1. Escaping Eurocentricity*

The concern with word classes, parts of speech, or, as they are referred to in this paper, syntactic categories, dates back to antiquity— for better and for worse. For better, since in linguistics, as in any other discipline, one sees further when standing on the shoulders of giants. But for worse, if it is the case that the giants themselves are standing in the wrong place.

Or, in the case at hand, in the wrong continent. My own interest in syntactic categories derives from ongoing attempts to obtain a better understanding of the major syntactic patterns of some languages whose syntactic structures appear to be very different from those of the classical languages of antiquity, and the well-known and well-studied languages of Europe. Increasingly, these efforts suggest that contemporary theories and frameworks do not provide the appropriate tools for a satisfactory description of such "exotic" languages. In general, available theories are of European origin, reflecting the peculiar properties of the particular European languages familiar to their progenitors. Often, their application to languages spoken in other parts of the world is an exercise in Eurocentricity, involving the unwarranted imposition of categories and structures that are simply irrelevant.

In the past, grammar books of English informed us that English nouns have six cases, which was what prompted Alice, in her adventures in Wonderland, to muse: "A mouse—of a mouse—to a mouse— a mouse—0 mouse!". Then it was Latin grammar that was being imposed on English; now it is English grammar that is being imposed on the rest of the world's languages, through theories based largely on English data, constructed for the most part by English-speaking linguists, and disseminated almost invariably in the world language of science English. As an illustration, consider the following garden-variety sentence in Tagalog:

(1)  

\[
\begin{array}{ccc}
\text{Manok} & \text{ang} & \text{kumakain}. \\
\text{chicken} & \text{TOP} & \text{PROGR-ACT.TOP.REAL-eat} \\
\end{array}
\]

'The chicken is eating.'
Since *manok* is the name of a thing, it is generally assumed to be a noun (or NP); similarly, since *kumakain* describes an activity and is inflected for voice and aspect, it is usually taken to be a verb (or *VP*). However, in the above sentence *manok* occurs in the sentence-initial position characteristic of predicates, while *kumakain* occurs in construction with the subject, pivot or topic marker *ang*. Accordingly, in their analysis of the above sentence, traditional descriptions of Tagalog characterise *manok* as a "verbalised or predicative nominal" and *ang kumakain* as a "nominalised verb". The problem, however, is that in Tagalog, just about any word (or phrase) can go anywhere; accordingly, if one's *theory* comes with a built-in distinction between nouns (or NPs) and verbs (or VPs), then almost *every* word (or phrase) in the language is going to end up being both a noun (or NP) and a verb (or *VP*). Or at best, having undergone a process of zero-conversion from one to the other. Which bears an uncomfortable resemblance to the six cases of Alice and her mouse.

A more perspicuous description of Tagalog syntax might acknowledge the fact that there is, quite simply, no viable distinction between nouns (or NPs) and verbs (or VPs). Rather, there is just a single open-class syntactic category, which contains almost all the words and phrases in the language: some arguments in support of this claim can be found in Gil (1993a, 1993b, 1995).2

Thus, cross-linguistic variation with respect to syntactic categories poses a serious challenge to conventional approaches to syntactic *theory*, underscoring the need to escape from the straightjacket of traditional Eurocentricity. However, breaking away from pre-conceived notions entails a reorientation of one's basic mode of investigation. Rather than asking "Is this form an X or a Y? (e.g. a noun or a verb, a NP or a *VP*, and so forth)", the more appropriate question to ask is "What are the significant syntactic patterns in the language, and what are the categories that must be posited in order to enable the necessary generalisations to be stated?". In doing so, the balance is tilted away from a deductive top-down mode of investigation and towards a more inductive bottom-up approach.

However, the bottom-up way of doing things confronts the researcher with a labelling dilemma. For example, no sooner does one arrive at the conclusion that Tagalog possesses a single open-class syntactic *category* than the question arises: "What to call it?". My first answer, in a gesture of anti-Eurocentricity, was to give it a Tagalog name, *pariral*, which means 'phrase' (Gil 1993a). But such an answer is unsatisfactory for two reasons. First, because it is just an arbitrary label that does not provide any information about the nature of the category in question. And secondly, because it does not lend itself to the typological enterprise and the comparison of syntactic categories across languages. lienee, what is called for instead is a principled and more revealing choice of terminology.

The issue, though, is not just one of labelling. Terms only have meaning within theories. In order to establish an adequate terminology, it is necessary to acknowl-
edge that every description, however inductive and bottom-up its orientation, in effect constitutes a theory about the phenomena under investigation. In other words, just as there can be no theory without description, there can be no description without theory—see Gil (1998, forthcoming b) for further discussion. As for the would-be dichotomy between theory and "mere" description which plays such a central role in the rhetoric of Generative Grammar; this is just a red herring often used to denigrate proponents of theories other than those favoured by the generativists. Accordingly, the quest for an appropriate terminology leads directly to another, broader enterprise, namely the construction of an explicit and well-articulated theory of syntactic categories, based on an adequate range of data spanning the typological diversity exhibited by the languages of the world.

This paper attempts to propose such a theory. Section 2, of a methodological and philosophical bent, sketches some of the basic principles upon which the theory is based. Section 3, of a rather formal nature, presents the outlines of a theory of syntactic categories. And section 4, more typologically oriented, provides some preliminary suggestions as to how the proposed theory may account for observed patterns of variation in the syntactic patterns of different languages.

2. Basic principles

Theories rest on principles. In the course of developing a theory, numerous assumptions, premises and presuppositions are brought to bear, some consciously, others less so. When presenting a theory, confusion and misunderstanding can therefore be minimalised by rendering these principles as explicit as possible—seven if this involves a certain amount of belabouring the obvious.

Following in (A1–4), are four principles which provide the foundation for the theory of syntactic categories presented herein.

A Basic principles:

1. Occam's Razor
A theory positing the existence of fewer entities is preferable to one positing the existence of more.

2. Semioticty
Language is a system of correspondences between sounds and meanings.

3. Formality
Sound-meaning correspondences are not direct, but rather are mediated by various intermediate forms.
4. **Multiplicity of Levels**

The forms of language are structured hierarchically, with distinct principles governing different levels of structure.

Probably most linguists are in agreement with most or all of the above principles. Still, it is worth spelling them out explicitly, if only to show how accepted premises can occasionally lead to unexpected results.

*Occam's Razor* says that simpler is better: a description making recourse to fewer entities is better than one that makes recourse to more. Everybody agrees with it—though it sometimes does not seem that way, when faced, for example, with a Government-and-Binding syntactic tree structure spread out over two pages of a linguistic journal.

*Semioticity* also states the obvious, namely that language uses sounds to express meanings in systematic, rule-governed ways. Again, almost everybody is in basic agreement—notwithstanding some linguistic traditions, such as later American Structuralism, as in Harris (1951), and its immediate successor early Generative Grammar, as represented by Chomsky (1957), which attempted to construct theories that would avoid recourse to meaning.

*Formality* underlies a crucial difference between human language and most other semiotic systems, such as, for example, traffic lights; it is also what makes the study of language so interesting and challenging. Basically, it says that the relationship between sounds and meanings is not direct, in the way that it is for, say, traffic lights, where *red* means 'stop' and *green* means 'go'. Rather, the relationship is mediated by various intervening entities: the linguistic forms—features, suffixes, adjectives, causatives, topics, and so forth—which constitute the basic building blocks of linguistic analysis. This principle is accepted by most or all linguists, though it is downplayed within those schools whose interests lie less in grammatical structures and more in meanings, for example the Columbia school of linguistics, as represented in Contini-Morava–Goldberg (1995).

Multiplicity of Levels characterises the way in which smaller linguistic forms group together to form larger ones. Segments combine to form morphemes, morphemes come together to create words, words combine to produce sentences, and sentences come one after another to result in discourses. Crucially, however, the rules and principles governing the groupings differ in numerous fundamental ways from level to level. It is these differences which underlie the division of the field into distinct disciplines associated with each level: phonology, morphology, syntax, and discourse analysis, a division which is generally accepted, notwithstanding various cross-domain parallels and analogies that have been observed—see Gil (1986, 1987) for prosody, Yip et al. (1987) for tiers, and Levin (1985), Shen (1985), and Anderson–Ewen (1987) for heads and X-bar structure.
Obvious though they may be, the above four principles point towards more specific and perhaps less generally accepted conclusions with regard to the nature of syntactic categories. The principles of Formality and of Multiplicity of Levels entail the Autonomy of Syntax, and its distinctiveness vis à vis semantics and morphology respectively:

B The Autonomy of Syntax

1. The ways in which words group together to form sentences differ fundamentally from the ways in which words and sentences are associated with their meanings.

2. The ways in which words group together to form sentences differ fundamentally from the ways in which morphemes group together to form words.

Formality entails that syntax, as one of the intermediate levels, is distinct from semantics—as specified in (A3). Evidence for this claim is overwhelming; suffice it to mention various arguments showing that the truth value of a sentence is undecidable within First Order Predicate Calculus, for example Hintikka (1979a, 1979b), Gil (1982b). Or, for those for whom mathematics speaks less strongly than inductive cross-linguistic generalisations, consider the virtually infinite number of potential semantic categories which have no effect on the syntax of any known language. For example, whereas clause structure is often affected by the animacy of one of the participants, no case is known of it being affected by whether one of the participants happens to be purple, or plastic, or perforated. Thus, syntax is fundamentally different from semantics: it is invariably much, much more frugal. The need to separate syntax from semantics is accepted by most schools of linguistics. However, in some approaches, the existence of an autonomous level of syntax is downplayed, as for example in the "West Coast" Functional school, as represented by Gi.6n (1979) and others; or even expressly denied, as for example in Generative Semantics, as represented by Bach–Hanns (1968)—cf. McCawley's (1977) title "The nonexistence of syntactic categories". In accordance with such approaches, syntactic structures may be viewed as being isomorphic to a restricted subset of all conceivable semantic structures. However, as argued below, there is ample reason to believe that there exist syntactic structures and categories that are completely devoid of any semantic motivation.

