

Furthermore the concept of "word order typology" as applied to any given language is only as strong as that of "basic word order," which is certainly does not have the same status in all languages. While in languages such as English or Mandarin a listener gets most basic information on syntactic and semantic rules from word order, in a language such as Russian or Latin equivalent information is mostly carried by inflection on the constituents, while basic constituent order is highly variable and depends largely on "pragmatic" or "discourse structure" phenomena. Apparently most if not all languages—including the "extreme" cases already cited—actually determine order by a mixture of syntax and pragmatics; and even in the most strictly syntactically ordering languages, the less syntactically essential an element of the sentence, the greater its freedom with respect to ordering. All this said, Greenbergian typology does describe (though often crudely and inadequately) real meaningful differences in syntactic structure between languages, even though "basic word order" is of cross-linguistically varying significance for surface structure.

There are also further obstacles to describing languages in these terms. One issue is the definition of "Subject" and "Object." It is controversial whether these terms have universal validity, or possibly validity for all languages meeting certain typological criteria. If valid, it is not clear that they refer to constituents at the same level of derivation. It is not clear how this typology is supposed to interact with the distinction

between basic and (?transformationally) derived sentence types, which will be discussed later, or with issues of "accusative"/"active"/"ergative" typology, which I propose to discuss now. In section 1.3, once these issues have been raised, I will show how they too interfere with attempts to reduce syntactic structure to the "dominant word order types" Greenberg advocates.

1.2 Subject, Object, Agent, Accusativity and Ergativity

The Greenbergian word-order types are stated in terms of "Subject" and "Object" which have clear definitions for the majority of languages, including English and most European languages. But some languages clearly should not be described in such terms, or at least not without certain modifications. In English "Subject," the most basic syntactic position for a nominal argument in relation to a verb and the position almost always accorded to the lone argument of a verb, is in prototypical transitive sentence given to the agent argument. The unity of the English category "Subject" is clear from word order, case marking and verbal agreement; based on the algorithm used to assign arguments subject or object status, English is considered an "accusative type" language. Some languages do not have an entirely corresponding category. These are languages assigned to a type, opposite to "accusative," known as "ergative." Traditionally ergativity has been most basically defined in terms of morphological case marking (Comrie 1989 p. 111): the morphological category that is the potential candidate for the label "subject" in an ergative language corresponds to the the English intransitive subject or to the English transitive object. A nominal case that indicates this role may be referred to as "absolutive" as an alternative to "nominative," while an "ergative" case means one reserved for the transitive agent (English transitive subject). However, it is not too early to point out that the morphological definition of ergativity is not the only one that has been proposed. In some languages, linguists have analyzed syntax as necessitating a definition of "subject" that does not correspond exactly to any of the categories defined by morphological case; hence it is sometimes necessary to distinguish "morphological ergativity" from "syntactic ergativity." Indeed, Dixon (1988) asserts that "ergativity may be recognized at three distinct levels: morphological,

syntactic, and discourse ergativity," the last of these being a controversial label the discussion of which lies outside the scope of this thesis.

While the concept of ergativity is necessary for the proper description of many of the world's languages, it does not follow that (as the preceding paragraph might suggest) ergative and accusative systems are equally common and basic in human languages. Linguists rightly or wrongly see many languages as showing only accusative type behavior; few if any languages exhibit a complete lack of accusative-type features. Many languages have been analyzed as having a morphosyntactically ergative surface structure but a more basic accusative deep structure. Since the definition of ergativity already given refers to morphosyntactic surface structure I should perhaps clarify how "accusative deep structure" is identified. The most common reasons for this analysis involve syntactic processes such as anaphora, reflexivization, and "equi-subject deletion" in subordinate clauses, which are often linked to the agent role even in otherwise ergative languages. Both "transformational" and "relational" grammar models have dealt with such situations by postulating an initial level in which the agent is "subject," so that these transformations can occur prior in the derivation to an obligatory transformation yielding ergative surface structure. Languages for which an ergative deep structure has been postulated are comparatively much rarer. Even Dyirbal, an Australian language that is the most typically cited example in the literature, has accusative case marking for personal pronouns, so even this case may be somewhat debatable.

If ergativity is a phenomenon produced by transformational processes, its role in syntax may be seen as more parallel to English passive than to English active. In the passive (like an ergative construction) THEME, not AGENT, take the essential external argument (subject) position, while an agent if present must be specifically marked for this role. There do remain differences, though, in that "passive" refers to a marked, less dominant surface structure, whereas "ergative" refers to a dominant or obligatory surface structure; if transformationally produced, that transformation is obligatory, either throughout the language or in a given domain. In a canonically "fully ergative" language a surface structure morphosyntactically comparable to English passive would be the only available strategy for forming transitive verbal sentences. For many languages which have a combination of active and ergative features, the set of

describable events may be divided into two "domains," within one of which only an ergative surface structure is permitted. Thus the "tense/aspect-conditioned accusative-ergative split" reported for many languages around the world, including many Austronesian and Indo-Iranian languages: in these languages the past tense or perfective aspect is an entirely ergative domain, while accusativity occurs elsewhere. There are, however, many different ways that ergativity and accusativity can co-occur.

Given "split ergativity," one might try to find criteria which will allow us to describe a given language as "more ergative" or "less ergative" than another. M. H. Klaiman's (1988) cross-linguistic study of split ergativity in North India gathers data which, despite the author's intentions, indicate that this is misguided. She attempts to establish criteria for ranking North Indian language according to degree of ergativity, but must select from conflicting possibilities. Although Klaiman does not say this outright, the most significant generalization to emerge from the vast amount of data she synthesizes is that among the languages surveyed, the more "ergative" a language with respect to nominal case marking, the more "accusative" it is with respect to verbal agreement, and vice versa. Only the languages that fall in the middle can be said to have prototypically ergative structures (simultaneously ergative in both case marking and verbal agreement); but then these are the languages with the clearest tense/aspect conditioned split, i.e. they also have fully accusative systems in non-perfective aspect.

Klaiman finds that a few southeastern Indo-Aryan languages exhibit no ergative features at all; these are the Dravidian and Munda languages of South and Central India, and certain neighboring eastern Indo-Aryan languages. In all the other surveyed languages, of various genetic affiliation (Indo-Aryan, Dardic, Iranian, Tibeto-Burman, Nuristani, Burushaski) Klaiman sees various mixtures of accusative and ergative features. For example, Dardic Shina and Indo-Aryan Nepali, Assamese and Bengali turn out to combine ergative case marking in the past tense with entirely accusative verbal morphology. Thus in the Nepali past tense one says, as it were, "By-me I-did the work" (Matthews 1984):

- 1) mai-le tyo ka:m gar-e~
 1sg erg/ins that work do- 1sg
 "I did that"

where the marking on the pronoun is not "nominative," as the verbal agreement might suggest, but ambiguously ergative or instrumental (an ambiguity that is resolved by the verbal agreement). Ergative case marking for agents is not found outside the past tense: compare the habitual present tense showing the real unmarked nominative case:

- 2) ma dinahu~ ka:m gar-ch- u
 1sg every day work do- hab.-1sg
 "I work every day"

Here there is no ergative case marking, while verbal agreement remains with the agent argument, as in 1). Neither is ergative case marking associated particularly with the agreement markers of the past tense (which the reader may have noticed are not identical to those of the present); in an intransitive past tense sentence, ergative marking is not found:

- (3) ma Nepa:l ga-e~
 1sg Nepal go-1sg
 "I went to Nepal"

In 3), verbal agreement is identical to 1), but case marking on the first person singular pronoun is "absolutive," i.e. identical to the "nominative" of the present tense.

The mix of ergative and accusative features exhibited by Nepali is not by any means the only type found in Klaiman's survey. There are Rajasthani dialects that exhibit the opposite phenomenon, i.e. Past tense verbal agreement with theme or other correlate of English direct object, but a complete lack of ergative case marking, as in the following Marwari examples (Klaiman 1988, examples credited to D. Magier 1983):

- (4a) Ra:m athe ka:le a:iy- o [put in diacritic for retroflex aspirated t]
 Ram here yesterday came-sg masculine
 "Ram came here yesterday"
- (4b) Ra:m la:psi: ji:ml-i:
 Ram wheat-gruel ate- sg feminine (concord with *la:psi:*)
 "Ram ate wheat gruel"

(4b) is in contrast to other dialects of Rajasthani and most related languages, in which the agent in a transitive past tense sentence receives ergative marking.

It should be obvious already that anyone wishing to rank North Indian languages in order of ergativity, as Klaiman concludes by doing, is faced with a

dilemma. There are at least two possible criteria on which she might rely, ergativity of case marking and ergativity of verbal agreement. Based on the initial definition of ergativity I gave above, one might take nominal case marking as a prime diagnostic of ergativity. A majority of the languages have tense-aspect conditioned ergative/accusative split, and for most of these case marking is of ergative type in the perfective "ergative domain," but not elsewhere. It turns out that the Tibetan languages as well as a few neighboring Indo-European languages (Dardic Shina in the Northwest, Indo-Aryan Assamese and non-standard forms of Bengali in the Northeast) lack the tense/aspect conditioned split and have ergative case marking throughout. These might then be considered the "most ergative" of the languages considered.

However, Klaiman takes conflicting criteria based on verbal agreement as the "true" indices of ergativity. Having first defined "sets of implicationally related ergative-accusative behaviors" in the various areas of the grammar surveyed—agent case marking, main verb agreement and auxiliary verb agreement, each of these as correlated to tense/aspect-conditioned ergativity split (TACS) and object/theme case marking ("Identified object marking" or IOM is a rule whereby definite or animate object or theme Nps receive dative case marking)—she attempts to collapse all her complex data into a single gradient of ergativity (Klaiman 1988 p.96):

Given the above implicational sets, one might propose tentatively that if a given South Asian language displays any ergative behavior at all, then the farther along the following scale it can be placed, the more ergative it is.

- (45a) The system has TACS (a prerequisite for ergative MV, AV).
- (45b) In the ergative domain, the system has MV with O for number (a prerequisite for having MV with O for other parameters)
- (45c) MV can occur with marked O (a prerequisite for having AV with marked O).
- (45d) The system lacks IOM in the ergative domain (a prerequisite for lacking IOM in the nonergative domain and for having verbal concord with O for person).

and indeed one might propose this. There are no languages in the survey which fail to fall somewhere along the scale, that is which violate the implicational relationships given here. In fact these four implicational relationships may be the strongest that could be abstracted from the wide range of typological data Klaiman gathers. But it is not actually clear that they are really the most important thing going on just because they

are simpler than analyses of the other issues involved would be. It is still less clear that this scale should be regarded as a scale of ergativity. The languages identified earlier as wholly ergative in terms of case marking—the Tibetan languages, Shina, Assamese, Bengali dialects—for this very reason emerge as the least ergative by Klaiman's (45a): they do not have a tense/aspect conditioned split. Since (45a) was probably designed to omit those languages with wholly accusative patterning, one might attempt to improve Klaiman's criteria by allowing languages with wholly accusative case marking to be granted some degree of ergativity. In this case we should expect the accusative verbal agreement of Shina, Assamese and Bengali (like that of Nepali) to keep these languages from moving further down the scale, and Klaiman's other criteria do take care of this. However the Tibetan languages, which are often considered canonical examples of ergativity, lack verbal agreement altogether, and Klaiman's criteria (45b-c), based on features of verbal agreement, also keep Tibetan as one of the least ergative grammars in North India.

The "most ergative" languages in the survey, in Klaiman's opinion, are Kashmiri and Pashto, because in these languages the common areal habit of expressing animate and/or definite theme/objects in dative case is confined to the non-ergative domain, these being treated identically to indefinite themes in the ergative perfective. But note that, like the majority of surveyed languages but unlike the Tibetan and neighboring languages, these languages have accusative domains outside the perfective which exhibit no ergative features at all. The same is also true of Klaiman's "runners-up": these are Rajasthani dialects including Marwari which, while case-marking definite/animate themes as dative, nevertheless maintain ergative verbal agreement even then. But note that in terms of case-marking the latter dialects are highly "accusative", using "nominative" case instead of the obsolete agentive case to mark the "ergative" agent. Despite Klaiman's proposal, I think the above discussion should make clear that there is no single cross-linguistic gradient of ergativity, but rather that a language may have various degrees of ergativity in various different areas of the grammar. Given all this complexity, how did Klaiman arrive at her few criteria for an ergativity gradient? I have already proposed that Klaiman found the few simple generalizations that would hold of all the data, without a theoretical position on which aspects of the data were intrinsically important. Klaiman says (p.97) "(45a-d) is proposed because to label one

language (compared with another) as more ergative—less accusative, or less ergative—more accusative, seems unmeaningful in the absence of some explicit metric which is valid for the particular language group or area, if not for ergative languages generally." While I agree with the reasoning, it seems to me that her (45a-d), while valid, does not constitute an ergativity metric in the conventional sense of "ergativity," and that to speak of languages as more or less ergative may still be unmeaningful.

