

## Language Variation and Linguistic Invariants<sup>†</sup>

Edward L. Keenan and Edward P. Stabler UCLA

May 25, 2009

We present a brief, informal overview of a formal approach to grammatical variation developed in K&S (Keenan & Stabler, 2003), to which we refer the reader for proofs and other formal statements. Our purpose here is to show how we can formulate structural universals of grammar given grammars that are structurally distinct in relevant respects.

### 1. The One and the Many

A central quest in syntactic theory is to reconcile the audible diversity of natural languages (NLs) with the claim that they have a common, biologically determined, form.

**1.1 The one** Mainstream syntactic theory (MST) from Chomsky 1957 to Bošković and Lasnik 2007 attempts this reconciliation by building it into the form of individual expressions which must satisfy general constraints on rules/derivations and representations. For a *core* expression X of L, MST asks “What is the structure of X?”. The initial response, now, is often a binary right branching tree in Spec-[Head-Complement] order, using language independent category symbols and structure building functions (Merge, Move). Regarding variation, some, perhaps much, is relegated to the *periphery* beyond *narrow* syntax and ignored; some is acknowledged in parameters with small ranges (e.g. question words remain in situ or front); and some lies in feature variation forcing slightly different patterns of movement/copying. Overt morphology is language specific, not determined by UG and not structurally autonomous (Bobjalik 2002) but a “reflection” of hierarchical constituent structure<sup>1</sup>. So MST focuses on the unity pole of the unity-diversity continuum, treating different languages as syntactically and semantically similar.

**1.2 The many** K&S in contrast, proposes a reconciliation that focuses on diversity. It provides a conceptual notion of structural **invariant** that can be satisfied by non-isomorphic structures. It formalizes the early Chomskyan desideratum that linguistic operations are *structure dependent* (Chomsky 1965:55-56; 1975:30 – 35; Radford 1997:11 – 15). Often invariants are not present in the grammar as conditions on rules or representations, and are not instantiated by single expressions. Given an expression X, we ask “What expressions have the *same* structure as X?” (not “What is *the* structure of X?”). Perhaps some feeling for our relational approach to linguistic invariants can be given with the following, very imperfect, anthropometric analogy: The absolute height of humans is certainly not invariant: it varies from about 2 to 8 feet. But the ratio of arm span to height is much closer to being invariant (the same). Also, in presenting our perspective we alert the reader to one difference in character between it and MST.

**MST focuses on the form of syntactic theory.** Minimalist principles such as inclusiveness and economy are notation based – they constrain how we derive expressions, not directly what can be derived. Kayne 1994 opens with the central role of notation (emphasis ours): “It is difficult to attain a restrictive theory of syntax. One

way... is to restrict the space of available syntactic *representations*, for example, by imposing a binary branching requirement,... The present monograph proposes further severe limitations on the range of syntactic *representations* ...". But restricting notation can be detrimental. Different notations may suggest different questions and generalize differently to new phenomena. For example, in the early 1960's context free grammars and categorial grammars were shown to define the same class of languages (Bar-Hillel, Gaifman and Shamir 1960). But it was categorial grammar that generalized to function-argument structure (Lewis 1970, Montague 1973), effectively creating the field of formal semantics as we know it today. The *significant* properties of objects are invariant under changes of descriptively comparable notation, notational artifacts aren't. The truth of *x is hotter than y* does not vary according as we measure temperature in Fahrenheit or Celsius. *x is twice as hot as y* does. So let us focus on regularities of linguistic nature, not their notational expression, adopting the slogan

**If you can't say something two ways you can't say it.**

**1.3 Back to the many** Here first is a semantic generalization in propositional logic based on notationally distinct objects. (1a) uses standard infix notation for conjunction and disjunction, (1b) uses prefix (Polish) notation.  $\models$  is "entails".

$$(1) \quad \text{a. } ((P \vee Q) \wedge \neg P) \models Q \qquad \qquad \qquad \text{b. } \wedge \vee P Q \neg P \models Q$$

The notational variation is non-trivial: prefix notation uses no parentheses and has no structural ambiguities. Eliminating all parentheses from infix notation yields structural and semantic ambiguity:  $((P \wedge Q) \vee R) \neq (P \wedge (Q \vee R))$ . On the other side, students of logic tend to find infix notation easier to understand, especially in long formulas. But the entailment fact in (1a,b) is *the same*: a disjunction of formulas conjoined with the negation of one entails the other. The entailments expressible in the two notations are the same. The synonymy of  $(P \wedge Q)$  and  $\wedge P Q$  is determined by independent compositional interpretation. We have no need to derive them from a common "deep structure" or map them to a common "LF".

## 2. On Structure

We treat a language as a compositionally interpreted set of expressions defined by a grammar  $G = (\text{Lex}_G, \text{Rule}_G)$ , where  $\text{Lex}_G$ , the *lexicon* of  $G$ , is a finite set of expressions called *lexical items* and  $\text{Rule}_G$  is a set of *structure building functions* called *rules*, which iteratively map tuples of expressions to expressions<sup>2</sup>. The *language*  $L_G$  generated by  $G$  is the set of expressions derived from  $\text{Lex}_G$  by finite iteration of the rules. So  $L_G$  is the closure of  $\text{Lex}_G$  under  $\text{Rule}_G$ . A *descriptively adequate*  $G$  for a NL  $L$  is *sound* (everything it generates is judged by competent speakers to be in  $L$ ) and *complete* (everything speakers accept is generated).