Multiplicity of Levels entails that syntax is distinct from morphology— as spelt out in (A2). Again, this is clear mathematically, given that syntax is recursive, whereas morphology, at least in languages that are not of the polysynthetic type, is non-recursive (with the exception of generally marginal constructions such as anti-missile-missile, anti-anti-missile-missile-missile, etc.). And it is also clear cross-linguistically, as vividly illustrated by the following example. Imagine a language in which a basic clause consists of subject plus predicate, but with the following
proviso: the subject must be exactly three words long, and the predicate is strung out before the subject, after it, but most importantly within it, between the first and second word, and between the second and third word. For example, using English words, and marking the subject in boldface: *Sat the on/ at the cat mat*. Clearly impossible syntax; but now what about a language in which a basic word consists of root plus inflection, where the root must be exactly three consonants long, and the inflection is strung out before the root, after it, and also within it, between the first and second consonant, and between the second and third consonant? This is an only somewhat idealised version of the standard word structure of Semitic languages such as Hebrew and Arabic. This and innumerable other examples support the generally accepted separation between syntax and morphology. Occasionally, models have been proposed which do away with this distinction, as for example Chomsky's (1965) "Aspects" framework, in which the terminal nodes of syntactic trees are filled by morphemes rather than words; however, even within Generative Grammar, the need to distinguish between morphological and syntactic structures soon reasserted itself, as argued for by Anderson (1982) and others.

Thus, in accordance with (A3), there is good reason to keep syntax apart from semantics on the one hand and from morphology on the other. These observations accordingly set the stage for the following Typology of Categories, characterising them with respect to the levels of structure in terms of which they are defined:

**C  A Typology of Categories**

1. *pure*
   a. semantic
   b. morphological
   c. syntactic

2. *mixed*
   a. semantic-morphological
   b. semantic-syntactic
   c. morphological-syntactic
   d. semantic-morphological-syntactic

The Typology of Categories in (C) above distinguishes between pure categories, defined in terms of a single level of structure, semantic, morphological or syntactic, and mixed categories, defined in terms of more than one level of structure. In Figure 1, each of the above types of categories is exemplified, with reference to a fixed set of eight English words:
Pure categories are of three types: in Figure 1, pure categories are represented by the areas enclosed by each of the three ovals. A semantic category is one which is defined exclusively in terms of meaning properties. An example of a semantic category in English is the category of words denoting dynamic activities, which, of the eight words in question, contains *speech, oversight, listen* and *overfeed* but not *cat, overseas, overhang* and *know*. A morphological category is one which is defined solely with reference to morphological properties, pertaining to the ways in which morphemes relate to each other and group together to form words. An example of a morphological category in English is the category of words that contain the prefix *over-* , which includes *oversight, overseas, overfeed* and *overhang* , but not *cat, speech, listen* and *know*. Finally, a syntactic category is one which is defined entirely in terms of syntactic properties, governing the ways in which words relate to each other and group together to form sentences; such properties include distributional privileges and participation in syntactic relations such as binding, government, agreement, and the like. An example of a syntactic category in English is the category of words which may occur in construction with *con, may, will* and *shall* ; this category contains *listen, overfeed, overhang* and *know*, but not *cat, speech, oversight* and *overseas*.

Mixed categories come in four types: in Figure 1 mixed categories are represented by four areas of intersection defined by the three ovals. A semantic-morphological
category is one which is defined with respect to semantic and morphological properties; for example, in English, the category of words which denote dynamic activities and contain the prefix over-, which includes \textit{oversight} and \textit{overfeed} but not \textit{cat}, \textit{speech}, \textit{overseas}, \textit{listen}, \textit{overhang} and \textit{know}. A semantic-syntactic category is one which is defined in terms of semantic and syntactic properties; for example, in English, the category of words which denote dynamic activities and may occur in construction with \textit{can}, \textit{may}, \textit{will} and \textit{shall}, which contains \textit{listen} and \textit{overfeed} but not \textit{cat}, \textit{speech}, \textit{oversight}, \textit{overseas}, \textit{overfeed}, \textit{overhang} and \textit{know}. A morphological-syntactic category is one which is defined with reference to morphological and syntactic properties; for example, in English, the category of words which contain the prefix over- and may occur in construction with \textit{can}, \textit{may}, \textit{will} and \textit{shall}, which includes \textit{overfeed} and \textit{overhang} but not \textit{cat}, \textit{speech}, \textit{oversight}, \textit{overseas}, \textit{listen} and \textit{know}. Finally, a semantic-morphological-syntactic category is one which is defined in terms of semantic, morphological and syntactic properties; for example, in English, the category of words which denote dynamic activities, contain the prefix over-, and may occur in construction with \textit{can}, \textit{may}, \textit{will} and \textit{shall}, which contains \textit{overfeed} but not \textit{cat}, \textit{speech}, \textit{oversight}, \textit{overseas}, \textit{listen}, \textit{overhang} and \textit{know}.

In the above examples, the semantic, morphological and syntactic properties are independent, each and every combination thereof defining a category with different members. In other instances, however, properties associated with different levels may define categories with the exact same membership. Two particular cases are of relevance here.

In the first case, semantic and syntactic categories may be coextensive. For example, in Gennan, the semantic category of expressions denoting natural numbers is coextensive with the syntactic category of expressions which may occur in construction with the distributive marker \textit{je}, containing \textit{eins}, \textit{zwei}, \textit{drei}, and so forth (see Link 1983, 1986a, 1986b for discussion of this construction). How then should this category be described, semantically or syntactically? Occam’s Razor, in conjunction with the principle of Semioticiry, provides the answer:

D \textit{Semantics before Syntax}

A theory positing the existence of semantic categories is preferable to one positing the existence of syntactic categories.

Since language is a system of sound-meaning correspondences, meanings are an inherent part of language. Hence, positing semantic categories does not add to the ontological complexity of the description. In contrast, positing linguistic forms, such as syntactic categories, increases the overall complexity of the description—it comes at an extra cost. Thus, when a semantic category turns out to be coextensive with a syntactic one, as in the above Gennan example, the category should be characterised
as semantic, not syntactic. In general, when constructing grammatical descriptions, a semantic analysis should always be attempted as the default; a syntactic analysis ought only to be adopted as a last recourse, when semantics fails.

In the second case, morphological and syntactic categories may be coextensive. For example, in Hebrew, the morphological category of infinitive is coextensive with the syntactic category of words which may head a phrase occurring in construction with a set of words including fasui 'may' and falul 'may', Again the question arises how to describe such a category; and again Occam's Razor points to a solution:

E  *Morphology before Syntax*

A theory positing the existence of morphological categories is preferable to one positing the existence of syntactic categories.

Unlike the preceding case, here both morphological and syntactic categories involve linguistic forms that are intermediate between sounds and meanings. In general, however, morphological categories are much more transparent than syntactic categories: they are independently motivated by readily observable morphological structures and paradigms. For example, the Hebrew infinitive constitutes an easily recognisable inflectional paradigm, consisting of the prefix ח plus a special non-finite form occurring in five out of the seven characteristic verb patterns (or "binyanim"). This paradigm is represented in Figure 2.

<table>
<thead>
<tr>
<th>verb</th>
<th>pa 'o1</th>
<th>nif'of</th>
<th>pi 'e/</th>
<th>hif'if</th>
<th>hUpo 'et</th>
</tr>
</thead>
<tbody>
<tr>
<td>infinitive</td>
<td>liCCoC</td>
<td>lehiCaCcC</td>
<td>leCaCeC</td>
<td>lehaCCiC</td>
<td>lehitCaCeC</td>
</tr>
</tbody>
</table>

*Figure 2, The Hebrew infinitival inflectional paradigm*

Thus, the morphological category of infinitive is needed anyway, to account for the above paradigm. Considerations of parsimony suggest that there is no reason to duplicate it with a coextensive syntactic category in order to account for the distribution of words in the infinitival construction.

So what, then, are syntactic categories like? Emerging from the preceding discussion are the two basic properties specified below:

F  *Syntactic Categories: Basic Properties*

1. Syntactic categories are defined exclusively in terms of syntactic properties,
2. Syntactic categories consist of words and of larger constituents.

In accordance with (F), syntactic categories are defined solely in terms of syntactic properties, such as distributional privileges and participation in syntactic relations such as binding, government and agreement: semantics is irrelevant, as is morphol-
ogy. And as specified in (F2), syntactic categories contain words, and also multi-word phrases, but they do not contain units smaller than the word—that is the realm not of syntax but of morphology.

It is not difficult to see how a large number of the syntactic categories posited within current theories of syntax fail to meet the above criteria. To cite just a few examples from one recent version of Generative Grammar, that represented by Stowell (1981), Chomsky (1986a), Abney (1987) and others: Categories such as Determiner and Quantifier, together with their phrasal projections DP and QP, are ruled out by (F1), since, as their names suggest, they are defined semantically. Similarly, categories such as Inflection and IP are ruled out by (F1), since syntactic categories cannot be defined morphologically. Moreover, to the extent that such categories contain elements that are smaller than single words, they are also ruled out by (F2). (Indeed, the very term "Inflection Phrase" is an oxymoron, suggesting an entity that is simultaneously smaller than a word and also larger.) Finally, categories such as Number and Tense and their projections NumP and TP are doubly bad, since these are defined in terms of semantic features and are manifest morphologically, thereby violating both (F1) and possibly also (F2).

Alongside the two basic properties in (F1-2) above, syntactic categories also exhibit the following additional properties:

**F  Syntactic Categories: Additional Properties**

3. Syntactic category membership is defined in terms of prototypes.

4. Syntactic categories exhibit different degrees of productivity.

In the preceding discussion it was assumed, for ease of exposition, that syntactic categories are unstructured sets—cf. the Venn Diagram in Figure 1 above. However, this is an idealisation: in actual fact, syntactic categories, like most other categories of a cognitive nature, are endowed with prototypical structure. In other words, whereas some elements are characteristic, clear cut, or prototypical members of a given category, other elements may be borderline, idiosyncratic, or less prototypical members of the same category. The prototypical nature of syntactic categories has been argued for by several scholars, including Ross (1973), Dixon (1977), Comrie (1981), Hopper-Thompson (1984), Croft (1991) and Taylor (1995). Finally, not all syntactic categories are of equal size. Open syntactic categories are productive, and contain a large, sometimes infinite number of members, whereas closed syntactic categories are non-productive, generally consisting of a small number of members.

This, then, in brief, is what syntactic categories are like. But what actually are the syntactic categories that different languages may have? This is the question that is addressed in the next section.
3. A theory of *syntactic* categories

The theory proposed herein is a more *elaborated* version of the framework first outlined in Gil (1993b, 1995).

The theory falls within the tradition of Caregorial Grammar. In Caregorial Grammar, one begins with a set of Initial (or Primitive) Categories, and a set of Category-Formation Operators. These Category-Formation Operators apply to simpler categories to derive more complex categories, and so on, recursively. Each category name spells out the history of that category's derivation, and how it is obtained by application of the Category-Formation Operators to the Initial Categories. In addition, each category name also provides a characterisation of the distributional privileges of members of that category.

