CHAPTER

19

Isolating-Monocategorial-Associational Language

David Gil

Max Planck Institute for the Science of Human History, Jena, Germany

OUTLINE

Abstract

Isolating-Monocategorial-Associational (IMA) Language is language with the following three properties: (a) *morphologically isolating,* without word-internal

Handbook of Categorization in Cognitive Science.
DOI: http://dx.doi.org/10.1016/B978-0-08-101107-2.00020-8

morphological structure; (b) *syntactically monocategorial*, without distinct syntactic categories; and (c) *semantically associational*, without distinct construction-specific semantic rules, compositional semantics relying instead on the association operator, which says that the meaning of a composite expression is associated with the meanings of its constituents in an underspecified fashion. IMA Language is present in the following five domains: (a) *semiotics:* some artificial languages are IMA Language; (b) *phylogeny:* at some stage in evolution, early language was IMA Language; (c) *ontogeny:* at some stage in acquisition, early child language is IMA Language; (d) *typology:* some languages are closer than others to IMA Language; and (e) *cognition:* IMA Language is a feature of general human cognition. The main part of this chapter is devoted to showing how one particular natural language, Riau Indonesian, comes close to displaying IMA Language.

19.1 INTRODUCTION

Imagine a hypothetical language, either natural or artificial, with the following three properties:

1.

a. *Morphologically isolating*
No word-internal morphological structure;

b. *Syntactically monocategorial*
No distinct syntactic categories;

c. *Semantically associational*
No distinct construction-specific rules of semantic interpretation (instead, compositional semantics relies exclusively on the Association Operator, defined in (2) below).

Such a language may be referred to as *Isolating-Monocategorial-Associational*, or for short, *IMA*.

Does IMA Language exist? Obviously, English is not an IMA Language, nor, to the best of my knowledge, has any other natural language been proposed to be in complete possession of the three defining properties in (1) above. Nevertheless, in this chapter, it is argued that the notion of IMA Language is of relevance to a variety of domains: semiotics, phylogeny, ontogeny, typology, and cognition. In particular, the main part of this chapter is devoted to showing that although no natural language *is* IMA Language per se, some languages may indeed come much closer to exhibiting the three properties in (1) than is commonly supposed.

19.2 WHAT IMA LANGUAGE IS LIKE

The three defining properties of IMA Language pertain to three different linguistic domains, morphology, syntax, and semantics;

logically, they are thus independent of each other. Accordingly, one may imagine various other kinds of hypothetical languages with different subsets of the three properties, e.g., a language that is isolating but not monocategorial or associational.

The defining properties of IMA Language represent the limiting points of maximal simplicity within each of the three domains, morphology, syntax, and semantics. Hence, for each domain, one may imagine languages approaching these end points along a scale of decreasing complexity. Accordingly, a language is increasingly isolating as it has less and less morphological structure, increasingly monocategorial as its syntactic categories decrease in number and importance, and increasingly associational as its construction-specific rules of semantic interpretation become fewer and less distinct. Alongside *Pure IMA Language*, as in (1) above, one may thus entertain the possibility of a range of *Relative IMA Languages*, approaching Pure IMA Language to various degrees within each of the three domains.

19.2.1 Isolating

The first defining property, *morphologically isolating*, is the one that is most familiar, since it pertains to a typology that has been the focus of considerable attention in the linguistic literature. As is well known, isolating languages such as Vietnamese have considerably less word-internal morphological structure than synthetic languages such as Russian, which in turn have considerably less morphology than polysynthetic languages such as Mohawk. However, no natural language is purely isolating, as per (1a); all known isolating languages still have *some* morphology—affixation, compounding, or other kinds of processes such as reduplication, stem alternation, and so forth.

In a purely isolating language, without any morphology whatsoever, there would be no distinction between words and morphemes: every word would contain exactly one morpheme, and every morpheme would constitute a word. There would thus be no need to maintain both concepts: one of the two could be discarded. In fact, Chomsky's (1965) *Aspects of the Theory of Syntax* makes a similar proposal for English; in this model, the terminal nodes of syntactic trees, called "formatives," are actually morphemes, and the notion of word is done away with entirely. More recently, the notion of word and the distinction between morphology and syntax has been called into question by Haspelmath (2011). In contrast, though, Anderson (1982) and others maintain that, in English and presumably all natural languages, the ways in which morphemes are put together to form words are fundamentally different from the ways in which words are grouped

together to form sentences. However, it is worth keeping in mind that although there are no purely isolating languages, this is a fact about the way languages are, not how they must necessarily be: it is not difficult to imagine a hypothetical language in which the smallest meaning-bearing units all behaved as independent words, grouping together according to syntactic principles.

19.2.2 Monocategorial

The second defining property, *syntactically monocategorial*, pertains to a domain within which the presence of cross-linguistic variation has only recently, and still only partially, been recognized. In the past, syntactic categories have generally been presumed to be universal, often in accordance with the eight parts of speech of traditional Latin grammar. Indeed, the assumption that syntactic categories must be the same in all languages has lingered on into much current linguistic work, in schools as diverse as linguistic typology and generative grammar; indeed, this assumption is in evidence whenever a linguist analyzing a language says that one word must be a noun because it means "chicken" while another word must be a verb because it means "eat."

However, in recent years an increasing body of literature has begun to examine the ways in which the inventories of syntactic categories may vary across languages; see, e.g., Gil (2000b), Rijkhoff and van Lier (2013), and others. One important issue that has attracted considerable attention has been the viability and nature of the category of adjective, the extent to which words denoting properties such as "big," "red," "good," and so forth exhibit distinct adjectival behavior, or, alternatively, are subsumed within larger categories of noun or verb; see, e.g., Dixon (1977), Wetzer (1992), Stassen (2005), and others. Another major focus has been on the universality of what is generally considered to be the most fundamental categorial distinction, namely that between noun and verb; such work has typically dealt with languages which seem, prima facie, to lack a noun/verb distinction, from families such as Munda (Bhat, 1997), Austronesian (Gil, 1993), Salish (Jelinek & Demers, 1994), and Wakashan (Swadesh, 1939). It is of course languages lacking a noun/verb distinction which come closest to being syntactically monocategorial. However, to the best of my knowledge, no language has ever actually been proposed to be purely monocategorial. In particular, most or all descriptions of languages without a noun/verb distinction still involve, at the very least, a distinction between a single open syntactic category (encompassing the equivalents of both nouns and verbs) and one or more closed syntactic category containing various "grammatical" or "functional" items.

19.2.3 Associational

The third defining property, *semantically associational*, although rooted in various common-place observations concerning the ways in which expressions derive their meanings, is nevertheless of a more novel nature. Consider the best translation of a basic transitive sentence such as "Mary hit John" into the language of your choice. How do you know who hit whom? If you chose Mandarin, then, like in English, the agent is differentiated from the patient by linear order: the agent precedes the verb while the patient follows it. However, if you chose Russian, then linear order provides no semantic information; instead, the agent is differentiated from the patient by its case marking, nominative as opposed to accusative, and by the fact that it triggers gender agreement on the past-tense form of the verb. The various rules according to which agents and patients are differentiated in English, Mandarin, Russian, and other languages constitute examples of *construction-specific rules of semantic interpretation*, as specified in (1c) above, in that they apply specifically to active transitive clauses. Most languages contain many such construction-specific rules, which, together, govern the compositional semantics of clauses, phrases, and other, more specific constructions, accounting for semantic features such as thematic roles, tense, aspect, number, definiteness, and numerous others.

Now imagine you are confronted with a three-word sentence in an unfamiliar language, armed only with a rudimentary dictionary. Somehow, you identify three word stems, meaning "Mary," "hit," and "John"; however, these three word stems bear rich additional morphological structure, and you know nothing about the grammar of the language. Can you figure out the meaning of the sentence? At first blush, the answer would seem to be no. With no information on thematic roles, tense, aspect, number, definiteness, and other such features, the sentence could mean anything from "Mary hit John" through "John will repeatedly try to hit Mary" to "John and Mary aren't hitting anybody," and so on. Still, the meaning of the sentence is hardly unconstrained: it is not very likely to mean "The rain in Spain falls mainly in the plains." Thus, although you have no knowledge of the grammar of the language, it is a safe bet, in fact a near certainty, that the meaning of the sentence, whatever it is, *has to do in some way* with "Mary," "hit," and "John."

The semantic relationship of "having to do with" may be formally represented by means of the *Association Operator*, defined as follows:

2. *The Association Operator A:*

Given a set of *n* meanings $M^1 \ldots M^n$, the Association Operator A derives a meaning A ($M^1 \ldots M^n$) read as "entity associated with M^1 and … and M^n."

Two subtypes of the Association Operator may be distinguished, the *Monadic Association Operator*, in which *n* equals 1, and the *Polyadic Association Operator*, for *n* greater than 1.

In its monadic variant, the Association Operator is familiar from a wide variety of constructions in probably all languages. Without overt morphosyntactic expression, it is manifest in cases of metonymy such as the often cited *The chicken left without paying*, where the unfortunate waiter uses the expression *the chicken* to denote the person who ordered the chicken. Using small upper-case letters to represent the meanings of individual expressions, we can represent the meaning of *chicken* in the above sentence by means of the Monadic Association Operator as A(CHICKEN), or "entity associated with chicken." The nature of the association between the entity and the chicken is left open by the Association Operator, to be filled in by context, which, in the case of a restaurant, is the obvious one involving a dishonest or perhaps forgetful customer. Similar examples are everywhere. In the *International Herald Tribune*, on 5–6 November 1994, a newspaper headline reads "Washington Turns Away From Japan Trade Fight," with a subheadline "Clinton Planning to Shift the Emphasis To Markets in Asia and Latin America." The continuation of the article makes it clear that the expressions *Washington* and *Clinton* are to be understood metonymically, as A(WASHINGTON), "entity associated with Washington," and A(CLINTON), "entity associated with Clinton," which, in the context of the article, both denote the Clinton administration.