A grammar  $G$  is *isomorphic* to a grammar  $G'$ ,  $G \approx G'$ , iff there is a bijection  $h: L_G \rightarrow L_{G'}$  matching the functions  $F$  in  $\text{Rule}_G$  with the  $F'$  in  $\text{Rule}_{G'}$ , so that when  $F$  derives an expression  $z$  from some expressions  $x_1, \dots, x_n$  then  $F'$  derives  $h(z)$  from  $h(x_1), \dots, h(x_n)$ ,

and conversely. An isomorphism  $h$  from a  $G$  to itself is called an *automorphism* (or *symmetry*) of  $G$ . They are the ways of substituting one expression for another within a language which do not change how expressions are built. So they *preserves structure*:  $F$  builds  $z$  from  $(x_1, \dots, x_n)$  iff  $F$  builds  $h(z)$  from  $(h(x_1), \dots, h(x_n))$ , all automorphisms  $h$ . The set  $\text{Aut}_G$  of automorphisms of  $G$ , represents the “structure” of  $G$ . It contains the identity map, is closed under function composition and inverses and so is a group, the *automorphism* (or *symmetry*) *group* of  $G$ , noted  $\text{Aut}_G$  (or  $\text{Sym}_G$ ).<sup>3</sup>

**2.1 Complexity** In cases of interest (natural languages)  $L_G$  is infinite. Since automorphisms map an infinite set to itself one might think that there are massively many of them, and thus hard to study and characterize in any given language. But in fact in cases of interest the number of automorphisms of  $G$  is finite. The reason is that the value an automorphism  $h$  assigns to a derived expression  $F(x_1, \dots, x_n)$  is uniquely determined by the values it assigns to the expressions it is built from, namely  $x_1, \dots, x_n$ . So once we have given the values of an automorphism on the lexical items – finite in number in cases of interest – we have determined its values at all expressions:  $h(F(x_1, \dots, x_n)) = F(h(x_1), \dots, h(x_n))$ . ( $\text{Aut}_G$  is *forced* to be finite if  $\text{Lex}_G$  is finite and none of its elements is also a derived expression).

**2.2 Invariants** An expression  $s$  *has the same structure as* an expression  $t$  iff  $h(s) = t$  for some automorphism  $h$ . We do not mention here “the structure” of an expression – an epistemological plus, as different syntactic theories agree more readily that *John sang* and *Bill danced* have the same structure, than they do about what the structure of *John sang* is. We define:

**Def 1** A relation  $R$  on  $L_G$  is **invariant** iff  $h(R) = R$ , all automorphisms  $h$ . That is, replacing all tuples  $(s_1, \dots, s_n) \in R$  by  $(h(s_1), \dots, h(s_n))$  leaves  $R$  unchanged. Whence  $w \in L_G$  is **invariant** iff  $h(w) = w$ , all automorphisms  $h$ .

**Thesis**  $w \in L_G$  is a *grammatical formative* (“function word”) iff  $w \in \text{Lex}_G$  and  $w$  is invariant

So function words are items that are isomorphic only to themselves. Replacing them with something else changes structure (usually destroying it yielding ungrammaticality). In the models of grammars in K&S, reflexive pronouns, case, voice, and agreement markers are provably invariant. We expect that lexical invariants correspond to heads of functional projections in more usual terminology. What is new here is a characterization of what counts as *functional* – namely, being a linguistic object (including bound morphemes) which can only be mapped to itself by the automorphisms, the structure preserving functions, of the grammar).

As an item that a function  $f$  maps to itself is called a *fixed point* of  $f$ , we characterize the functional expressions in a grammar as the fixed points of the syntactic automorphisms. We turn now to some generalizations built on this notion of invariant.

**2.3 Degrees of Invariance & Language Change** *Invariance* generalizes to a scalar by:

**Def 2 a.**  $Inv(x)$ , the *degree of invariance* of a linguistic object (expression, property, relation)  $x$  is the proportion of automorphisms that *fix*  $x$ , that is, that map  $x$  to itself. So if  $x$  is invariant per Def 1, then  $Inv(x) = 1$ , as all automorphisms map  $x$  to itself.

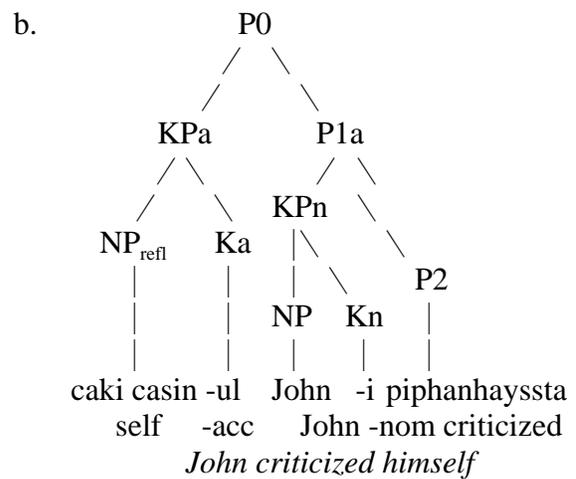
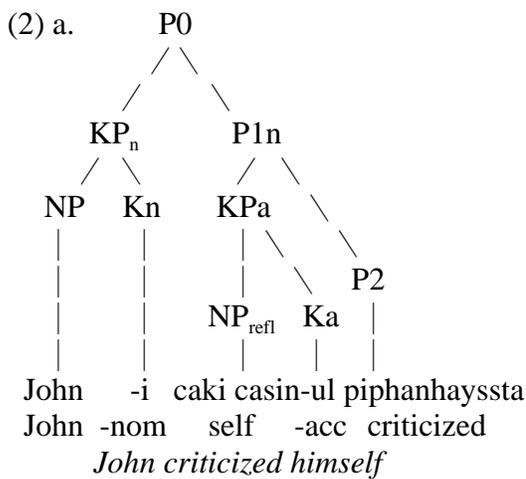
**b.**  $x$  is *more invariant than*  $x'$  iff  $Inv(x) > Inv(x')$