The theory presented in this paper posits one Initial Category and two Category-Formation Operators; these are spelt out in (G) below. After that, in (H), the ways in which categories group together to form syntactic structures are indicated:

---

**G Category Formation (Paradigmatic)**

1. **Initial Category**: SO
2. **Category-Formation Operators**:
   a. **Slash**:
      
      for any two categories X and Y, \(XfY\) is a category, called 'X slash Y',
   b. **Kernel**:
      
      For any category X" , X·⁻¹ is a category, called 'the Kernel Category of X".

**H Category Combination (Syntagmatic)**

1. **Identity Combination**: \(X \leftrightarrow IX, X, X \ldots J\)
2. **Slash Combination**: \(X \leftrightarrow (Y, XIY, XIY \ldots)\)

As specified in (G1), the theory makes use of a single Initial Category, So. The letter "S" may be construed as a rough mnemonic for "Sentence", while the superscript "0" reflects the basic nature of the category in question. Indeed, the category SO may be thought of as corresponding approximately to the traditional category of sentence.

The theory accordingly assigns privileged status to the sentence, characterising it as the most basic or fundamental syntactic category. In this respect it follows in the footsteps of a long tradition, encompassing, among others, American Structuralism and Generative Grammar. Thus, for example, Boas (1911: 23) writes: "Since all speech is intended to serve for the communication of ideas, the natural unit of expression is the sentence; that is to say, a group of articulate sounds which convey a complete idea." Echoing this insight, phrase structure grammars such as that of
Chomsky Gil (1965) typically begin with a formula such as "s → ...". However, in this regard, the present theory differs from previous versions of Categorial Grammar, which generally posit more than one Initial Category. Thus, Ajdukiewicz (1935) posits two Initial Categories, corresponding to 5 and NP; Montague (1970a, 1970b) and many others opt for three, corresponding to 5, NP and N; while some, for example Morrill (1994), even postulate four, corresponding to S, NP, N and PP.

We are now in a position to examine the first of the two Category Combination rules, namely Identity Combination, as specified in (H1). What this rule says is that an expression of category X may consist of any number of daughter expressions also of category X, in what amounts to a conjunction, apposition or juxtaposition of coordinate elements. For example, an S° may consist of two, three, four, or any number of daughter S°'s, in a construction of the form S° ↔ (S°, S°, S°...).

In order to enrich the inventory of syntactic categories, recourse is required to the two Category-Formation Operators, as specified in (G2). The first of these, in (G2a), is the familiar binary Slash Operator, which is at the heart of all previous theories of Categorial Grammar. What it says, quite simply, is that if X and Y are both syntactic categories, then so is X\/. For example, from S° we can form the category S°/S°, from these two categories we can form the categories S°/(S°/S°), (S°/S°)/S°, and (S°/S°)/(S°/S°); and so forth.

In its appearance, the Slash Operator is reminiscent of the division sign in elementary algebra-and for good reason. To see why, let us turn our attention to the second Category Combination rule, namely Slash Combination, in (H2). What this says is that an expression of category X may consist of an expression of category Y in construction with one or more expressions of category XIV. For simplicity, assume for the moment that the number of X/Y expressions is just one. Then what Slash Combination is saying is that a Y expression in construction with an X/Y expression results in an expression of category X: X ↔ [y:X/Y]. Which looks just like the elementary algebraic equation X = y. To take a real example now, an expression of category S° may consist of an expression of category 5 in construction with, say, three expressions of category S°/S°: S° ↔ IS°, S°/S°, S°/S°. S°/S°).

As noted above, the Slash Operator is common to most or all versions of Categorial Grammar. However, whereas most versions, such as Montague (1970a, 1970b) and Keenan-Faltz (1985), allow only for binary branchings, the present version permits multiple branchings, thereby reflecting a body of evidence that has accumulated to the effect that at least some constructions, in some languages, are endowed with flat rather than hierarchic syntactic structure-s-see, for example, Hale (1982, 1983). Gil (1983), Austin-Bresnan (1996). Also, whereas some versions, such as Bar-Hillel (1953) and Lambek (1958), build linear order into the theory, typically by introducing distinct slash symbols, the present framework follows Ajdukiewicz (1935), Keenan-Faltz (1985) and others in positing syntactic struc-
Syntactic categories, cross-linguistic variation and universal grammar

This is motivated by the observation—see, for example, Saumjan (1965), Sanders (1975) and Keenan (1978)—that a large proportion of the generalisations governing the syntactic structures of languages and the ways in which these structures are interpreted do not require recourse to linear order.

The second Category-Formation Operator, in (G2b), is the unary Kernel Operator. The effect of this operator is, quite simply, to add 1 to the value of the superscript of the category to which it applies. For example, it applies to $S^0$ to yield $S_1$, to $S^1$ to yield $S^2$, and so forth. In addition, it can apply to categories resulting from the application of the Slash Operator, in which case the category produced by the Slash Operator is understood to bear the default superscript "0". For example, the Slash Operator may produce the category $S_0/S_0$ actually $(S^0/S^0)^0$; the Kernel Operator then may apply to $(S^0/S^0)^0$ to yield the category $(S^0/S^0)^1$.

The Kernel Operator is an innovation within the tradition of Categorial Grammar, but elsewhere it is almost familiar. Almost, but not quite: in fact, it is an upside-down version of the bar operator of Xcbar theory, as proposed by Chomsky (1970), Jackendoff (1977) and others. Within X-bar theory, words are associated with lexical categories, which are taken to be basic, and are accordingly assigned the superscript "0". These categories, or Xczeros, then project upwards, resulting in phrasal categories with ascending indices, X-bar, Xdouble-bar, etc., until an arbitrary limit, usually taken to equal two, is reached, at which point the resulting category is renamed as an XP (and also referred to as the "maximal projection" of X). X-bar theory thus characterises the word, rather than the sentence, as the most fundamental linguistic unit. However, as is suggested below, there are good reasons to believe that it is the sentence that is the more basic of the two. Or, more generally, that for any X, it is XP that is more basic than X. Accordingly, the Kernel Operator turns the tables, starting at the top, with what corresponds, very roughly, to the XPs, assigning these the superscript "0", and then working its way down, as far as may be necessary.

In addition to the Initial Category and the two Category-Formation Operators, it is necessary to introduce one more primitive into the theory, namely headedness. Given an expression X consisting of daughter expressions XI...X, the head of X is that expression X, which is characteristically associated with a range of properties which include the following: (a) obligatoriness: the head cannot be emitted; (b) hyponymy: the construction as a whole is a hyponym of its head; (c) percolation of features: the construction as a whole acquires grammatical features from its head; (d) agreement: the head controls agreement of the other elements in the construction; and (e) government: the head determines the morphological form of other words in the construction. The notion of head is well supposed within linguistic theory; see for example Tesniere (1959), Zwicky (1985) and Hudson (1990). In particular, the notion of head plays a central role within X-bar theory, where, for any X, X is taken...
to be the head of the Xcbar and XP containing it. Nevertheless, the notion of head is logically independent of the mechanisms of Xcbar theory, and is of greater generality. Thus, for example, the notion of head has been argued to be relevant in a variety of cognitive domains where the notions of Xcbar structure are not applicable, such as the theory of tonal music proposed by Lerdahl-Jackendolf (1983).

In the present theory, beadedness correlates with the rules of Category Combination in the following ways:

1. **Headedness (Syntagmatic)**
   1. Identity Combination:
      In a construction of the form $X \leftrightarrow [X, X, X ... J$, one of the daughter X's may be head.
   2. Slash Combination:
      In a construction of the form $X \leftrightarrow [Y, X/Y, XIV ... J$, $Y$ is head.

Rule (II) says that in a construction of the kind formed by Identity Combination, one of the daughter expressions may be head, but this is not a necessity: the construction may remain headless. Such freedom, however, is not the case for constructions formed by Slash Combination: here rule (12) specifies that in a construction of the form $[Y, X/Y, XIV ... J$, it is invariably $Y$ that is the head.

Among the syntactic categories generated by the Category Formation rules in (G), two particular kinds stand out as worthy of mention:

1. **Syntactic Caugones:** Two kinds
   1. **Modifier Categories**
      A Modifier Category is a category of the form XIX, for some category X.
   2. **Argument Categories**
      An Argument Category is a category of the form $X/Y$, for some categories X and Y, where Y is the Kernel Category of X.

Modifier Categories are referred to as such because, in accordance with the rule of Slash Combination in (H2), one or more XIX expressions combine with an X expression to yield another superordinate X expression: $X \leftrightarrow [X, XiX, X/X, XIX ... ]$. In such cases, the X/X expression(s) may be characterised as the modifier(s) of its/their sister X expression, which, in accordance with (12), is the head of the construction. Some examples of Modifier Categories are $S^o/S^o, 5/5^1, (S^0/S^0)/(S^0/S^0), (SoS I)/(SoS I)$, and so forth. Argument Categories are given this name because, also in accordance with the rule of Slash Combination in (H2), one or more X/Y expressions combine with an X expression to yield a superordinate Y expression: $Y \leftrightarrow [X, X/Y, XIY, X/Y ... ]$. In such instances, the XIV expression(s) may be characterised as the argument(s) of its/their sister X expression, which, again in ecor-
dance with (12), is the head of the construction. Some examples of Argument Categories are $SO_S$, $S^1/S^2$, $(S^0/S^0)/(S^0/S^0)^1$, $(SOSI)/(SdSI)1$ and so forth.\(^{12}\)

The rules of Category Formation in (G) generate an infinite number of syntactic categories from the Initial Category $So$. This set may be visualised in terms of a tree structure. Since the actual tree is boundless, it is unrepresentable; however, a very small subset of it is shown in Figure 3:

$$\begin{align*}
(SO/sO)/(SO/sO) \\
(SO/sO)/(S0/S0) \\
(S0/S0)/(S0/S0)^1 \\
S^1/S^2/(S^1/S^2) \\
(S^1/S^2)/(S^1/S^2) \\
(S0/SI)/(S0/SI) \\
(S0/SI)/(S0/SI)1 \\
(S0/SI)/(S0/SI) \end{align*}$$

*Figure 3. The Syntactic Category Tree (Partial)*

In Figure 3, the root node of the tree, at top centre, is the Initial Category $So$. Arrows lead from nodes representing syntactic categories to other nodes representing other syntactic categories that are derived from them by application of a Category-Formation Operator.

With its arrows leading from category to category, the above tree diagram points towards the following definitions:

**K Parenl and Ancestor Symocotic Categories**

1. For any syntactic category $X$, the parents of $X$ are the categories from which $X$ is formed by a single application of a Category-Formation Operator.
   a. If $X$ is the Kernel Category of $Y$, for some $Y$, then $Y$ is the parent of $X$.
   b. If $X$ is of form $Y/Z$, for some $Y$ and $Z$, then $Y$ and $Z$ are the parents of $X$.