Even more often than the above cases, the Monadic Association Operator is overtly expressed via a specific form, which is commonly referred to as a *genitive*, *possessive*, or *associative* marker. Consider, e.g., the English possessive enclitic *'s*. Application of *'s* to *John* yields the expression *John's*, which has the interpretation A(JOHN), "entity associated with John," where the nature of the association is unspecified. Some idea of how unconstrained the association is can be obtained by comparing the obvious meanings of phrases such as *John's father, John's nose, John's shirt, John's birthday, John's suggestion*, and so forth, or by considering the range of meanings of a single phrase such as *John's book*, which could denote: the book that John owns, the book that John wrote, the book that's about John, or, in more specific contexts, the book that John was assigned to write a review of, and so forth. Another example of an association marker is provided by the Mandarin form *de*. In Mandarin, the expression *Yüehàn de* has more or less the same range of interpretations as the English *John's*. However, unlike English *'s*, Mandarin *de* can apply to expressions belonging to other syntactic categories, as e.g., in *Yüehàn mâi de*, where *Yüehàn mâi* means "John buy." Many descriptions of Mandarin characterize constructions such as these as relative clauses, and assign them translations such as "One that John bought." However,

the same expression can also mean "the manner of John's buying," "the extent of John's buying," and so forth. Moreover, characterizing *de* as a genitive marker in *Yüehàn de* but at the same time as a relative clause marker in *Yüehàn mâi de* misses an obvious generalization. Specifically, in both cases, *de* shares the function of an associative marker. While in the former case, *Yüehàn de* has the interpretation A (JOHN), "entity associated with John," in the latter case, *Yüehàn mâi de* is assigned the interpretation A(JOHN.BUY), "entity associated with John's buying," an interpretation which underlies all of the available readings of the expression in question. As suggested by the two examples considered above, association markers may differ from each other in their syntactic properties. Indeed, even within English, association markers may differ from each other syntactically and also semantically, as evidenced by the numerous and sometime subtle contrasts between the enclitic *'s* and the other associative marker *of*. Such contrasts suggest that the denotations of such markers, although based on the Association Operator, may involve additional and sometimes more idiosyncratic semantic components.

In its polyadic variant, the Association Operator provides for a basic mechanism of compositional semantics in which the meaning of a complex expression is derived from the meanings of its constituent parts. In accordance with the Polyadic Association Operator, whenever two or more expressions group together to form a larger expression, the meaning of the combined expression is associated with, or has to do with, the meanings of each of the individual expressions. Obviously, polyadic association applies in a default manner throughout language; it is hard to imagine how things could be otherwise. Thus, in the little thought experiment described above, it is what made it possible to be sure that in an unfamiliar language, in the absence of any specific grammatical information, in a sentence with three words whose meanings were based on "Mary," "hit," and "John," the meaning of the sentence would still be associated in some way with "Mary," "hit," and "John," or, in terms of the Polyadic Association Operator, A(MARY, HIT, JOHN), "entity associated with Mary, hitting, and John."

One grammatical domain in which the Polyadic Association Operator is overtly visible is in genitive constructions. In many languages, genitive constructions are formed by the bare juxtaposition of the two expressions, in which case the derived meaning may be represented by means of the Polyadic Association Operator applying without any overt morphosyntactic expression. For example, in Yagua, a language isolate of north-eastern Peru, *Tomáása rooriy* has a range of interpretations resembling that of its English translation "Tom's house" (Payne and Payne, 1990:348); its meaning may thus be represented as A(TOM, HOUSE), "entity associated with Tom and house." A similar mechanism generates the set of potentially available meanings of nominal compounds in

English and other languages, though their actual meanings are usually the product of further arbitrary conventionalization. Thus, e.g., a *magnifying glass* is an instrument of magnification, a *looking glass* is a goal of looking, a *sherry glass* is a receptacle of sherry, and so on. However, even in the above examples, the actual meanings are somewhat more specific than those derived by the Polyadic Association Operator; they thus fail to be purely associational. Thus, in Yagua, genitive constructions are right-headed, hence the "entity associated with Tom and house" has the further property of being, itself, a house; a more accurate translation would have been "house associated with Tom." Similarly, English compounds such as the above ones are also right-headed, accordingly, in these cases too, the first element is construed as modifying the second one. (Some of the ways in which semantic representations based on the Polyadic Association Operator are narrowed down by the imposition of headedness are discussed in more detail in Section 19.4.2.3.)

More generally, the Polyadic Association Operator may be considered as a universal default mechanism for semantic interpretation, but one that is in most cases overridden and narrowed down substantially by the application of additional construction-specific rules. A purely associational language would be one in which there were no such further construction-specific rules of semantic interpretation, and in which, therefore, the compositional semantics were effected exclusively by the Polyadic Association Operator. It is almost certainly the case that no natural language is purely associational; however, as argued in Section 19.4, some languages may come closer to being purely associational than is generally assumed.

Thus, Pure IMA Language represents a limiting case of maximal simplicity within the domains of morphology, syntax, and semantics. One may indeed wonder whether IMA Language is capable of fulfilling the multifarious functions associated with human language in the diverse contexts in which it is used. Nevertheless, as we shall now see, IMA Language is in fact more widespread than might be expected, and can indeed fulfill a wider range of functions than might seem, prima facie, to be the case.

19.3 WHERE IMA LANGUAGE IS FOUND

IMA Language, or a system that comes close to IMA Language, is manifest in the following five distinct ontological realms:

3. a. *Semiotics*
Some artificial languages are IMA Language;
b. *Phylogeny*
At some stage in evolution, early language was IMA Language;

c. *Ontogeny*
At some stage in acquisition, early child language is IMA Language;
d. *Typology*
Some languages come closer than others to IMA Language;
e. *Cognition*
IMA Language is a feature of general human cognition.

The first three domains, semiotics, phylogeny, and ontogeny, are considered briefly in this section; the fourth, typology, is discussed in greater detail in the Section 19.4; while the fifth, cognition, is examined in Section 19.5.

19.3.1 Semiotics

IMA Language may be observed in a variety of artificial semiotic systems. One such system is the language or languages of pictograms, those familiar iconic signs that can be seen in airports, railway stations, and many other places, including in particular those that have the specific function of traffic signs. To see how pictograms instantiate IMA Language, let us consider a typical example of pictogram usage: the juxtaposition of two signs, one consisting of an arrow, the other depicting a bicycle, as represented below:

Clearly, the language of pictograms is compositional, since we can take simple signs and combine them to form more complex signs. Nevertheless, there would seem to be no evidence for any distinction between different compositional systems corresponding to the morphology and syntax of natural languages. Under the most obvious analogy, the arrow and the bicycle picture are the equivalents of words, while the combination of the two signs belongs to syntax; however, neither of the two signs has any internal meaning-bearing structure of the type that might then be characterized as morphological. (Whatever internal structure the bicycle sign may exhibit does not qualify, since these are inherent to the iconic nature of the sign; it would not make sense to characterize, say, the line depicting the handlebar as an individual morpheme, since it occurs in no sign other than the bicycle icon.) Accordingly, in the absence of anything corresponding to inflection or polysynthesis, the language of pictograms may be considered to be morphologically isolating.

Similarly, the language of pictograms would appear to be devoid of any evidence for distinct syntactic categories. In the above example,

the arrow and bicycle signs belong to the same "part of speech," in fact the only one in the language of pictograms. More generally, there are no noun signs, adjective signs, verb signs, or any other syntactic categories of signs. All signs have the same distributional privileges: any two or more signs may be juxtaposed without any constraints of the kind that are reflected in the familiar grammaticality judgements of ordinary natural languages. Thus, the language of pictograms may be also be viewed as monocategorial.

But what about the meaning of our pictogram example in (4)? In many European cities, the most common meaning of such a collocation, one that has undergone a certain degree of conventionalization, is to denote a special bicycle lane, "bicycles go thattaway." However, in at least one case, I have observed a similar combination used to point the way to a bicycle shop, "go thattaway for bicycles." Thus, the example would seem to be vague or ambiguous. If we were thinking in terms of the categories of natural language, the arrow would seem to denote an activity which may assign a thematic role to the bicycle: agent in the former meaning, goal in the latter. However, there is no reason internal to the language of pictograms to posit the existence of thematic role assignment of any kind. Rather, by means of the Polyadic Association Operator, we may represent a general unified meaning underlying the two more specific ones, with the formula A(BICYCLE, THATTAWAY), "entity associated with bicycle and with thattaway," where the details of the association are filled in by the context. In general, whenever we encounter two signs in close proximity, we assign the combination a meaning that has to do in some way with the meanings of the individual signs, in accordance with the Polyadic Association Operator. Accordingly, the language of pictograms may also be viewed as associational

In sum, then, the language of pictograms satisfies the three properties of IMA Language. Clearly, the language of pictograms does not have the entire range of expressive power associated with ordinary natural languages. Nevertheless, one can still say quite a lot with pictograms, and their functionality is boosted by a substantial reliance on context: whether our example pictogram is intended to mean "bicycles go thattaway," "go thattaway for bicycles," or perhaps something different again, can readily be inferred by the location of the sign, supported by various other contextual cues.

19.3.2 Phylogeny

Although we have preciously little direct evidence of any kind concerning the evolution of natural language, it is reasonable to suppose

that early human language was IMA Language. More precisely, the following two logically distinct hypotheses may be formulated:

4. a. *Evolution of linguistic abilities*
At some stage in evolution, the cognitive abilities of humans or prehumans were limited to the representation of IMA Language;
b. *Evolution of actual languages*
At some stage in evolution, all natural languages were IMA Language.

While hypothesis (4a) is about the evolution of cognition, or, more specifically, mental grammar, sometimes referred to as I-language, hypothesis (4b) is about the evolution of actual languages, also known as E-languages.

A commonly held position, most often associated with Chomsky and his followers, is that contemporary human linguistic abilities emerged ex nihilo in a single gigantic leap, presumably associated with a unique genetic mutation. Such a view is clearly inconsistent with hypothesis (4a); however it is agnostic with respect to hypothesis (4b), since even if human linguistic abilities went straight from nothing to what they are now, actual languages might have taken a variety of incremental paths over the course of time in order to make use of such abilities (indeed this process may still be far from complete); and one of those possible paths could easily have involved IMA Language as an evolutionary way station.

A more refined position is put forward by Bickerton (1990), who argues that man's linguistic abilities evolved into their contemporary shape through an intermediate stage which he refers to as *protolanguage*. Structurally, Bickerton's protolanguage is a form of IMA Language; however, it embodies at least one significant further restriction that is not part of IMA Language, namely that it does not permit syntactic recursion. Ontologically, too, Bickerton's protolanguage is akin to IMA Language, in that he considers it to be manifest in a variety of realms, including three of the five listed in (3) above, phylogeny, ontogeny, and cognition. Notably, however, Bickerton has nothing to say about the other two domains, semiotics and typology. Moreover, he expressly denies the existence of any "interlanguage" between protolanguage and contemporary linguistic abilities; thus, like Chomsky, his position is inconsistent with hypothesis (4a), though in the case at hand, what is at issue is a single, albeit very important structural feature, namely, syntactic recursion. Conversely, hypothesis (4a) is consistent with, but does not necessarily entail, the existence of a stage, prior to IMA Language and the evolution of recursion, corresponding to Bickerton's protolanguage.

So how might we seek support for the two evolutionary hypotheses in (4)? Although we cannot go back in time, we can jump across the branches of our evolutionary tree to see what our nearest relatives, the various primates, have accomplished in the realm of language. Many species have a lexicon of predator cries; however, since these usually involve individual cries in isolation, there is no compositionality, and hence nothing near the possible richnesses of IMA Language. A somewhat more interesting case, reported by Zuberbuhler (2002), is that of the male Campbell's monkeys, who appear to be able to juxtapose two different calls, a predator cry preceded by a "boom" sound, to produce a complex cry whose meaning seems to involve some kind of attenuation or even negation of the predator-cry meaning. However, to this point at least, no clear examples of productive compositionality of meaning-bearing signs have been attested in the naturally occurring repertoire of nonhuman primates, or any other animals.