These definitions provide a rigorous way to say that conjunctions and prepositions are more grammaticized than intransitive verbs. And we can use this notion to represent the grammaticization (Hopper and Traugott 1993) of an expression  $w$  by saying that  $Inv(w)$  increases over time. If  $Inv(x)$  reaches 1 then  $x$  is fully grammaticized.

**2.4 Relation invariants** K&S prove that for all  $G$  (not just  $G$  for NLS), *is a constituent of* and *c-commands* are invariant relations. These notions require a more general than usual definition since we have not limited ourselves to rules whose action is modeled by standard labeled trees. Such functions basically just derive expressions by concatenation, not, for example, substitution (widely used to derive consequences from premisses in logical deductive systems for example). Here are some sample definitions:

**Def 3**  $u$  is an *immediate constituent* of  $v$  iff  $v$  is the value of a rule  $F$  at a tuple  $(s_1, \dots, s_n)$  and  $u$  is one of the  $s_i$ .  $u$  is a *constituent* of  $v$  iff  $u$  is  $v$  or  $u$  is an immediate constituent of a constituent of  $v$ .  $s$  is a *sister* of  $t$  in  $v$  iff  $s \neq t$  and for some constituent  $u$  of  $v$ ,  $s$  and  $t$  are immediate constituents of  $u$ .  $s$  *c-commands*  $t$  in  $u$  iff  $s$  is a constituent of a sister of  $t$  in  $u$ . All these relations are provably invariant in all  $G$ .

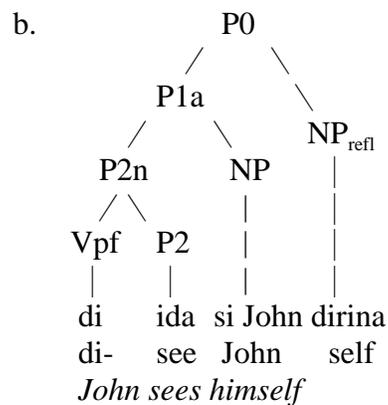
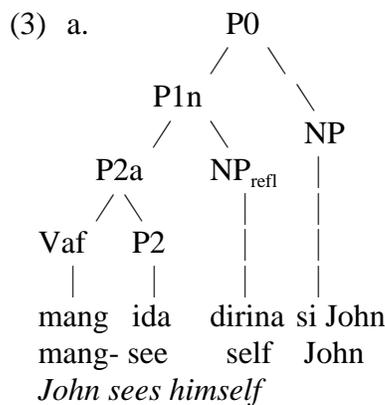
In contrast, the property of being an *anaphor* and the relation  *$x$  is a possible antecedent of an anaphor  $y$  in  $z$*  are not invariant in all  $G$ , but they are for all  $G$  we have constructed to model NLS. This illustrates invariants among non-isomorphic structures. K&S's grammar for minimal clause structure in Korean generates (2b) where anaphors like *self-acc* asymmetrically c-command their antecedents. To see that *self-acc* is an anaphor one must check its semantic interpretation in K&S (see below): it maps a binary relation  $R$  to the property  $\{a|aRa\}$ .



Each of these Ss is interpreted compositionally, and they receive the same interpretation: True iff  $(j,j) \in \text{CRITICIZE}$ ,  $j$  the denotation of *John*. This fact is not different in kind from the fact that  $(P \wedge Q)$  is logically equivalent to the structurally distinct  $\neg(\neg P \vee \neg Q)$  in propositional logic.

**Lovers of trees Beware!** (2a,b) are isomorphic qua ordered labeled trees (same branching structure, each has distinct labels just where the other does). But the *expressions* are not isomorphic in our grammar: if an automorphism  $h$  mapped *john* to *self* and *-nom* to *-acc* it could map *John -nom laughed* to *self -acc laughed*, which is not in our model language, contradicting that  $h$  is an automorphism. See K&S p.50.

(3) is from Toba Batak (W. Austronesian; N. Sumatra (Schachter 1984, Cole & Hermon 2008. The distribution of anaphors in Toba, as in other W. Austronesian languages (Tagalog, Malagasy, Balinese) is conditioned by verb voice, the reflexive often occurring as what we thought was a “subject”. And as in the Korean case, lexical items occurring in both Ss have identical interpretations. So again anaphors may asymmetrically c-command their antecedents.