2. For any syntactic category $X$, the ancestors of $X$ are the categories from which $X$ is formed by one or more applications of Category-Formation Operators.

In terms of the Syntactic Category Tree in Figure 3, the parents of a category $X$ are the nodes immediately dominating it, while the ancestors of $X$ are the nodes simply
David Gil

dominating it. For example, in accordance with (KIa), $S^0$ is the single parent of $S_I$; $S_I$ is the single parent of $S^2$; $S^0/S^1$ is the single parent of $(SOSI)_I$; and so forth. Similarly, in accordance with (KIb), $S^0$ and $S_I$ are the two parents of $S^0/S^1$; $S_I$ and $S^2$ are the two parents of $S^1/S^2$; $S^0/S^1$ and $(SOSI)_I$ are the two parents of $(SOSI)/(SOSI)_I$; and so on.

Since the set of syntactic categories generated by the Category Formation rules in (G) is infinite, only a very small proportion of these categories will be of relevance to linguistic theory. In particular, the actual syntactic category inventory of any given language will be finite, and in fact very small. Accordingly, constraints on syntactic category inventories need to be imposed.

The most important constraint on syntactic category inventories across languages makes reference to the definition of Ancestor Category in (K) above; this constraint is given in (L) below:

$L$ \textit{The Ancestral Constraint on Syntactic Category Inventories}

For any language $L$:

If $X$ is a syntactic category in $L$, then all $X$'s ancestors are syntactic categories in $L$, of equal or greater productivity.

The Ancestral Constraint says that each individual language constructs its inventory of syntactic categories in accordance with (G), beginning with the Initial Category $S^0$, and then forming additional syntactic categories by means of the Slash and Kernel Operators. That is to say, a language selects its inventory by starting at the top of the Syntactic Category Tree and working its way down to a certain point, and then stopping. Or, to be more precise, tapering off. This is because once the inventory of open categories is established, the Category Formation rules may continue to apply, to produce closed categories of lesser and lesser degrees of productivity.

Some examples of syntactic inventories permitted by the Ancestral Constraint are given in (M) below:

$M$ \textit{Some Category Inventories permitted by the Ancestral Constraint}

1. \{S\}
2. \{S^0, S^0/S^0\}
3. \{S^0, S_I\}
4. \{S^0, S^1, S^0/S^1\}
5. \{S^0, S^0/S^1, S'\}
6. \{S^0, S^0/S^0, S^1, SOSI\}

However, many other inventories are excluded by the Ancestral Constraint. For example, \{S^0, SOSI\} is not a possible inventory, since it does not include $S_I$, which is an ancestor to $SOSI$. Similarly, \{SoSo\}, \{S_I, SOSI\}, and \{S^0/S^0, S_I, SOSI\} are not possible inventories, since they do not include $S_0$, which is an ancestor to each.
of the categories in each of these inventories. In fact, since SO is the Initial Category, it is an ancestor to all syntactic categories, which leads to the following corollary of the Ancestral Constraint

N Corollary of the Ancestral Constraint

The Initial Category $S^0$ is contained in the syntactic category inventory of every language.

However, $S^0$ is the only syntactic category that is universal in this sense."

The Ancestral Constraint establishes a correlation between the complexity of a syntactic category and its cross-linguistic distribution. The complexity of a syntactic category can be measured by the length of its derivational history, as reflected in the number of symbols in its name, and the height of the numerical indices. Equivalently, the complexity of a syntactic category can be gauged by its distance from the root node SO in the Syntactic Category Tree. Thus, the Ancestral Constraint makes an intuitively appealing statement about the relationship between complexity and cross-linguistic distribution, namely: simpler categories will be more widespread, while more complex categories will occur less frequently in the languages of the world.14

However, the Ancestral Constraint alone is insufficiently restrictive; further constraints are required to restrict the class of possible syntactic category inventories. Imagine a language with the inventory in (M3), namely \{SO, SI\}. In such a language, no construction could contain expressions belonging to both of these categories: given the rules of Category Combination in (H), there is no way that expressions from these two classes could come together in a single construction. This clearly does not make sense. A similar problem arises with the inventory in (M5), \{SO, S^0/S^0, SI\}, as well as many other inventories permitted by the Ancestral Constraint. Hence the following additional constraint:

O The Syntagmatic Constraint on Syntactic Category Inventories

For any language L:

If X and Y are syntactic categories in L, then L must have constructions containing expressions belonging to both X and Y.

The Syntagmatic Constraint thus rules out syntactic category inventories that are disjoint, that is to say, contain categories which can never enter into any kind of construction together."
The Open-Category Constraint on Syntactic Category inventories

For any language L:

If $XN$ is an open category in $L$, then $X/Y$ is either a Modifier Category (i.e., $Y$ equals $X$) or an Argument Category (i.e., $Y$ is the Kernel Category of $X$).

The Open-Category Constraint says that Modifier Categories and Argument Categories are the two most important types of categories, in the sense that they alone can be open categories in any given language. All other categories can only be closed categories. So far, all of the examples of categories that have been discussed, or portrayed in Figure 3, have been either Modifier or Argument Categories. Some examples of categories belonging to neither of these two types are $S'/S^0$, $S0(S0sj)$, $(S0sj)/S0$, and many others; such categories, if present in a language, can only be closed.

Together, the Ancestral, Syntagmatic, and Open-Category Constraints define a set of permissible syntactic category inventories which different languages may choose from. The four simplest inventories permitted by these three constraints are listed in (Q) below:

Q. The four simplest Syntactic Category Inventories
1. $\{S\}$
2. $\{S0, S0/S0\}$
3. $\{S0, S', S0S1\}$
4. $\{S0, S0/S0, S', S0/S1\}$

And these same four inventories are plotted onto the Syntactic Category Tree, by means of ovals, in Figure 4:
4. Syntactic categories and linguistic *typology*

At this point, it *would* be appropriate to look at some real languages. Unfortunately, this is easier said than done. Languages do not *wear* their syntactic category inventories on their sleeves. An hour or two with a grammar book, or a few sessions *with* a native speaker, will reward the investigator with a pretty good indication of a *variety* of things such as inflectional categories, kinship terms, and other similarly visible features. Not so, *however*, syntactic categories.

As suggested in section I, the appropriate question to ask is "What are the significant syntactic patterns in the language, and *what* are the categories that must be posited in order to enable the necessary generalisations to be stated?". Thus, in order to determine the syntactic category inventory of a language, it is *necessary*, essentially, to construct a comprehensive description of the major syntactic patterns of the language in question. Only after one has "done" the language pretty much exhaustively can one then *take* one's labour in hand and say: Here is a comprehensive description of the major syntactic panems of the language, and these *are* the categories—some semantic, some morphological, others perhaps syntactic— which must be posited in order to capture the *necessary* generalisations. Conversely, in the absence of such a comprehensive description, it is very difficult to come to any meaningful conclusions with regard to the syntactic category inventory of a given language.

This leads to *what* might be dubbed the "typologist's dilemma", Whereas some features— such as inflectional categories and kinship terms---can, without too much difficulty, be compared *with* a reasonable degree of consistency and reliability *across* a sample of tens or even hundreds of languages, other features are simply much less amenable to such large-scale cross-linguistic comparisons.

Consider, for example, Greenberg's (1963) classic study titled "Some universals of grammar with particular *reference* to the order of meaningful elements". Greenberg's results provided the empirical basis for a whole *new* enterprise *within* linguistics, devoted to *refining* the *word-order* correlations that he observed, and explaining them in terms of more general principles—see, for example, Hawkins (1983, 1994), Li (1975), Tomlin (1986), Dryer (1992), Siewierska (1998), and many others. Yet in spite of their proven worth, Greenberg's generalisations rest on shaky foundations. Take two languages, French and Malay, both characterised by Greenberg as belonging to the same "basic order type", his "type 9", with SVO word order, prepositions, noun-genitive and noun-adjective order. The similarity of these two languages with respect to Greenberg's typology can be illustrated by means of a pair of *parallel* sentences in the two languages, *in which* the equivalent content words line up neatly one under the other, in the same order.
Now obviously, there is a myriad of superficial and easily observable differences between French and Malay which Greenberg's typology ignores; for example the presence, in French, of subject-verb agreement, noun-adjective agreement, definite articles, and an overt possessive particle, all of which are lacking in Malay. But this is of necessity so, since all typologies, by their very nature, choose to focus on some parameters while ignoring others. The problem, however, is with the more profound, albeit less readily visible differences between French and Malay, which call into question the very underpinnings of Greenberg's typology. Consider, for example, the fact that both French and Malay are characterised as having N-A and N-G order. Whereas for French this constitutes a substantive observation with regard to the "serialisation" or "harmonisation" of what are, quite obviously, two fonnally distinct construction types involving adjectival and genitival attribution respectively, for Malay this observation is, arguably, of no substantive import whatsoever, since a reasonable case can be made that, say, buku hijau 'green book' and rumah Bobak 'Bobek's house' in (2) above are but two instantiations of a single more general type of construction involving nominal attribution. In other words, Malay may not distinguish between the categories which Greenberg labels as "A" and as "G". Or, consider the fact that both French and Malay are characterised as having SVO order. Whereas for French this seems relatively uncontroversial, for Malay, questions have been raised regarding the viability of the grammatical relations of subject and object—see, for example Alsagoff (1992) and Gil (1994, 1999, forthcoming c). The basic problem is this: Greenberg's universals are essentially a catalogue of translations, into various languages, of sentences such as Allan was reading the green book in Bill's house, and a tabulation of the order of constituents exhibited within these translations. However, such a large-scale cross-linguistic study cannot do otherwise than to rely on semantic categories. By default, the "Allan" word will be labelled "S", the expression meaning 'was reading' will get the label "V", the phrase corresponding to 'red book' will be characterised as an "0",

\[ \begin{array}{cccccccc}
 & tis-\text{au} & l, & \text{livre} & \text{vert} & \text{dans} & to & \text{manan} & d, \\
\hline
\text{Alain} & \text{read-IMP:3SG} & \text{ART:DU} & \text{SG.M} & \text{book} & \text{green-sc.a} & \text{m} & \text{ART:DEF:SG.F} & \text{house of} \\
\hline
\text{S} & \text{V} & \text{ON} & \text{A} & \text{Prep} & \text{N} & \text{G} \\
\hline
\text{Aeen} & \text{read} & \text{book} & \text{green} & \text{LOC} & \text{house} & \text{Bobok} \\
\text{Aeen} & \text{baca} & \text{baku} & \text{hl'au} & \text{kai} & \text{rumah} & \text{Bobok} \\
\end{array} \]
Syntactic categories. cross-linguistic variation and universal grammar

Greenberg could only have made use of the best grammatical descriptions that were available to him; however, most grammatical descriptions automatically assign words and phrases to syntactic categories on the basis of their meanings. (This point has already been made by several scholars, including Croft (1991) and Dryer (1992).) Thus, when the Malay sentence emerges with the order of meaningful elements as shown in (2) above, it is classified together with French as having SVO word order, prepositions, N-G and N-A order. This in spite of the fact that a more adequate and less Eurocentric description of the language might well reveal it to have neither S, V or O, and neither N, G or A.