However, amongst primates in captivity, there is an increasing body of evidence suggesting that they can be taught to master compositionality, and with it also IMA Language. Two of the more celebrated cases are those of the bonobo Kanzi (Greenfield & Savage-Rumbaugh, 1990) using lexigrams, and the orangutan Chantek (Miles, 1990) using American Sign Language. Some examples of Kanzi's spontaneous linguistic production are given below:

5. a.	LIZ HIDE	*agent—HIDE*
b.	WATER HIDE	*patient—HIDE*
c.	HIDE AUSTIN	*HIDE—agent*
d.	HIDE PEANUT	*HIDE—patient*

Kanzi's usage of lexigrams provides no evidence for morphological structure or for distinct syntactic categories; it is thus isolating and monocategorial. Moreover, as suggested by examples such as the above, it is also associational. The above examples form a miniature paradigm (schematized to the right) in which the same sign HIDE is either preceded or followed by a participant, which, as indicated by the context of the utterance given by the authors, may, in either position, be understood as either the agent or the patient. Thus, there would seem to be no evidence for any grammatical assignment of thematic roles in the lexigram usage of Kanzi. Rather, the semantic relationship between the two signs is vague. Like in the language of pictograms and example (4) above, the juxtaposition of lexigrams has a single general meaning that may be represented in terms of the Polyadic Association Operator as, for (5a), A(LIZ, HIDE), "entity associated with Liz and with hiding." Thus, the bonobo Kanzi's use of lexigrams satisfies the three properties of IMA. Similar observations hold also for the orangutan Chantek's usage of ASL.

It would seem, then, to be the case that both bonobos and orangutans are endowed with the cognitive abilities to represent IMA Language, even though they apparently have not made any use of these abilities to create any actual IMA Languages in the wild. Given that the common evolutionary ancestor of bonobos and orangutans is shared also by humans, it is thus likely that this common ancestor also had the cognitive abilities to represent IMA Language without having any actual IMA Languages. (The alternative, less parsimonious scenario would involve positing the independent development of IMA Language abilities in at least two separate evolutionary lineages.) Quite obviously, however, no primates, even in captivity and with the dedicated efforts of their caregivers, are capable of acquiring the full-blown complexities of natural human language. Thus, the linguistic capabilities of captive apes support the reconstruction of a stage in human evolution, perhaps 8 or 10 million years ago, in which the abilities to represent IMA Language were already present, in accordance with hypothesis (4a). Alongside the above, the linguistic capabilities of captive apes also increase the plausibility of hypothesis (4b), though the alternative logical possibility remains that prehuman cognitive abilities may have developed past IMA Language before actual languages ever reached the IMA stage.

It should be noted, though, that since, to the best of my knowledge, the linguistic behavior of captive apes does not provide any evidence for the mastery of syntactic recursion, the abilities of Kanzi, Chantek, and other such captive apes may equally well be characterized in terms of the more restrictive protolanguage of Bickerton. In order to provide specific support for the existence of an evolutionary stage of IMA Language, either in addition to or instead of protolanguage, evidence of a different kind is called for; at present I am not familiar with any such evidence.

19.3.3 Ontogeny

Following the dictum whereby ontogeny recapitulates phylogeny, IMA Language may also be observed in early child language. Again, whereas a nativist position, associated with Chomsky and his supporters, holds that all the complexity of adult language is, in some form or guise, present from the outset, alternative approaches to first-language acquisition point towards the more commonsensical position that language does indeed develop as the child grows older. And indeed, it would seem to be the case that children pass through a stage in which they have acquired a system resembling that of IMA Language.

In the domain of morphology, there is ample evidence that children acquiring a language with rich morphology start out by treating individual words as unanalyzable wholes, only later becoming aware of their internal structure. Thus, early child language may be characterized as isolating.

In syntax, it would seem to be the case that early child language lacks distinct syntactic categories. If such categories are defined distributionally, then of course at the one-word stage, early child language is monocategorial by definition, since all words occur in the same one-word construction. However, there is reason to believe that monocategoriality may extend also into the two- or multiword stage. In Gil (2000b) a categorial-grammar-based theory of syntactic categories is proposed which, among other things, suggests specific hypotheses about the order in which syntactic categories are acquired. And in Gil (2006b), empirical support for this theory is provided from a study of the acquisition of Jakarta Indonesian, in which, it is argued, children pass through a monocategorial but multiword stage before acquiring an additional distinct syntactic category.

Semantically, it has also been suggested that early child language is lacking in many or all of the construction-specific rules of semantic interpretation characteristic of adult language. Consider the following two examples, cited by Bloom (1973), from the spontaneous speech of Allison, at age 20 months, who is playing with a pig inside a toy truck; the pig is hurt by a sharp corner of the truck:

6. a. hurt truck — *HURT − cause*
b. hurt knee — *HURT − patient*

Like Kanzi's lexigram examples in (5), the above utterances form a mini-paradigm (indicated to the right) in which *hurt* is followed by a participant, which, as suggested by the context of the utterance, may be understood as either the cause or the patient. Accordingly, Bloom argues that there is no justification for reading into utterances such as these any kind of grammatical structure involving relations such as subject and object which determine thematic roles. Rather, the semantic relationship between the two words is underspecified. Like in the language of pictograms and the signs of captive apes, the juxtaposition of words in early child language may thus be attributed a single general meaning represented in terms of the Polyadic Association Operator as, for (6a), A(HURT, TRUCK), "entity associated with hurt and with truck." Accordingly, early child language may also be characterized as associational.

Thus, early child language passes through a stage in which it exhibits the properties of IMA Language, before moving on to develop further complexity. Note that at the two-word IMA stage, early child language

is syntactically nonrecursive; at this stage, then, the child's linguistic abilities instantiate the more restricted system of protolanguage, as indeed is pointed out by Bickerton. Whether the child retains IMA Language after acquiring syntactic recursion is an issue that is in need of further exploration; in fact, it is possible that the answer to this question may vary with the choice of target language, in accordance with the language's own typological profile.

19.4 TYPOLOGY

As suggested above, the structural properties of IMA Language are shared by three quite distinct ontological realms, artificial semiotic systems, early evolutionary stages of language, and the language of young children. However, as noted earlier, most natural languages exhibit a much greater degree of complexity than is characteristic of IMA Language.

From a typological perspective, natural languages may vary independently in the extent to which they exhibit each of the three defining features of IMA Language. Of these three features, however, only the first, pertaining to morphological structure, is readily observable in a relatively theory-neutral way; the remaining two, pertaining to syntactic and semantic structure, presuppose in-depth linguistic analysis, which may vary in its conclusions in accordance with the theoretical persuasions of the linguist conducting the analysis. Accordingly, in the present state of the art, we cannot really compare languages, but only descriptions of languages, each as seen through the eyes of a different linguist armed with different theoretical assumptions and using different research methodologies. Further compounding the problem, most linguistic descriptions, of whatever orientation, exhibit a bias in favor of positing more syntactic categories and construction-specific semantic rules than are actually warranted by the facts of the language in question; see Gil (2001b) for a discussion of the pervasive Eurocentrism that underlies this bias. Accordingly, existing descriptions of languages tend towards a systematic underestimation of the degree to which the properties of IMA Language are approximated by individual languages.

With these qualifications in mind, one may nevertheless engage in some elementary comparisons of the IMA properties of different languages. Russian, under any standard description, is as far from IMA Language as one can get: it has rich morphological structure, well-motivated syntactic categories, and lots of construction-specific rules of semantic interpretation. Vietnamese, in accordance with most descriptions, e.g., Thompson (1965), is strongly isolating, however, it

is characterized as possessing distinct syntactic categories and construction-specific rules of semantic interpretation. Tagalog, as argued in Gil (1993), comes close to being monocategorial, with but a single open syntactic category; however, it clearly has rich morphological structure and a variety of construction-specific semantic rules. Given the logical independence of the three IMA properties, collapsing the scalar nature of the three properties into an idealized binary "high/low" distinction would yield a total of eight different language types. Adopting a functionalist perspective, one might speculate that only languages that were both isolating and monocategorial would also be associational: it would seem pointless for a language to have rich morphology and a large inventory of syntactic categories and not make use of these resources for the purpose of construction-specific semantic rules. However, a formalist might reasonably retort that much grammatical structure has no obvious communicative functions, and we should therefore not have any a priori expectations concerning cooccurrences of the three defining IMA features. In fact, some tentative observations offered in Gil (2013) suggest that there may indeed exist languages instantiating all eight of the locally possible combinations of the three IMA properties.

As we shall now see, some natural languages may come surprisingly close to exhibiting the three properties characteristic of IMA Language. Following is a more detailed exploration of one particular exemplar of a Relative IMA Language: the Riau dialect of Indonesian.

19.4.1 Riau Indonesian: Overview

Riau Indonesian is the variety of Malay/Indonesian spoken in informal situations by the inhabitants of Riau and Kepulauan Riau provinces in east-central Sumatra, Indonesia. The population of these two provinces is linguistically and ethnically heterogeneous. Although the indigenous population is mostly Malay, a majority of the present-day inhabitants are migrants from other provinces, speaking a variety of other languages. Riau Indonesian is acquired as a native language by most or all children growing up in these two provinces, whatever their ethnicity. It is the language most commonly used as a lingua franca for inter-ethnic communication, and in addition, like other colloquial varieties of Indonesian, it is gradually replacing other languages and dialects as a vehicle for intra-ethnic communication.

Riau Indonesian is quite different from Standard Indonesian, familiar to many general linguists from a substantial descriptive and theoretical literature. It is one of a number of regional varieties of colloquial Indonesian, which, although different from each other in numerous

details, nevertheless share the same typological ground plans. One such regional variety is Jakarta Indonesian, referred to briefly in Section 19.3.3; see Conners, Bowden, and Gil (2015) for further discussion. The characterization of Riau Indonesian as a Relative IMA Language is this probably applicable to a wide range of colloquial varieties of Indonesian, totaling tens of millions of native speakers.

