

## Linguistic Typology and Formal Grammar

Maria Polinsky

University of California, San Diego

### Introduction

The goal of this chapter is to provide an overview of the relationship between linguistic typology and formal grammar—a relationship that has existed for several decades now and is unlikely to disappear any time soon. As the reader will see, the two orientations differ in a number of respects, but they share the custody of language, and that motivates the need for communication between the two. More importantly still, the field of linguistics as a whole is beginning to study language as a dynamic system operating simultaneously on multiple levels of representation, rather than as a disparate assemblage of discrete levels of analysis (lexicon, phonology, syntax), or as a collection of particular linguistic phenomena. This common challenge to both theory construction and typology is motivated by the increasing integration of linguistics with more technically sophisticated disciplines that also investigate human cognition and consciousness.

The study of language is no longer solely the prerogative of introspective investigation (since linguistics purports to be a social science, and not a branch of philosophy or literature) and/or generalizations made on the basis of individual grammars. Instead, language is something that can be measured using standard scientific methodology, and modeled on the basis of rigorously established data. The challenge no longer lies in bringing typology and theory construction closer together—the challenge lies instead in the continued survival of both approaches in the face of an ongoing paradigm shift.

In what follows, I will examine several areas where typology and formal grammar diverge and where they most need to establish a viable dialogue: general goals, the nature of primary data, the structure of theory, and the significance of methodology. It would be counterproductive to

provide a list of differences and similarities between the two orientations, so I will conclude with some suggestions about bridging the gap between the two orientations; these suggestions include a possible shift in the research strategies used in typology and a significant shift in existing methodologies, of which all the orientations in linguistics need to be cognizant.

## **2. General goals and the nature of data**

The initial questions asked by typology and formal grammar are quite different. By virtue of asking very different questions at the outset, the two orientations have a very different approach to primary or preferred data.

A crucial difference between typology and formal grammar construction resides in their differing views on language diversity and linguistic theory. With respect to the former, all linguists agree, without much hesitation, that natural languages share a number of intriguing similarities and also show fascinating differences. This agreement does not extend very far beyond this clause, however. Typologists ask why (and how) languages differ, while grammar construction takes “the apparent richness and diversity of linguistic phenomena [to be] illusory and epiphenomenal, the result of interaction of fixed principles under slightly varying conditions” (Chomsky 1995: 8). Such an approach leads one to ask why and in what ways languages are similar, at least at some deep level. Thus, typology’s vested (albeit not always consciously recognized) interest is in making languages appear more different, whereas formal grammar wants them all to look alike. Of course this is an oversimplification, verging on a caricature, but it is striking how much the two different questions in (1) and (2) shape the overall approach to linguistic data.

(1) *Typology*: What makes natural languages so different from each other?

(2) *Formal grammar*: What makes natural languages so similar to each other?

Typology sees its goal in finding and explaining correlations between various aspects of meaning and structure among the languages of the world. There are few if any constraints on the range of such possible correlations, which immediately creates an impression that typology looks at a

smorgasbord of phenomena and has no theory at all (or has a very fragmented, almost balkanized approach to language).

And indeed, typology does not aspire to an articulated theory of the language faculty, simply because it does not have the tools (or desire) to impose limits on possible correlations that may be found or make predictions about what else one might expect to find. Such an unconstrained approach has both positive and negative consequences. On the positive side, it allows typologists to venture bravely where no one has tread before and uncover new data. In the 1970s, when linguists disagreed on whether or not it was necessary to examine a wide range of languages, the quest for unconstrained correlations saved typology and helped it flourish, to the point that formal grammarians also started appealing to wide-ranging cross-linguistic differences (Comrie 1993: 3-4). One could counter (and researchers often do, cf. Anderson 1999 or Steele 1997, among others) that this leads to open-ended “fishing expeditions” that often fail to distinguish between what is theoretically interesting and what is trivial.

Starting out with the similarity assumption (2), formal grammar instead sets as its goal the construction of a theory of language (not languages). A formal grammarian is more or less interested in knowing how the building blocks of language are assembled to form its structure. It is fair to say that rules are of primary importance: a linguist seeks “an explanation for the general process of projection by which speakers extend their limited linguistic experience to new and immediately acceptable forms” (Chomsky 1955/1975: 519). Because rules rule, the building blocks are assumed, without much empirical evidence, to be the same across all languages. As a consequence, not all data are treated as equally relevant; the truly relevant data are those that allow the analyst to test his or her favorite theory. Under the best of circumstances, this can result in the useful streamlining of the empirical investigation to include only those data points that are absolutely necessary to the analysis. For example, it may not really be necessary to take into consideration all imaginable influences of discourse context in the discussion of the purely structural properties of a particular syntactic construction. Under the worst of circumstances, such an approach runs the risk of potential oversanitization of what may be a more complex data set than the analyst would prefer to admit—and this then becomes the source of criticism from the side of typology that formal analyses do not take into account the full range of relevant data.