In an ideal world, the typologist would have at his or her disposal a library full of grammatical descriptions, each adhering to a standard of rigour in which each and every grammatical category that is invoked—word, noun, agent, subject, topic, whatever—is explicitly motivated within the language being described. However, until the state of the art of linguistic description attains those very high standards, there will remain certain linguistic features which, because of their abstract and non-obvious nature, the typologist will be unable to compare in an adequate fashion across a wide range of languages. And among such features are the syntactic categories which form the topic of this paper.

At the present time, it is thus simply not feasible to examine the syntactic category inventories of a large sample of languages, and on such basis to propose and support an empirically-grounded linguistic typology pertaining to syntactic category inventories. Instead, one has to make do with much less. Accordingly, in what follows, a number of hypothetical, albeit hopefully realistic abstract language types are proposed, to which actual languages may, in future work, be shown to conform.

Each language type is defined in terms of three properties, listed from left to right. In the first column, the syntactic category inventory is specified, in a vertical list. In the second column, each syntactic category is characterised as either open or closed. And in the third column, some of the syntactic categories are related, via prototypical rules of association, to one or more semantic categories, indicated in small caps.

The first three language types are presented in (RI-3):

R 1. **Type I**  
   a. SO  
      open

2. **Type 2**  
   a. S^0  
      open  
      ACTIVITY

   b. S^0/S^0  
      open  
      TIIING
3. Type 3
   a. S: open
   b. S′ open ACTIVITY
   c. SΩS′ open TILLING

Types 1, 2 and 3 above instantiate the three syntactic category inventories indicated in (Q1-3) and represented with three of the four ovals in Figure 4. Type J possesses the smallest inventory, with just a single syntactic category, while Types 2 and 3 are associated with increasingly larger, albeit still rather small inventories.

In Type I languages, all words and phrases belong to So that is to say, all words and phrases exhibit the same syntactic behaviour. At present I am not familiar with any realistic candidates for Type J languages. However, as suggested below, Type I is closely approximated by Tagalog and Malay/Indonesian, which deviate from it only minimally, by the possession of an additional closed syntactic category.19

In Type 2 languages, there is a basic distinction between two open syntactic categories: one, So, which typically contains expressions denoting activities, the other, So/So, which generally consists of expressions referring to things. Type 2 is perhaps that which most closely reflects the properties of languages such as Warlpin, Lakhota, and Mohawk, which have been characterised by Jelinek (1984), Van Valin (1985), Baker (1996) and others as "pronominal argument" languages.

In Type 3 languages, there is also a dichotomy between activity expressions and thing expressions; however, this dichotomy manifests itself in a rather different way. In Type 3 languages there are three distinct open categories. One, 51, typically contains expressions denoting activities; the second, 5′/5′', usually consists of expressions denoting things; while the third, So, generally comprises complex expressions formed from expressions of categories 51 and 5′/5′'. Type 3 perhaps comes closer than the preceding two types 10 capturing some of the essential properties of many European languages.

Languages of Types 1, 2 and 3 may be distinguished from each other with respect to a number of salient features. One is the ability of single words to stand alone, as a complete, non-elliptical sentence, in a wide range of contexts:

S 1. Type I languages: all words can stand alone
   2. Type 2 languages: some words, typically denoting activities, can stand alone
   3. Type 3 languages: few or no words can stand alone

In languages of Type I, all words, as Sαs, can stand alone as a complete, non-elliptical sentence. The ability of words denoting things, in particular, to stand alone, would appear to be rather uncommon cross-linguistically, though it has been observed for some languages, including Tagalog (Gil 1993b, 1995), Riau Indonesian (Gil 1994) and Singlish (Gil forthcoming a). In languages of Type 2, activity words typically are Sαs, while thing words generally are not; hence, in
Type 2 languages, activity words can usually stand alone as a complete sentence while thing words generally cannot. This is in fact the typical situation in languages that have been characterised as "pro-drop", such as Mandarin (Huang 1984) and Hebrew (Borer 1984). Finally, in languages of Type 3, neither activity words nor thing words are generally members of $S^0$; hence, the class of words that can stand alone as a complete sentence is either small or non-existent. This is the typical situation for languages that have been characterised as not exhibiting "pro-drop", for example English. Note that in accordance with the present analysis, "pro-drop" is actually a misnomer: rather than containing an empry syntactic position occupied by a phonologically null pronoun, the constructions in question simply have nothing."

A second important feature distinguishing languages of Types I, 2 and 3 is the structure of basic constructions combining a thing expression with an activity expression. (T) below shows how a simple sentence such as The chicken is eating—recall Tagalog sentence (1) at the beginning of the paper—might be rendered into languages of these three types:

(T) 1. **Type 1 languages:** Identity Combination

2. **Type 2 languages:** Slash Combination: modifier plus modified

3. **Type 3 languages:** Slash Combination: argument plus kernel
In Type I languages, all expressions are So's; therefore, the only available construction type is that formed by Identity Combination. In Type J languages, then, there is no dedicated syntactic structure for expressions with meanings such as 'the chicken is eating', which would distinguish them from expressions with meanings such as 'the chicken which is eating', 'the chicken and the eating', and many others. All these different meanings are expressed via the single available construction type, that indicated in (T1) above.

In Type 2 languages, expressions referring to things are prototypically associated with the Modifier Category S0Sb. Accordingly, in constructions such as (T1), the thing expression is a modifier of the activity expression, which, in accordance with (12), is the head of the construction. In such languages, then, sentences such as 'the chicken is eating' are of similar structure to sentences such as 'here (it) is eating' and 'today (it) is eating', in which the head activity expression is modified by other adjunct expressions, denoting a place or time.

In Type 3 languages, expressions referring to things are prototypically associated with the Argument Category S0bS, while expressions referring to activities are prototypically associated with the category S1. Thus, in constructions such as (n), the thing expression is an argument of the activity expression which, once again, is the head of the construction. In languages such as these, then, the structure of sentences such as 'the chicken is eating' bears a closer resemblance to that generally posited for such sentences within most traditional theories of syntax.

In conjunction, the three structures represented in (T) above provide a vivid reflection of the Autonomy of Syntax, and the independence of syntactic and semantic categories. Take the expression EAT. Whereas in Type I and 2 languages it is an So in languages of Type 3 it belongs to a different category, namely S'. This corresponds to the fact, noted previously in (S), that an activity word can stand alone as a complete sentence in languages of Types I and 2 but not 3. Now look at CHICKEN. This same word belongs to three different categories in each of the three language types: in Type I languages it is an So in Type 2 languages an So'So, and in Type 3 languages an S0Sb. These three categories reflect its different syntactic behaviour, as a simple juxtaposed expression in Type 1 languages, a modifier in Type 2 languages, and an argument in Type 3 languages. Given the Autonomy of Syntax and the independence of syntactic and semantic categories, it is only natural that words with the same meanings will, in different languages, exhibit different syntactic behaviour, and hence be associated with different syntactic categories.

However, once the autonomy of syntactic categories is duly acknowledged, it is possible to focus on the ways in which syntactic categories are actually related to semantic categories, by means of prototypical rules of association. We begin with a definition:
Semantically associated Syntactic Categories

A semantically associated syntactic category is a syntactic category which, in a certain language, is prototypically associated with a particular semantic category.

Although the term "semantically associated syntactic category" is unfamiliar and somewhat awkward, it sets the stage for some new definitions of two more traditional terms:

V 1. A Verb-Phrase (or VP) is a syntactic category which, in a certain language, is prototypically associated with the semantic category of activity.

2. A Noun-Phrase (or NP) is a syntactic category which, in a certain language, is prototypically associated with the semantic category of thing.

In accordance with the above definition, VPs and NPs are not syntactic categories per se: rather, they are semantically associated syntactic categories, that is to say, syntactic categories which happen to be prototypically associated with semantic categories in a certain language.

Since Type I languages have only one syntactic category, they obviously do not have dedicated syntactic categories for activities and things; hence, such languages do not have VPs and NPs. In contrast, Type 2 and 3 languages do have distinct syntactic categories prototypically associated with activities and things; accordingly, they can be said to have VPs and NPs. However, the VPs and NPs are not the same in these two language types. Specifically, whereas VPs are $S^0$s in Type 2 languages, they are $S"$s in Type 3 languages; similarly, whereas NPs are $S^0/S^0$s in Type 2 languages they are $SOS"$s in Type 3 languages. As this shows, then, VPs and NPs are not always the same syntactic categories: this reflects the fact that in different languages, VPs and NPs may exhibit different syntactic behaviour.

The definition of semantically associated syntactic categories such as VP and NP is reminiscent of the way in which grammatical relations such as subject and object are defined in terms of clusterings of features. Some syntactic (e.g. order, control of agreement, licensing of reflexives), others semantic or pragmatic (e.g. thematic roles, referentiality, topicality) - see, for example, Keenan (1976), Comrie (1981) and Croft (1991). More specifically, the way in which languages of Types 2 and 3 with distinct NPs and VPs differ from languages of Type 1 (and its variant Types I· and I++) with no NP-VP distinction is analogous to the way in which languages with well-defined grammatical relations of subject and object differ from languages in which such grammatical relations have been argued not to be viable - see, for example, Schachter (1976), Gil (1984) and Kibrik (1997). Specifically, in both cases, features belonging to different syntactic and semantic levels cluster together and undergo grammaticalisation in some languages while remaining disassociated in others.
To this point, we have considered only language types which consist entirely of open categories. Let us now examine two further types of languages containing combinations of open and closed categories:

W1. Type I*  
   a. S' open 
   b. so-so closed

2. Type I**  
   a. S' open 
   b. S' closed NUMBER
   <. S0S' closed UNIT

As suggested informally by their names, Types I' and I** closely resemble Type I. Like Type I, they contain a single open syntactic category, S°. However, they differ from Type I in the presence of one or two additional closed syntactic categories. In fact, the closed syntactic categories in Types I' and I'' are the same as the additional open syntactic categories in Types 2 and 3 respectively. Thus, abstracting away from the distinction between open and closed categories, Type r has the same inventory of syntactic categories as Type 2, and Type 1'' the same inventory as Type 3.