The first IMA property, morphologically isolating, clearly applies to a very great extent to Riau Indonesian; see Gil (2002a, 2004a, 2006a) for a description and analysis of various aspects of Riau Indonesian word structure. Inspection of any text will reveal a low word-to-morpheme ratio, as well as substantial stretches in which the word-to-morpheme ratio is actually one-to-one. Riau Indonesian has no inflectional morphology whatsoever, and little in the way of derivational morphology. Only three items are of clearly affixal nature and productively used: the prefixes *se-* "one," *(s)i-* marking names of persons, and *N-* marking agent-orientation. (Actually, the latter form is prefixal in only some of its allomorphs, it is otherwise sometimes realized as a proclitic *me-*.) In addition, there are a handful of items, probably less than 10, whose nature is intermediate between affixes and clitics; among these are the forms *ber-* marking nonpatient orientation, *ke-* marking direction, and *-an* with a variety of usages which may or may not be related to each other. Some other items whose cognates in Standard Indonesian are written joined on to their hosts are clearly clitics rather than affixes in Riau Indonesian; these include the forms *di-* marking patient orientation, *ter-* marking nonagent orientation, and *-kan* marking end-point orientation. In fact the most commonly occurring bound morphemes in Riau Indonesian are actually suprasegmental rather than linear. The most important of these is reduplication, usually complete though sometimes partial, which has a variety of usages; see Gil (2005a) for detailed discussion and analysis. Another is truncation, used productively to create familiar forms from names, e.g., *Ril* from *Kairil*, or other terms of address, e.g., *bang* from *abang* "elder brother." Finally, like most or all languages, Riau Indonesian makes use of compounding, though with two important qualifications. First, the two terms of the compound are less strongly bound to each other than in many other languages; in fact, there would appear to be no phonological grounds to distinguish between compounds and phrasal collocations (corresponding to, say, the stress shift that is evident in English, or the construct-state inflection that is characteristic of Hebrew). Secondly, compounds would appear to be less common than in many other languages. This is particularly striking in comparison with the isolating languages of mainland Southeast Asia: Chinese, Thai, Vietnamese, and so forth. Although the latter languages are traditionally thought of as monosyllabic, a substantial body of phonological literature suggests

that they are also characterized by a bisyllabic minimal word; see, e.g., Bao (1990), Yip (1991), and Feng (2002) for Sinitic languages. And indeed, one of the most productive devices for achieving the canonical bisyllabic word in such languages is that of compounding. However, in Riau Indonesian, the canonical monomorphemic word is already bisyllabic, and possibly for this reason, compounding occurs much less frequently. Indeed, the scarcity of compounding in Riau Indonesian in contrast to the monosyllabic languages of mainland Southeast Asia suggests that Riau Indonesian may actually have a better claim to represent the most extreme case of an isolating language.

But what about the other two IMA properties? In order to evaluate the extent to which these apply, we need to take a deeper look into the syntactic and semantic patterns of Riau Indonesian. (Much of what follows in Sections 19.4.1 and 19.4.2 is an abridged version of a more detailed discussion and analysis presented in Gil, 2005b.) As a point of departure we shall take the following English sentence:

7. The chicken is eating

How might one go about translating the above sentence into Riau Indonesian? Following are two natural and idiomatic translations:

8. **a.** Makan ayam
eat chicken
b. Ayam makan
chicken eat
"The chicken is eating"

Sentences (8a) and (8b) each consist of two monomorphemic words, "eat" and "chicken"; the only difference between them is with respect to word order. They mean the same thing, and they are of equal naturalness, though their pragmatic appropriateness conditions differ somewhat.

How similar are (7) and (8)? A pedagogical grammar of Malay (quoting an anonymous source) has the following to say:

> . . .the Malay and English sentence structures are so similar that one scholar has even remarked that 'Indonesian (in this case Malay), is a western language using Indonesian (Malay) words'. ***Liaw (2002: iv)***

This view is implicit in typologies such as Greenberg (1963), who classifies the world's languages into 24 different word-order types, and puts Malay into the same cell as many European languages, including the Romance languages and Modern Greek. And it is explicit in much recent work within the generative framework on Malay/Indonesian, such as Guilfoyle, Hung, and Travis (1992), who propose syntactic structures that are well-nigh indistinguishable from those of English.

It takes a novelist, albeit one with great linguistic sensitivity, to see the obvious differences between Malay and English, which have escaped the sight of so many linguists. Here is the perspective of Anthony Burgess:

> What strikes the learner of Malay is the complete lack of those typically Indo-European properties—gender, inflection, conjugation. It is like diving into a bath of pure logic. Everything is pared to a minimum. [...] If one digs deeply enough into Malay, one comes to the conclusion that the Western concept of 'parts of speech' is alien to it.
>
> ***Burgess (1975: 183,184)***

Anthony Burgess was right on. Let us now take Burgess up on his suggestion and begin digging.

Table 19.1 summarizes some of the more salient differences between sentence (7) in English and its two translations into Riau Indonesian in (8).

The first difference in Table 19.1 is a formal one. English sentence (7) exhibits numerous structural asymmetries. Two morphosyntactic asymmetries are easily visible on the surface: the Noun Phrase *the chicken* controls agreement of the auxiliary *is*, and the auxiliary *is* in turn governs the *-ing* ending on the verb. Lurking beneath these

TABLE 19.1 A Contrastive Analysis of (7) and (8)

	English ***The Chicken is Eating***	**Riau Indonesian** ***Makan Ayam/Ayam Makan***
Symmetry	asymmetric: agreement: *The chicken* → *is* government: *is* → *-ing*	symmetric
Number (on CHICKEN)	marked: singular	unmarked: also... "The chickens are eating"
Definiteness (on CHICKEN)	marked: definite	unmarked: also... "A chicken is eating"
Tense (on EAT)	marked: present	unmarked: also... "The chicken was eating" "The chicken will be eating"
Aspect (on EAT)	marked: progressive	unmarked: also... "The chicken eats" "The chicken has eaten"
Thematic role (on CHICKEN)	marked: agent	unmarked: also... "Someone is eating the chicken" "Someone is eating for the chicken" "Someone is eating with the chicken"
Ontological type (on CHICKEN EAT)	marked: activity	unmarked: also... "The chicken that is eating" "Where the chicken is eating" "When the chicken is eating"

morphosyntactic asymmetries are a host of syntactic asymmetries, providing the motivation for grammatical analyses of sentences such as (7) as involving subject and predicate, Noun Phrase and Verb Phrase, or whatever. In contrast, the Riau Indonesian sentences in (8) are completely symmetric, their two constituent parts being totally balanced. There is no morphological agreement or government, either in (8) or anywhere else in Riau Indonesian. Moreover, the lack of morphological asymmetries mirrors the absence of any deeper syntactic asymmetries. As argued in detail in Gil (2013), words such as *makan* "eat" and *ayam* "chicken" have the same distributional privileges and, more generally, identical syntactic behavior. They thus belong to the same syntactic category, in fact the only open syntactic category in Riau Indonesian, namely S. Structurally, then, the two Riau Indonesian sentences in (8) are instances of sentential coordination, with a structure of the form $[_S$ S S], as represented in (10) below.

The remaining differences in Table 19.1 are semantic. In English (7), the subject Noun Phrase is marked for number and definiteness, like most other Noun Phrases in English. In contrast, in Riau Indonesian (8), *ayam* "chicken" is unmarked for number and definiteness; in Riau Indonesian, number marking is almost completely absent, while definiteness marking is optional. Thus, Riau Indonesian (8) has a wider range of interpretations than its English counterpart, as suggested by the additional translations of (8) back into English in Table 19.1. Similarly, in English (7), the verbal phrase *is eating* is marked for tense and aspect, like most other verbal phrases in English. In contrast, in Riau Indonesian (8), *makan* "eat" is unmarked for tense and aspect; in Riau Indonesian, these two categories are expressed by optional periphrastic devices which are for the most part only weakly grammaticalized. Once more, Riau Indonesian (8) has a wider range of interpretations than its English counterpart, as suggested by the additional translations of (8) back into English in Table 19.1.

Whereas the absence of number, definiteness, tense, and aspect marking are familiar areal features of Southeast Asian languages, the remaining two characteristics of Riau Indonesian are perhaps somewhat more exceptional from a cross-linguistic point of view. In English (7), the Noun Phrase *the chicken* is marked as bearing the thematic role of agent. In general, thematic roles are central to the grammatical organization of English and of many other languages. In contrast, in Riau Indonesian (8), the expression *ayam* "chicken" is not marked for thematic role; as suggested by the alternative translations of (8) back into English in Table 19.1, *ayam* "chicken" could also be interpreted as patient, or, given an appropriate context, as benefactive, as comitative, or as standing in any other thematic role whatsoever. The indeterminacy of thematic roles in Riau Indonesian is exemplified and discussed

in detail in Gil (1994, 2002b), and is argued in Gil (2001b, 2005b) to be an instance of vagueness rather than ambiguity.

The final difference between (7) and (8) presented in Table 19.1 is perhaps the most fundamental one; it pertains to the ontological type of the expressions. Whereas English (7) denotes an activity, Riau Indonesian (8) is unmarked for ontological type. Again, Riau Indonesian (8) has a wider range of interpretations than its English counterpart, as evidenced by the additional translations of (8) back into English in Table 19.1. As suggested by these translations, the sentences in (8) could also denote a thing ("The chicken that is eating"), a place ("Where the chicken is eating"), a time ("When the chicken is eating"), and so on. Again, the indeterminacy of ontological types in Riau Indonesian is exemplified and discussed in detail in Gil (2001b, 2005b), where it is also argued that such indeterminacy is an instance of vagueness rather than ambiguity. In turn, such vagueness is then argued, in Gil (2012), to support an analysis of Riau Indonesian making little or no reference to the notion of predication and the distinction between predication and attribution.

Thus, as summarized in Table 19.1, the Riau Indonesian sentences in (8) differ in fundamental ways from their English counterpart in English (7); in fact, they bear a much greater resemblance to the pictogram example in (4), Kanzi's lexigram examples in (5), and the early English child-language examples in (6). Formally, the Riau Indonesian sentences lack any evidence of asymmetrical structure; semantically, they are unmarked and in fact vague with respect to the categories of number, definiteness, tense, aspect, thematic role and ontological type. Anthony Burgess was thus right; this is indeed a language "pared to a minimum." In fact, this minimum is one that provides a relatively close approximation to IMA Language.

19.4.2 Riau Indonesian: Analysis

We shall now sketch the outlines of Riau Indonesian syntax and semantics, proposing explicit representations for the observations made above. In doing so, we shall continue to use the two sentences in (8) as a convenient point of reference.

19.4.2.1 Syntax

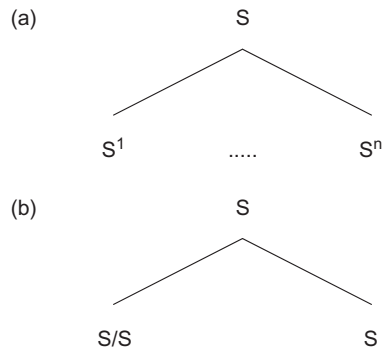
As argued in Gil (1994, 2000b, 2001b), in Riau Indonesian there is but a single open syntactic category S, or sentence. All members of S exhibit the same syntactic behavior, including the same distributional privileges. In particular, all members of S can stand alone as complete nonelliptical sentences. The category S includes *makan* "eat" in (8), and practically all other words whose translational equivalents into English

are verbs, and also *ayam* "chicken" in (8), and just about all other words whose translational equivalents into English are nouns. In addition, the category S includes most words whose translational equivalents into English are adjectives, prepositions, and determiners, plus a variety of words whose closest English counterparts are function words or morphemes: among such words are *tak*, marking negation; *udah*, denoting the perfect; *sendiri*, which expresses a variety of notions including restrictive focus, intensification, and reflexivity (Gil, 2001c); and *sama*, whose usages range over categories such as nonabsolutive, conjunction, togetherness, reciprocity, and sameness (Gil, 2004b). Alongside individual words, the category S includes all multiword expressions in the language, among which are *makan ayam* and *ayam makan* in (8). However, in addition to the open syntactic category S there is also a closed syntactic category S/S, which contains a couple of dozen semantically heterogeneous words, including *kalau* marking topics, *tiap* "every," *dengan* "with," "and," and others. Thus, Riau Indonesian comes close to being purely monocategorial; it is only the existence of the closed syntactic category S/S which prevents it from actually being so.