Not only can a language show a mixture of accusative and ergative features, it may also exhibit neither. Such languages are sometimes assigned to an intermediate "active" syntactic type. One such language (Durie 1988a, b) is Acehnese, an Austronesian language which is spoken in the northwest extremity of Sumatra. The reason that Acehnese cannot be fit into either "accusative" or "ergative" typologies is that there is no single most basic syntactic position that is filled for all verbs with a single argument. Acehnese verbs may be divided into two classes, those that require an actor argument and those that have none, and "intransitive" verbs are found in both categories. That is, if a verb has a single argument, that argument may be encoded as either of the syntactic positions Mark Durie calls "actor" and "undergoer." Note that these are names for syntactic positions, not of underlying semantic roles.

As in English "subject" assignment, the assignment of (syntactic) "actor" argument status in Acehnese is dependent on (semantic) thematic roles not on ("pragmatic") information structure. But the algorithm is not identical: In English an ACTOR thematic role must take syntactic "subject" position unless demoted by passivization, but if it is not present other arguments (INSTRUMENT, or if this is absent, THEME) are assigned "subject" position. In Acehnese the syntactic position Durie calls "actor" may only be filled by an argument with thematic role ACTOR. An "actor" must be animate and must be ascribed the volition/intention to carry out the action described by the verb. (Occasionally, Durie notes, inanimate objects may be given this role, but only if volition is metaphorically being ascribed to them.)

Verbs that do not have an "actor" argument use a different syntactic structure, where the "actor" position is absent. These are often different formations from "active" verbs: for example an active(volitional) verb may be rendered inactive/non-volitional by the prefix *teu-*, or a non-volitional verb becomes volitional by the addition of the prefix *meu-*. With some verbs, however, a simple stem occurs with both volitional and

non-volitional syntactic structures, depending on whether there is an argument that semantically is an intentional actor. Such a distinction is shown in the following sentences. The optional initial position can be used for arguments with either syntactic role, but for an active verb an agent **must** appear in immediately pre-verbal position—this is either a pronominal clitic as in (5b) or a full NP (or pronoun) as in (5c-d). A post-verbal agent is marked by *le-*.

- (5a) (gopnyan) ka maté (-geuh) (gopnyan)
 (3sg-polite) PERF die (3sg-polite-UNDERGOER) (3sg-polite)
 "he/she (polite) died"
- (5b) (gopnyan) ka geu- maté (lé- gopnyan)
 (3sg-polite) PERF 3sg-polite-AGT die (AGT- 3sg-polite)
 "he/she (familiar/polite/reverential) died [intentionally]"
- (5c) ka gopnyan-maté (ditto)
- (5d) ka ureung-maté "A person has died [intentionally]"

Durie gives the example of dying as a martyr in holy war as a situation where the active/volitional structure would be used with *maté* 'die'. This shows the deep roots in semantics of the differing syntactic structures used with 'active' and 'non-active' verbs.

Durie subtitles his article (1988b) "Arguments Against the Category 'Intransitive Subject.' The reason for doing this, he explains, is that two earlier authors Dixon (1979) and Comrie (1978) treat 'intransitive subject' as a universal syntactic category, precisely in order to use it in definitions of "ergativity." According to Durie, Acehnese is sufficient to show that it is not. Durie does prove the distinct syntactic as well as morphological status of actors versus undergoers, and thus I accept his point. On the other hand he leaves it somewhat unclear which part of the structure of Acehnese is morphology and which syntax, or what part of syntax the various parts of the Acehnese sentence are. The next section will delve into these ambiguities, but I bring this up here principally because Durie's (1988b) arguments mostly deal with differences between the semantics and discourse structure properties of actors and undergoers, whereas his claims for the wider implications of his article—the absence of a category 'intransitive subject'—depend on syntax. There is a question, then, of when a language's distinguishing two 'intransitive subjects' warrants classification as an 'active language'

and extricates it from questions of ergativity and accusativity. In order to answer this it is necessary to draw a clear line between syntax and morphology, as one could have a language where volitionality makes a clear difference to morphology and little to syntax. Tagalog, a relative of Acehnese, is such a language, as we will show in section 3.3.

We have shown that, at least sometimes, ergativity and accusativity are both inapplicable labels. However the data from North India show that ergativity and accusativity are also real features of human languages, or else the highly diverse behaviors observed by Klaiman would have no way of differing. It is clear from the various data we have seen that ergativity and accusativity cannot be simple, either/or phenomena that operate on a single level of a language's grammar. Klaiman's data show varying and not directly correlated degrees of ergativity in case marking and in verbal agreement, without even considering at what level of generation these phenomena are produced or whether either corresponds to the syntactic structure on which transformations operate. The meaning of "ergativity" appears elusive and perhaps there is no single useful meaning. Therefore to classify a language as "ergative" appears a highly questionable move, one that should not be undertaken without close analysis both of syntax and of the kinds of data used by Klaiman, and one that may not necessarily point to any meaningful commonalities between that language and other "ergative" languages.

Later in this thesis I will focus on attempts linguists have made to solve the ergativity debate for Tagalog. Without claiming to settle this possibly spurious issue, I will look for a meaningful analysis that will explain the behavior seen as "ergative" and at the same time hopefully show the kind of information syntactic typology should really look for.

1.3 Review

It is instructive at this point to backtrack and look at the interaction in a single language, Acehnese, of the issues just introduced with those of dominant word order. Acehnese word order is not fixed, as is shown by the capacity for various arguments to appear optionally in both initial and final position. According to the typologists, this is

not a problem; there must still be orders describable as "dominant" or "basic." However they face a dilemma: Acehnese has been called both SVO (Lawler 1977) and predominantly V-initial (Durie 1988, p.107). Why this is, and why the choice of terminology depends on the theoretical underpinnings of the syntactic analysis can now be demonstrated. Essentially there are strict ordering constraints for only part of the sentence, which might be characterized as the VP, or alternatively as the syntactic core of a sentence which may also contain a preposed topic. Within this part of the sentence, "agent" is proclitic to the verb, either as a full NP or as a pronoun/agreement marker (it is unclear to me how we would decide between these two labels), and "undergoer" may be enclitic to the verb or follow as a separate phrase. But it is not necessarily obvious whether we can treat these as "S" and "O" positions, or whether we must consider the entire complex as only "V." If we take the latter position there is no clear statement of basic word order. The initial, pre-VP position² is optional, and arguments in this syntactic position may actually play any (semantic)thematic role—actor, undergoer, oblique, adverbial etc. This syntactic position must thus be related to (pragmatic) information structure; it is apparently a topic position. According to Durie (1988b) topicality is more commonly a feature of actors than of undergoers, so even if we call the strictly ordered domain "V" but still call actor "S" and undergoer "O," we still obtain a dominant "SVO" ordering. On the other hand, more often than not (specifically, in 85% of clauses: Durie 1988a, p.107) the "topic" it is absent. Durie (1988b) explains this phenomenon in noting that animate actors tend to be topical, whereas undergoers are less likely to be so, and that if topical, actors tend to be referenced only by a proclitic. Thus for the majority of topical sentences, the "topic" is expressed only by zero anaphora—Durie considers the actor prefix part of the V, not as anaphora, so for this analysis, even in an active sentence, the V tends to come first. Hence, Acehnese is also

² This position is distinguished syntactically from the "actor" position in that it precedes an auxiliary instead of being directly prefixed to the verb stem. If an actor NP appears in initial position, there must be a prefix agreeing with it on the verb. Thus the difference between *peulandôk ka ji-jak* and *ka peulandôk-jak* "A/the mousedeer went" lies in the topicality of the mousedeer: in the first sentence *peulandôk* appears in topic position, in the second in agent position.

"primarily V-initial," but this is interpreted as a consequence of the phenomena just discussed and not itself a constraint on what sentences Acehnese can generate.

Consequentially we can see that the assignment of Acehnese to one or another word order type is contingent on the following interrelated theoretical issues: How do we make the division between morphology and syntax? Acehnese has a category that might be seen as subject or as a subject agreement markers. What is the division between core syntax and peripheral syntax? That is, how do we deal with cleft topics, focuses or other elements whose position is determined by discourse structure and has nothing to do with semantic thematic roles? Are they simply floating syntactically, given meaning only by their proximity to syntactically well-formed sentences they relate to semantically? Or are they a grammatical part of the sentence at the highest level of syntactic organization? What is, according to Durie, a fronted "topic" in Acehnese is initially considered by Lawler (1977) as equivalent to "subject," until this analysis led him to results he claims rule out the existence of a category "subject" for Acehnese altogether. If Lawler is correct, prospects for determining the basic ordering of subject, verb and object in Acehnese look pretty grim. Later on we will look at a number of treatments of potential "subject" categories in Tagalog, and review some radically different conceptions of "subjecthood" which, until these issues are resolved, appear to render Greenberg at least temporary hors de combat.

Durie has pointed out that his analysis of Acehnese word order and verbal morphology had to be completed before the correct identification of Acehnese as an "active type" (as opposed to accusative or ergative) language could be made. Formerly linguists such as Paul Postal and John Lawler (1977), not realizing the limitations of the characterization of Acehnese as SOV, had incorrectly seen a distinction between "accusative" and "passive" or "ergative" structures (Lawler attributes the starting point of his original to Postal, personal communication). Thus sentence (6),

- (6) buku nyan ka ku- blòë (lé- kèë)
book that PERF 1sg-AGT buy (AGT 1sg)
"I bought that book(top)" ("That book, I bought it")

which Lawler would formerly have treated as an example of Acehnese "passive," turns out to have the same syntactic structure as its active counterpart: the domain of syntactic ordering is confined to the VP, where the agent marking is the same as for a sentence where *kèë* "I" precedes the VP.

- (7) *kèë ka ku- blòë buku nyan*
1sg PERF 1sg-AGT buy book that
"I(top) bought that book"

The error initially made by Lawler (1977) indicates what may result from attempts to treat issues of "passive" or "ergative" in isolation from a broader analysis of syntactic structure, as well as the dangers of over-simplified statements of word-order type. Not only will theoretical issues of syntax determine how we classify a language in terms of word order as illustrated in the last paragraph, we see here that if we are not thorough in syntactic analysis, misuse of word order typology may lead to false conclusions about syntax.

2. Syntactic Typology and Western Malayo-Polynesian Languages

The initial reason for Postal's misinterpretation "Acehnese passive" is that the sentence type in question is very similar to a construction in Malay (Indonesian), a closely related Austronesian language, that is almost universally regarded as either "passive" or "ergative." This demonstrates a further potential pitfall of syntactic typology, the possibility of incorrectly treating superficially similar and historically equivalent constructions as synchronically syntactically equivalent. I propose in this section to look briefly at the interaction of syntactic typological with historical and comparative linguistics, specifically to examine what existing schemes of syntactic typology have to say about the Western Malayo-Polynesian languages, the family to which Acehnese and Malay belong. This section will then further serve to introduce data on syntax and grammatical relations in Tagalog, another less closely related member of this family.

The Western Malayo-Polynesian (WMP) languages are, as the name, suggests, a subfamily of the Malayo-Polynesian group, which also includes the huge Oceanic family

of the Pacific. The latter is regarded as constituting a branch of the Austronesian languages; Malayo-Polynesian forms a single subgroup coordinate to the many divisions found in the aboriginal languages of Taiwan. The Western Malayo-Polynesian languages cover a huge geographical range: They include essentially all the languages of the Indonesian archipelago, the Malay peninsula, and the Philippines, along with the Malagasy language of Madagascar and the Cham languages of Indochina. Within such a large family, it is natural to expect some degree of syntactic diversity, as we indeed find.

First of all, the WMP family as a whole exhibits a typological split in terms of dominant word order typology. This may be crudely put as a division between VSO in the east versus SVO in the west, although in none of these languages do such descriptions (to my knowledge) reflect unvarying patterns. There is actually considerable pragmatic variation in word order for most languages. However, it is true that a great many of the languages, particularly in the Philippines and Kalimantan and also including Malagasy, are dominantly (and usually considered to be underlyingly V-initial). On the other hand a number of Western members, notably Malay and Javanese, are with some justification considered dominantly SVO or V2. V2 ordering means, as with German, that there is a tendency for the verb to remain in the second position, despite variation in what precedes the verb. The difficulties already noted in trying to assign a word-order type to Acehnese further complicate the picture. Leaving aside issues of defining kinds of NP arguments—"subject" and "topic" are controversial labels in essentially all these languages—we can see that there really is a strong and noticeable distinction between V-initial and non-V-initial ordering in these languages. Any historical account of syntax for the family will have to account for this.