Formal assumptions about the cross-linguistic uniformity of linguistic building blocks make typologists very uncomfortable, as they are very concerned about the notion of engineering from sparse parts, not all of which appear comparable across languages. Another sign of typology's wariness with building blocks is its long-standing interest in categorization and classification: in order to theorize about the engineering aspects of language, it is important to know the inventory of existing materials. But this preoccupation with classification likewise comes at a cost. An overemphasis on classification can easily turn into a case of overclassification, as when stative predicates are classified differently solely on the basis of whether they are suppletive or not, or underclassification, as when classificatory schemes of verbal complementation options fail to distinguish predicates that involve control from those that do not. Merely classifying types of expressions into categories does not necessarily solve the underlying analytical problem.

Overall, it is striking and somewhat disconcerting how much particular orientations in modern linguistics differ with respect to what constitutes "proper" data and what does not. Depending on orientation, it may be acceptable or unacceptable to use only naturalistic data, only discourse data, or only minimal pairs and grammaticality judgments.

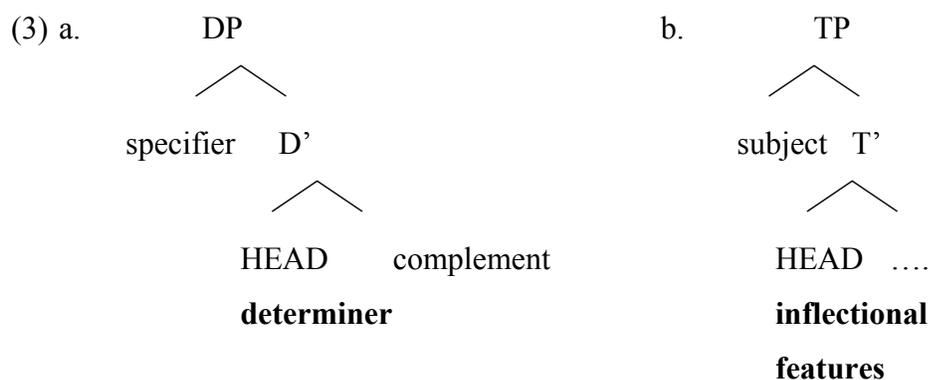
An interesting consequence of different approaches to data lies in the way the relevant data are elicited. Much has been written about typologists' extreme reliance on descriptive grammars, some of which may be too brief or inexplicit. This reliance points to another difference, namely that which arises between reliance on naturally occurring data (in typology and functionalism) versus potentially unnatural, controlled elicitations (in formal grammar). The analogy could be drawn from physiology: one can observe a number of people in natural running environments, or test a set of subjects on a treadmill in a lab. In those two conditions, the generalizations are different: natural observations would yield generalizations about preferred patterns, treadmill studies would show what a human body can do when pushed to its limits. Physiologists seldom argue whether one method of observation is better than the others—they have long learned how to combine the data from both. Linguists, though, are just now barely entering that stage in the development of the field.

### 3. Theory and methodology

This difference in initial questions leads to significant differences in theory and methods. Although the preceding section dealt mainly with primary data, I have already identified some theoretical differences there. Typology generally eschews uniform, all-questions-answered theorizing in favor of general constructs (markedness, iconicity, grammaticalization) and methodological devices, such as semantic maps (Croft 2001, and many others), often developed in reaction to more formal approaches. Its allegiance to large samples and “superficial” generalizations is simply one of the consequences of casting the net wide and looking for differences in a quick and easy way: testing for possible placement of a negative marker should reveal more variation across two hundred languages than it would across ten.

The absence of an articulated theory and the general fragmentation of typology at first blush come across as a severe handicap. The lack of a theory makes it more difficult to determine which data are relevant to the investigation at hand and which are of less importance. But this handicap may also be typology’s strongest asset: the absence of a unified theory and entrenched formalism has sometimes been liberating to typologists, allowing them to come up with genuine cross-linguistic generalizations that challenge existing theories for an adequate explanation.