The syntax of Riau Indonesian can be stated very simply. Syntactic structures are hierarchic but unordered trees, in which each node is labeled with one of the two syntactic categories, S and S/S. Of course, in any physical representation of such trees on a page, it is impossible not to introduce a linear order; however, it is important to keep in mind that such order is not part of the actual representation. A number of scholars working within different theoretical frameworks have provided arguments in support of unordered tree structures and the representational separation of hierarchic structure and linear order; see, e.g., Sanders (1975), Keenan (1978), Keenan and Faltz (1986), Kayne (1994), and Bury (2005).

Syntactic tree structures are formed from subtrees of the following two kinds:

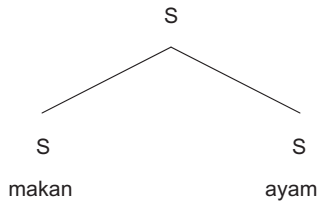
9. *Two kinds of subtrees*



In (9a), *n* expressions belonging to S combine with each other to constitute a superordinate S. Although there is no strict upper limit on the size of *n*, branching is most commonly binary, and rarely goes beyond ternary. Formally, (9a) has the structure of a coordination, in which each of the constituent parts is equally ranked. In (9b), a single word belonging to the category S/S combines with a single expression belonging to the category S to yield a superordinate expression of category S. The category name S/S reflects this fact, making use of the familiar "slash" operator from categorial grammar. Unlike (9a), the structure in (9b) is asymmetric. Complex hierarchic structures are built up recursively from the two kinds of subtrees represented in (9).

For present purposes, we shall be concerned only with the former of the two kinds of subtrees. Setting *n* equal to 2, (9a) yields a representation for the syntactic structure of the two sentences in (8):

10. *Syntactic structure of (8)*:



Since the structure in (10) is unordered, it provides an equally appropriate representation for both *makan ayam* in (8a) and *ayam makan* in (8b). To say that the two words *makan* and *ayam* belong to the same syntactic category S is to say that they exhibit the same syntactic behavior, including, specifically, the same distributional privileges as each other and as all other members of S. In addition, as indicated above, they can combine with each other to yield the superordinate S expressions *makan ayam* and *ayam makan*, which, once more, share the same syntactic behavior and distributional privileges. As suggested in (10), the two sentences *makan ayam* and *ayam makan* are of identical syntactic structure, that of a sentential coordination. This reflects the fact, discussed in the previous section, that the two constituent words are equally ranked, lacking in any structural asymmetries such as agreement, government, and the like.

This, then, in a nutshell, is the syntax of Riau Indonesian. So far, I have found no evidence for syntactic categories other than the open category S and the closed category S/S, and no evidence for syntactic structures other than those that can be built up recursively from subtrees such as those in (9). In particular, I have found no evidence for any kinds of empty syntactic positions, or for any kinds of structural

dependencies of the type commonly expressed by rules of movement. Thus, on the available evidence, there is indeed ample reason to characterize the syntax of Riau Indonesian as very simple indeed, in fact close to monocategorial.

19.4.2.2 Semantics

The fundamental semantic structure of Riau Indonesian is also simple. Every expression in Riau Indonesian has a basic semantic structure in the form of an unordered tree that is isomorphic to that of its syntactic structure: each node of the semantic structure represents the interpretation of the corresponding node of the syntactic structure of the expression. Whereas the interpretation of terminal nodes is specified in the lexicon, that of nonterminal nodes is derived by compositional principles from that of their constituent nodes.

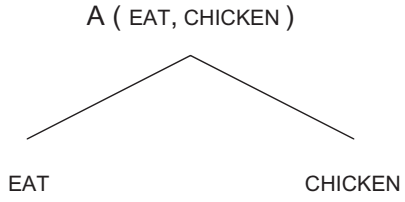
In fact, the overwhelming majority of the compositional semantics of Riau Indonesian can be captured in a single simple rule making reference to the Polyadic Association Operator:

11. *Polyadic Association Rule of Semantic Compositionality*:
Given a syntactic structure $[_X X^1 \ldots X^n]$ $(n > 1)$ where $X^1 \ldots X^n$ have interpretations $M^1 \ldots M^n$, respectively, $[_X X^1 \ldots X^n]$ is assigned the interpretation $A(M^1 \ldots M^n)$.

The Polyadic Association Rule says, quite simply, that whenever two or more expressions are combined, the meaning of the combination is obtained by applying the Association Operator to the meanings of the individual expressions. In other words, when X^1 to X^n, with meanings M^1 to M^n, are put together, the resulting meaning is $A(M^1 \ldots M^n)$, or "entity associated with M^1 to M^n." Since the constituent meanings M^1 to M^n are unordered and equally ranked, the resulting meaning may be characterized as a conjunction. The Polyadic Association Rule thus provides an unified semantic representation for Riau Indonesian sentences, reflecting their characterization as vague with respect to thematic roles and ontological types.

The way in which the Polyadic Association Rule works may be illustrated through the semantic representation that it provides for the sentences in (8):

12. *Semantic structure of (8)*:
"entity associated with eating and with chicken"



In (8), *makan* means EAT and *ayam* means CHICKEN. The Polyadic Association Rule applies to the collocation of these two meanings, and assigns them the interpretation A(EAT, CHICKEN), "entity associated with eating and with chicken." Since EAT and CHICKEN are unordered and equally ranked, the interpretation A(EAT, CHICKEN) is thus a completely symmetric conjunction. The above structure constitutes a single unified meaning, encompassing the entire range of interpretations of the sentences in (8), including, among others, those expressed by the various translations of (8) into English provided in Table 19.1. In particular, it accounts for indeterminacy with respect to thematic roles, allowing for the chicken to assume any role whatsoever in relation to the eating; and for indeterminacy with respect to ontological types, permitting *makan ayam* and *ayam makan* to denote activities, things, places, times, and so on.

The Polyadic Association Rule thus constitutes the basic mechanism governing semantic compositionality in Riau Indonesian. In doing so, it provides a way to represent the semantic indeterminacy that is so prevalent in the language. The basic semantic structures produced by the Polyadic Association Rule are of an absolutely minimal degree of specificity, adding nothing substantive to the combination of the constituent meanings other than to say that they are related in some way, left open to context. The central role that the Polyadic Association Rule plays in the compositional semantics of Riau Indonesian thus supports the characterization of Riau Indonesian as an associational language.

19.4.2.3 Further Analysis

In the preceding pages, we have seen how Riau Indonesian exhibits each of the three properties of IMA Language to a substantially greater extent than many other languages, and perhaps also to a greater degree than is often supposed to be possible in a natural human language. Nevertheless, Riau Indonesian is still a considerable way off from the limiting case of Pure IMA Language.

The Polyadic Association Rule produces basic semantic structures forming a skeleton which may be subsequently fleshed out by further more specific semantic rules applying whenever appropriate to produce more elaborate representations, involving domains such as coreferentiality, quantifier scope, conjunctive (focus) operators, and many others. One of the most important kinds of semantic enrichment is that of head-modifier structure.

Headedness is a property that is not limited to syntax, or even to language, but rather is characteristic of a wide range of cognitive domains.

Headedness may apply wherever hierarchical tree structure is present, in accordance with the following rule:

13. *Headedness assignment rule*:
Given a structure $[_X X^1 \ldots X^n]$, one of its constituents, X^j, may be coindexed with the entire structure for similarity: $[_X X^1 \ldots X^j{}_i \ldots X^n]_i$.

In a structure X consisting of X^1 to X^n, one of the constituents, X^j, is singled out as bearing a resemblance to the entire structure, X, with respect to a certain unspecified feature. In such a case, X^j is said to be the *head* of the structure X, and all the other X^i (where $i \neq j$) are the *modifiers* of X^j.

Headedness, as defined above, is present in a variety of cognitive domains; see Gil (1985) for discussion. Imagine a plate on which, in roughly equal proportions, are a piece of chicken, some beans, and a mound of rice; now imagine that, having just consumed the food on the plate, somebody asks you what you had for dinner. You are unlikely to answer by simply listing the three items on the plate. If you are a European, you are likely to respond with the simple answer "chicken." And if you are a Southeast Asian, you are equally likely to answer "rice." What these responses show is that both Europeans and Southeast Asians assign headedness to dishes of food, though the specific choice of head varies. Europeans think of the chicken as the head, and hence of the dish as a whole as a chicken dish, not a bean dish or a rice dish. And Southeast Asians conceptualize the rice as the head, and the dish as a whole as a rice dish, not a bean dish or a chicken dish. Thus, although the details differ, both Europeans and Southeast Asians assign headedness to dishes of food. Such assignment of headedness is reflected not just in the ways people talk about their food, but also in a variety of behavioral patterns. For example, whereas for a European, missing out completely on the modifying rice might be of relatively little import, for many Southeast Asians, every single mouthful of food must contain at least some rice, otherwise it would be headless and hence ill-formed and uneatable. This suggests that the headed hierarchic structures of food dishes are extra-linguistic, part of general cognition. Of course, headed hierarchic structures are not specific to food dishes; in one guise or another, headedness lies at the heart of many theories of particular domains of cognition. For example, in tonal music, heads and modifiers form the basis of Lerdahl and Jackendoff's (1983) theory of time-span reductions, which accounts for the ways in which a complex melody may be successively peeled of its less-important modifying elements, retaining at each stage a smaller and smaller skeletal melody consisting entirely of heads. Within language, too, headedness, as defined in (13) above,

is present in a variety of domains, ranging from narrative discourse, as suggested by Shen (1985), through syntax, as in X-bar theory proposed by Jackendoff (1977), all the way to syllable structure, as argued by Anderson and Ewen (1987).