According to the other typological scheme we are examining, Durie's analysis shows Acehnese to be "active type," i.e. impossible to label as either ergative or accusative. Neither Malay nor the Philippine languages, the two syntaxes with which the author is somewhat familiar, have syntactic positions restricted to volitional actors or to involitional non-actors, in the way Acehnese has (though there are sometimes features of morphology and subcategorization of individual verbs that reflect a similar concept). Rather, agent and patient arguments for the most part occur in similar positions. Without an Acehnese-type split in intransitive argument structure, we might expect for the ergativity-accusativity debate to be easily resolveable one way or another for these

languages: with a single intransitive argument structure, it should be easy to tell whether the position corresponding to intransitive subject is taken prototypically by transitive agent or patient. However, linguists have spent a huge amount of time debating ergativity v. accusativity both for Malay and for Philippine languages.

The reason is that both syntaxes permit multiple surface forms for transitive sentences. In Malay (Indonesian) the problem is simpler: transitives only have a clearly secondary, "passive" construction which, nevertheless, often defies comparison to English passive, being not nearly so marked (or, at the very least, marked in a different way), and often failing to "demote" an agent and function as intransitive in the way English passive does (these terms will be explicated later on in discussions of English and of Tagalog). While the non-occurrence of "passive" morphology for Malay intransitives makes it clear that we cannot call Malay either morphologically or syntactically ergative, the use of Malay "passive" perplexes linguists enough to spark a debate over "discourse ergativity" (In favor see Cartier (1979), which I have not been able to obtain, and without reference to a definition of ergativity, Cartier (1985) and Hopper (1983); opposed see Cumming and Wouk (1988); for a detailed syntactic analysis of Malay passive constructions, see Chung (1976), and for an interesting treatment of apparently non-demote agents in Malay passives, see Myhill (1988)).

In Tagalog the controversy over ergativity has been much more intense. According to Gibson and Starosta (1990), Tagalog is a "purely ergative" language, and they cite analyses done in four different theoretical models—lexicase, relational grammar, Government and Binding and Categorical Grammar—that agree with this conclusion, all from the 1980s. Therefore Gibson and Starosta assert that "Four quite different frameworks have agreed on the basic nature of the data." However it may not be so much the frameworks agreeing as individual linguists' applications of these frameworks: there is definitely dissent. An analysis I will refer to frequently is that of Sweetser (1980), who sees Tagalog as a "split-ergative" language, i.e. having both ergative and accusative structures in different, semantically distinct "domains." Also, Gibson and Starosta are entirely contradicted by the relational grammarian Bell (1983), and more recently by the GB approach of Guilfoyle et al. (1992), both of whom analyze Tagalog as fundamentally "accusative," similar to English in assigning a privileged "subject"-like role to agent arguments at a deep stage of derivation. It is this ergativity

debate that I wish to focus on most thoroughly, to see where a more in-depth attempt to solve such issues leads my thoughts on syntactic typology.

3. Is Tagalog "ergative"? If not, what is it?

I would now like to examine the data underlying the ergativity/accusativity controversy in Tagalog. Linguists have used arguments of various kinds both to argue for the basic "ergativity" and for the basic "accusativity" of this language. There are really two controversies that make up this debate. Firstly it is not obvious which Tagalog syntactic position to regard as equivalent to "subject." Whatever position is treated as "subject," we should expect that in transitive sentences with the language's "basic" structure, this position will be assigned to the agent if Tagalog is basically "accusative," and to theme, experiencer or other non-agent thematic roles if the language is really basically "ergative." However as we shall see, it is also somewhat unclear which if any of several common sentence structures we can call "basic." Since these two issues are so inseparably interlinked, it will not be possible to discuss them separately and sequentially. Instead what I would like to do is first to introduce one possible candidate for "subjecthood" in Tagalog, and then address the role of this constituent in the varying sentence structures found in Tagalog, none of which can safely be labeled as "basic" to the exclusion of the others. Once the diversity of sentence structures has been shown, we can return to a more detailed examination of what "subjecthood" might mean for Tagalog.

In Tagalog, one nominal argument is usually treated as external to the VP (or IP in some interpretations). Arguments in this syntactic position are given various labels by linguists: "subject," "topic," "focus" and "theme" have all been used at one time or another; according to current usage only "subject" or "topic" may be appropriate. The choice of which of these terms to employ may not be significant for the language learner, but it becomes crucial in some theoretical debates: if we call it "topic," we do not have to treat it as "subject" for purposes of resolving the ergativity debate; using the term "topic" allows the linguist to treat this position as really lying outside the core part of the sentence where we look for analogs of "subject," "object," "absolutive" or "ergative" syntactic positions.

The subject/topic argument must be definite. It need not be topic in the strictest sense of topic of an entire conversation or discourse; it need not even, in answering a question, be an element presupposed by the question. However it is said to be of necessity the "most referential" of the NPs in its clause, that which is most firmly grounded in discourse structure, and thus the label of 'topic' is warranted. It is apparently a highly local topic in discourse, and the speaker generally has considerable choice of which argument to make the subject or topic of the sentence without a very major shift in emphasis, as this may only entail a slight shift in the way the existing discourse structure is viewed.

3.1 Case

An NP in subject/topic position is generally preceded by the case marker *ang* (*si* for proper names, while separate case forms exist for pronouns), and is found canonically in sentence-final position though there is considerable word order variation. As already mentioned, only a single NP generally appears with the case marker *ang*; such an NP should be analyzed as occurring outside of the VP, as will become clear later on. The case marking of subjects/topics distinguished them absolutely from other NPs, which receive distinct case markings. There are only two case markings for non-topic NPs: these are *nang* (written <ng> in standard orthography; the marking for proper nouns is *ni*) for direct arguments and *sa* for oblique arguments. The label "direct" arguments covers a number of different theta-roles played by nouns as complements of a verb, the "voice" morphology making clear which of these is entailed: the same case is used to mark theme, experiencer, agent, and possibly others. As any more specific label would be misleading, I will refer to this simply as "complement case." There is a further case marker, *sa* (*ka* or *sa ka* with proper names) which is used for less central arguments, it may be used for a location as well as for an essential 'indirect object' argument, and also appears often after prepositions.

3.2 Voice

Assignment of the subject/topic role is, of course, simplest with intransitive sentences. If a verb has a single theme argument (such verbs usually have the prefix *ma-*), the syntax treats this as subject/topic (example 8-9 from Sweetser 1980).

- (8) n-a- patay ang tigre "the tiger died"
 REAL-*ma-die* TOP tiger

For transitive verbs with two arguments, however, two options are available. The same transitive verb typically has at least two forms, one in which the semantic actor is syntactic subject/topic, and one in which the subject/topic is taken by the theme or other non-actor argument.

- (9a) p-um-atay nang tigre ang lalaki "the man killed a tiger"
 um-die COMP tiger TOP man
- (9b) p-um-atay ang lalaki nang tigre (ditto)
 um-die TOP man COMP tiger
- (9c) p-in-atay nang lalaki ang tigre "a/the man killed the tiger"
 in-die COMP man TOP tiger ("the tiger was killed by a/the man")

The controversy, then, depends on which of these two constructions is considered more "basic," and thus whether the second of these constructions should be regarded as basic and "ergative" or as derived and "passive."

3.2.1 "Active" Voice

Tagalog has a number of affixes that attach to a verb root to form an "active voice" verb. When such a verb is used as the main verb of a sentence, the agent NP must appear in the subject/topic case outside the VP (as in (9a-b) above). Such a verb can also be used as a modifier in an NP (a property of all verbs that take a subject/topic as main verbs); in this usage it modifies an NP that is the agent of the action described, as follows:

- (10) matalino ang lalaki-ng bumasa nang diyaryo
 intelligent TOP man-LK um-read COMP newspaper
 (LK = linker between an N and its modifier)

"The man who read a newspaper is intelligent"

Tagalog has several morphemes that create a verb exhibiting this behavior, principally the infix *-um-* and the prefixes *mag-* and *mang-*. There are also less basic formations in the active voice, like "abilitative" *maka-* which shows that the subject/topic has the ability to perform the action indicated by the verb stem, i.e. is a potential agent. All these formations are referred to as "active voice" because of the special external position given to the actor, and because of the similarity of this syntactic behavior to "active voice" in European languages.

However, use of the Tagalog "active voice" is not necessarily as basic and unlimited as we would expect for an unproblematically "accusative language." This is because not just any NP is allowed to appear as subject/topic of an "active" sentence in Tagalog. Only a definite NP may do so, and generally only an animate NP capable of intentional action. Further, while there exist intransitive "active" verbs—verbs of motion, in particular, fall into this category—the use of transitive active verbs is limited to those cases where the patient argument, appearing as complement of the verb, is *not* definite. Hence the only translation of (9a-b) above is '*The man killed a tiger.*' Lastly, there are some verbs which, though they can appear in other voices with a patient argument and an obligatory agent (see below, "transitive passive voice"), can never appear in the active voice. The reasons for this are unclear, but *takot* 'fear/frighten' is one such root that is never "active."

The diversity of morphology within the active voice, and the occurrence here of both transitive and intransitive verbs, have both been used to argue for the "basic" status of this voice in Tagalog morphosyntax, hence for the "accusativity" of Tagalog. However, the restrictions on its use—definiteness, animacy and volitionality for its subject, indefiniteness for its object, and further lexical restrictions—all contribute to make the use of this voice more limited, say, then that of the "transitive passive," which may be used as an argument that this voice is either not "basic," or at least not "dominant" in Tagalog.

3.2.2 "Transitive Passive"

Essentially any transitive verb that occurs in the Tagalog "active voice," and many that do not, also has at least one "voice" form in which a patient, not agent, argument appears as subject/topic. (9c) above is an example of such a formation. In such a "transitive passive" construction, the agent argument is obligatorily present, and appears in the same complement-of-V position as the "active" direct object. As with the "active voice," there are several different morphemes used to form "transitive passives": these are most basically *i-/(-in-)*, *-in- (-in)*, and *(-in-) -an*. (The *-in-* infix represents realis aspect for this voice only; however the *i-* morpheme drops in this aspect, so *-in-* is often the only marker of voice and aspect.) These three affixes have been analyzed as representing prototypically different kinds of patient argument appearing as subject, *(-in)* representing theme most generally, *i-* an instrument or other patient that is put in motion, and *-an* a motionless patient or especially a location, goal or direction. There are also more specific "transitive passive" morphemes, such as *ipag-*, often topicalizing a beneficiary.

The diversity of "intransitive passive" morphemes takes on more significance than the diversity of "active" morphemes for the following reason: while we normally conceive of there being a single agent (or group of agents acting together) for a single event, the same event may have numerous passive arguments participating in highly varied ways. So it is perhaps unsurprising that Tagalog allows the same verb stem to take multiple "transitive passive" morphemes, each allowing a different patient morpheme to appear as grammatical subject/topic. Despite the prototype meanings of the different morphemes already given, in practice the occurrence of these morphemes with specific verbs are lexically determined, as are the specific thematic roles of topics appearing with the verbs with these morphemes. Schachter (1976) states that syntactically, not morphologically, there are really at least three different transitive passive voices, which he calls "goal-topic" (meaning really, apparently, something like "theme topic"), "direction topic" and "beneficiary topic"; he also alludes mysteriously to others. This analysis is based on the marking of the topic arguments of these voices when they appear in other voices, i.e. whether the topic argument would, if not topic, take "direct" *nang*, "indirect" *sa*, or, as beneficiary, the preposition-case combination *para sa*. The same morpheme may create different "voices" for different verb stems, and crucially, all three of the basic "transitive passive" morphemes form "goal-topic" verbs.

It is only Schachter's "goal-topic voice" that I would like to discuss in this section. It is only this voice that takes as topic an argument clearly as central to a proposition as an agent, and whose case marking reflects a position as basic as that of the agent. It is only this voice that can be profitably compared to psasives or ergative constructions in most languages. The reader will note that I hesitate between using an active and a passive English sentence to translate (9c). There are indeed parallels between such "goal-topic" Tagalog formations and the English passive, namely that in both constructions a patient argument is syntactically the external argument, and that both constructions are more likely to be used the more central to the discourse the patient is as opposed to the agent. However there are key differences.