Two examples of such generalizations come to mind: the consistency in headedness and the Accessibility Hierarchy. The headedness generalization is quite robust: languages *tend* to avoid arbitrary combinations of different word orders and *tend* to linearize their heads in a consistent fashion. The recognition of this tendency has played a major role in generative analysis, which usually adheres to categorical primitives. For instance, headedness figured prominently in the development of the DP hypothesis (Abney 1987 and many others), which allowed linguists to see that languages are even more consistent in terms of headedness than the NP structure would have pointed to. If a determiner is viewed to be the head of what was traditionally thought of as a noun phrase (3a), by analogy with clause structure (3b), it turns out that such heads tend to appear on the same side of their complements as inflectional (verbal) heads do.



Typological research made it possible to satisfy the drive for symmetry, which has long been strong in the formal approaches to grammar, but it has also helped to keep this drive in check. The actual richness of language data shows that symmetry may be desirable but not always attainable and that linguists can gain a lot from examining the apparent cases of asymmetry—compare Kayne’s radical antisymmetry approach (Kayne 1994) and on the more empirical side, a wide and diverse range of phenomena discussed in Di Sciullo (2003).

The well-known Accessibility Hierarchy (Keenan and Comrie 1977) is another example: the ease of relative clause formation depends on the grammatical function of the head noun inside the relative clause. All languages are expected to relativize subjects; beyond subjects, the accessibility of grammatical positions to relativization obeys the following hierarchy:

(4) Subject > Direct Object > Indirect Object > Object of preposition > Genitive > Comparative Object

Since the hierarchy has been proposed, many of its apparent violations have been shown to be due to the misanalysis of particular data; once the data were reanalyzed, they were typically in line with (4). For example, some cases of apparent relativization of genitives are actually fed by possessor raising, which promotes the possessor into the subject or object position, from which it can then relativize. Imagine a language where the relativization of possessor is possible from (5), which instantiates possessor raising, but not from (6). If the linguistic analysis of a particular language that allows structures like (5) is carried out properly, this language simply shows the

relativization of direct object and becomes rather unremarkable from a typological perspective (cf. Payne and Barshi 1999).

(5) She patted the cat on the back > the cat that she patted (on the) back

(6) She patted the cat's back > the cat whose back she patted

The Accessibility Hierarchy applies to other types of extraction phenomena (*wh*-question formation and topicalization, which are subsumed under A-bar phenomena in formal grammar). The hierarchy seems to generalize beyond extractions as well. Thus it reflects the special status of subjects which surfaces in other grammatical phenomena, such as the interpretation of anaphors, which are also more likely to select a subject antecedent than a grammatical function lower on the scale. This is a manifestation of an even more general principle: subjects have a special status in providing referential identification for “missing” (unexpressed) constituents, as in the establishment of coreference across clauses. And in some phenomena, the missing constituent itself must be the subject. This is the case of control and raising, for which the missing element in the embedded clause is the subject (see Davies and Dubinsky 2004 for a detailed overview). While all these phenomena may have different explanations, they still point to the generalization that subjects are privileged across languages. Yet the question remains as to why the preference for subjects exists, and what makes them “special” (cf. Kluender 2004).

Headedness and the Accessibility Hierarchy are perhaps the best known examples of genuine typological discoveries that have had a major impact on the entire field: no linguist, regardless of persuasion, can willfully choose to ignore either one. However, typologically inclined linguists have produced many other descriptive accounts of equally rich and intriguing data that would, if recognized, present challenges to any linguistic approach, and could perhaps stimulate thought leading to scientific advancement. The sad fact is that since typologists and formalists tend not to talk to each other (or read each other's literature), many of these veritable riches remain buried in descriptive grammars such as those in the Mouton Grammar Library series. This is a very general point of course, and illustrating it would constitute a research project in its own right, so at the risk of overgeneralizing, I will illustrate with a single example—the mystery of Kayardild redundant marking (Evans 1995, 2003). In Kayardild, morphological markers of

case appear on every subconstituent of a DP, and on top of that, *some* NPs also get inflected for the categories of the main predicate of their clause such as negation, modality, or tense. This happens if and when these DPs contain a so-called verbalizing suffix (V\_DAT in (7)). The suffix appears on the head noun of the phrase, as well as on its modifiers. For example,

(7) a. ngada      waa-jarra    wangarr-ina                      ngijin-maru-tharra  
       1SG.NOM    sing-PAST    song-MODAL\_ABLATIVE    my-V\_DAT-PAST  
       thabuju-maru-tharra  
       brother-V\_DAT-PAST  
       ‘I sang a song for my brother.’