In a number of previous publications, I have argued that Type r provides the most adequate characterisation of the syntactic patterns of two Austronesian languages, Tagalog (Gil 1993a, 1993b, 1995) and Malaylndonesian (Gil 1994). And work in progress on various mainland Southeast Asian languages such as Thai and Vietnamese tentatively suggests that they may approximate Language Type l**.

Let us now take a closer look at the Riau dialect of Indonesian. As a Type i' language, it has one open category, S°, and one closed syntactic category, S0S°. Following, in (3) and (4) below, is a fragment of the lexicon of Riau Indonesian:

(3) A partial lexicon of Riau Indonesian: S° words
   a. orang person
   b. buku book
   d. hijau green/blue
   e. besar big
   g. lari run
   h. baca read
   J. Kairil [name of person]
   k. Pekanbaru [name of place]
   m. ebang elder.brother
   o. aku ls
   <. rumah house
   f. lapar hungry
   l. kasi give
   l. Lebaran [name of holiday]
   o. INI

   ...
Almost all words in Riau Indonesian belong to So; (3) presents just a very small sample thereof. As evident from the above, SO words in Riau Indonesian may denote things, as in (3a-c); properties, as in (3d-t); or activities, as in (Jg-f). In addition, SO words may be names of people, places or times, as in (3j-l); deictic, as in (3m-r); quantificational, as in (Js-u); or interrogative, as in (Jv-x). Finally, S-words may have a variety of meanings which, in most other languages, are typically
expressed by means of various grammatical markers; these include existence, as in (3y); possession, as in (3z); ability, as in (3aa); aspect, as in (3bb); superlativity, as in (3cc); negation, as in (3dd); and various macrofunctional words, as in (3ee), (3ft) and (3gg).

As \( S^6 \)'s, almost all words in Riau Indonesian exhibit identical syntactic behaviour. In particular, there are no syntactic differences between words referring to activities and words referring to things. As \( S^6 \)'s, almost all words can occur readily as a complete non-elliptical sentence in a wide range of contexts; moreover, any SO word can combine with any other SO words, in accordance with the rule of Identity Combination in (H1), to yield a multi-word SO expression, and so on recursively. Thus, any string of SO words, with any associated constituent structure, is syntactically well-formed—though it may turn out to be semantically anomalous.

Whereas a Type I language would stop here, Riau Indonesian goes one step further, introducing a closed Modifier Category of SO/So expressions. While the list in (3) represents a tiny proportion of the SO words in the language, the list in (4) actually includes a large proportion of the existing SO/So words—work in progress suggests that the total number of SO/So words in Riau Indonesian will not exceed a few dozen. As evident from the list in (4), SO/So expressions are a very mixed bag semantically: if any generalisation can be made, it is that their meanings are all of an abstract nature. Interestingly, many SO/So expressions in (4) correspond closely in their meanings to other SO expressions in (3). For example, the wide range of functions of dengan in (4i) is subsumed within the even wider range of functions of sama in (3gg); the universal quantifier Nap in (4j) is the distributive counterpart of the non-distributive universal quantifier semua in (3t); the operator pun in (4t) overlaps in its range of meanings with the operator lag; in (3ee); and the reflexive diri in (4u) is a hyponym of the macrofunctional sendiri in (3ft). The only motivation for the SO/So category is thus syntactic: as suggested by their category name, members of SO/So cannot stand by themselves as complete sentences; rather, they must occur in construction with expressions belonging to the category of \( S^6 \).25 In Riau Indonesian, then, the category SO/So may be characterised as the category which contains the "grammatical markers" of the language.

Recall, now, that the Modifier Category SO/So is one that is shared by languages of Types 1' and 2. However, whereas in Type 1' languages this category contains a small set of semantically heterogeneous grammatical markers, in Type 2 languages it is an open category, prototypically associated with things—in other words, an NP category. Thus, the contrast between the category SO/So in Type 1" and Type 2 languages further underscores the independence of syntactic categories, underlying the syntactic behaviour of linguistic forms, from semantic categories, pertaining to their meanings.
Type \(1^{++}\) languages resemble their Type \(1^+\) counterparts in that almost all words and larger expressions belong to the single open syntactic category \(S^0\), while just a small residue of expressions do not. However, in Type \(1^{++}\) languages, this residue in turn divides into two distinct closed syntactic categories, \(S^1\) or \(SOS1\). Expressions belonging to \(S1\) can only occur in construction with expressions belonging to \(S^0/S^1\), which function as their arguments; in such cases, the result is an expression belonging to the category \(S\). In principle one could imagine a variety of different closed-class categories fining the above bill. As defined in (W2) above, \(S1\) contains a set of quantificational and/or determinative elements, which can only occur in construction with another set of elements, commonly referred to as numeral classifiers. Work in progress suggests that this language type may provide an adequate characterisation of various mainland Southeast Asian languages such as Thai and Vietnamese.

The language types considered so far have all been rather frugal in their inventories of syntactic categories. However, it is possible to imagine languages with much richer inventories. A full exploration of the possibilities lies beyond the scope of this study. One illustrative example may, however, be briefly considered:

<table>
<thead>
<tr>
<th>X</th>
<th>A more elaborate language type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>(SO) open</td>
</tr>
<tr>
<td>2.</td>
<td>(S^1) open</td>
</tr>
<tr>
<td>3.</td>
<td>(SOS') open</td>
</tr>
<tr>
<td>4.</td>
<td>((SOS1)/(SOS1)) open</td>
</tr>
<tr>
<td>5.</td>
<td>(SoSo) open</td>
</tr>
<tr>
<td>6.</td>
<td>(S^0/S^1) open</td>
</tr>
<tr>
<td>7.</td>
<td>((S^0S^0)/(S^0/S^1)) closed</td>
</tr>
<tr>
<td>8.</td>
<td>((S^0/S^1)/(So/S1)) closed</td>
</tr>
<tr>
<td>9.</td>
<td>(S1/(SOS1)) closed</td>
</tr>
</tbody>
</table>

The above language type begins with the three categories of Type 3, \(S0\), \(S1\), and \(SOS^1\). Unlike Type 3, however, it has an additional syntactic category, \((S^0S^1)/(S^0/S^1)\), prototypically associated with the semantic category of property. This points towards the definition of another semantically associated syntactic category, beyond the two defined previously, in (V):

| Y | An Adjective-Phrase (or AP) is a syntactic category which, in a certain language, is prototypically associated with the semantic category of property. |

In languages of the above type, then, \((SoS')/(SoS')\) is the syntactic category of AP. Such languages thus contrast with the previous types, in which there is no category of AP. 26
The next two syntactic categories in the above language type, $S_0/S_0$ and $S_1/S_1$, are prototypically associated with the semantic categories of time, place and manner, thereby suggesting the definition of a fourth semantically associated syntactic category:

\[Z\]

An Adverb-Phrase (or AdvP) is a syntactic category which, in a certain language, is prototypically associated with the semantic category of time, place or manner.

The two categories, $S_0/S_0$ and $S_1/S_1$, differ with regard to the constructions into which they may enter: whereas $S_0/S_0$ contains sentential AdvPs, which modify $S_0$ expressions, $S_1/S_1$ consists of VP AdvPs, which modify $S_1$ expressions.

The next two syntactic categories in (X) above, $(S_0^0/S_0^0)\,(S_0^1/S_1^1)$ and $(S_1^1/S_1^1)\,(S_0^0/S_1^1)$, are closed categories, containing expressions whose function is to convert $S_0S_1S_2$, which in this language type are NPs, to either $S_0/S_0$ or $S_1/S_1$, which, as mentioned in the preceding paragraph, are AdvPs. These two categories contain a small class of expressions of the kind commonly referred to as "adpositions". Finally, the last syntactic category, $S_1^1/(S_0^0/S_1^1)$, is a closed category, containing expressions whose function is to convert $S_0S_1S_2$ or NPs, to $S_1S_2$ or VPs. This category contains expressions which are usually referred to as "copulas".

Many additional, even more elaborate language types can readily be defined, possessing a variety of different syntactic category inventories, associated in different ways with semantic categories. Such language types, then, would hopefully provide the basis for a more adequate account of the diverse syntactic patterns exhibited by the languages of the world.

\[S.\] Towards universality

In conclusion, let us now return to the Tagalog sentence in (I) with which this article commenced: *manok ang kumakain* 'The chicken is eating'. As argued in the preceding section, words with similar meanings may belong to different syntactic categories in different languages. Thus, as suggested in the introduction to this paper, even though *manok* 'chicken' is the name of a thing, it is not an NP; similarly, even though *kumakain* 'is eating' denotes an activity, it is not a VP. Rather, as claimed in section 4, Tagalog is a Type I" language, with but a single open syntactic category $S_0$ containing *inter alia* both items, *manok* and *kumakain*. Accordingly, the syntactic structure of sentence (I) involves the juxtaposition of two $S_0$'s, in accordance with the rule of Identity Combination, as represented in (T1).
The theory of syntactic categories proposed in this paper thus provides the necessary tools for escaping Eurocentricity, and breaking out of the straightjacket of approaches within which "If it refers to a thing, well then it must be a noun". In doing so, it shows how languages may differ with respect to their inventories of syntactic categories to a much greater degree than is usually assumed or than is generally allowed for within current frameworks.

But can they differ without bound? Sapir (1921: 119) seemed to be edging towards an affirmative answer when he "To the that "no logical scheme of the parts of speech— their number, nature and necessary confines—is of the slightest interest to the linguist. Each language has its own scheme." A similar view would appear to be held by Croft (1991: 42), who argues that within a "structural" (i.e. syntactic, as opposed to semantic) approach "no adequate cross-linguistic definition of a syntactic category is possible since the grammatical manifestations of syntactic categories are so varied across languages."

However, the theory of syntactic categories put forward in this paper points to a different answer: languages cannot differ in unlimited ways with respect to their syntactic categories. Although highlighting cross-linguistic variation, the theory provides a unitary framework within which syntactic categories can be meaningfully compared across languages. To cite just one example, it is possible, within the present theory, to claim that a small set of semantically heterogeneous items in Riau Indonesian (Type 1) exhibit the same distributional privileges as an infinite set of expressions referring to things in, say, Warlpiri (Type 2); and, on this basis, to assign both sets of expressions, notwithstanding their different sizes and prototypical meanings, to the same syntactic category, namely SO/SO. Thus, the theory provides for a universal set of syntactic categories from which all languages must draw. Moreover, although different languages may draw from this set in different ways, possible inventories of syntactic categories are governed by universal constraints to which all languages, without exception, must adhere.