First of all, it appears that unlike English passive, the use of Tagalog "transitive passive" is actually less marked than that of the active voice. While as in the "active," subject/topic position is restricted to a definite NP, there is no further restriction on subject/topic and no restriction on the definiteness of the agent argument. Hence we say (9c) to translate both "The man killed the tiger" and "A man killed the tiger," as well as the passive equivalents of these sentences. We can see, then, that use of "active" and "passive" in Tagalog coincides more clearly with definiteness than with voice in English. Another instance of non-equivalency of "voice" for the two languages is that such verbs as *takot* which do not have active voice forms, occur freely as "transitive passives":

- (11) *t-in-akot nang lalaki ang anak*
in-fear COMP man TOP child
"The child was afraid of/frightened by a/the man"

The second major difference has already been pointed out; it is the obligatory appearance of the agent in these passives. English passive removes agent as a core argument of the verb, so that may either be omitted entirely, or expressed in a prepositional phrase, the same syntactic position granted to other non-obligatory modifiers such as locations. The agent in these Tagalog constructions has the same central syntactic position as is given to "direct object" in a transitive active construction; clearly agent remains a core argument. If we wish to omit the agent as in English and say "the tiger has been killed," we have no choice but to say (8) above, which we translated "the tiger died." The *ma-* construction, which has sometimes been labeled

"intransitive passive," is also used for many intransitive constructions that cannot possibly be called "passive." We will examine the use of this morpheme momentarily.

Before turning to the various *ma-* constructions, however, I wish to take a moment to refine the terminology in use. Based on the non-correspondence between Tagalog "transitive passive" and English-type passive, the obligatory presence of agent in this construction, and the apparent lesser markedness of this construction as opposed to Tagalog "active," it seems only natural to distinguish this construction from passive in our terminology, and relabel it as "ergative." This does not necessarily entail accepting the proposal that Tagalog is a full-blown ergative language, but it does prevent us from making assumptions about this construction based on our understanding of what "passive" means, and reserves the label "passive" for those instances of *ma-* which actually do parallel English passive.

3.2.3 "Intransitive Passive"

The verbs that are covered under this heading, while they all share the same morphology, are not all placed by traditional grammars of Tagalog in the same "voice" category. However I believe doing so is warranted. Sweetser (1980) has analyzed the different types of *ma-* construction in detail, and I will here follow her findings. Sweetser finds these verbs to be of the following types: simple intransitives, normal verbs of perception, "intransitive passives" correlating with active verbs from the same stem, and an anomalous group of "verbs of active perception."

First of all, while we have already noted that certain intransitive verbs take "active" morphology, there are a great many that take *ma-* as well. We can see a semantic difference between these, parallel to that seen between Acehnese intransitives taking an "actor" and those taking an "undergoer." Among Sweetser's examples of basic *ma-* intransitives are *mapatay* "die," *matulog* "sleep," *magutom* "be hungry" and *magalit* "be angry." Like the active intransitives, these verbs also require an animate subject, yet, unlike the active intransitives, not a volitional one. According to Sweetser, analyses previous to hers, including Schachter's and the standard tradition of Tagalog grammars, call these verbs "actor focus" (or "actor topic"), perhaps because the theory assumes that an obligatorily intransitive verb, which can therefore have only a single voice, must thereby only be "active voice." She argues that such a theory would be based on a

misconceived category of "actor." Schachter and other grammarians of Tagalog report a syntactic category of "actor" corresponding neither to the thematic role AGENT nor to Tagalog verbal morphology. (The category cannot be based on case marking, because actors take no case markings they do not share with other categories.) Rather "actor" seems to combine the thematic roles of AGENT and of PERCEIVER/EXPERIENCER, or to correspond roughly with the prototypes of English "active subject."

I believe the demonstrable semantic difference between *ma-* and other "active intransitives" is sufficient to show that the traditional "active" category should be divided; these verbs are not "active" in that their subject is not AGENT, but rather EXPERIENCER (we might also call it THEME, but this term is so broad it seems preferable to use the more concise term). Doing this allows Sweetser a more unified analysis of *ma-* as simply not active and not transitive. This is a broader characterization than is possible at this stage in the analysis of *ma-*, but the other option "experiencer topic" is eliminated by the next category to be examined, verbs of perception.

Verbs such as *makita*, *madinig*, *marasa*, *maramdam* (corresponding in broad meaning, but not in argument structure, to English 'see,' 'hear,' 'feel,' 'taste' take only a single obligatory argument, appearing in subject/topic position; that argument corresponds not to the English (PERCEIVER/EXPERIENCER) subject of English verbs of perception, but to the English object of perception (here we are forced to use THEME). Thus it is possible to say

- (12) *n-a- ki- kita ang manga bulaklak*
REAL-*ma*-RED-see TOP plural flower
"The flowers are being seen"

(In this example, realis *n-* and reduplication are both aspect markers which combine to suggest a present tense reading) This usage, as compared with English, suggests the "intransitive passive" label (see below), except that in Tagalog this can not be regarded as the passive of any particular formation. It is the normal construction for verbs of

perception, which do not take basic "active voice" morphemes³. In this construction the verb optionally takes a complement, understood as PERCEIVER/EXPERIENCER:

- (13) n-a- ki- kita nang baba'e ang bulaklak
REAL-*ma*-RED-see COMP woman TOP flower
"The woman sees the flower(s)"

The behavior of these verbs seems exactly parallel to that of the "intransitive passive" use of *ma*-, an example of which has already been seen in (8). The only difference is that intransitive passives by definition must correspond to an active and/or ergative construction, or else there will be no sense in which we can call them passive at all, and they will be simply intransitive. For all their intransitivity, "transitive passives" can take an optional complement of the verb, which we might be tempted to identify with Schachter's "actor":

- (14) n-a-patay ni Juan ang tigre
perf-*ma*-die COMP J. TOP tiger
"Juan killed the tiger"

However we should not be too hasty to call Juan "actor" in (14). According to Sweetser, in (9a-c) "the agent, Juan, is represented as having deliberately killed the tiger"; in (14) "he need not have intended to do it (Sweetser 1980 p332)". The distinction between Juan's role in the different constructions accords with the narrower, volition-based definition of "actor" that Sweetser argues for, one paralleling Durie's Acehnese requirements for actor. Juan in (14) would then have to be some other kind of complement, although not a perceiver/experiencer as in (13). It seems Juan in (14) is some kind of abstract cause, the operation of which is unspecified, that is "The tiger died of Juan." Yet how the hearer arrives at a causal reading for (14) but not (13) is unclear; it may be that the semantics of VP complements is guessed at based on the lexical meaning verb.

³ Such verbs can take secondary active morphemes, such as *maka*-, where the perceiver is also an abilitative agent: *makakita nang bulaklak si Maria* 'Maria is able to see the flower(s).'

The volitional definition of "actor" allows us to exclude "Juan" as "actor" in (14), and unite the three categories of *ma-* seen already as a single category, defined as non-active intransitive. The only problems Sweetser sees to this classification are a fourth class of *ma-* verbs, verbs of active perception such as "look at" or "listen to." According to Sweetser, this class has features both of "active transitives" and of "inactive transitives," not unsurprising for a category whose subject topic unites the features EXPERIENCER/PERCEIVER and AGENT. However, Sweetser is not very explicit in pointing out which features she means, so I will accept the broad categorization she gives of *ma-* verbs and avoid discussing this subclass.

3.3 Comparison of Tagalog and Acehnese

Having addressed the most central and dominant constituents of Tagalog's voice system, we may be ready to return to the issue of "ergativity" and "accusativity." If we accept Sweetser's analysis of two separate classes of intransitives, based on the presence or absence of an AGENT thematic role, might we expect the issue to disappear and yield an "active" type analysis as a similar split did in Acehnese? In Acehnese, which lacks a voice system, Durie found that "intransitive subject" is not a single category, but rather two different syntactic positions; therefore it is impossible to say whether transitive agent or transitive patient (each of which has a single syntactic position) has the position of "transitive subject."

Actually Tagalog is not like Acehnese in this respect. Though "intransitive subject" must be broken down into volitional and non-volitional (if you like, active and passive) categories, these yield not different syntactic positions as in Acehnese, but rather different verbal morphology (voice) and a single syntactic position for "intransitive subject." Thus we are left with the question of whether this same position is used more prototypically or more basically for the agent or patient argument in transitives, and this question still looks equally complex for Tagalog. For while in Acehnese "intransitive subject" is two different syntactic positions while transitive agent and patient have one unvarying syntactic position apiece, Tagalog's voice system means there is a single syntactic position for intransitive subject, there is no single syntactic position for either transitive agent or transitive patient. Both of these occur

sometimes as sentential subject/topic, with *ang* case, and sometimes as VP complement, with *nang* case.

Might we say that Tagalog simply transcends the ergativity/accusativity split? We might, though we cannot say that it stands in between like Acehnese. Rather, we would have to say that it stands on both sides of the fence. And we will, later on. But we are not ready to argue this yet.

3.4 Review

What Sweetser has really done in separating the "active," volitional actor and the pseudo-actors of *ma*-intransitives is to recognize a different, more direct linkage of theta-roles to morphology and argument structure than is suggested by the older model. Sweetser's analysis does help toward a correct understanding of the ergativity issue to the following extent: it shows that transitives must be defined, as in Acehnese, as those verbs requiring an agent and at least one patient. It does not include those verbs which allow the interaction of two non-agent arguments, as in the verbs of non-active perception. Schachter, with his definition of "actor" including some non-agents, might have allowed these to be transitive, and we would then have cases of transitives permitting the omission of "actor." We can now be certain that only the status of (narrowly) active and ("transitive passive") ergative constructions are at issue in a discussion of ergativity.

On the other hand, we are not much closer to a resolution; neither of the constructions really looks too much more basic, (It would be good to have data on their statistical frequency, but I never thought of getting that before now) and all the evidence really on the table so far is the lesser restriction on the definiteness of arguments in the ergative construction and the obligatory occurrence of the ergative in verbs like *takot*. Both these pieces of evidence seem to point towards the "basicness" or "dominance" of the ergative construction, and towards calling Tagalog an "ergative language." However I think "Tagalog is ergative" is really an overly hasty judgement: there is other evidence that linguists have adduced that seems to point in a different direction.

In cautioning against the "ergative language" analysis we should first remind the reader of the multiple semantic and syntactic range of the transitive passive/ergative

constructions. Really active and ergative are not just two sentence structures with subject/topic playing two different roles. For verbs with more arguments than two, there will often be more than one "ergative" voice of the verb. Recall that multiple ergatives from the same root have different characteristic morphology and different arguments as subject/topic, but that the linkage between the two levels of difference is lexically idiosyncratic.

(15a) i-b-in-igay nang Juan ang kape sa ka-n-ako
ERG-REAL-give COMP Juan TOP coffee IND DAT-LK-1sg
"Juan gave me the coffee"/"The coffee was given to me by Juan"

(15b) b-in-igy-an nang Juan nang kape ako
REAL-give-ERG COMP J. COMP coffee (top)1sg
"Juan gave me the coffee"/"I was given the coffee by Juan"

If we are to say that, for verbs with multiple basic arguments, an ergative construction is most "basic" (for this is what the label "ergative language" essentially means), we will have to resolve which ergative construction the most basic construction is. Presumably we would want the most basic to be that where THEME is topic, but we have no actual evidence to show any of the different ergatives to be less basic than others, still less any to show that other ergatives are somehow derived from a THEME topic ergative. In fact, the most noticeable way utilized in English to derive one argument structure (passive) from another (active), the "demotion" of the most essential (subject) argument of the more basic (active) construction, cannot be invoked in Tagalog to derive different ergatives from each other or to derive either active from ergative or vice versa: all these constructions will have the same number of essential arguments.

Not only can we not argue for "basicness" of ergativity by arguing for the derivation of other voices from it, we can not even get this from the pro-ergativity evidence we have accepted as such. Remember that "dominant" at surface structure need not be the same as "basic"; for German, we saw that a "basic" SOV structure was obscured by the dominant operation of a transformation creating a V2 and predominantly SVO structure. Recognizing this distinction, there is no reason we could not see a "dominant" ergative and a "recessive" active construction as equally "basic," or alternatively, as equally derived.

3.5 Voicelessness

There is another issue we need to explore in analyzing active versus ergative constructions. The very terminology we have been using in labeling these constructions turns around the argument in subject/topic position. Thus the "ergativity" hypothesis relies on treating the status of subject/topic as resolved, and as essentially equivalent to "subject," while I have already intimated that this view may be problematic. We have already seen that more different kinds of arguments can take this position than can be "subject" in a language like English; however maybe this shouldn't be worrisome. What is definitely worrisome about treating this position as "subject," that is the basic external argument position to take into account when discussing the ergativity or accusativity of Tagalog, is that there are also a number of sentence types where subject/topic cannot occur.

I should clarify that there is a difference between sentences where no subject/topic is present and those where a subject/topic is grammatically impossible. The former is by no means rare in Tagalog but it shows nothing in particular to aid our understanding. Zero anaphora is, apparently, common enough in Tagalog for subject/topic, and thus if a human actor is already under discussion it is grammatical to say

(16) pumatay nang tigre 'He killed a tiger.'