b. ngada    waa-nangku                      wangarr-u                      ngijin-maru-nangku  
       1SG.NOM sing-NEGATIVE.POTENTIAL    song-MODAL\_PROP    my-V\_DAT-NEG.POT  
       thabuju-maru-nangku  
       brother-V\_DAT-NEG.POT  
       ‘I won’t sing a song for my brother.’ (Evans 2003: 215)

What is going on here? Does the beneficiary turn into a secondary predicate, which then gets inflected for tense? That would be a reasonable assumption; judging by Evans’ very detailed grammar, most clausal arguments, except subject, can undergo such verbalization. Evans himself takes a different route—he argues for a rather radical reconceptualization of agreement in light of Kayardild data: “In many ways, Kayardild is a language with a great deal of agreement, in the familiar sense of grammatically stipulated featural compatibility between different words. But it just happens to manifest agreement by unfamiliar semantic categories, on unfamiliar targets, over unfamiliar domains, in unfamiliar directions, with unfamiliar patterns of nested multiple agreement, ... and with unfamiliar functions” (Evans 2003: 232). Here the data definitely address the theory, theory of agreement in particular, and such work, done by a careful researcher, points to further probing questions, whose answers may change our conception of agreement or bring into being a different view of Kayardild grammar. The latter solution is hinted at by Corbett (2006), who notes that a radical reconceptualization of agreement would lead to a rather skewed typology in which most known languages would make up one class, and

Kayardild alone would form another. Whatever the solution, consistent verbalization of clausal arguments is an unusual phenomenon definitely calling for further investigation.

While typology (at least in its functional instantiation) has done its best to consciously ignore generative grammar, much typological research has nonetheless unconsciously been driven by the generative engine—but in the opposite direction. Because of an unfortunate confluence of historical circumstances, typological research has, by way of reaction to the purportedly misguided generative emphasis on structure, become equally obsessed with structure in a negative sense, namely with demonstrating that the structures proposed by generative analyses cannot possibly be right (cf. Polinsky and Kluender 2006 for more discussion). Both parties, formal theoreticians and typologists alike, stand a lot to gain from breaking the cycle of arguing and trying to prove the other side wrong. A better theory of language may emerge in the process.

#### **4. Explanatory tools**

Nowhere have typology and formal grammar been more distant than in their approach to the explanation for language phenomena. This difference stems from the difference in initial premises discussed above but it becomes much more prominent once we move from facts to interpretations.

The difference in explanatory tools has many guises: the opposition between formal and functional explanations (Hyman 1984; Hall 1992: Ch. 1), between internal and external explanations (Newmeyer 1998: Ch. 3), or between synchronic and diachronic motivations (Bybee 2001 and Lightfoot 1999 are good examples of the opposing views here). I will follow Newmeyer in casting this opposition in terms of internal *vs.* external explanation.

Internal explanation is one that exists within the context of a given theory of language, and is based on the principles of that theory. Under this type of explanation, linguistic phenomena are accounted for by principles (i) built into the theory (hence the term ‘theory-internal’) or (ii) inherent in the level of linguistic representation to which the phenomena belong. Because formal grammars are very focused on theory and theory-construction, it should come as no surprise that they give more weight to internal explanation. But a successful internal explanation has to be

based on a descriptively accurate account. This requirement is shared by any serious theory committed to the scientific method, and this is where the two orientations may successfully be united (see Moore and Polinsky 2003 for an extensive discussion).

The essence of external explanation is in relating linguistic and extralinguistic phenomena as manifestations of the same principles. Thus, external explanation is often defined negatively, as one where a given linguistic phenomenon is linked to a phenomenon or principle that either (i) is not particular to language or (ii) is outside the level of linguistic representation where the phenomenon belongs. The broad and narrow conceptions of external explanation run in parallel to the two conceptions of internal explanation.

Linguists have explored several domains that may provide us with external principles motivating linguistic phenomena. General cognitive ability is the first of those domains. Under this approach, language is viewed as part of human cognitive ability. If language is similar to other cognitive systems (vision, attention), then linguists need to identify principles that are shared by language and such systems. Significant advances towards identifying such cognitive principles have been made by Cognitive Grammar (Langacker 1991). Within Cognitive Grammar, many general principles motivating language structure are taken to be by-products of the structure of human cognition and attention. For example, an important property of language is construal, defined as the observer's interpretation with regard to a particular scene. Construal includes the distinction between figure and ground, which goes back to Gestalt psychology, and the notions of perspective and focus, also recognized in cognitive psychology. However, in order for this approach to achieve a truly *external* mode of explanation, there needs to be independent corroboration of the relevant principles as belonging to language-external components. That is, experimental cognitive psychology and psycholinguistics must build upon the descriptive apparatus of Cognitive Grammar and test its hypotheses in such a way that would show their external explanatory power. Otherwise, being part of theory, and lacking independent corroboration through experimental methodologies, these notions run the risk of qualifying as internal, not external, factors.