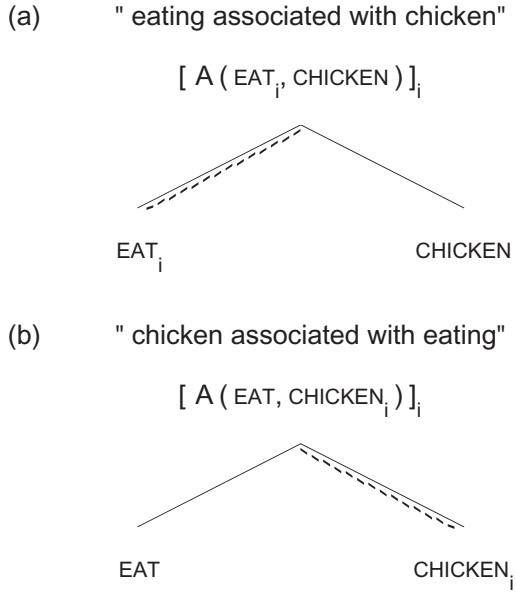
The following rule, a particular case of the Headedness Assignment Rule in (13) above, assigns headedness in the domain of basic semantic structures built up by the Polyadic Association Rule:

14. *Headedness assignment rule for associative interpretations*:
Given an associative interpretation $A(M^1 \ldots M^n)$, one of its constituent substructures, M^j, may be coindexed with the entire semantic structure for coreferentiality: $[A(M^1 \ldots [M^j]_i \ldots M^n)]_i$.

In (14), the general notion of similarity referred to in (13) is replaced by a more specific kind of similarity, namely, coreference. In a headed semantic structure, the head constituent projects its referential identity up to the entire meaning, whose range of interpretations is accordingly narrowed down. Thus, a headed semantic structure is more specific than the corresponding headless structure.

The effect of assigning headedness to basic semantic structures may be illustrated through the application of headedness to the interpretation of the two sentences in (8) shown in (12) above:

15. *Semantic structure of (8) enriched with headedness*



In (15) above, headedness is depicted twice: by coindexation, in accordance with the definition in (14), and, redundantly but for purposes of greater clarity, by an extra dashed line running up from the head constituent to the root node, tracing the projection of referential identity. The two structures in (15) represent the two possible assignments of headedness: in (15a) EAT is assigned headedness, while in (15b) CHICKEN is the head.

By projecting referential identity, headedness narrows down the range of possible interpretations of the superordinate meaning. In (15a), head EAT projects its identity up to [A(EAT, CHICKEN)]. Accordingly, the superordinate interpretation no longer denotes an arbitrary "entity associated with eating and chicken" but rather the actual eating, or, more specifically, "eating associated with chicken." The resulting interpretation retains its indeterminacy with respect to number, definiteness, tense, aspect, and thematic roles, but loses its indeterminacy with regard to ontological type, which is now identical to that of the head EAT, namely, activity. Thus, when EAT is assigned headedness, the interpretation necessarily denotes eating. Some of its possible translations into English might include "The chicken is eating," "Someone is eating the chicken," and so forth. Conversely, in (15b), head CHICKEN projects its identity up to [A(EAT, CHICKEN)]. As before, the superordinate interpretation no longer denotes an arbitrary "entity associated with eating and chicken" but instead denotes the actual chicken, "chicken associated with eating." Again, the resulting interpretation retains its indeterminacy with respect to number, definiteness, tense, aspect, and thematic roles, while losing its indeterminacy with regard to ontological type, which is now identical to that of the head CHICKEN, namely, thing. Accordingly, when CHICKEN is assigned headedness, the interpretation necessarily denotes chicken. Some of its potential translations into English might include "The chicken that is eating," "The chicken that someone is eating," and so forth. Thus, as illustrated above, headedness reduces the range of available interpretations of the superordinate meaning, limiting them to ones that are consistent with properties projected upwards from the head constituent.

How prevalent is headedness in the syntax of Riau Indonesian? As evidenced by the widespread occurrence of vagueness with respect to ontological categories, the Headedness Assignment Rule frequently fails to apply. If the semantic structures of such examples were headed, the head would project its ontological type

up to the superordinate interpretation, and, in doing so, restrict its meaning to one ontological type to the exclusion of the other. Nevertheless, in many other cases, the superordinate interpretation is more limited, in ways that suggest that the semantic structure may indeed be headed. In particular, work in progress suggests that the presence of headed structures underlies certain observable effects involving long-distance dependencies bearing a superficial resemblance to island constraints in other languages; see Gil (2000a). So from the availability of both possibilities, it must be concluded that application of the Headedness Assignment Rule is optional.

It should be noted, however, that in real life, i.e., to say when analyzing naturalistic data, it is often the case that although two different interpretations are available, one is more prominent, and the other more difficult to obtain. In order to represent such states of affairs, it is necessary to introduce an element of fuzziness into the description. This is easily achieved by attributing to each potentially available assignment of headedness a scalar figure representing the degree to which it is conceptually salient. An essentially identical mechanism, involving "preference rules," plays a central role in Lerdahl and Jackendoff's (1983) theory of tonal music, accounting for, among other things, assignments of headedness in the form of time-span reductions.

Headedness is what underlies a large proportion of the word-order constraints that are in evidence in Riau Indonesian. The relevant principle is as follows:

16. *Head-initial order*:
Heads precede modifiers.

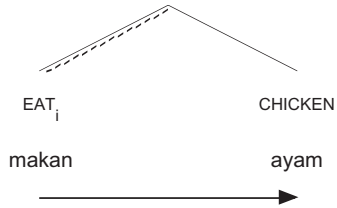
Since headedness is optional, this principle applies only in those cases where headedness is assigned. Nevertheless, head-initial order alone accounts for much of the "word-order typology" of the language.

To see how this works, let us examine the application of head-initial order to (8) above. Some of the apparent properties of the respective constructions, those that might perhaps be attributed to them within conventional grammatical descriptions, are indicated, on the right-hand side, within scare quotes.

17. *Semantic structure of (8) enriched with headedness and linearized*

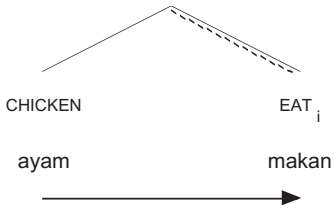
(a) "eating associated with chicken" (8a) head-initial

[A (EAT_i, CHICKEN)]$_i$ "predicative," "verb-initial"



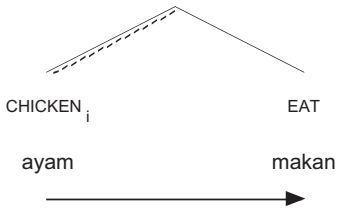
(b) # "eating associated with chicken" (8b) head-final

[A (EAT_i, CHICKEN)]$_i$ "predicative," "verb-final"



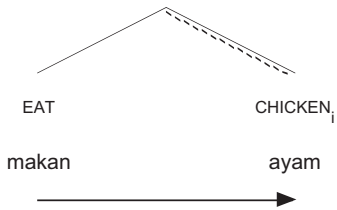
(c) "chicken associated with eating" (8b) head-initial

[A (EAT, $CHICKEN_i$)]$_i$ "attributive," "noun-initial"



(d) # "chicken associated with eating" (8a) head-final

[A (EAT, $CHICKEN_i$)]$_i$ "attributive," "noun-final"



In (17a) and (17b), EAT is the head, as per the semantic structure in (15a), and therefore the expression as a whole denotes eating. In contrast, in (17c) and (17d), CHICKEN is the head, in accordance with the semantic structure in (15b), and hence the expression as a whole denotes chicken. In a conventional grammatical description, (17a) and (17b) would be characterized as having a "sentential" interpretation in which the eating is "predicative," while (17c) and (17d) would be associated with a "nominal" interpretation in which the eating is "attributive."

Within each of the above pairs, the preferred head-initial linearization, shown first, is contrasted with the dispreferred head-final linearization, shown beneath it, and marked with a "#." Comparing (17a) and (17b), we see that in (17a) head-initial order creates the appearance of a "verb-initial" order, while in (17b) head-final order creates the opposite appearance of a "verb-final" order. Thus, the preference for head-initial order, in accordance with (16), results in an apparent preference for "verb-initial" order, as in sentence (8a), over "verb-final" order, as in sentence (8b). Moving on to (17c) and (17d), we see that in (17c) head-initial order creates the appearance of a "noun-initial" order, while in (17d) head-final order creates the opposite appearance of a "noun-final" order. In this case, then, the preference for head-initial order results in an apparent preference for "noun-initial" order, as in sentence (8b), over "noun-final" order, as in sentence (8a). Thus, if the speaker wishes to convey a specifically "predicative" meaning, with EAT as head, head-initial order will entail a preference for sentence (8a), *makan ayam*, over sentence (8b), *ayam makan*. Conversely, if the speaker wants to express a specifically "attributive" meaning, with CHICKEN as head, head-initial order will dictate a preference for sentence (8b), *ayam makan*, over sentence (8a), *makan ayam*.

The above observations may be restated from the hearer's perspective, by taking the sentences in (8a) and (8b) as the starting point. In sentence (8a), *makan* "eat" precedes *ayam* "chicken." Head-initial order, as in (17a), creates the appearance of a "verb-initial" order, whereas head-final order, as in (17d), creates the appearance of a "noun-final" order. Thus, the preference for head-initial order entails a preference for the "verb-initial predicative" interpretation over its "noun-final attributive" counterpart. In contrast, in sentence (8b), *ayam* "chicken" precedes *makan* "eat." Head-initial order, as in (17c), creates the appearance of a "noun-initial" order, whereas head-final order, as in (17b), creates the appearance of a "verb-final" order. Thus, the preference for head-initial order entails a preference for the "noun-initial attributive" interpretation over its "verb-final predicative" counterpart.

Thus, head-initial order, as formulated in (16) above, creates the appearance of a language with an array of word order correlates

characteristic of a "verb-initial" language, including, in particular, "noun-attribute" order within Noun Phrases. Moreover, it does so on the basis of the grammatical description presented above, making reference solely to syntactic structures involving a single open syntactic category S, plus polyadic association and headedness. Thus, even a Relative IMA Language, with an extremely impoverished inventory of grammatical categories, can still appear to display a typologically conventional pattern of word-order preferences.

Still, the above account is not entirely complete. Consider the predictions of head-initial order with regard to the sentences in (8), as represented in (17). Head-initial order makes the correct prediction that (17a) and (17c) will be preferred, and also the correct prediction that (17d) will be dispreferred; however, its prediction that (17b) will be dispreferred is in need of further qualification. Let us take a closer look at (17b), representing sentence (8b), *ayam makan*, with a head-final semantic structure characterizing *makan* "eat" as head, and therefore denoting eating. In actual fact, the acceptability of (17b) is dependent on the thematic role of *ayam* "chicken." For most potential thematic roles, (17b) is indeed dispreferred relative to (17a), in accordance with head-initial order. However, for the single case in which *ayam* "chicken" is the agent of *makan* "eat," (17b) is more acceptable, and, in fact, is as readily available as its head-initial counterpart in (17a). In other words, in order to say "The chicken is eating," (8b) *ayam makan* is every bit as good as (8a) *makan ayam*: Riau Indonesian appears to be at least as "verb medial" as it does "verb initial." In order to account for such facts, additional principles of linearization making specific reference to thematic roles are thus required. More generally, as argued above, many semantic structures in Riau Indonesian are unheaded, and for them, of course, the principle of head-initial order is inapplicable. Nevertheless, all such structures end up underlying strings of words which occur one after another in nonrandom fashion, thereby revealing the need for additional principles of linearization. Two such principles governing word order in Riau Indonesian, making reference to iconicity and information flow, are proposed in Gil (2005b).