This is, however, not grammatically any different from (9a-b); the construction is still one that takes a subject/topic, and we might even say that there is a null subject/topic present in the sentence, as one is at least understood. (16) is still active voice. Much the same may be said of such impersonal sentences as the equivalent of English 'It's raining.' For such sentences, where English uses a "dummy subject" 'it,' Tagalog again shows no surface subject, but still uses active voice morphology that requires the presence of a subject, whether explicit or null. While these are examples of, in a sense, subjectless sentences, they are still active voice. There are other constructions, however, that truly are "voiceless" and whose syntax is incapable of accommodating a subject/topic.

3.5.1 Existentials

Earlier I mentioned that the subject/topic of a Tagalog sentence must be definite, that is already known, either from earlier in the discourse in which the sentence occurs, or from outside the discourse in question. There are, however, some sentences which can contain no definite arguments, as when new entities are introduced into a discourse. The principal sentence type that does this is the existential sentence, which in English uses the dummy subject 'there.' Schachter (1977) has questioned the traditional analysis of English existentials, in which the NP argument introduced into the discourse by such sentences is treated as "underlying subject." Rather he feels that the NP whose existence is affirmed is demonstrably semantically a predicate, and I agree.

According to Sean Lewis (personal communication) only those English sentences he calls "categorical" are semantically (as opposed to syntactically) divisible into subject (also, "topic") and predicate. For such sentences, subject/topic is "presupposed," specifically identifiable either from earlier in the discourse or from knowledge external to the discourse; this is the old information or "topic" serving to anchor the statement in the discourse. Meanwhile a semantic predicate (equivalent also to "comment," "focus") is new information introduced into and advancing the discourse. Lewis asserts that a number of factors show the semantic division of "categorical" sentences into subject and predicate, notably that the semantic scope of negation, emphasis and questioning all seem to be restricted to the predicate.

On the other hand Lewis sees an opposed class of English sentences which, while they share with "categorical" sentences the syntax of subject and predicate, lack the semantic division into subject and predicate. This occurs when an entire sentence consists of elements being introduced into the discourse, no part of the sentence being presupposed and "topical." Such a sentence is called "thetic," and is distinguishable from the "categorical" kind in that here the semantic scope of questions, emphasis and negation appears to be the entire semantic content of the sentence. If we accept this division of English sentences on a semantic level, we can see that the argument whose existence is asserted by an "existential" sentence cannot be semantic subject; it is introduced into the discourse by the existential sentence, and cannot be presupposed. Therefore existentials must be "thetic," and their arguments semantically predicates.

This actually became a foregone conclusion at the point where we accepted that, on a semantic level, "predication" is simply equivalent introducing new information into the discourse. To follow out the consequences of this definition a little bit, the difference between "categorical" and "thetic" sentences is that, in the former, new information is predicated *of* a presupposed argument. In a thetic sentence, according to Lewis, new information is predicated of a "default subject" consisting logically of a context—i.e. principally a time and place—either known from the discourse or assumed to be the context in which speaker and hearer actually find themselves. The idea of a "context" or "time and place" as semantic subject of an existential sentence explains the surface "dummy subject" 'there' of English existentials much better than an analysis considering the NP being introduced as "underlying subject." Rather, Lewis's and Schachter's work combine to yield an analysis where the NP introduced is underlyingly a predicate, and the contextual "default subject" is expressed by the existential "dummy subject" 'there.' (In a variation of the standard English rule, BEV (Black English Vernacular) uses the same "dummy subject" 'it's' for existentials as for impersonals.)

Tagalog shows the same "categorical" : "thetic" distinction as English, and indeed shows it much more clearly at surface structure. This is because "thetic" sentences in English imitate the structure of "categorical" sentences to the point of syntactic indistinguishability. This is done in either of two ways: first, an argument that is not semantically a subject (topic) may be treated as syntactically a subject by putting it in the same position as the semantic subject of a "categorical" sentence. Alternatively, English uses "dummy subjects" which take the syntactic position of the "categorical" subject. Tagalog "thetic" sentences cannot be camouflaged in either of these ways. Recall the definiteness requirement for the Tagalog syntactic category subject/topic, stated in the introduction to section 3. I hold that this requirement is identical to the requirement for English semantic category "categorical subject," also called "topic," which I defined not so long ago as "'presupposed,' specifically identifiable either from earlier in the discourse or from knowledge external to the discourse." While English allows "dummy subjects" and non-topic arguments to stand in syntactically for "categorical subject" if this is absent, no dummy subject or non-topic can take topic position in Tagalog. Thus it is precisely the "thetic sentences" in Tagalog that do not grammatically permit

subject/topic, and likewise do not use any of the "voices" enumerated in section 3.2, which all predicate something of a subject/topic.

Not surprisingly, Tagalog existentials are among those sentences that are "thetic," "voiceless" and lacking a subject/topic. Since the argument introduced by the existential sentence cannot be presupposed, it clearly cannot be subject/topic, and indeed it does not take subject/topic case. On the other hand, it does not take either of the other two Tagalog case markings introduced in section 3.1.

- (17) may aksidente
EX accident
"There's been an accident"

The structure of this sentence looks dissimilar to anything we have encountered thus far, and the lack of case marking on the NP *aksidente* is also new. As the argument introduced is semantically a predicate, it is probable that the existential construction should be seen as parallel to the predicate portion of the "categorical" sentences so far. Unfortunately given the limited range of categorical sentence types examined it may not be easy to convince the reader of this.

A predicate in the sentences examined so far has consisted of a verb with "voice," followed by the verb's arguments excepting that argument which the voice of the verb specifies as "external." However there are several other types of predicate used in "categorical" sentences (ie these are predicated of a subject/topic). A predicate may consist of a modal with its complements. Modals, like existential *may*, do not have voice marking, but Sweetser (1980) argues that, in their argument structure, modals function analogously either to actives or to intransitive passives, depending on whether their semantics include volition. In any event, modals do case-mark arguments like verbs, so they do not serve as a good analogy for *may*. A categorical predicate may also consist simply of an adjective (which is a category not perhaps really distinguishable from an intransitive passive verb: most "adjectives" I have encountered appear to be *ma-* forms not taking complements, but on the other hand I haven't seen them with aspect markers either. Anyway, I'm not the one to answer this question). Finally, a categorical predicate may consist of a predicate NP not marked for case, the best "categorical" parallel we can find for the existential construction.

- (18) abogado ang malaki- ng lalaki
 lawyer TOP (?ma-)big LK (?RED-)man
 "The big man is a lawyer"

Since, outside the existential construction, an NP only occurs without case marking when it forms a predicate by itself, it stands to reason that the NP within the existential should be some kind of predicate. The syntactic role of the existential marker *may* remains a little unclear, but I can live with that.

There is a further argument for the predicate nature of existentials, but one that might conceivably lead to the abolishment of the category. This is that the construction in Tagalog corresponding to English 'to have' appears to consist of the juxtaposition of an existential and a subject/topic.

- (19) may libro ang abogado
 EX book TOP lawyer
 'The lawyer has a book'

This clearly shows a construction identical to the existential used as a predicate. It is unclear though if we want to call this an existential, which we define as a kind of "thetic" predicate, or a "categorical predicate" of the NP *abogado*. While there is still a sense of existentiality, i.e. the sentence does assert the existence versus the nonexistence of a book belonging to the lawyer, I still think this is not a true existential. In fact I think it is best to think of the true existential as a "thetic" form of the above: instead of a "presupposed" topic possessing a book, in the true existential the "default subject" or context at large is said to possess a book. Although there is no "dummy subject," it is possible to think of the true existential as containing zero anaphora equivalent to "dummy subject" for subject/topic, as in impersonals and many normal "categorical" sentences. This is not unlike the Spanish existential

- (20) hay libro
 (3sg)have book
 '(It) has a book' = 'There is a book'

On the other hand it is not desirable to treat *may* as a verb, as it is then the only verb not to case-mark its complement. In any case we have established that the existential is

an entirely predicate sentence, but we have not definitively severed it from sentences with a subject/topic. Fortunately for the direction I have chosen to explore, there is another sentence type where we can be certain topic plays no role whatsoever, which will be seen in the next section.

As an aside, I want to return briefly to impersonals, which I dismissed earlier from the discussion of "voicelessness." We have just proposed the possibility that true existentials both are genuinely "thetic" and at the same time bear a voice-like relation to the "default subject" of time and place. This suggests the following parallel analysis of certain impersonal constructions, exemplified by the Cebuano and English sentences in (21) (Cebuano is a close relative of Tagalog; see Bell 1983)

(21) n- ag- ulan
REAL-mag-rain
'It's raining'

As with Tagalog impersonals, (21) combines voice morphology with the obligatory absence of a subject, with the same effect as the dummy "it" of English. I submit that what we see here is the metaphorical (or real?) attribution of volition (shown by the use of "active voice" mag-) to the circumstantial "default subject" (it). This would then, while clearly not voiceless, actually would be a special case of the "thetic" scenario, in which the context of discourse is seen as itself volitional, which is a good a way as any to describe the weather.

3.5.2 "Recent past"

According to Sweetser(1980, p.338), "in Tagalog it is very difficult to say a sentence with no definite NPs—the pragmatic structure of the language requires an *ang*-phrase in every non-existential sentence." Sweetser's "sentence without definite NPs" would, by definition, be a "thetic" sentence. As it turns out, Sweetser is not quite correct, as there is one other type of sentence which does not just allow the absence of an *ang*-phrase, it requires it. In his discussion of existentials, Schachter (1977) also ignores this other class of "thetic," subjectless sentence. According to Guilfoyle et al.(1992), the construction known traditionally as the "recent past" takes all the

arguments of the verb as complements within the VP; no "external" argument is allowed:

- (22) Ka-ka-ka'in nang leon nang tigre
ka-RED-eat COMP lion COMP tiger
'A lion has eaten a tiger'(unlikely, but helpful)

Perhaps the reason Sweetser and Schachter both missed this crucial evidence is that it seems to be conventionally regarded as a "tense," not a "voice." Actually linguists generally recognize that "tense" in Tagalog is an inaccurate interpretation based on some correspondence of complex Tagalog "aspect" forms with European "tenses." While I have been ignoring aspect for the most part, Tagalog verbs in the voices we saw in section 3.2 are marked for two aspect features: [\pm DURATIVE], shown by the presence or absence of pre-verb stem reduplication, and [\pm REALIS], shown either by a prefix *n-* or, in the ergative, an infix *-in-*. [+REALIS, -DURATIVE] verbs often are called "past tense" and [+REALIS, +DURATIVE] verbs often get called "present tense." A future continuous "tense" may be expressed by [-REALIS, +DURATIVE], and so on.

The "recent past" construction cannot be easily placed in this system. In fact, it appears to be both a distinct aspect and a distinct "voice." Since a [+REALIS, -DURATIVE] active or ergative verb may equally express "recent past" time, the difference between the "recent past" construction and both ergative and accusative realis constructions must not really one of aspect, but of the roles played by its arguments in discourse structure. From my perspective, the most remarkable thing about the "recent past" is that no NP is topical; it would be helpful if we could regard the other, aspectual, characteristics of this construction as secondary to the "voice" characteristics, and not vice versa.

I suggest that the pragmatic distinction between this and subject-assigning voices is precisely equivalent to the distinction Sean Lewis is examining between "thetic" and "categorical" readings of homophonous English active sentences. In "thetic" statements, no argument within the sentence is "topical"; the entire sentence is "focal," that which is "at issue." Because no argument is topical, a "null topic" is assumed, by default the current time, place etc., whence the "recent past" aspectual implication of this voice.

3.5.3 The Meaning of Voicelessness

I have been arguing for the interpretation of "recent past" as principally a distinct "voice" for Tagalog. Really, though, I want to argue that "recent past" is the absence of "voice." I say this because all the other "voices" we have examined involve verbal affixes carrying features of argument structure, and all specify the "externality" of some argument. The "recent past" construction shows that an "external argument" is not a necessary for all verbs, and that all arguments can appear within the VP. It is thus unlike all the other "voices" in not privileging a certain NP. The same thing is shown by the fact that both agent and patient arguments in (22) have the same complement case marking. The real significance of this is that it shows once and for all the futility and needlessness of calling one of the surface "voice" forms of Tagalog "basic." If the "ergativity" debate is really one of which NP argument is most prototypically "external," or most clearly privileged, the recent past construction shows that the answer may be "none of them." If "recent past" preserves the most "basic" argument structure of Tagalog verbs, as I would like to argue, the only clue as to which argument is more privileged is the word order, and it is really a metaphysical question which of two successive identically marked arguments of a transitive verb is really more equivalent to the single argument of an intransitive verb.