**Theorem 1** The case markers *-nom* and *-acc* and voice markers *mang-* and *di-* are invariant in their grammars in K&S (p. 49 and 70 respectively).

Thus is morphology “structural” in the *same* sense as constituent structure: both are *preserved by all the automorphisms*.

**Theorem 2** a. The property of being an anaphor is invariant in K&S’s model grammars for English, Korean and Toba Batak (pp. 36, 50 and 70 respectively)  
b. The Anaphor-Antecedent relation is invariant in K&S’s models (p.53 and p.71 respectively), as it is in their model of English, not illustrated, in which the usual c-command conditions hold.

Theorem 2 builds on two properties which distinguish our approach to anaphora from more standard ones:

- [1] *anaphor* is semantically defined, so we are not free to stipulate that *himself* in English, *caki casin* in Korean, and *dirina* in Toba Batak are anaphors. Rather, the

anaphoric expressions in a language are those whose semantic interpretation satisfies our definition (the same for all languages). Often the anaphors in a language cannot be listed as there are infinitely many (all but finitely many being syntactically complex).

- [2] The structural means used to identify anaphors and their antecedents are not structurally the same across languages. Nonetheless the anaphor-antecedent relation is, we claim, invariant in each NL grammar (and so universally invariant). We illustrate:

**Referentially Autonomous Expressions** (RAEs) such as *Zelda*, *most men who Zelda dates*, etc. combine with  $n+1$ -ary predicates ( $P_{n+1}$ s), to form  $n$ -ary ones ( $P_n$ s). For simplicity we limit ourselves here to  $P_n$ s,  $0 \leq n \leq 2$ . Semantically  $P_n$ s denote  $n$ -ary relations over whatever domain  $E$  is under consideration<sup>4</sup>. RAEs denote *Type-1* functions – they map  $n+1$ -ary relations to  $n$ -ary ones, thus reducing arity by 1. But not just any Type-1 function is a possible RAE denotation. The value such a function assigns to a binary ( $n+1$ -ary) relation is determined by the values it assigns to the unary relations (subsets of  $E$ ). Thus *Dana praised every student* is true iff Dana is in the denotation of *praised every student*, the set of objects that praised every student, that is, the set of objects  $b$  such that (EVERY STUDENT) holds of the *set* of things that  $b$  praised. For  $R$  a binary relation write  $bR$  for  $\{d \mid (b,d) \in R\}$ . So  $bR$  is the set of objects  $d$  that  $b$  stands in the relation  $R$  to. Then the possible RAE denotations  $F$  are those Type-1 functions satisfying (4):

- (4) For  $R$  a binary relation,  $F(R) = \{b \mid F(bR) = 1\}$ <sup>5</sup>.

**Theorem 3** The Type-1  $F$  satisfying (4) are just those which satisfy the AEC (Keenan 1988):

### Accusative Extensions Condition (AEC)

For all  $x, y \in E$ , all binary relations  $R, S$ , if  $xR = yS$  then  $x \in F(R)$  iff  $y \in F(S)$

So *most men who Zelda dates* satisfies the AEC since, for example, whenever Sue distrusts just the people that Ann likes then *Sue distrusts most men Zelda dates* and *Ann likes most men Zelda dates* must have the same truth value (both true, or both false).

Now observe that expressions like *himself*, *herself*, etc. as they occur in *No model hates herself* fail the AEC. If Sue distrusts just Jean, Robyn, Amy, Ann, Mary, and Pat and those are exactly the people that Ann likes then *Sue distrusts herself* is false and *Ann likes herself* is true. Such expressions do however satisfy a weaker invariance condition (Keenan 1988):

### Accusative Anaphor Condition (AAC)

For all  $x \in E$ , all binary relations  $R, S$  if  $xR = xS$  then  $x \in F(R)$  iff  $x \in F(S)$

So if  $x$  bears  $R$  to the same things that  $x$  bears  $S$  to then  $x$  bears  $R$  to himself/herself iff  $x$  bears  $S$  to himself/herself. And we may, on first pass, define *anaphoric expression* by:

**Def 4** An expression  $\alpha$  in  $L_G$  is an *anaphor* iff the interpretation of  $\alpha$  satisfies the AAC in all models and fails the AEC in some<sup>6</sup>.

One checks by example that the underlined complex expressions in (5) are anaphors:

- (5) a. John criticized every student but himself / no student but himself  
b. John criticized only himself and the teacher / neither himself nor the teacher  
c. John knows many people smarter than himself / no one as talkative as himself  
d. He nominated someone other than himself  
e. A prime number has no divisors except itself and 1.

**4. Structural Universals** We consider some candidates **U1** – **U5** as structural universals of human language. Our purpose is to illustrate how such claims can be formulated in our framework. We think they are plausible, but much empirical investigation is needed.