In order to arrive at a universal theory of syntactic categories, it is first necessary to construct comprehensive syntactic descriptions of a wide range of languages, each dealt with on its very own terms. Such descriptions will, for each language, provide motivation for a set of syntactic categories, facilitating the formulation of generalisations governing the observable syntactic patterns within that language. Once these descriptions are complete, the inventories of syntactic categories posited for each language can be put side by side and compared, in the quest for differences and commonalties. Thus, the escape from Eurocentricity becomes the first essential step towards the discovery of a truer universality.
Notes

- It is a pleasure to be able to thank Sharon Armcn-Lotem and Peter Cole for their continual encouragement and intellectual support in the course of my work on syntactic categories. The major ideas contained in this paper were developed in informal seminars in the Department of Linguistics at Tel Aviv University in 1989; the Department of English Language and Literature at the National University of Singapore in 1992; and several meetings of the EUROTYPO project, in Valletta, Donostia-San Sebastian, Konstanz and elsewhere between 1990 and 1995. I am grateful to participants at all of these venues for their many stimulating suggestions. Bits and pieces of this paper were also presented at the "Agreement by Suffixaufnahme", Franz Nikolaus Finck Memorial Symposium [Konstanz, Federal Republic of Germany, 23 September 1991]; the Third International Symposium on Language and Linguistics: Pan-Asiatic Linguistics (Bangkok, Thailand, 9 January 1992); the Eighth Annual Meeting of the Societas Linguistica Europaea Sodalicum Israelense (Haifa, Israel, 3 May 1992); the Eighth Annual Conference of The Israel Association for Theoretical Linguistics (Ramat Gan, Israel, 10 June 1992); the Nineteenth Annual Meeting of the Berkeley Linguistics Society (Berkeley, CA, USA, 15 February 1993); the Third Annual Meeting of the Southeast Asian Linguistics Society (Honolulu, HI, USA, 8 May 1993); the Pre-Inaugural Meeting of the Association for Linguistic Typology (Konstanz, Germany, 29 November 1994); Dialogue '95, International Seminar on Computational Linguistics and Its Applications (Kazan, Tatarstan, Russia, 2 June 1995); Workshop on Voice in Austronesian. Twenty-Eighth Annual Meeting of the Societas Linguistica Europaea (Leiden, The Netherlands, 1 September 1995); the Fourth Annual Meeting of the Austronesian Formal Linguistics Association (Los Angeles, CA, USA, 26 April 1997); the Workshop on the Typology of Parts-of-Speech Systems, Second Meeting of the Association for Linguistic Typology (Eugene, OR, USA, 13 September 1997); and the Second Symposium on Malay/Indonesian Linguistics (Ujung Pandang, Indonesia, 12 July 1998)—again. I am grateful to participants at all of these conferences for their numerous helpful comments. Finally, this paper would not have been possible without the numerous speakers of Tagalog and Malay/Indonesian, who opened my eyes to a different perspective on language.

1. Note that even Alice didn't get it quite right, listing only five of the six forms—missing the ablative.

2. Although the distinction between nouns (or NPs) and verbs (or VPs) is generally considered to be of a fundamental nature, claims have occasionally been made to the effect that this distinction is lacking in certain languages, most famously Nootka—see Swadesh (1939). However, such claims have been disputed, for example by Jacobsen (1979), and it would probably be true to say that most syntacti-
clans now take the noun-verb (or NP-VP) distinction to be universal, indeed a necessary design-feature of language.

3. Although the term "autonomy" has been co-opted by Generative Grammar, cf Chomsky (1986b), the substance of the autonomy claims, as outlined in (AJ), effectively cross-cuts any division of the field into "generative and non-generative" camps, with theories on either side of the wcut-d-be divide identifying with these claims to varying degrees.

4. The three-way division between semantics, morphology and syntax is of course an oversimplification, motivated by ease of exposition; in a more detailed account, each of these three levels might be decomposed into several others. In particular, the term "semantics", as used here, is understood very broadly, to include also aspects of meaning which are more properly subsumed under pragmatics.

5. Although it does not always seem that way in practice, this is in fact the orthodox generative position on the relationship between semantics and syntax. Thus, for example, Kiparsky (1968: 48) writes that "progress in linguistics should consist in reducing the abstract part of language, the part consisting of the various theoretical constructs which must be set up to mediate between the concrete levels of phonetics and meaning, the only aspects of language which can be directly observed." Indeed, the only way to really show that there exists an autonomous syntax in the classical generative sense is try to do away with as much of it as is possible, by providing ahemanve semantic explanations. What is left (if anything), after that is done, will then be the real autonomous syntax. Unfortunately, many practicing generative syntecncians misinterpret the autonomy hypothesis, taking it instead as a license to construct syntactic explanations of phenomena whose nature is clearly semantic rather than syntactic.

6. In Figure 2, the letters "C" represent the three consonants of the root morpheme: the inflectional paradigm consists of the prefixes, the affixes, and the intercalated vowels. Figure 2 abstracts away from various formal complications involving so-called "defective" roots, those whose consonants belong to particular classes triggering a variety of morphophonemic rules.

7. In addition to the above two cases, there are additional ways in which properties belonging to different levels may define coextensive sets. One additional case involves coextensive semantic and morphological categories. For example, in Tagalog, the semantic category of expressions denoting natural numbers is coextensive with the morphological category of expressions which may occur in construction with the distributive prefix ug-, containing iso 'one', dalawa 'two', tallo 'three', and so forth (see Gil 1982a for discussion of this construction, a morphological counterpart of the periphrastic German je construction considered above). Another additional case involves coextensive semantic, morphological and syntactic properties. For example, in Japanese, the morphological class of expressions containing classifier suffixes (-ko, -hon, -nin etc.) such as itiko 'one-ct.', nihon 'twe-et.', san-nin 'three-ct.' and so forth is coextensive with the semantic class of expressions denoting natural numbers, which in turn is coextensive with the syntactic class of
expressions which may occur in postmodifying position, as, for example, in *hana itiko* 'flower one-ct.', *pen nihon* 'pen two-ct.', *otoko sannin* 'child three-ct'.

8. More nneresung, however, are the basic categories Noun, Adjective and Verb and their phrasal projections NP, AP and VP. Most commonly, such categories are defined by a combination of semantic and syntactic properties. The status of such categories is considered in section 4 below.

9. In (H), and subsequently, square brackets denote unordered sets; that is to say, syntactic constituents whose internal parts are unspecified for linear order.

10. It is probably not a coincidence that a Polish-speaking logician opted for two Initial Categories while the English-speaking tradition posits a third. Specifically, the additional Initial Category in Montague's system, N, is in fact one for which there is considerably more prima facie support in English than in Polish. In English, count nouns cannot readily occur in argument position without a determiner; this fact is often argued to support a distinction between the syntactic categories N and NP. In contrast, in Polish, all nouns can occur by themselves, in argument position, as complete NPs, thereby eliminating the most obvious motivation for the distinction between Ns and NPs. Thus, Ajdukiewica's logic follows his native language, while Montague's logic reflects his-even though both were proposing purely formal systems, and not attempting to describe actual languages, their own or others'.

11. A corollary of the above definitions is that within the present theory there is no distinction between lexical and non-lexical, or phrasal categories. Specifically, all categories can contain both words and longer, multi-word expressions.

12. Thus, in constructions such as \( Y \leftrightarrow IX, X/Y, XN, XJY \ldots J \), the head of the \( Y \) expression is a member of the Kernel Category of \( Y \). This underscores the affinity between the Kernel Category-Formanon Operator and the mirror-image bar operator of Xsbar theory.

13. The Ancestral Constraint thus provides the primary mouvauon for turning Xsbar theory upside down, as is done with the Kernel Operator. Assume \( X \) and \( Y \) are categories such that \( X \) is the Kernel Category of \( Y \). wnr the Ancestral Constraint says is that there can be languages with \( Y \) b\( \bar{u} \)l n\( \bar{O} \)l \( X \). but not languages with \( X \) but not \( Y \). Or, translated into X-bar theory: there can be languages with \( X \cdot 1s \) but not \( X^{\#} s \), but not languages with \( X \cdot s \) but not \( X^{+}h^* s \). And as a corollary: there can be languages with XPs but not Xs, but not languages with Xs but not XPs. Thus, beginning at the top, as the Kernel Operator does, rather than at the bottom, as per X-bar theory, reflects the cross-linguistic distributional facts. with more Widespread syntactic categories associated with lower indices than less commonly occurring categories.

14. It may be speculated that the complexity of syntactic categories is a reflection also of ontogeny, simpler categories being acquired before more complex ones. Thus, for example, at the one-word stage, all words would belong to the category \( S^0 \), this being the only syntactic category in the child's inventory at this stage (see also Note 19 below). Somewhat more adventurously, one might also conjecture that the com-
plexity of syntactic categories mirrors phylogeny, and the evolution of human language from simpler systems spoken by ancient humans or hominids.

15. The Syntagmatic constraint also rules out may other inventories that are, independently, ruled out also by the Ancestral Constraint, such as (S₀, S₀S₁) and {S₀S₀, S₁, S₀S₁}. However, other inventories, such as {S₀/S₀} and rS', S₀S₁), satisfy the Syntagmatic Constraint but are ruled out by the Ancestral Constraint, thereby showing that these two constraints are logically independent.

16. Using the same abbreviations as Greenberg, "S" stands for "subject", "V" for "verb", "O" for "object", "N" for "noun", "A" for "adjective", "G" for "genitive", and "Prep" for "preposition". The Malay example is in the colloquial dialect of Kuala Lumpur, the capital city of Malaysia.

17. The various language types discussed in this paper do not constitute a privileged set in any theoretical sense; they are merely chosen from a much larger set of possible language types, for expository purposes. The absence of any deep significance to the choice of language types discussed is reflected by their rather unrevealing names; "Type 1", "Type 2", and so forth.

18. To say that a syntactic category and a semantic category are related via a prototypical rule of association is not equivalent to positing a mixed, semantic-syntactic category. In the case at hand, each of the two categories, the syntactic one and the semantic one, may have different extensions, resulting from the interplay of different prototypical properties. Nevertheless, the syntactic and semantic categories may overlap to a considerable degree, and the intersection of the two categories may contain the most prototypical members of each category. Such, in fact, may be argued to be the case for the English syntactic and semantic categories in Figure I. Although the syntactic category of words which can occur in construction with can, may, will and shall and the semantic category of words denoting dynamic activities are independent, the intersection of these two categories arguably contains the forms that are the most prototypical members of both categories—in which case it might be concluded that these two categories are related by a prototypical rule of association. Accordingly, whereas the intersection of these two categories is a mixed semantic-syntactic category, the syntactic category of words which can occur in construction with can, may, will and shall is a pure syntactic category, which just happens to correlate, imperfectly and via a rule of prototypical association, with the semantic category of words denoting dynamic activities. Similarly, the categories listed in the first column are indeed pure syntactic categories, which also correlate prototypically with the distinct semantic categories listed in the third column.