The "basic" status of the argument structure of the "voiceless," "thetic," "recent past" construction seems likely for a number of reasons. We have already seen how hard it is to make convincing arguments for the "basicness" of one or another of Tagalog's "voices." I believe that any such argument is not merely difficult, it is inherently unlikely. The pragmatic definiteness restrictions on arguments in subject/topic position for each of the voices show that voice is really a matter of discourse structure, hardly related to the lexical semantics of the verb stem. To have an answer to the ergative/accusative question at the level of voice and subject/topic assignment would be to claim that a given verb stem takes one "voice" construction more "basically" than another, and also "basically" prefers to be generated when a particular relationship obtains between discourse structure and the verb's argument structure. It would be much simpler to claim that verb stems do not inherently prefer one voice, and that voice is a feature added to the verb when it comes up in a particular

situation in discourse structure. The fact that voiceless forms such as (22) appear, and seem to appear in the most neutral of situations in terms of discourse structure, i.e. when all arguments are equal untopical, is strong evidence for the non-basic nature of all "voice" morphology in the generation of Tagalog verbs.

If voice is not lexically basic and essential to the Tagalog verbs, most recent theories hold that it must be generated as some kind of "transformation." A common assumption of now-defunct Transformational Grammar, Postal and Perlmutter's Relational Grammar, and the Government and Binding model is that structurally different sentences with identical lexical contents and identical semantic truth value must, at some stage in the derivation, have identical syntactic structures. Although the vocabularies of the different theoretical models differ somewhat, I will refer to the level at which these structures are identical as "deep structure," and the processes by which identical deep structures are converted into varying surface structures as "transformations." According to the common assumptions of all three of these theories, then, Tagalog verbal sentences, with or without "voice," must at "deep structure" have identical structures in the VP.

If we are certain "voice" operates as a "transformation" on a verb or VP, it is economical and tempting to propose that the difference of different "voice" forms of otherwise equivalent sentences is due entirely to the "voice" transformation. Voice would then would create the surface topic-comment structure by moving one argument of the verb to "subject/topic" position, outside the VP; the argument structure seen in the "recent past" construction, then, is simply the failure of any such "transformation" to apply. This is the approach taken by Guilfoyle et al. (1992), who share my predisposition to see all "voice" in Tagalog as a non-basic transformation. In this Government and Binding approach, "voice" is seen as a transformation whereby one argument from within VP is "raised" to a position Spec of IP, thereby becoming subject/topic.

On the other hand, there are a few things which make me uncomfortable with this approach. The topic-comment structure of "categorical," voiced sentences in Tagalog is something semantically very basic. We may see speech as the process of "advancing" the information structure we call "discourse"; if we are not advancing the discourse, we do not speak at all. The "topical" part of a sentence is the point of

departure within existing discourse that we advance from, the comment the direction in which we are advancing. The special status of the "topic" already has a psychological reality when we begin to generate a sentence. Furthermore, the topic-comment structure of "voiced" verbal sentences is really identical to the structure of other "categorical" sentences, including those with nominal predicates. We may see nouns as having argument structure like verbs (at the very least, a noun has the argument structure <REFERENT>(Bobaljik 2000)), but given that they do not have different "voice" forms, it seems ludicrous to see them as needing a "voice transformation" to generate their argument structure and the syntactic structure of copular sentences. Therefore, if we are to look at "voice" as a transformation, we should look at it as changing the structure of the VP, but not as actually creating the topic-comment structure of the "categorical" sentence.

Another thing to notice about the the operation of "voice" is its highly lexically idiosyncratic nature. Voice involves both the addition of verbal morphology and a change in the argument structure: for example the verb stem *ká:'in* 'eat' generates "recent past" *ka-ka-ká:'in* <<AGENT, THEME>>, active *k-um-á:'in* <<THEME>AGENT> and ergative *k-in-á:'in* <<AGENT>THEME>. Yet, as mentioned earlier, the change in argument structure that accompanies different morphemes is not entirely predictable, and which morpheme should be added to produce a desired change in argument structure depends entirely on the lexical idiosyncracies of the verb stem. Thus the best we can do to describe "voice" on a syntactic level is to see it as a process giving features to the verb; the grammar would then look to the lexical entry of the verb to obtain a verb with the desired features of argument structure.

I propose the following model for the derivation of "voice" in Tagalog: all arguments must be generated within the VP, in order to allow for the "recent past" argument structure. At the same time, the semantic topic is generated in its surface external position. The "voice" rule would then operate by eliminating an argument corresponding to the subject/topic from the VP and adding a voice feature to the verb; the surface manifestation of the voice feature is obtained from the lexical entry to the verb. This model would allow a single deep structure for the verb phrase in different voice forms, while also allowing the special position of the semantic topic (which I believe to play a very basic special role in the generation of language) to be basic in

erg-frightened COMP man TOP DAT-LK-3sg-LK self
"The man was frightened of/by himself"

The "control" of reflexives exhibited above has been taken as an argument for the underlying "subject"-ness of agents even in sentences with an "ergative" surface structure. In fact agents turn out to retain other canonical "subject" properties in "ergative" sentences, such as equi-subject deletion, and address of imperatives (Schachter 1975, 1977). For this reason many linguists feel that, if Tagalog is "ergative," it is one of the many languages that is not "basically" but only "superficially" so. Ergative surface structure is then, like English passive, regarded as derived from the active structure by "transformations" or, in the language of Relational Grammar, "advancements" (Bell 1983).

For reasons such as the above, Schachter proposes that there is not just one syntactically "privileged" kind of NP argument (as we have thus far assumed subject/topic to be) but two, subject/topic and actor. He thus refuses to settle on one definition of "subject" for Tagalog. On the basis of Tagalog data, Paul Schachter divides canonical "subject properties" into two sets: "role-related" and "reference-related" properties. The first, roughly speaking, determine a syntactic "subject" based on the semantic properties now generally called "thematic roles," such as ACTOR, EXPERIENCER, THEME, DATIVE, BENEFACTIVE, INSTRUMENT, etc. "Reference-related" properties determining subject assignment, on the other hand, derive from "pragmatics" or "discourse structure," based on such factors as definiteness/indefiniteness, old v. new information, and possibly other factors such as animacy or person which affect the relative prominence (as viewed by the speaker) of different arguments in a proposition. Crucially to Schachter, the two sets of properties may each determine a different "subject" for their respective purposes, and each is relevant to determining the scope of different syntactic processes. Thus in Tagalog such processes as reflexivization and equi-subject deletion are controlled by the "actor" argument, which Schachter accordingly treats as "subject" from the standpoint of "role-related properties." "Actor" retains this role in all of the various "voices"/"topic" or "focus" constructions of Tagalog, in which "actor" appears in different syntactic positions, and various arguments—several types of oblique argument as well as "actor"—may appear as syntactic "subject" or "topic" of the

sentence. It is this latter surface syntactic position which Schachter calls "subject" from the standpoint of "reference-related properties."

3.7 Syntactic Status of Two Subjects

Schachter's division of "subject properties" in two is not inconsistent with the analysis I proposed earlier for Tagalog voice. In this proposal, the "reference-related subject" (subject/topic) held a privileged position in syntax and was not created by a transformation. On the other hand its role with respect to the VP was not considered part of "deep structure"; rather, I proposed that VPs do not initially privilege any argument, but that they came to when "voice" occurred through interaction of the VP with a given subject/topic. Now that it turns out to be the case that Tagalog privileges two different kinds of NPs and that there is a "role-related subject" unrelated to the subject/topic, my model can easily accommodate it. Since the selection of "role-related" subject appears to have everything to do with thematic roles and argument structures and nothing to do with discourse roles, it stands to reason that the privileged status of the "role-related subject," unlike that of the subject-topic, is base-generated within the VP. The selection of "actor" for such treatment would appear to make Tagalog "accusative" with respect to this kind of "subject" only.

Schachter refuses to take any position on which of his two candidates for "subject" is really better or has a "more basic" status in order of generation. Rather he concludes that the notion of "subject" is actually somewhat illusory, or rather the result of two different universal categories being lumped together in languages like English while they are underlyingly two quite distinct entities. By contrast, Bell (1983) has analyzed the two different classes of "subject" and "subject properties" as operating on separate, ordered strata of derivation within a Relational Grammar framework. (Bell's work deals largely with Cebuano, a closely related language, but she refers to Schachter and reassures readers that the facts of syntax do not appear to differ significantly between the two languages.) Sentences with "reference-related" surface "subjects" which are not actors are by this model transformationally derived from a lower stratum (corresponding to the "deep structure" of classical transformational grammar) in which the actor is subject, and on which the syntactic processes having to do with "role-related

subject properties" occur. On the other hand, sentences where a single argument is both "role-related" and "reference-related" are treated as not having undergone transformation. In the Relational Grammar framework she uses, basic and derived sentences are analyzed in terms of the grammatical relations of their NP arguments, syntactic structure being ignored. A "subject" is taken to be the most important, privileged argument, and thus "subject" status is denoted as "1," direct object status as "2," and so on. A transformation whereby an argument is raised to a higher level in this hierarchy is called an "advancement," and this is how Bell treats non-active voices in Philippine languages.

It is interesting to notice that in the relational grammar analysis of Tagalog, initial subject assignment according to thematic roles operates very comparably to English. The different "voices" then become comparable to the transformational passive of English, with some crucial differences—namely that the various "passives" are less marked than that of English, can transform more arguments into surface subject, and even are obligatory under certain circumstances (if there is a definite patient argument) or with certain lexically specified verbs e.g. *takot*). These differences are not all the cautions we need if we are to regard "active" as basic and ergative as derived in Tagalog. If Tagalog ergative is, like English passive, a derived structure, it is not one that is really parallel to English passive. English passive "demotes" an "initial subject," so that it is either omitted from the passive sentence altogether, or else appears in a non-core argument position within a prepositional phrase. Secondly, when English "advances" an underlying complement of VP to "subject" position, all subject properties are accorded to the new subject. In Tagalog an ergative agent is not "demoted"; it is still required, and in fact is subject to the most stringent ordering requirements of any argument, always appearing after the verb. It appears to be cross-linguistically the case that the more "central" a constituent to syntactic structure, the more restricted its linear position is, and vice versa, adverbials and other non-essential constituents having the most freedom to move around syntactically; the near total-freedom of such "adjuncts" as adverbs of time and place in English as compared with the strict ordering of subjects and objects exemplifies this. Clearly then the Tagalog ergative agent remains a "core" argument, not a "demoted" argument like the English passive agent. And, as already mentioned, the ergative agent in Tagalog retains some apparent "subject" properties

which are not transferred, as in English, to the new "subject"(topic) created by "advancement."

Bell's Relational Grammar analysis is problematic in that, at least at the time she was writing, this framework explained all "advancements" in terms of what is known as either the Relational Annihilation Law or the Chômeur Law. The essence of this principle is that all advancements will work as in English: during a transformation, a slot in the relational hierarchy (Subject > Direct Object > Indirect Object > Oblique Object) can only be filled by an advancing argument if the previous occupant of the slot is removed from the hierarchy altogether: thus its grammatical relations are "annihilated,"and it becomes a "chômeur" (French = 'unemployed person'). This adequately describes the demotion of English underlying subject during the passive transformation. Clearly this does not happen to non-topical agents in Tagalog (or in Cebuano). These continue to function as core arguments and have the syntactic properties that constitute Bell's arguments for "actor as initial 1"; it appears that they could have none of these properties were "advancement" to "annihilate" their grammatical relations.

I am sure the Relational Grammar analysis could be rescued by a modification of this principle, whereby instead of vacating the premises entirely, initial subjects could be only demoted a rank or two. Yet we might then expect them to take on the grammatical properties of a lesser argument. Rather what we see is that they maintain properties all their own when not in subject/topic position. We could alternatively have an "ergative" model in which they maintain their properties in this position because this in fact is their initial position—but we would then have the opposite issues: why would a patient not be "chômeur" in an "active transformation," and why would an agent maintain its special properties if advanced? I think this whole avenue is likely to be fruitless. The evidence of the "recent past" should be again brought out to assert that no "voice" is neutral or basic in Tagalog; rather all bear the same relationship to some other underlying verbal structure, whose argument structure at least is reflected in "recent past."

If we are to have two levels of subject assignment and a transformational derivation of voice, we should expect all voice to be a transformation on an initially voiceless VP. We would expect the role-related subject assignment to be already

present in the untransformed argument structure of the "recent past"; this would be interesting but I have no evidence for this. "Voice" by such a model would entail the "advancement" to a position—subject/topic—that was not previously filled at all. This would explain the lack of any need to "demote" an argument when a voice transformation applies. Guilfoyle et al. (1992) take precisely such an approach to voice. In their model, not only there are two distinct sets of "subject properties" for Tagalog, there are two corresponding "subject" positions in syntax. In this Government and Binding analysis the two different candidates for "subject"—agent and topic—correspond to parallel positions at different levels of syntactic organization, as spec of VP and of IP respectively, and all sentence-level topics are generated by "raising" from positions within the VP.