Despite their shared conceptual premises, Cognitive Grammar and functional typology have had little interaction, and Cognitive Grammar has remained largely Anglocentric. This lack of dialogue between the two frameworks which share some general principles may seem puzzling at first sight, but on closer scrutiny it appears to be simply an extension of the relationship between formal grammar and typology. Like formal grammar, Cognitive grammar is an articulated theory of language, intent on deriving all language forms from a set of well-defined principles. This premise is at odds with the more free-wheeling, open-ended, bohemian approach of functional typology which, as noted earlier, is uncomfortable with the notion of rule-based structure building, using components that do not even appear to be comparable across languages.

The second source of external explanation is in the function of language as a communication device. Such external explanations have long been associated with functional approaches to language; they establish a causal relationship between the form of language and the need to express a proposition/concept efficiently and coherently (hence the well-known conversational implicatures), to facilitate the identification of referents (hence pragmatic and information structural principles), and to rule out ambiguity as much as possible (cf. Gricean principles or iconicity, which has long played an important role in typology—Haiman 1983).

The cross-linguistic distribution of overt reflexive pronouns across grammatical persons, discussed by Comrie (1993), offers a good example of such causal principles at work. There are languages that have no overt reflexive pronouns distinct from non-reflexives; Old English was one such language (Keenan 2003). There are languages such as modern English that have distinct reflexives across all grammatical persons, and there are languages such as Romance where the reflexive/non-reflexive distinction is overtly expressed only for third person. There seem to be no languages where this distinction is overtly expressed for first or second person only. Is this gap in the distribution accidental? Comrie (1993) provides a functional external explanation by noting that the reflexive/non-reflexive contrast does not distinguish between different possible referents in the first and second persons ((8a) and (8b) mean the same), whereas it does make a difference in the third person (cf. (9a) and (9b)). Only in the third person does the overt non-reflexive/reflexive distinction allow for a more efficient tracking of referents.

- (8) a. I hurt me  
 b. I hurt myself
- (9) a. She<sub>i</sub> hurt her<sub>j</sub>/\*<sub>i</sub>  
 b. She hurt herself

Many complaints about functional explanations arise from cases where an external explanation is taken to replace an internal one or where a description of the function is taken to constitute both the description of a phenomenon and the motivation for that phenomenon. However, it is important to notice that Comrie's account of the cross-linguistic distribution of reflexive distinctions does not have any direct bearing on the binding principles and should not be taken as their replacement but rather as a corroboration of internal principles of language modeling by general functional strategies.

Still another domain of external explanation is found in diachrony. *Sensu stricto*, a diachronic or “evolutionary” explanation may not be external because it pertains to the immediate domain of language and to language structure in particular. However, even assuming that languages are the way they look because of historical developments, we still face the question of whether a child learning a given language can access the relevant diachronic information—does this child know that the current state of the language input arose under competing pressures from an earlier state? This prior state may not be accessible to the learner, and as a result, the appeal to grammaticalization of a prior state into the new grammar has the distinct flavor of an external explanation. I am certainly cognizant of the many important discoveries that have been made in trying to relate the current shape of language grammars to a prior grammar. It is still unclear, however, what the limits of such a historical explanation are and how to determine them (where should a historical explanation apply?). One of the lessons evident from historical explanations is that they cannot constrain the data in such a way that would predict an unambiguous outcome. Two lines of research seem possible and promising at this stage of our knowledge: first, the understanding of how smaller scale, local changes can affect language subsystems (Moore and Polinsky 2003: 15-17), and second, a better understanding of such fundamental mechanisms as

frequency or extension. As our understanding of such mechanisms progresses, historical explanation has a chance of moving away from a post-hoc, causal approach to a more predictive, teleological theory.