19.4.3 Riau Indonesian: A Relative IMA Language

As shown in the previous pages, Riau Indonesian comes closer than is commonly thought possible to displaying the three properties of IMA Language: morphologically isolating, syntactically monocategorial, and semantically associational. In comparison to languages such as Russian, Riau Indonesian bears a closer resemblance to other instances of IMA Language such as the language of pictograms, captive apes,

and infants. Of course, Riau Indonesian is still a long way from instantiating Pure IMA Language: it does have some morphology, a second closed-class syntactic category, and various construction-specific rules of semantic interpretation. However, as suggested in Section 19.4.2 above, the bulk of the language is indeed pure IMA: nothing beyond IMA structure is necessary for the morphological, syntactic, and compositional-semantic analysis of basic sentences such as (8). Similarly, more extensive investigations show that nothing beyond IMA structure is required for the representation of the most important properties of sentences whose translational equivalents into English involve complex constructions such as questions, reflexives, relative clauses, and sentential complements. Indeed, examination of the various non-IMA items in Riau Indonesian, the bound morphemes and the members of the closed syntactic category S/S, suggests that they form a heterogeneous set, with no specific characteristic functions of importance to the overall organization of the language.

As a Relative IMA Language, Riau Indonesian is thus simpler than many other natural human languages. In previous eras, there was a widespread belief that, in comparison with their European counterparts, the languages of Africa, Asia, and the Americas were simpler, or more primitive, or plain inferior; in many cases these assumptions about the languages were coupled to other assumptions about their speakers which today would be judged as morally reprehensible. With the advent of modern linguistics and greater familiarity with the world's languages, such beliefs were duly discarded; however, their place was taken not by serious empirical investigation of the issues involved, but rather by another dogma, to the effect that all languages are of roughly equal overall complexity. In part, this dogma stems from extraneous considerations having to do with "political correctness"; but there are other, more substantive motivations: linguistics over the course of the last century has simply chosen to concern itself with a different range of issues, and besides, perhaps most importantly, complexity of linguistic structure is a notion that is extremely difficult to formalize in an explicit and quantitative manner. None of the above, however, should be reasons not to try and address the issue of complexity, as indeed has been suggested in studies such Comrie (1992), Romaine (1992), and McWhorter (1998, 2000, 2001, 2002), and more recently in three edited volumes, Miestamo, Sinnemäki, and Karlsson (2008), Sampson, Gil, and Trudgill (2009), and Newmayer and Preston (2014).

Given the efficiency with which Riau Indonesian fulfills the multifarious functions of a natural human language with so little grammatical machinery, one can only wonder why it is not the case that all languages are like Riau Indonesian. When a speaker of Riau Indonesian

asks how you say in English something like (8b) *Ayam makan* and is answered with something like (7) *The chicken is eating*, the speaker's next question is likely to be: So if *chicken* is *ayam* and *eat* is *makan*, then what are *the*, *is*, and *-ing* all about? Rather than adopting the stance of an English teacher or grammarian and explaining definiteness, tense, aspect, agreement, and government, it is actually worth trying to enter into the mind of the Riau Indonesian speaker in order to share some of his or her awe and bewilderment at the profusion of structure that is evident in even the simplest of English sentences. Why, indeed, does English need all this stuff? The existence of Riau Indonesian, in which basic sentences such as *ayam makan* are simple coordinations of two words belonging to the same syntactic category, interpreted associationally, thus poses a serious challenge to functionally oriented theories of language which attempt to explain grammatical structures in terms of communicative functions. In Gil (2009), it is argued that IMA language is all that is needed in order to sail a boat; indeed, as suggested by Riau and other similar colloquial varieties of Indonesian, it is pretty much enough to run a modern country of over 200 million inhabitants.

In fact, one might go one step further and wonder why it is not the case that all natural human languages are Pure IMA Languages. One reason for this might be diachronic. McWhorter (1998, 2000, 2001a,b) argues that languages tend to accrue grammatical complexity over the course of time. In support, he claims that newly created creole languages are invariably characterized by lesser overall complexity than most other "older" languages with their lengthy continuous and uninterrupted histories. Accordingly, if a language started off as a Pure IMA Language, processes of grammaticalization would soon endow it with morphological structure, syntactic categories, and construction-specific semantic rules, and it would thereby lose its IMA characteristics. So the reason there are no natural Pure IMA Languages may be simply that they have been around for too long a time. However, the simplicity of Riau Indonesian shows that the accretion of complexity cannot be construed as a inexorable monotonic process. Riau Indonesian is not a creole and has no recent history of radical restructuring of any kind, and yet it is at least as simple in its overall grammatical structure as many creole languages; see Gil (2001a) for detailed argumentation. Comparison of Riau Indonesian with related Austronesian languages suggests that their common ancestor, Proto-Austronesian, spoken perhaps 5000 years ago, was of substantially greater complexity than Riau Indonesian in many grammatical domains. Thus, at some stage between Proto-Austronesian and Riau Indonesian, the accretion of complexity must have been reversed, in order for Riau Indonesian to emerge, gradually over time, as a Relative IMA Language. Gil (2015) provides an extensive overview of

how such processes of simplification might have applied to the Austronesian languages ancestral to Malay/Indonesian as they spread south from the Philippines into the Indonesian archipelago some 4000 years ago. Given that such simplification happened at least once, the question then arises once again why more languages in other parts of the world could not have taken the same path, and indeed why some languages could not have gone further along the path of simplification to end up as Pure IMA Languages. I have no answer to this question. Perhaps the diachronic forces that produce complexity are just "stronger" than those that operate in the opposite direction towards simplicity. Or perhaps the greater complexity of most other languages is due to entirely different factors.

19.5 COGNITION

The presence of IMA Language in artificial semiotic systems, in early child language, and to a considerable degree in some natural languages, suggests that IMA Language is an important feature of human cognition. However, to see how this is so, we need to adopt a somewhat different way of looking at IMA Language. The definition in (1) and much of the subsequent discussion were of an essentially *negative* nature: IMA Language was taken to be language *without* particular features: morphological structure, syntactic categories, and construction-specific semantic rules. This choice of perspective reflected our presuppositions regarding what languages are like: we expect them to possess these particular features, and it requires mental effort to entertain the possibility of languages lacking them. However, in order to appreciate the role of IMA Language in human cognition, we shall adopt an alternative *positive* perspective, focusing instead on particular features that IMA Language is endowed with.

Three such features are the following:

18. **a.** Recursive tree structure
b. Sign-meaning pairings
c. Associational compositional semantics

While all IMA Languages possess the above features, they are not definitional of IMA Language, since they are also exhibited by languages of greater complexity. On the other hand, Bickerton's protolanguage fails to exhibit the first property, syntactic recursion, though it possesses the latter two, together with IMA Language. Of the above three features, none is specific to natural language,

and therefore none should be considered to be part of a domain-specific Universal Grammar. Rather, each of the above three features is manifest in a variety of cognitive domains, and may thus be attributed to general human cognition.

Recursive tree structure refers to the ubiquitous human cognitive ability to construct groupings. Presented with a collection of objects, we view them as clustering together into groups according to various criteria: spatial or temporal organization, size, shape, color, quality, whatever. The groups then constitute new objects which, in turn, cluster into larger groups, and so on over and over again. Following is a simple illustration:

19. ★★★●★★★★ ●●★★★★●●★★★ ★★★●●●★★

In the above example, we form groups in accordance with shape: three stars, two circles, four stars, three circles, four stars, two circles, three stars, three stars, four circles, two stars. The resulting groups in turn form larger groups in accordance with spatial orientation: three stars plus two circles plus four stars, three circles plus four stars plus two circles plus three stars, three stars plus four circles plus two stars. The cognitive ability to form groupings has been the focus of a considerable amount of investigation; in particular, the ability to form groupings is essential to all higher-level cognitive capacities, as shown, e.g., in Lerdahl and Jackendoff's (1983) theory of tonal music. It is this general cognitive ability that is also manifest in the tree structures that represent the syntactic structures of IMA Language, as present in an array of pictograms, a sequence of signs by a captive ape, an utterance by a young child acquiring English, or a sentence in Riau Indonesian.

Sign-meaning pairings reflect another general human cognitive ability, one that lies at the heart of semiotic theory. Although the quintessential realization of sign-meaning pairings is that evident in the lexicons of natural languages, similar pairings occur in many other domains: consider a commercial logo such as that of Apple computers, a red traffic light, or the individual pictograms in examples such as (4) above. In semiotic theory, a major concern is to demonstrate that sign-meaning pairings are everywhere around us; in doing so, the notion of conventionalized sign-meaning pairing is sometimes extended well beyond its central domain of applicability, and consequently watered down considerably. Nevertheless, even under a conservative construal of the notion of sign-meaning pairing, it is clear that this feature of IMA Language too is not specific to language, but rather part of general human cognition.

Associational compositional semantics pertains to the cognitive ability to assign interpretations to combinations of signs in accordance with the Polyadic Association Operator as defined in (2) above. Again,

this feature of IMA Language is clearly of a general, nondomain-specific nature. Whenever we encounter a collocation of two or more sign-meaning pairings, we assume that the collocation is intentional, and attribute it a meaning that has to do in some way with the meanings of the individual constituent signs. Imagine the rather unlikely juxtaposition of an Apple logo with a red traffic light. When encountering such a combination, we assign it the meaning A(APPLE.COMPUTERS, STOP), "entity associated with Apple computers and with stopping," in accordance with the Polyadic Association Operator, and then seek for some contextually plausible interpretation, perhaps "Stop here for Apple computers," or alternatively "Don't use Apple computers." A substantial recent literature in cognitive psychology may be construed as dealing with the mechanisms with which the broad meanings assigned by the Polyadic Association Rule are narrowed down in particular contexts; see, e.g., Murphy (1990), Estes and Glucksberg (2000), and Wisniewski (2000); see also Shen and Gil (this volume) for discussion of similar processes in the context of visual hybrids.

Thus, IMA Language, with its three properties presented in (18), is a characteristic feature of general human cognition. In its most transparent form, IMA Language is evident in semiotic systems such as the language of pictograms. IMA Language may also be considered as a foundation on which the more elaborate and domain-specific structures of natural human languages are constructed. Moreover, just as houses are built from the foundations up, so children acquire natural languages beginning with IMA Language. As did our prehuman ancestors, thereby bequeathing IMA Language to us as an evolutionary relic from our distant past, but one that is very much alive and with us in the present.