I understand and approve of the motivations for this model. I am uncertain about the specifics. The category IP clearly has been suggested by comparison with GB analyses of English. The motivation for this category in English was, I believe, to be able to treat verbs and their inflections as syntactically different constituents; Infl may appear at surface together with the verb or with the auxiliary depending on what transformations have taken place. Guilfoyle et al. do not explicitly state what they see as surface evidence (if any) for Infl in Tagalog. For what they see as parallel structures in Malagasy, voice morphology on the verb is treated not as Infl but as added by a filter to case-mark arguments that remain in the VP. If we assume this is also the case for Tagalog, there is still aspect morphology that may be treated as the surface manifestation of Infl. The clearest reason Guilfoyle et al. use an IP as distinct from VP is purely in order to have two parallel Spec positions at different levels of syntactic organization which can correspond to often distinct agent and topic positions. Guilfoyle et al. obtain two different syntactic positions for two subjects as follows: they posit an underlying SVO word order, with the stipulation that in all sentences, without exception, V must raise to Infl, putting it in a position before underlying subject (Spec of VP in this analysis). The notion of V raising to Infl is well known from English, but since Guilfoyle et al. cannot point to any cases of V not raised to Infl, I find this a little weak.

Overall, however, I find this to be the most satisfying generative account of Tagalog Grammar I have thus far encountered. My concerns with it are still those I raised at the end of section 3.6.3: firstly, the subject/topic in verbal sentences should not

be underlyingly empty and filled only by transformation in verbal sentences any more than they are in copular sentences, or sentences of possession. Especially in the latter, it is unclear what syntactic position the subject/topic could have raised from (although I should do more research before I say this). It might be possible to create models of all these sentences whereby raising is necessary to generate subject/topic, but I am concerned that this may be a needless complication.

The second concern is a bit more philosophical; it may not be an issue, depending on what we mean when we talk about a generative model of grammar. Clearly we do not simply generate sentence structures, we generate the sentence structures that express what we mean, those that are capable of holding the lexical contents of our sentences in the right relationship. I'm not very interested in theories claiming that utterances are produced by first generating all the possible sentences of English (which are of course infinite) and then rejecting those that aren't the right sentence. It seems to me that the intended meaning of the sentence must guide the process of language generation to the desired result. Furthermore, it seems to me that if anything in semantics is fundamental and given in generating a sentence, it is the semantic subject/topic, the information that is the given within discourse on which the sentence builds. In Guilfoyle et al.'s model of generation, the last thing that happens is the selection of some NP out of another structure to become subject/topic. This is understandable, given that while topic and comment are coordinate at the highest level of syntactic organization, the topic also has a role to play within the comment: it must occupy a specific place in the argument structure and often assign features within the comment, which is much easier to model if we assume it actually is present somehow in the comment. On the other hand I feel that the syntactic division in Tagalog between topic and comment corresponds with something semantically very fundamental: topic and comment are coordinate in the crudest semantic division of the sentence as well. I hold that syntax must be generated as a translation of semantics, not an independent sentence factory from which semantics simply takes what it likes, and thus I want the topic-comment structure to come early in the generation of syntax. This is why earlier I proposed that "voice" results not by raising an argument from the VP to become subject/topic, but by eliminating an argument from the VP if that argument has

already been generated as subject/topic, and the addition of morphemes to the VP which indicate what slot in the argument structure is taken by the subject/topic.

This model could accept that in non-active sentences, an actor argument is still syntactically privileged within the VP. Tagalog would then become in a way "accusative," but only with the qualification that "accusativity" is a feature only of the VP, not of the sentence as a whole, and that there exists a higher level of syntactic organization which is organized entirely by discourse structure, making issues of ergativity versus accusativity irrelevant at this higher level.

4. Syntactic Typology Revisited

After the preceding discussion we are left with the conclusions that semantic argument structure and pragmatic discourse structure both act to privilege certain NPs syntactically in Tagalog, but that these processes work at different levels of syntactic organization. We know that "role-related subject properties" belong to a particular argument—agent, at least prototypically—and that the assignment of these properties happens in the VP. On the other hand, "reference-related subject properties" belong to the subject/topic, which is only found at a higher level of syntactic organization, according to Guilfoyle et al. (1992) the IP. We do not know for certain which of these happens earlier or is more "basic," but the author personally feels that role-related subject assignment may be in a sense arbitrary, whereas reference related subject assignment is not, or rather that there may not even be reference-related subject assignment in that the reference-related subject, as "most referential" NP, is already topical when the generation begins, serving as the semantic point of departure for the sentence.

Also, we know that semantic argument structure and pragmatic discourse structure are also the factors in determining what NP plays the subject role in an English sentence. Yet syntactically at least we do not think of these two factors as determining two different subjects in English. We are urged to accept by Guilfoyle et al. that the two levels on which the two different subjects are privileged are the VP and IP, also found in English. The question for syntactic typology is then, why do the same

initial factors and parallel syntactic structures produce different results in Tagalog and English? Guilfoyle et al. have done a very convincing job of arguing that the clear differences between the grammars of English and Tagalog *need not* be primarily differences in syntax proper at all. Not counting variation in left-right branching, they use the same syntactic tree structures to generate English and Tagalog. They also uses similar raising rules. All that differs between the two languages in their model are the semantic and pragmatic conditions under which certain raising rules apply.

I do not mean to simplify and imply that syntactic typology is only a matter of the conditions that call for the application of identical syntactic transformations. For another kind of difference that is simple enough to be stated concisely, recall the clear distinction made both in Acehnese and Tagalog between volitional actors and experiencers. Acehnese made this difference, depending one one's analysis, into either one of verbal morphological class or a difference of syntactic structures, though not at the highest, discourse-driven level which includes topics. Tagalog clearly made this a difference of verbal morphological class, and thereby also allowed these kinds of arguments different slots in verbal argument structure. Neither English nor Spanish, so far as I know, does not make this distinction a matter of grammatical rules, such that (according to Sweetser) traditional-(read European-)minded grammarians have preferred to distort the morphological categories of Tagalog rather than recognizing a clear semantic distinction manifested in linguistic rules.

There must be all kinds of different problems in syntactic typology, but I feel I should try to make a broad statement of these. Linguists have gotten better and better in showing cross-linguistic parallels in syntactic organization and syntactic rules in the strictest sense. However these issues are not usually what "syntactic typology" is about. Broadly, I assert that "syntactic typology" is really usually about the procedures a language uses to map lexical items into syntactic positions based on their thematic roles and discourse roles; occasionally it attempts to capture the results of the combination of these widely differing algorithms with comparatively trivially differing syntactic structures. By 'algorithm' is meant a formula for a procedure consisting of an ordered set of rules. My contention is that what syntactic typological categories latch on to are really just side effects of the complex algorithms mapping thematic and discourse roles into syntax, and we should not be too convinced that similar side effects mean similar

algorithms. Of course at this stage in our understanding we might like to call different languages' algorithms "similar" purely because they produce similar side effects. But if we are able to set out the algorithms in detail, then we will have a much more nuanced and theoretically useful view of their similarities and differences.

As an example let us take ergativity. Ergativity is an issue for syntacticians and for syntactic typologists. But really "ergativity" need not necessarily refer to a phenomenon of syntactic structure at all, and can never refer to a purely syntactic (as distinct from semantic) phenomenon. First of all, it is well known that many languages are morphologically ergative while revealing an underlyingly accusative syntactic structure, possibly through word order or through the operations of syntactic transformations According to Anderson (1976) Basque is such a language.

In her survey of ergative features in North Indian languages, Klaiman does not address syntactic transformations or ordering phenomena at all. In fact, even for those languages in her survey which, like Hindi, show the clearest relationship between case marking and verbal agreement in the perfective "ergative domain," in terms of word ordering ergative agent tends to have the same position as the subject of a non-perfective nominative/accusative sentence, and likewise a theme argument tends to have the same position whether "nominative" (unmarked and triggering agreement), "accusative" (or "dative" (marked as "indirect object" because definite or animate, and not triggering agreement).

(24a) lark-a roti kha:t- a
boy-nom bread eat- pres-sg masculine
'The boy eats (the) bread'

(24b) lark-a roti- ko kha:t- a
boy-nom bread-IOM eat- pres-sg masculine
'The boy eats the bread'

(24c) lark-e- ne roti kha:i
boy-obl-erg bread eat- sg feminine
'The boy ate (the) bread'

(24d) lark-e- ne roti- ko kha:-ya
boy-obl-erg bread-IOM eat- sg masculine
'The boy ate the bread'

In other words actual word ordering would be more closely tied to theta-roles than to either of the morphological phenomena addressed by Klaiman, case marking and verbal agreement. If—though I make no definite claims about this—only one of these

can count as "the syntax," it should probably be the word order, as the other two can be considered "the morphology." So we might, for these reasons, call even the most "prototypically ergative" of Klaiman's languages not ergative from a syntactic point of view. (This is my take on the "most prototypically ergative" language, based on Comrie's (1989) definition; it is quite different from Klaiman's.)

By contrast, let us presuppose a completely ergative language according to a simplistic application of Comrie's definition. In such a language there would be a syntactic "subject" position, used for any single argument of an intransitive verb, but (in stark contrast to English subject) never used for agent of a verb with multiple arguments but rather for another core argument. In this theoretical most prototypically ergative of languages, the prototypically non-agent subject position would be uncontroversially "subject" from all applicable points of view, e.g. case marking, verbal agreement, syntactic ordering, syntactic transformations such as subject-to-object raising and equi-subject deletion, as well as such interactions of semantics and syntax as control of reflexivation and address of imperatives.

In such a case, there would be nothing *syntactic* whatever to distinguish such an ergative "subject" from the uncontroversial "subject" of a highly "accusative" language such as English. "Ergativity," for such a language, would be *only* a part of the algorithm for assigning syntactic positions to NP arguments based on their thematic roles. "Ergativity" for such a language would be a much smaller, simpler, and less significant part of the grammar than in those languages which seem to mix canonically "ergative" and "accusative" features.

(Since this kind of prototypical ergativity seems to yield no syntactic distinction from accusativity, it might not be unreasonable to propose another, nearly opposite definition of ergativity: this would be one of inconsistency within morphosyntax as to which argument is subjectlike, as both accusative languages and our hypothetical most ergative of languages both have no inconsistency, while most ergative languages have inconsistency. Yet given the number of competing usages of "ergativity" already in existence, if one were going to introduce such a measure one should probably call it something else.)

Unlike ergativity, Greenberg's typologies do to some extent reflect the idiosyncracies of actual syntactic structure. More precisely, they reflect a combination of

variations in the ordering of syntactic structure with various variations in mapping thematic role structure to syntactic structure, to wit the kinds of issues that play into what becomes a "subject" and what becomes an "object," assuming such categories to be unproblematic. This is what I mean by syntactic "types" being emergent from semantics-and-discourse-to-syntax algorithms.

In the next section, I will take a stab at the kind of syntactic comparison I advocate, which is not really strictly syntactic, but rather compares algorithms for mapping thematic as well as discourse roles into syntactic positions. I should warn the reader that as my own approach is undeveloped insofar as it differs from Guilfoyle et al.'s, and insofar as they produce identical results, I will assume Guilfoyle et al.'s model for generating Tagalog sentences.

5. Conclusion—Comparison between English and Tagalog

The model proposed for Tagalog D-structure and voice transformations by Guilfoyle et al. is much more similar to English than the surface facts of the language might suggest. The crucial difference is that Tagalog has the option to transform or not to transform, while English has an obligatory initial subject assignment rule which always raises a role-related subject to the highest-level subject in syntax. In English AGENT, as the highest member in a subject hierarchy, must always raise to subject if present (and it is often required by the argument structure of a verb to be present at deep structure); if it is not found in the argument structure, the subject slot goes to PERCEIVER/EXPERIENCER, then INSTRUMENT, then perhaps THEME. Examples (25a-e) illustrate this rule:

(25a) I opened the door (with a key)

(25b) *Opened me the door etc. (Initial subject assignment cannot fail to apply; of course we can't be sure this is what we'd get if it did)

(25c) *The door opened me with a key etc. (Doors cannot be AGENTS)

(25d) The key opened the door (INSTRUMENT can be subject if agent is omitted)

(25e) *Opened the door with the key

(INSTRUMENT cannot fail to raise if it is present and agent is not)

Only after the operation of this rule does English have an option, that of whether to "demote" agent and have a further transformation. For English, passivization really does work like the Relational Grammar model of "advancements" in that raising of another argument to subject position is dependent on the demotion of all arguments higher in the hierarchy. Passivization does occur as a result of discourse pressures, but I believe it requires stronger discourse pressures to force initial subject demotion and passive transformation in English than to trigger any of the voice transformations in Tagalog. Specifically passivization occurs when the discourse referentiality of the advancing argument is sufficiently higher than that of the initial subject that to not passivize would cause startling discontinuity at the beginning of the sentence. This statement may cover cases of unspecified agents, as an unspecified agent is lowest in the discourse referentiality hierarchy. In Tagalog, on the other hand, since there is no initial subject in subject/topic position that must be dislodged, much slighter discourse pressures will result in a voice transformation.