#### **U1 The property of being an anaphor is a structural invariant of human language**

**U1** says that anaphors are always mapped to anaphors by the syntactic automorphisms. Note that even if each lexical anaphor is mapped to itself, each complex anaphor typically will not be. An automorphism of English might map *himself* to *himself* but map *no doctor but himself* to *no lawyer but himself*. Nonetheless, **U1** says that in each  $L_G$  anaphors have some syntactically distinctive properties, ones which may fail to be comparable across languages. In nuclear clauses in Batak for example they concern distributional constraints with respect to *mang-* vs *di-* prefixed verbal roots. In Korean they concern case marking, and in English it is Principle A that distinguishes the distribution of anaphors: *John laughed* is fine, *Himself laughed* is not. These same observations support:

**U2** The relation *x is a possible antecedent of an anaphor y in z* is a linguistic invariant.

**4.1 Theory Design** Merely representing a NL as a pair  $(Lex_G, Rule_G)$  is not a *theory* – at best it is a common denominator of theories such as Minimalism, HPSG, LFG, RG, ... But two features of our approach do have a liberating effect on theory design: (1) the requirement of a compositional semantics, and (2) the theorem that morphology may be structural. From (2), in designing a grammar we are free to condition the distribution of anaphors (semantically defined) directly in terms of case or voice, it is not necessary to derive them from, or reconstruct them to, forms c-commanded by their antecedents.

From (1), semantic representations – on our view compositionally interpreted audible structures – differ from language to language, whereas in MST (Higginbotham 1985; Chomsky 1986:156; Hornstein 1995:Ch 1) LFs for different languages are assumed or argued to be roughly the same. They claim that primary semantic data do not suffice for the child to infer the structure of LF parameters (in distinction to primary

phonetic data which do permit the inference of diverse phonologies). Since we do learn to use language meaningfully the semantic module must be innate, hence essentially the same across speakers of different languages.

*The argument does not convince.*

No characterization of primary semantic data is given, nor are reasons why they are insufficient to set “LF parameters”. In fact children learn to use language meaningfully as they learn to pronounce it. Simple situations in which they follow commands, make requests, answer questions, disagree, ... are easily seen to contribute to learning the meaning of expressions – reference (count and mass), basic argument structure and theta roles, modification (*Hand me the red crayon. – No, no, not the BLUE one, the RED one!*), etc. And of course learners of different languages are interpreting different expressions, an unproblematic fact as our example with  $((P \vee Q) \wedge \neg P)$  and  $\wedge \vee PQ \neg P$  shows. (6) and (7) below exhibit pairs of logically equivalent sentences with different internal structures – showing that different structures may determine the same (logical) meaning.

- (6) a. Between a third and two-thirds of Americans brush their teeth regularly  
b. Between a third and two-thirds of Americans don't brush their teeth regularly
- (7) a. All but two students read at least as many poems as plays  
b. Exactly two students read more plays than poems

(6) is surprising as they differ just in that the predicate in (6b) negates that in (6a). The pattern is general as long as the fractions sum to 1. Keenan (2004).

We should note that LF is not designed to represent meaning in general. It doesn't even support a definition of entailment (Chomsky 1986:67, fn11) and several of the semantic notions it does represent are esoteric and likely learned later than the notions we mentioned above. For example, relative scope of quantifiers only arises with two quantified arguments of a given predicate. Languages provide many means for distinguishing the arguments of transitive verbs (word order, agreement, case marking, chain of being orders) but do not systematically disambiguate scope (Keenan 1988). T. Lee 1986 supports experimentally that both English and Chinese children understand some basic quantifiers in intransitive Ss by age 4, but even by age 8 do not have adult competence on their relative scope judgments in transitive Ss. Even in formal logic it was only with Henkin 1961 that we learned to construct formulas with universal and existential quantifiers lacking scope dependencies.

**4.2 Property invariants** Is the property of being a lexical item invariant in all G for NL? That is, do automorphisms always map lexical items to lexical items? This is a possible empirical truth (not a theorem) even though the lexicons of different NLs differ. Similarly, is the property of having a given grammatical category invariant? That is, do all automorphisms map each phrase of category C to a phrase of category C? Taking this as axiomatic would provide a universal structural role for categories. But K&S show this

fails, only supporting the weaker **U3**. Still **U3** does provide a universal structural role for categories which allows grammars of different languages to have different categories. Embarrassingly the field has no purely syntactic definition of *category* that allows us to infer that category C in language L and category C' in language L' are the same. See Baker 2003 for some discussion.

**U3** For all stable automorphisms h, all expressions x, if x has category C so does h(x).

An automorphism of a grammar G is **stable** iff<sub>def</sub> it extends to an automorphism of each **lexical extension** of G.  $G_n$  is a *lexical extension* of G iff there is a sequence  $(G_1, \dots, G_n)$  of grammars such that each  $G_{i+1}$  differs from  $G_i$  just by the addition of a single lexical item isomorphic to one in  $\text{Lex}_{G_i}$ . (The set of stable automorphisms in  $\text{Aut}_G$  is always a subgroup of  $\text{Aut}_G$ ). K&S's model of Spanish for example has two gender classes of nouns,  $N_m$  and  $N_f$ , and overt agreement of adjectives and dets with Nouns. There are automorphisms h which interchange the lexical  $N_m$ s and  $N_f$ s, provided these two sets have the same number of members. If we add one new member to just one of the classes we can no longer interchange them by an automorphism, so such h are *unstable*: their status as automorphisms is not preserved under trivial additions to the lexicon. A fourth candidate for a structural universal is:

**U4** Theta role assignment is invariant

So if x bears theta role  $\tau$  to y in z then h(x) bears  $\tau$  to h(y) in h(z), h any automorphism. That is, theta role assignment  $\Theta$  is a function of structure. If *Ed* has different theta roles in *Ed ran* and *Ed arrived* then these Ss must be non-isomorphic (per current theories). But *Ed* may occur in structurally distinct environments and still be assigned the same theta role. **U4** is strictly weaker than UTAH, which requires that  $\Theta$  be (structurally) one to one: so same theta role  $\Rightarrow$  isomorphic sources. But functions may take the same values at different arguments:  $(2+3) = (1+4)$ . So active subjects and their corresponding passive agent phrases may originate in non-isomorphic configurations. We close with a much more contentious candidate universal, **U5** below.