19. However, as foreshadowed in Note 14 above, Type I provides a plausible characterisation for the earliest, one-word stage of child language, in which all words, by default, are associated with the same syntactic properties. Conceivably, it might also characterise the second, or two-word stage, in which words are put together in structures that may be much simpler and less differentiated than those of the corresponding target adult language. Of course, a possible Type I adult language would
differ from, say, Type I child English, in numerous other respects, one obvious one being the recursive application of the rule of Identity Combination in (HI).

20. The correlation between languages of Types I, 2 and 3 and the ability of words to stand alone, as represented in (S), is probably an oversimplification: in reality, many other factors may play a role in determining the ability of a word to stand alone as a complete, non-elliptical sentence in a particular context. In particular, it is conceivable that in languages of Types 2 and 3, there might exist real "pro-drop" constructions which are analysable in the usual way, as involving a null element occurring in an empty syntactic position. However, the existence of such zero entities would have to be well-motivated on language-internal grounds, for example as a product of paradigmatic pressure.

21. However, these meanings may be distinguished, at least in part, through different assignments of headedness. Recall that in accordance with (II), in a structure formed by identity Combination, such as that in (Tl), either of the two daughter constituents may be head, or the construction may remain headless. Thus, in (T), whereas assigning headedness to EAT yields the desired interpretation 'the chicken is eating', assigning headedness to CHICKEN results in the interpretation 'the chicken which is eating', while a headless version ends up with the interpretation 'the chicken and the eating'. These alternative assignments of headedness may or may not be reflected in the actual forms of the respective sentences in different languages. In some languages they are, corresponding to different morphological markings on each of the two expressions; this is in fact the case in Tagalog, which belongs to a similar language type, Type I' - see (W) below. But in other languages they are not, as a result of which most sentences exhibit a very wide range of possible interpretations; this is the case in another Type I' language, Riau Indonesian.

22. The above definitions differ in two important ways from the superficially similar definitions of "noun", "adjective" and "verb" proposed in Croft (1991). First, although Croft also defines his categories in terms of prototypical associations of features at different levels, his two levels are semantic (objects, properties, activities) and pragmatic (reference, modification, predication)-syntactic properties seem to play no role in his approach. Secondly, whereas Croft's categories are basically lexical, the semantically associated syntactic categories defined in (V) are-like the pure syntactic categories defined in the preceding section-neutral with respect to the distinction between lexical and phrasal.

23. From a quite different perspective, Carstairs-McCarthy (1998) argues convincingly that there is no a priori reason why languages should associate semantic categories with syntactic ones. Characterising the S-NP distinction as a grammaticalisation of the distinction between making statements and referring to things, he suggests that a language without such a distinction could fulfil the necessary communicative functions equally efficiently. To this end, he defines a hypothetical "Uniformitarian" language without distinct syntactic categories. However, he goes on to claim that such languages do not exist, and proposes an evolutionary explanation
for their absence; whereas in actual fact Tagalog and Riau Indonesian appear to provide **two** real examples of his hypothesised "Uniformitarian" language.

24. Riau Indonesian is the dialect of Malay/Indonesian spoken in informal situations by the inhabitants of Riau province in east central Sumatra. It is one of a wide range of colloquial varieties of Malay/Indonesian which are not mutually intelligible and which also differ substantially from the standardised versions of Malay and Indonesian that many general linguists have a passing familiarity with. Work in progress suggests that the characterisation of Riau Indonesian as a **Type J** language holds true, mutatis mutandis, for most or all of the colloquial varieties of Malay/Indonesian. including Kuala Lumpur Malay, ...hich is the specific dialect of example (2) above.

25. The S0S0 expressions in Riau Indonesian may be further classified with regard to their linear order in relationship to their heads. Whereas the forms in (4a--o) obligatorily precede their head S6's, those in (4p--u) invariably follow it. Only the forms in (4v--x) may occur either before or after their head S6's.

26. One may now go back to Types 2 and 3 and ask what syntactic category property words belong to in these languages. Within each type, there are **two** possibilities: either property words group with activity words, as So's in Type 2 or S1's in Type 3, or else they group with thing words, as S0S0s in Type 2 or S0'S's in Type 1. Accordingly, in the former case property words are VPs, while in the latter case they are NPs. These two possibilities thus yield a well-known typology, usually referred to as governing whether "adjectives are verbly or nounly"- see Weuer (1992, 19%) and Stassen (1997) for discussion. Of course, in addition to these two basic possibilities there is a third **intermediate** possibility, in accordance with which some property words group with activity ...ords while others group with thing words.

27. In many languages, these **two** categories. So/S0 and S1/S1, may turn out to be coextensive. In general, there is no reason why intensionally distinct categories should **not be** extensionally equivalent. An abbreviatory notation may be introduced ...hich will provide a unitary **name** for a **set** of coextensive categories- by collapsing the similar symbols and listing the distinct indices. Thus, for example, in instead of saying that in Type 4 languages. AdvPs are S0S0 and S1/S1, one may say. simply, that they are S0/S0.

28. As in the preceding case, these **two** categories. (S0S0)'(S0S0) and (S1S1)/(S0S0), are coextensive, and **can** accordingly be abbreviated as (S0/S0)(S0S0).

References

Abney, Steven
1987  
Ajdukiewicz, Kazimierz

Alsagoff, Lubna

Alves, Mark (ed.)

Anderson, John M.-Colin I. Ewen

Anderson, Stephen R.

Austin, Peter- loan Bresnan

Bach, Emmett-Robert T. Hanns (eds.)

Bailey, Charles-James N.- Roger W. Shuy (eds.)

Baker, Mark C.

Bauerle, Rainer-Christoph Schwarze-Arnim von Stechew (eds.)

Bar-Hillel, Yebosbua

Boas, Franz
1911 "Introduction", in: Franz Boas [ed.], 1-84.

Boas, Franz (ed.)

Borer, Ilagit

Carstairs-McCarthy, Andrew
Chomsky, Noam A.


Comrie, Bernard


Ccnuni-Morava, Ellerri–Barbara Goldberg (eds.)


Croft, William


Dixon, Robert M. W.

1977  "Where have all the adjectives gone?", *Studies in Language 1*: 19-80.

Dryer, Matthew S.


Efrat, Barbara S. (cd.)


Gil, David


1982b  "Quantifier scope, linguistic variation, and natural language semantics", *Linguistics and Philosophy 5*: 421–472.


1999 "Riau Indonesian as a pivotless language", in: Ekaterina V. Rachilina- Yakov G. Testelec (eds.), 187-211.

fome. a "English goes Asian: Number and (in)definiteness in the Singhilian Noun-Phrase", in: Frans Plank [ed.].

fome. b "Escaping Eurocentrism; Fieldwork as a process of unlearning", in: Paul Newman- Martha RatliT (eds.).

forthc. c "The prefixes di- and N- in MalayIndonesian dialects", in: Fay Wouk-Malcolm Ross (eds.), *Papers From the 8-ICAL Workshop on Focus*, Pacific Linguistics, Canberra.

Given, Talmy


Greenberg, Joseph H.

1963 "Some universals of grammar with particular reference to the order of meaningful elements", in Josephs H. Greenberg (ed.j, 73-113.

Greenberg, Joseph H. (ed.)


Guenter, Joshua S.- Barbara A. Kaiser-s-Cheryl C. Zoll (eds.)


Hale, Ken


Harris, Zellig S.


Hawkins, John A.


Hintikka, Jaakke

1979a "Quantifiers vs. quantification theory", in Esa Saarinen (ed.), 49-80.

Hopper, Paul J.– Sandra A. Thompson

Huang, C.-T., James

Hudson, Richard A.

Hurford, James R.- Michael Studdert-Kennedy--Chris Knight (eds.)

Jackendoff, Ray S.

Jacobs, Roderick A.- Peter S. Rosenbaum (eds.)

Jacobsen, William H., Jr.

Jelinek, Eloise
1984 "Empty categories, case, and configurationality", Natural Language and Linguistic Theory 2: 37-76.

Keenan, Edward L.
1976 "Towards a universal definition of 'subject'", in Charles N. Li (ed.). 303-333.

Keenan, Edward L.- Leonard M. Fahz

Kefer, Michel- Johan van der Auwera (cds.)

Kibrik, Aleksandr E.

Kiparsky, Paul

Lambek, Joachim

Lerdahl, Fred- Ray S. Jackendoff
Levin, Juliette

Liu, Charles N. (ed.)
1975  *Word order and word order change*. Austin: University of Texas Press.


Linguaggi nella societa e nella tecnica

Link, Godehard


1986b  *Je drei Äpfel—three apples each: Quantification and the German je*. [Manuscript, Center for the Study of Language and Information, Stanford.]

Luksaneeyanawin, Sudaporn (ed.)

McCawley, James D.

Modgil, Sohan-s-Celia Modgil (eds.)

Montague, Richard
1970a  "English as a formal language", in: Linguaggi nella societa e nella tecnica 189-224.


Morrill, Glyn V.


Nichols, Johanna-Anthony C. Woodbury (eds.)
1985  *Grammar inside and outside the clause: Some approaches to theory from the field*. Cambridge: Cambridge University Press.
Plank, Frans (ed.)


Pustejovsky, James–Peter Sells (eds.)

Ross, John R.

Rachilina, Ekaterina V.- Yakov G. Testelec (eds.)

Saarinen, Esa (ed.)
1979 *Game-theoretical semantics*. (Synthese Language Library 5.) Dordrecht: Reidel.

Sanders, Gerald A.

Sapir, Edward

Schachter, Paul
1976 "The subject in Philippine languages: Topic, actor, actor-topic, or none of the above?", in Charles N. Li (ed.), 491-518

Saumjan, Sebastian K.
1965 *Strukturnaja lingvistika* [Structural linguistics]. Moskva: "Nauka".

Shen, Yeshayahu

Siewierska, Anna (ed.)

Stassen, Leon

Stowell, Tim

Swadesh, Morris
Taylor, John R.

Tesniere, Lucien

Tomlin, Russell S.

Van Valin, Robert D., Jr.

Harnen, Harne
1996  *The typology of adjectival predication*. (Empirical Approaches to Language Typology 17.) Berlin: Mouton de Gruyter.

Wouk, Fay–Malcolm Ross (eds.)
forthc.  *Papers From the 8-/CAL Workshop on Focus*. Canberra: Pacific Linguistics.

Yip, Moira–loan Maling–Ray S. lackendoff

Zwicky, Arnold M.