In this analysis of English passive, for the ditransitive verb 'give,' only AGENT can be subject if no arguments are demoted (i.e.. either deleted or exiled to a prepositional phrase). There can be no instance of AGENT that has not been removed from the VP by either raising or raising followed by demotion: unlike other arguments, AGENT is never present at surface structure within VP, which being an argument of V is probably how it is generated.

(26a) John gave me the book

(26b) *Gave John me the book

If actor is demoted the next argument in the hierarchy, BENEFICIARY (or RECIPIENT or GOAL or whatever you want to call it) can raise; this is accompanied by passivization of the verb.

(27) I was given the book (by John)

Only if BENEFICIARY, the second-ranked argument in the hierarchy, is also demoted can theme raise to subject:

(28) The book was given to me by John

In Tagalog, according to Guilfoyle's model, subject (topic) assignment is a parallel raising process that extracts arguments from the VP. However, this process is not constrained by a rigid hierarchy as in English; Tagalog has no initial requirement that a given argument *must* raise, as AGENT must in English. Consequently it is not necessary for any argument to be demoted for another argument to raise to subject. This fact of Tagalog can now be seen not as an anomaly to be explained, requiring a more complex theory for Tagalog than for English. Linguists should no longer be surprised to find agents in Tagalog "ergative" constructions that remain core, not demoted, arguments. The phenomena can now simply be regarded as an absence of the kind of subject assignment hierarchy we have found in English. Tagalog "ergative" agents are thus simply those that remain in their original D-structure position. There is still a question I will leave unresolved of whether English generates any argument at Spec of VP, as Guilfoyle et al. claim Tagalog generates agent. If it does, it probably generates AGENT here, then raises it obligatorily. Tagalog (in Guilfoyle's model) would then be completely parallel apart from the absence of the obligatory agent-raising rule.

Another difference falling out of this slight difference in rules has to do with the surface visibility of "categorical" subject/topics. In the ideal case of an English "categorical" sentence, (local) discourse topic is also syntactic subject, however it became so. The semantic scope of negation and questions is then the syntactic predicate.

(29) My name is Pete. I like golf. I'm not married. I really like golf. Am I boring?

In all but the first sentence of (28), 'I' is clearly categorical subject/topic. A little thought will confirm that the semantic scope of negation, emphasis, and question in the last three sentences is the syntactic predicate, as we predicted. However it is possible for the obligatory subject raising rule prevents a categorical subject/topic from appearing in syntactic subject position.

(30) Well, I don't like golf. Do you like golf? Pete really likes golf more than me.

In these sentences, liking golf is the old information continued from earlier in the discourse (or just 'golf,' if we confine ourselves to relative topicality of NPs). However, in all these sentences, obligatory role-related subject raising has put an argument I believe to be PERCEIVER/EXPERIENCER into subject position for all these. For these

sentences it is actually the argument raised to subject that is in discourse terms the "comment," and it is over the syntactic subject that negation, questioning and emphasis have scope. It is because of the obligatory nature of subject assignment in English and its strictly following a theta-role hierarchy that we get identical forms for "thetic" sentences and actor-topic "categorical" statements in English; nevertheless semanticists notice the difference. Because Tagalog doesn't have the obligatorily initial subject assignment rule, or at least not at the highest level of syntactic organization which includes subject/topic, the two types of statements in Tagalog take on distinct syntactic structures: only categorical subject/topics appear in subject/topic position. Actually, in Guilfoyle's analysis, there must be an initial subject assignment rule in order to account for the syntactic properties of Schachter's 'role-related subject,' but since this only places an initial subject in Spec of VP position, leaving Spec of IP available, the point still holds.

This is the most concise statement I have been able to formulate of the difference between the rules governing subject assignment in Tagalog and English. By this statements, "subject" can hardly be said to mean one and the same thing for the two languages. In Tagalog we have not been able to avoid seeing two levels of 'subject' at two levels of syntactic organization. We could not follow the relational grammarian's lead in seeing this as equivalent to English active and passive. Passive is a level ordered after active in derivation, but it annihilates the grammatical relations of an initial subject. In Tagalog both levels of organization are simultaneously present and both 'subjects' simultaneously retain different syntactic privileges. English subject can be seen as fusing the two properties of the two Tagalog subjects, usually bound by semantics to one, but sometimes accorded to the other either by coincidence of the two or by unusually strong discourse pressures.

Bearing in mind Comrie's overview of typology, we may want to look at the same data again from the other direction, and point out what it shows English and Tagalog have in common. According to my premises, initially both languages work from two linked structures, a semantic event structure and a discourse structure of the same event. The ways in which these levels are translated into linguistic structure can be laid out as remarkably parallel: we have first an operation of 'initial subject assignment,' wherein an argument of the verb is selected according to a hierarchy of theta-roles and placed in a privileged position. The cross-linguistic correspondence of this hierarchy

needs to be examined: for both languages agent is clearly highest in the hierarchy, and I suspect that universal semantic criteria may be established for the operation of these criteria. On the other hand I don't suspect the hierarchy is entirely universal: we have already seen that Tagalog and Acehnese distinguish agent from perceiver/experiencer, while I don't know that English makes this distinction grammatically.

In English, this initial stage of subject assignment raises its subject to the highest level of syntactic organization within a sentence, and thus generally determines the final subject assignment already. In Tagalog we only see the residue of this level of derivation in what Schachter calls the "role-related 'subject' properties" of agents in Tagalog sentences. There 'initial subject' very often does remain in the middle level of syntactic organization where this operation places it, however its behavior there is largely docile and unremarkable.

Second, both languages optionally permit a second level of subject assignment, one that selects a topical subject from the existing structure based on pragmatic discourse information. This operation is highly restricted in English, at least partly as a result of the different operation of the initial subject assignment algorithm. English can assign a new subject only in the face of discourse pressures great enough to warrant the 'demotion' of an initial subject. In Tagalog this constraint does not exist, the first level of subject assignment still leaving an empty spot at the highest level of syntactic organization; it seems essentially any argument can climb to this position by virtue of a slight advantage in discourse referentiality. There is a puzzle at this juncture, which is why only an argument that is so to speak 'next in line' for subject raises under discourse pressures in English (we saw on p.64 that theme becomes a passive subject of 'give' only if both agent and beneficiary have been demoted), while numerous arguments can raise from the Tagalog VP. It may be that the theta-role hierarchy, less powerful for Tagalog than English in the initial subject assignment phase, is less powerful here as well.

Apparently in Tagalog every sentence that has a "topic/subject" has gone through stage 2, and we can reiterate the point that neither "active" or "passive" structures are more "basic" than the other. We have found that the debate on "ergativity" and "accusativity" has meaning for Tagalog at the level of the VP (where it is ergative), but that the process to which this is relevant is far less powerful than the

parallel operation in the derivation of English. We need to be able to have schemes of syntactic typology that recognize such nuances, which we cannot do if we try to work as closely as possible from surface features of language. I feel that a better approach to syntactic comparison will work in the opposite direction, as I have rudimentarily attempted to do in this final section. We should start neither with surface forms nor with simple abstract syntactic structures; rather we must begin with the semantic and discourse structures of events, which we recognize form the pre-linguistic basis of the linguistic expression of those events. As much as possible, we should try to follow as closely as possible the generative progress from start to finish, and always see semantic and discourse structures as guiding the generation of linguistic structure. If we stick to an approach that tries to answer questions in trying to penetrate as little as possible beyond surface forms, we are not only working in the wrong direction to correctly model the generation of language, we may be attempting to answer unimportant or spurious questions and we will never know it. Furthermore, it is likely that we would miss the kind of significant parallels as well as significant divergences in generative algorithms that I hope I have been able to bring out in this conclusion.

Anderson, Stephen (1976) "On the Notion of Subject in Ergative Languages." In Charles N. Li, ed., *Subject and Topic* (Symposium on subject and topic held at the University of California, Santa Barbara, March 1975) New York: Academic Press.

Bell, Sarah J. (1983) "Advancements and Ascensions in Cebuano." In David M. Perlmutter, ed., *Studies in Relational Grammar 1*. University of Chicago Press.

Bobaljik, Jonathan D. (2000) (Unpublished lecture notes for an introductory class in morphology)

Cartier, Alice (1979) "De-voiced transitive Verb Sentences in Formal Indonesian." In Frans Plank, ed., *Ergativity: Towards a Theory of Grammatical Relations*. New York: Academic Press.

Cartier, Alice (1985) "Discourse Analysis of Ergative and Non-Ergative Sentences in Formal Indonesian." In Frans Plank, ed., *Relational Typology. Trends in Linguistics Studies and Monographs 28*. Berlin/New York/Amsterdam: Mouton de Gruyter.

Chung, Sandra (1976) "On the Subject of Two Passives in Indonesian." In Charles N. Li, ed., *Subject and Topic* (Symposium on subject and topic held at the University of California, Santa Barbara, March 1975) New York: Academic Press.

Comrie, Bernard. (1989) *Language Universals and Linguistic Typology* (2nd ed.) University of Chicago Press.

Cumming, Susanna and Fay Wouk (1988) "Is There 'Discourse Ergativity' in Austronesian Languages?" In R. M. W Dixon, ed., *Lingua 71: Studies in Ergativity*. Amsterdam: Elsevier Science Publishers (North-Holland)

Dixon, R. M. W. (1988) "Introduction" in R. M. W Dixon, ed., *Lingua 71: Studies in Ergativity*. Amsterdam: Elsevier Science Publishers (North-Holland)

Durie, Mark (1988a) "The So-Called Passive of Acehnese." *Language*, Vol. 64, No. 1.

Durie, Mark (1988b) "Preferred Argument Structure in an Active Language: Arguments Against the Category 'Intransitive Subject.'" in *Lingua 74*. Amsterdam: Elsevier Science Publishers (North-Holland)

Gibson, Jeanne D. And Stanley Starosta "Ergativity East and West." In Philip Baldi, ed., *Linguistic Change and Reconstruction Methodology. Trends in Linguistics Studies and Monographs 28*. Berlin/New York/Amsterdam: Mouton de Gruyter.

Greenberg, Joseph H. (1966) "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In J. H. Greenberg, ed., *Universals of Language* (2nd ed.). Cambridge, Mass: MIT Press.

Guilfoyle, Eithne, Henrietta Hung, and Lisa Travis (1992) "Spec of IP and Spec of VP: Two Subjects in Austronesian Languages." *Natural Language and Linguistic Theory* V.10 No.3, August 1992. Dordrecht, Netherlands: Kluwer Academic Publishers.

Hawkins, John A. (1983) *Word Order Universals*. Series: Quantitative analyses of linguistic structure. New York: Academic Press.

Hopper, Paul J. (1983) "Ergative, Passive, and Active in Malay Narrative." In Flora Klein-Andreu, ed., *Discourse Perspectives in Syntax*. New York: Academic Press.

Klaiman, Miriam H. (1987) "Mechanisms of Ergativity in South Asia." In R. M. W Dixon, ed., *Lingua 71: Studies in Ergativity*. Amsterdam: Elsevier Science Publishers (North-Holland).

Lawler, John M. (1977) "A Agrees with B in Acehnese: A Problem for Relational Grammar." In Peter Cole and Jerrold M. Sadock eds., *Syntax and Semantics Volume 8: Grammatical Relations*. New York: Academic Press.

Matthews, David (1984) *A Course in Nepali*. University of London School of Oriental and African Studies. New Delhi: Heritage Publishers.

Myhill, John (1988) "Nominal Agent Incorporation in Indonesian." *Journal of Linguistics* 24, p.111-136. London: Cambridge University Press (published for the Linguistics Association of Great Britain)

Schachter, Paul (1976) "The Subject in Philippine Languages: Topic, Actor, Actor-Topic, or None of the Above." In Charles N. Li, ed., *Subject and Topic* (Symposium on subject and topic held at the University of California, Santa Barbara, March 1975) New York: Academic Press.

Schachter, Paul (1977) "Reference-Related and Role-Related Properties of Subjects." In Peter Cole and Jerrold M. Sadock eds., *Syntax and Semantics Volume 8: Grammatical Relations*. New York: Academic Press.

Siewierska, Anna (1988) *Word Order Rules*. Croom Helm linguistics series. Worcester(UK): Billing & Sons Limited and New York: Methuen.

Sweetser, Eve E. 1980. "Tagalog Subjecthood Reexamined." In Jody Kreiman and Almerindo E. Ojeda eds., *Papers From the Sixteenth Regional Meeting of the Chicago Linguistic Society April 17-18, 1980*. Chicago Linguistic Society, University of Chicago.