**4.3 Greenberg Duality.** Two languages are word order *duals* if the expressions of one are the mirror images of those of the other. Lexical items are self dual. A rigid SXOV L, like Turkish, is (isomorphic to) the dual of a rigid VOXS language, like Malagasy. Formally, the dual  $v^d$  of a sequence  $v = \langle v_1, v_2, \dots, v_n \rangle$  of lexical items is just its mirror image,  $\langle v_n, \dots, v_2, v_1 \rangle$ . The dual  $K^d$  of a set K of expressions is the set of  $w^d$  for  $w \in K$ . If F is in  $\text{Rule}_G$  then its *dual*  $F^d$  is that function with domain  $\text{Dom}(F)^d$ , that maps  $\langle w_1^d, \dots, w_n^d \rangle$  to the dual of  $F(\langle w_1, \dots, w_n \rangle)$ . We define  $G^d$  to be that grammar with the same lexical items as G and whose rule set is the set of duals of rules of G. We have:

**Theorem 4.1**  $(L_G)^d = L_{G^d}$

**4.2**  $G \approx G^d$ , the map sending each  $w \in L(G)$  to  $w^d$  is an isomorphism.

Theorem 4.1 just says that the dual of the language is the language generated by the dual grammar, so we defined  $G^d$  right. Theorem 4.2 says that G and  $G^d$  are isomorphic.

**U5** The set PHG of possible human grammars is closed under isomorphism

**U5** just says that if  $G \in \text{PHG}$  and  $G \approx G'$  then  $G' \in \text{PHG}$ . Our justification for **U5** is that UG only selects for structure not content and thus cannot distinguish between isomorphic variants (though other constraints, say ones on possible phonological systems, might rule out some isomorphic images as being non-natural on other grounds)..

**Corollary (Duality): PHG is closed under duals** (From **U5** and Theorem 4.2)

The Duality Corollary makes us hesitant to accept Kayne's Antisymmetry axiom, which forces right branching structures. If only such grammars were acceptable then PHG would not be closed under duals. But since there are left branching Ls (Toba Batak, Malagasy) the effect of the Antisymmetry axiom must be weakened.

But **U5** is also problematic. The Duality Corollary suggests an equal distribution of word order types and their duals, which is not the case. Right branching (SXOV) Ls are the most common across areal and genetic groupings whereas VOXS Ls are a clear minority (but include Malagasy (Keenan 1976) and several other Austronesian languages, and Tzotzil (Aissen 1976) and several other Mayan languages). Worse, the OVS duals of SVO languages have (to our knowledge) just Hixkaryana (Carib; Brazil; Derbyshire 1977) as a well attested exemplar, while the OSV duals of VSO Ls are just barely attested: *Ethnologue* (Gordon 2005) cites Jamamadi, an Arawakan language in Brazil.

But immediate rejection of **U5** would be short sighted. Much work in the physical sciences, esthetics, mathematics, and the philosophy of science supports both the fundamental role of symmetry and also the presence of asymmetries and symmetry breaking in the phenomena those fields study. Acknowledging and studying these asymmetries and symmetry failures has been a significant stimulus to deeper understanding. "The study of anomalies now plays an important role in our search for the symmetries of nature". (Zee 1986:300).

We can not summarize here the basic, provocative, at times dazzling, work in this area. We just note a few highlights that have influenced our thinking and point the reader to several accessible and enlightening introductions to symmetry and symmetry breaking. Weyl (1952) is a classic and still one of the best. It discusses symmetries and asymmetries in physics, biology, and art. Bunch (1989) and Gardner (1990) are more recent and very informative. Darvas (2007) focuses more on symmetry in art, and Zee (1986) on symmetry in physics. Mathematically we recommend Stewart and Golubitsky (1999), Ash and Gross (2006), and a basic textbook on group theory, the language of symmetry and invariance, for example Rotman (1999:Chs 1 – 3). On the more philosophical and epistemological side we have found van Fraassen (1990) and Wigner (1979) enlightening.

Concerning the conceptually fundamental nature of symmetry and invariants (what remains unchanged under the action of the symmetries) the first author must

acknowledge his awe at Felix Klein's Erlangen dissertation (Klein, 1872) in which Klein stood Euclidean geometry on its head. The objects of study shift from points, lines and triangles to whatever was invariant under the action of the operations – translations, rotations, and reflections. Other geometries are invariants of other transformations. Later, topology, which grew out of geometry, became the study of continuous transformations and topological invariants – those objects, properties, ... which remain invariant under these transformations. In 1918 Emmy Noether (Bunch 1989:95, Brewer and Smith: 1981; Cole 1997:183), at Erlangen, proves that symmetry principles in physics (including relativity theory) imply conservation laws. Indeed Hermann Weyl (cited in Bunch 1989:144) claims “The entire theory of relativity...is but another aspect of symmetry”. Gardner (1989:337) quotes Einstein to the effect that *invariant theory* would have been a better name for his achievement than *relativity*. More recently symmetries have been used to predict new elementary particles (Sternberg 1995). And symmetry breaking is the exclusive subject of a recent monograph (Strocchi 2005). Still, “Why”, asked the great physicist Richard Feynman (1918 – 1988), “is nature so nearly symmetrical?” (Bunch p.189).