The range of explanations sought in formal grammar and typology is quite broad, and it is almost ironic that certain phenomena remain unexplained in spite of this embarrassment of riches. Two outstanding examples were brought up earlier in this chapter: the headedness parameter and the Accessibility Hierarchy. Despite its clear empirical validity, a full explanation of headedness is still beyond the reach of either orientation. Likewise, the Accessibility Hierarchy, while empirically robust, evades a true explanation in the most ironic way. The original explanation, proposed by Keenan and Comrie, was that the referent of the subject is most salient and hence easier to access, and thus ultimately an effect of processing. While this explanation is rather vague, it has been confirmed by several studies of relative clause processing in languages as diverse as English (King and Just 1991, and many others), Japanese (Miyamoto and Nakamura 2003), and Korean (Kwon et al. 2005). However, psycholinguists have now been placed in the strange position of providing processing evidence for a phenomenon that was originally attributed, if only vaguely, to processing—an obvious case of circularity. Despite the richness of explanatory tools, we are still left with a puzzle as to why subjects are privileged in a number of respects.

## **5. Building bridges**

So far, our discussion of two different orientations has pointed to a fragmented field, where one orientation refrains from pushing for deeper generalizations, while the other desperately tries to derive every observable phenomenon from structural principles. One of the proposed solutions to the fragmentation problem involves changing the methodology of typology in particular in such a way that both orientations can interact in a more productive fashion (Polinsky and Kluender 2006; Baker and McCloskey 2006). The proposed change would involve comparative investigation of closely related languages. A good illustration of such an approach comes from comparative Romance linguistics which helped create a key testing ground for research on language variation in the early 1970's, with work by Richard Kayne and his students, comparing

and contrasting French, Italian, and Spanish. Comparative Romance has demonstrated the enormous benefits of in-depth research on closely related languages and dialects. Such research allows us to uncover subtle distinctions and fine details of grammar that often remain unnoticed in a coarse-grained approach to language typology (see Comrie 1993 for similar observations). Comparative Germanic syntax followed suit, creating a large, vibrant field with numerous research projects underway.

Maybe because I am personally invested in this subfield, I think a similar moment has arrived for the birth of comparative Austronesian syntax. Because of the sheer number of Austronesian languages, such a field could provide an excellent testing ground for linguistic theory—one larger and typologically more diverse than Romance or Germanic. To give just one example, many Austronesian languages exhibit the uncommon word orders verb-subject-object (VSO) or verb-object-subject (VOS). These word orders pose an apparent challenge to theories of word order in which universally, all sentences have the underlying word order SVO (Kayne 1994). The abundance of heretofore unknown languages with VSO and VOS orders allows us to further test the classical Greenbergian correlations against new empirical data. Another example that comes to mind is in keeping with this chapter’s preoccupation with subjects: many Austronesian languages favor just one syntactic argument in extractions and clause linkage (Keenan 1976, Pearson 2005, Aldridge 2005, and many others). The following examples from Malagasy illustrate this restriction—in order to be accessible to relativization, the head noun must be the highest syntactic argument in the relative clause. The semantic role of the privileged argument (underlined below) is indicated by voice morphology on the verb (AF, TF, and CF stand for agent, theme, and circumstantial “focus” respectively):

- (10) a. n-i-vidy      ny    kadoa    ho    an-dreni-ny      ny zaza  
          PAST-AF-buy    DET    gift      for    OBL-mother-3SG    DET    child  
          ‘The child bought a gift for his mother.’
- b. ny zaza      (izay)nividy      ny    kadoa    ho    an-dreniny  
          DET    child    that    bought      DET    gift      for    OBL-his\_mother  
          ‘the child that bought the gift for his mother’
- c. \* ny kadoa    (izay)    nividy    ho    an-dreniny      ny    zaza

DET gift that bought for OBL-his\_mother DET child  
 ('the gift that child bought for his mother')

d. \*ny reny (izay) nivity ny kadoa ny zaza  
 DET mother that bought DET gift DET child  
 ('the mother for whom the child bought the gift')

(11) a. no-vid-in'ny zaza ho an-dreniny ny kadoa  
 PAST-buy-TF'DET child for OBL-his\_mother DET gift  
 'The gift, the child bought for his mother.'

b. ny kadoa (izay)novid-in'ny zaza ho an-dreni-ny  
 DET gift that PAST-buy-TF'DET child for OBL-his\_mother  
 'the gift that the child bought for his mother'  
 (lit.: the gift that was bought by the child for his mother)

(12) a. n-ividi-an'ny zaza (ny) kadoa ny reni-ny  
 PAST-buy-CF'DET child DET gift DET mother-3SG  
 'The mother, the child bought her a gift.'

b. ny reny (izay) n-ividi-an'ny zaza (ny) kadoa  
 DET mother that PAST-buy-CF'DET child DET gift  
 'the mother whom the child bought a/the gift'  
 (lit.: the mother who was bought the gift for by the child')

The privileged argument, indexed by the voice morphology on the verb, is often referred to as the "external argument". While the single external argument restriction is well-documented and by now is almost expected for a "well-behaved" Austronesian language, it is much less clear what the grammatical status of the external argument is. The external argument could be the subject, as was originally proposed by some researchers, most notably Keenan (1976), but it could also be the topic, and thus associated with a non-argument position (A-bar position in derivational frameworks). Furthermore, it is also conceivable that some Austronesian languages could treat the external argument as subject, and others, as topic.