Acknowledgments

I would like to thank the numerous speakers of Riau Indonesian who presented me with a new perspective on language, and my many colleagues at the Max Planck Institute for Evolutionary Anthropology who provided the intellectual environment that made it possible to develop that perspective into the ideas presented in this paper.

References

Anderson, J. M., & Ewen, C. J. (1987). *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry*, *13*, 571–612.

Bao, Z. (1990). Fanqie languages and reduplication. *Linguistic Inquiry*, *21*, 317–350.

Bhat, D. N. S. (1997). Noun-verb distinction in munda languages. In A. Abbi (Ed.), *Languages of tribal and indigenous peoples of India, the ethnic space* (pp. 227–251). Delhi: Motilal Banarsidass.

Bickerton, D. (1990). *Language and species*. University of Chicago Press.
Bloom, L. (1973). *One word at a time, the use of single word utterances before syntax*. Mouton: The Hague and Paris.
Burgess, A. (1975). *Language made plain* (Revised ed.). Glasgow: Fontana/Collins.
Bury, D. (2005). Preverbal particles in verb-initial languages. In A. Carnie, S. D. Collberg, & H. Harley (Eds.), *Verb first: papers on the syntax of verb initial languages* (pp. 135–154). Amsterdam: John Benjamins.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
Comrie, B. (1992). Before complexity. In J. A. Hawkins, & M. Gell-Mann (Eds.), *The evolution of human languages, proceedings of the workshop on the evolution of human languages, held August, 1989 In Santa Fe, New Mexico* (pp. 193–211). Redwood City: Addison-Wesley.
Conners, T., Bowden, J., & Gil, D. (2015). Valency classes in Jakarta Indonesian. In A. Malchukov, & B. Comrie (Eds.), *Valency classes in the world's languages*. Berlin: DeGruyter Mouton.
Dixon, R. M. W. (1977). Where have all the adjectives gone? *Studies in Language*, *1*, 19–80.
Estes, Z., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory and Cognition*, *28*, 28–34.
Feng, S. (2002). *Minimum Word in Mandarin Chinese*, ROA #507-0302.
Gil, D. (1985). What does grammar include? *Theoretical Linguistics*, *12*, 165–172.
Gil, D. (1993). Syntactic categories in Tagalog. In S. Luksaneeyanawin (Ed.), *Pan-asiatic linguistics, proceedings of the third international symposium on language and linguistics* (pp. 1136–1150). Bangkok: Chulalongkorn University, January 8-10, 1992, Volume 3.
Gil, D. (1994). The structure of Riau Indonesian. *Nordic Journal of Linguistics*, *17*, 179–200.
Gil, D. (2000a). Island-like phenomena in Riau Indonesian. *Paper presented at the seventh annual meeting of the Austronesian formal linguistics association*. Amsterdam, The Netherlands, 12 May 2000.
Gil, D. (2000b). Syntactic categories, cross-linguistic variation and universal grammar. In P. M. Vogel, & B. Comrie (Eds.), *Approaches to the typology of word classes* (pp. 173–216). Mouton, Berlin and New York: Empirical Approaches to Language Typology.
Gil, D. (2001a). Creoles, complexity and Riau Indonesian. *Linguistic Typology*, *5*, 325–371.
Gil, D. (2001b). Escaping eurocentrism: Fieldwork as a process of unlearning. In P. Newman, & M. Ratliff (Eds.), *Linguistic fieldwork* (pp. 102–132). Cambridge: Cambridge University Press.
Gil, D. (2001c). Reflexive anaphor or conjunctive operator? Riau Indonesian *sendiri*. In P. Cole, C.-T. J. Huang, & G. Hermon (Eds.), Long distance reflexives, syntax and semantics 33 (pp. 83–117). New York: Academic Press.
Gil, D. (2002a). Ludlings in Malayic languages: An introduction. In B. K. Purwo (Ed.), *PELBBA 15, Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya: Kelima Belas* (pp. 125–180). Jakarta: Unika Atma Jaya.
Gil, D. (2002b). The prefixes *di-* and *N-* in Malay/Indonesian dialects. In F. Wouk, & M. Ross (Eds.), *The history and typology of western austronesian voice systems* (pp. 241–283). Canberra: Pacific Linguistics.
Gil, D. (2004a). On the notion of the word in isolating languages: The case of Riau Indonesian. *Paper presented at the workshop on word domains, theory and typology*. Universität Leipzig, 8 April 2004.
Gil, D. (2004b). Riau Indonesian *sama*, Explorations in macrofunctionality. In M. Haspelmath (Ed.), *Coordinating constructions*, Typological Studies in Language 58 (pp. 371–424). Amsterdam: John Benjamins.
Gil, D. (2005a). From repetition to reduplication in Riau Indonesian. In B. Hurch (Ed.), *Studies on reduplication*, Empirical Approaches to Language Typology 28 (pp. 31–64). Berlin: Mouton de Gruyter.

Gil, D. (2005b). Word order without syntactic categories: How Riau Indonesian does it. In A. Carnie, H. Harley, & S. A. Dooley (Eds.), *Verb first: On the syntax of verb-initial languages* (pp. 243–263). Amsterdam: John Benjamins.

Gil, D. (2006a). Intonation and thematic roles in Riau Indonesian. In C. M. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus, cross-linguistic perspectives on meaning and intonation,* Studies in Linguistics and Philosophy 82 (pp. 41–68). Dordrecht: Springer.

Gil, D. (2006b). The acquisition of voice morphology in Jakarta Indonesian. In N. Gagarina, & I. Gülzow (Eds.), *The acquisition of verbs and their grammar: The effect of particular languages* (pp. 201–227). Dordrecht: Springer.

Gil, D. (2009). How much grammar does it take to sail a boat? In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 19–33). Oxford: Oxford University Press.

Gil, D. (2012). Where does predication come from? *Canadian Journal of Linguistics, 57,* 303–333.

Gil, D. (2013). Riau Indonesian: A language without nouns and verbs. In J. Rijkhoff, & E. van Lier (Eds.), *Flexible word classes, typological studies of underspecified parts of speech* (pp. 89–130). Oxford: Oxford University Press.

Gil, D. (2015). The Mekong-Mamberamo linguistic area. In N. J. Enfield, & B. Comrie (Eds.), *Languages of Mainland Southeast Asia, The State of the Art* (pp. 266–355). Berlin: Pacific Linguistics, DeGruyter Mouton.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elementsIn J. H. Greenberg (Ed.), (pp. 58–90). Cambridge: Universals of Language, MIT Press.

Greenfield, P. M., & Savage-Rumbaugh, S. (1990). Grammatical combination in Pan paniscus: Processes of learning and invention in the evolution and development of language. In S. T. Parker, & K. R. Gibson (Eds.), *"Language" and intelligence in monkeys and apes, comparative developmental perspectives* (pp. 540–578). Cambridge: Cambridge University Press.

Guilfoyle, E., Hung, H., & Travis, L. (1992). Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory, 10,* 375–414.

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica, 45,* 31–80.

Jackendoff, R. S. (1977). *X-bar syntax: A study of phrase structure*. Cambridge: MIT Press.

Jelinek, E., & Demers, R. A. (1994). Predicates and pronominal arguments in Straits Salish. *Language, 70,* 697–736.

Kayne, R. (1994).) The antisymmetry of syntax. Cambridge: MIT Press.

Keenan, E. L. (1978). On surface form and logical form. In B. B. Kachru (Ed.), *Linguistics in THE Seventies: Directions and prospects* (pp. 163–203). Champaign-Urbana: University of Illinois.

Keenan, E. L., & Faltz, L. M. (1986). *Boolean Semantics for Natural Language*. Dordrecht: Reidel.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Liaw, Y. F. (2002). *Malay grammar made easy, a comprehensive guide*. Singapore: Times Media Private Limited.

McWhorter, J. (1998). Identifying the creole prototype: Vindicating a typological class. *Language, 74,* 788–818.

McWhorter, J. (2000). Defining "Creole" as a synchronic term. In I. Neumann-Holzschuh, & E. Schneider (Eds.), *Degrees of restructuring in creole languages* (pp. 85–123). Amsterdam: John Benjamins.

McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology, 5,* 125–156.

McWhorter, J. (2002). The rest of the story: Restoring pidginization to creole genesis theory. *Journal of Pidgin and Creole Languages*, *17*, 1–48.
Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.), (2008). *Language complexity: Typology, contact, change* Amsterdam: John Benjamins.
Miles, H. L. W. (1990). The cognitive foundations for reference in a signing orangutan. In S. T. Parker, & K. R. Gibson (Eds.), *"Language" and intelligence in monkeys and apes, comparative developmental perspectives* (pp. 511–539). Cambridge: Cambridge University Press.
Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*, 259–288.
Newmayer, F. J., & Preston, L. (Eds.), (2014). *Measuring grammatical complexity* New York: Oxford University Press.
Payne, D. L., & Payne, T. E. (1990). YaguaIn D. C. Derbyshire, & G. K. Pullum (Eds.), *Handbook of amazonian languages* (Vol. 2, pp. 249–474). Berlin and New York: Mouton de Gruyter.
Rijkhoff, J., & van Lier, E. (Eds.), (2013). *Flexible Word Classes, Typological Studies of Underspecified Parts of Speech Oxford* Oxford University Press.
Romaine, S. (1992). The evolution of linguistic complexity in pidgin and creole languages. In J. A. Hawkins, & M. Gell-Mann (Eds.), *The evolution of human languages, proceedings of the workshop on the evolution of human languages, held August, 1989 In Santa Fe, New Mexico* (pp. 213–238). Redwood City: Addison-Wesley.
Sampson, G., Gil, D., & Trudgill, P. (Eds.), (2009). *Language complexity as an evolving variable* Oxford: Oxford University Press.
Sanders, G. (1975). *Invariant ordering*. The Hague: Mouton.
Shen, Y. (1985). *Action Structure in the Narrative Text*, PhD Dissertation, Tel Aviv University, Tel Aviv.
Shen, Y., & Gil, D. (this volume). How language influences the way we categorize hybrids.
Stassen, L. (2005). Predicative adjectives. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *World Atlas of language structures*. Oxford: Oxford University Press.
Swadesh, M. (1939). Nootka internal syntax. *International Journal of American Linguistics*, *9*, 77–102.
Thompson, L. C. (1965). *A vietnamese grammar*. Seattle: University of Washington Press.
Wetzer, H. (1992). 'Nouny' and 'Verby' adjectivals: A typology of predicative adjectival constructions. In M. Kefer, & J. van der Auwera (Eds.), *Meaning and grammar, cross-linguistic perspectives* (pp. 223–262). Berlin: Mouton de Gruyter.
Wisniewski, E. J. (2000). Similarity, alignment, and conceptual combination, comments on Estes and Glucksberg. *Memory and Cognition*, *28*, 35–38.
Yip, M. (1991). Prosodic morphology of four Chinese dialects. *Journal of East Asian Linguistics*, *1*, 1–35.
Zuberbuhler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, *63*, 293–299.