Cotton (1990), a standard textbook, exemplifies the utility of symmetry in chemistry by using group theory to structurally classify molecules. “...the number and kinds of energy levels that an atom or molecule may have are rigorously and precisely determined by the symmetry of the molecule or of the environment of that atom.” (Cotton: p.3). In mathematical domains group theory, as noted, the mathematics of symmetry and invariants, has become a major subfield in mathematics beginning with Galois (1811 – 1832) who initiated the study of symmetries among solutions to polynomial equations. In logic, Tarski (1966, only edited and published posthumously, by J. Corcoran 1986), presents informally the idea that “logical operations” are simply the most general ones, in distinction to translations, continuous maps, etc. which must obey constraints in addition to being bijective. See Keenan (2001). And of linguistic interest Roman Jakobson (1963) pushed the notion of linguistic invariant in the famous 1963 Universals conference: “Naive attempts to deal with variations without attacking the problem of invariants are condemned to failure” (Jakobson 1963:272).

Let us return now to the unequal distribution of word order duals – which, being isomorphic but distinct, we might have expected to be equally distributed (perhaps assuming a Leibnizian “Sufficient Reason” basis in Nature). But they are not. Nor are right and left spiraling DNA. As far as we know, all animals are built from right spiraling DNA, though some left spiraling DNA has been found in nature, and it can be synthesized, so it is not incompatible with nature in principle. What accounts for the statistical disparity? Why do righthanders outnumber lefthanders? Why can humans distinguish caraway seeds and spearmint by taste, though the molecules (d-carvone and l-carvone respectively) which underlie them are “duals” (*mirror images, enantiomers*). In some cases at least we feel the choice among symmetric options was arbitrary and “random”, but once established it self-perpetuated. Driving on the left or the right is an arbitrary convention, entailing the construction of symmetric cars (steering wheel on the left or the right). But as one came to dominate in continental Europe Sweden was pushed to “walk in step”, that is, drive on the right. Britain is still holding out. (Japan, Australia,

and New Zealand continue to drive on the left).

Now, how good is the comparison between linguistic symmetries and invariants and those in physical or mathematical fields? Answer: “Excellent”. In both cases we study structure preserving functions and what they preserve. So we are not making an analogy, we are using linguistic science as another field where symmetries and their invariants can be studied. Will this approach lead to enlightening results, as it has in the fields discussed above? We won’t know until we pursue it detail to see where it leads. But we do have a positive prognosis.

First, already a strong reason to pursue our approach, we have a methodology that permits the description of structural regularities across structurally non-isomorphic languages. And second, we can draw on the massive amount of work already done on the study of symmetries and the mathematics (group theory) needed to describe them. Can we classify human languages by their automorphism groups (recall Cotton’s work above)? Are there characteristic properties that distinguish the automorphism groups for natural languages from other structured systems? A very simple example here are the automorphism groups of regular n-gons (n sided polygons). They are just the *dihedral* groups –  $2n$  automorphisms:  $n$  rotations and  $n$  reflections. To pursue answers to such a big question we first need to know much more about the group structure of well studied grammars. And this builds on precisely the sort of formal analysis of languages that linguists have been engaged in for the past half century.

For example, merely modeling the simplest type of agreement us taught us that some automorphisms are unstable – not preserved under linguistically trivial additions to the lexicon. The negative connotation of *unstable* is unfortunate and unintended. To the extent that agreement is characteristic of natural language we expect “unstable” automorphisms in their automorphism groups. How much other linguistic structure is coded in the automorphisms? Can we reconstruct grammatical category distinctions among lexical items from them? Perhaps lexical items of the same category as a lexical item  $d$  are just those in  $\text{Orbit}_{\text{stable}}(d)$ , the set of expressions that stable automorphisms can map  $d$  to.

Do any of the following structural notions from grammar force some distinctive property on the automorphisms of  $G$ : *paradigm, clitic, inflectional morphology, allomorph, subcategory, extraction and copy rules* (Kandybowicz 2008; Kobele 2006), *deletion rules, category changing operations, case and voice marking*? Are locality constraints or cyclic domains (or phases) characterizable in terms of automorphisms? We don’t know. And more fundamental perhaps, does the requirement of expressing predicate+argument structures impose any lower bound on the complexity of automorphisms? What about the existence of Modifier systems (adjectives, PPs, adverbs, even adjective modifiers like *a very tall student, a light-green car*)?

Lastly, in the other direction, are there concepts from invariance theory which apply to yield novel observations about human language? Here is one candidate. We agree that NLS have infinitely many expressions and that a learner projects a grammar on the

basis of finite exposure.