Identifying the range of variation in Austronesian languages would not only provide further insights into the syntax of these languages but could also inform general linguistic theory, which

has long struggled with the notion of subject. On the one hand, subjects are assumed to be universal: pretty much every theory, from functional grammar to Relational Grammar to Minimalism, assumes that all clauses have subjects, even the impersonals, where that subject is silent (but see Babby 1989, McCloskey 2001 for a different view). On the other hand, the universal principle that clauses must have subjects often comes as a stipulation (the Extended Projection Principle is just one notable example: see Chomsky 1981) and neither the actual status of this crucial constituent, nor the motivation for its presence, are well understood.

Although “micro-typology” is highly desirable (and within reach), it is also useful to keep the achievements of large scale, sampling typology in sight. The latter approach is often useful in uncovering first-pass, coarse-grained generalizations, which can then be tested in more subtle microstudies or subjected to computational or experimental testing. It is hard to see why a particular approach has to be advocated at the expense of all the others—as long as a reasonable dialogue between the orientations is possible, there is no loss in combining several fields of inquiry with respect to a particular phenomenon, be it clitic climbing, word order, or pharyngeal spreading.

I also think that a more fundamental change in mindset is needed. The starting point of this proposal is quite simple: we need to stop pretending that linguistics is mathematics or physics, which has long been the underlying desire of many formal grammarians, and at least entertain the idea that language may operate like other natural systems. If so, linguistics has a lot to learn from biology. Biology has long since moved beyond classificatory schemes that have little to say about the underlying mechanisms of natural systems, and this carries an important lesson for typology. In order to understand how natural language works, it is worth considering whether it might better be assessed as an entire system, much in the way that biology has learned to study the entirety of interactions at various levels of the organism. This is where subfields like neurolinguistics and computational linguistics can be of service: neural imaging techniques allow us to look at the neural composite of interactions at all levels of linguistic analysis, and computational models provide us with the means to determine how those interactions can create composite patterns at the systems level (e.g., Kirby 1999; van Everbroeck 2003). Even within linguistics proper, we should be taking all available data sources into account in constructing our

theories of language: not just principles of structure building, but their processing correlates and functional properties as well.

The bad news is that continuing to base our linguistic inquiry on partial data sets (derived from introspection, observation of limited though naturally occurring data, incomplete elicitation of minimal pairs, etc.), is more likely than not an exercise in inevitable obsolescence, planned or unplanned. Otherwise, we seem doomed to continue along the path of scholastic disputes over insufficient, albeit preferred, data. The good news, on the other hand, is that many of the components of such an enterprise are already in place: formalists are good at deducing principles of structure building, while typologists are good at recognizing their functional properties.

One approach within typology that moves in this direction is that pursued by Hawkins in various publications (Hawkins 2004 and references therein). Hawkins proposes quantifiable and falsifiable hypotheses about why languages exhibit the structural properties that they do, and how those structures might relate to facts of language processing. He does explicitly theorize relying on phrase structural configurations, but that theorizing is not invoked merely in the service of structure *qua* structure. Instead, Hawkins makes attempts to relate what we can find out about structure from large sampling to what we know about language processing from numerous experimental studies—and suggests further possibilities for experimental validation of the theory in the process. This approach truly offers a synthesis of the best of both worlds. In other words, typology and theory construction can indeed coexist to good effect.

## **Conclusion**

Despite numerous differences, some of which were outlined above, the typological and formal orientation are united in their commitment to language and in their ability to enlighten other disciplines studying consciousness in the intricacies of language structure and linguistic diversity. Typology and formal grammar are actually much closer to each other than might initially be assumed: over time, theory construction has learned to be sensitive to cross-linguistic details, and good typology has of course always depended on structural generalizations. In one sense, the

challenge faced by typology and formal grammar is essentially the same: both can choose to step boldly into the 21<sup>st</sup> century by changing their approach to data and by forging connections with specialists in sister disciplines like psychology, cognitive neuroscience, and computer science, or they can choose to run the risk of being relegated to history and (re)subsumed by the humanities. This is a common challenge faced by linguists of all stripes. If the mindset of the field as a whole doesn't change, debates such as those between typologists and formalists will continue to have religious rather than scholarly overtones.

## References

- Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Ph. D. Dissertation, MIT.
- Aldridge, Edith. 2005. Ergativity and word order in Austronesian languages. Ph. D. Dissertation, Cornell.
- Anderson, Stephen R. 1999. A formalist's reading of some functionalist work in syntax. In M. Darnell, E. Moravcsik, F Newmeyer, M. Noonan, and K. Wheatley (eds.). *Functionalism and Formalism in Linguistics*, vol. 1, 111-135. Amsterdam: John Benjamins.
- Babby, Leonard. 1989. Subjectlessness, external subcategorization, and the projection principle. *Zbornik Matice srpske za filologiju i lingvistiku* 32: 7-40.
- Baker, Mark, and Jim McCloskey. 2006. On the relationship of typology to theoretical syntax. To appear in *Language*
- Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Chomsky, Noam. 1955. *The logical structure of linguistic theory*. MS., Harvard University and MIT (Revised version published by Plenum, New York, 1975).
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, Mass.: The MIT Press.
- Comrie, Bernard. 1993. Language universals and linguistic typology. *Sprachtypologie und Universalienforschung* 46: 3-14.
- Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Croft, William. 2001. *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Davies, William, and Stanley Dubinsky. 2004. *A Grammar of Raising and Control*. Oxford: Blackwell.
- Di Sciullo, Anna-Maria (ed.) 2003. *Asymmetry in grammar*. Amsterdam—Philadelphia: John Benjamins.
- Evans, Nicholas R. 1995. *A grammar of Kayardild*. Berlin: Mouton.
- Evans, Nicholas R. 2003. Typologies of agreement: Some problems from Kayardild. *Transactions of the Philological Society* 101: 203-234.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59: 781-819.
- Hall, Christopher. 1992. *Morphology and mind: A unified approach to explanation in linguistics*. London: Routledge.

- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hyman, Larry M. 1984. Form and substance in language universals. In B. Butterworth, B. Comrie, and Ö. Dahl (eds.). *Explanation of language universals*, 67–85. Berlin: Mouton.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge: MIT Press.
- Keenan, Edward L. 1976. Remarkable subjects in Malagasy. In C. N. Li (ed.). *Subject and Topic*. New York: Academic Press, 247-301.
- Keenan, Edward L. 2003. An historical explanation of some binding theoretic facts in English. In J. Moore and M. Polinsky (eds.). *The Nature of Explanation in Linguistic Theory*, 153-190. Stanford: CSLI.
- Keenan, Edward L., and Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8: 63-99.
- King, Jonathan, and M. Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30: 580-602.
- Kirby, Simon. 1999. *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press.
- Kluender, Robert. 2004. Are subject islands subject to a processing account? *WCCFL* 23: 475-499.
- Kwon, Na-Young, Maria Polinsky, and Robert Kluender. 2005. Trace or pro? CUNY 18, University of Arizona, April 2005.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar*, vol. 2. *Descriptive application*. Stanford: Stanford University Press.
- Lightfoot, David. 1999. *The development of language: Acquisition, change, and evolution*. Malden, MA: Blackwell.
- McCloskey, Jim. 2001. On the Distribution of Subject Properties in Irish. In W. D. Davies and S. Dubinsky (eds.). *Objects and Other Subjects*, 157-192. Dordrecht: Kluwer.
- Miyamoto, Edson, and Michiko Nakamura. 2003. Subject/Object asymmetries in the processing of relative clauses in Japanese. *WCCFL* 22: 342-355.
- Moore, John and Maria Polinsky. 2003. Explanations in linguistics. In J. Moore and M. Polinsky (eds.). *The Nature of Explanation in Linguistic Theory*, 1-30. Stanford: CSLI.
- Newmeyer, Frederick. 1998. *Language form and language function*. Cambridge, MA: MIT Press.

Payne, Doris, and Immanuel Barshi (eds.). 1999. *External possession*. Amsterdam—Philadelphia: John Benjamins.

Pearson, Matthew. 2005. The Malagasy subject/topic as an A'-element. *Natural Language and Linguistic Theory* 23: 381-457.

Polinsky, Maria, and Robert Kluender. 2006. Linguistic typology and theory construction: Common challenges ahead. To appear in *Language*

Steele, Susan. 1997. The future of typology. *CLS* 37 (2): 287-302.

van Everbroeck, Ezra. 2003. Language type frequency and learnability from a connectionist perspective. *Linguistic Typology* 7: 1-50.