**Query** What must a finite sample of  $L_G$  be like to express all its structure types?

Certainly iterated application of the  $F$  in  $\text{Rule}_G$  will lead to new expressions at each “level” of application. But how can we tell whether we have enough to represent the language as a whole, beyond which further application of rules just “iterates” patterns already present? Consider iterative children’s texts such as:

- (8) a. This is the house that Jack built
- b. This is the malt that lay in the house that Jack built
- c. This is the rat that ate the malt that lay in the house that Jack built
- d. This is the cat that killed the rat that ate the malt that lay in the house that Jack built

By (8c) we see the pattern. We confirm it in (8d) and giggle as we start adding in our own clauses, repeating the core operations with new lexical items.

Here is a way of characterizing a structurally representative finite sample of an  $L_G$ . We first define the natural complexity hierarchy on  $L_G$ , for  $G = (\text{Lex}_G, \text{Rule}_G)$  as above. We write, standardly,  $K^*$  for the set of finite sequences of elements of  $K$  and we assume for simplicity that no lexical item is also a derived expression.

**Def 5** For  $G$  a grammar, the Complexity Hierarchy on  $L_G$  is  $\text{Lex}_0, \text{Lex}_1, \dots$  where:  $\text{Lex}_0 =_{\text{def}} \text{Lex}_G$  and for all  $n$ ,  $\text{Lex}_{n+1} =_{\text{def}} \text{Lex}_n \cup \{F(d) \mid F \in \text{Rule}_G, d \in \text{Lex}_n^* \cap \text{Dom}(F)\}$

**Theorem 2** For each  $n$ ,  $\text{Lex}_n$  is a finite subset of  $L_G$  and  $L_G =$  the union of the  $\text{Lex}_n$

Recall that  $\text{Lex}_G = \text{Lex}_0$  is finite by the definition of  $G$ . So each  $\text{Lex}_{n+1}$  is finite as it is just a finite addition to the finite set  $\text{Lex}_n$  (recall:  $\text{Rule}_G$  is finite and each  $F \in \text{Rule}_G$  has bounded arity). Obviously each  $\text{Lex}_n$  is a subset of  $\text{Lex}_{n+1}$ , in fact a proper subset as long as  $L_G$  is infinite: if  $\text{Lex}_n = \text{Lex}_{n+1}$  for some  $n$  then  $\text{Lex}_n = L_G$  since applying any rule  $F$  to any tuple of elements in  $\text{Lex}_n$  just yields an element of  $\text{Lex}_n$ , so  $\text{Lex}_n = L_G$ , contradicting that  $L_G$  is infinite.

Now we say what properties a  $\text{Lex}_k$  must have to contain all the “structure types” in  $L_G$ . Let us think of each  $\text{Lex}_n$  as a (finite) language. Its structure is given by *its* automorphisms: bijections from  $\text{Lex}_n$  to  $\text{Lex}_n$  which commute with the  $F$  in  $\text{Rule}_G$  in so far as they apply to tuples from  $\text{Lex}_n$  (which includes  $\text{Lex}_{n-1}$ ) to yield values in  $\text{Lex}_n$ . Write  $\text{Aut}_n$  for the set of such “local” automorphisms from  $\text{Lex}_n$  to  $\text{Lex}_n$ . Note that  $\text{Aut}_0$  is the entire set of bijections from  $\text{Lex}_0$  to  $\text{Lex}_0$  since any rule  $F$  that applies to tuples from  $\text{Lex}_0$  yields elements only of  $\text{Lex}_1$  as values, so automorphisms of  $\text{Lex}_0$  are not constrained to commute with anything. But suppose that  $\text{Rule}_G$  has a CaseMark rule combining lexical NPs with case markers, say *-nom* and *-acc*, to form KPs (Kase Phrases). Now all automorphisms  $g$  of  $\text{Lex}_1$  must commute with CaseMark and must (provably) fix  $\text{Dom}(\text{CaseMark})$ , thus requiring in effect that NPs get mapped to NPs and case markers to case markers. But  $\text{Aut}_1$  still has much “junk” in it, as a bijection of  $\text{Lex}_1$

that behaved nicely on NPs and case markers might still interchange lexical intransitive verbs and lexical transitive ones. Reasoning in this way we can see that it is only automorphisms on  $Lex_3$  that are forced to fix transitive and intransitive verbs. And we see that as we move up the complexity hierarchy at some  $n$  the automorphisms of  $Lex_n$  will be just the automorphisms of  $G$  restricted to  $Lex_n$ . (For  $h \in Aut_G$ ,  $h$  restricted to  $Lex_n$ , noted  $h \upharpoonright Lex_n$ , is  $h \cap Lex_n \times Lex_n$ ). And this yields, from K&S (p.168 – 170):

**Theorem 3 (Bounded Structure)** For  $G = (Lex_G, Rule_G)$  as above, there is an  $n$  such that the  $Aut_n$  is exactly the set of automorphisms of  $G$  restricted to  $Lex_n$

Given Bounded Structure, the *structure index*  $k(G)$  of a grammar  $G$  is<sub>def</sub> the least  $n$  such that  $Aut_n = \{h \upharpoonright Lex_n | h \in Aut_G\}$ . Then the structure maps on  $L_G$  are represented by those on the finite subset  $Lex_{k(G)}$ . So the *structure* of  $G$  is determined by the structure of a finite subset of  $L_G$ .

### Footnotes

<sup>‡</sup> This paper is a modestly augmented version of an LSA plenary session presentation by E.L. Keenan January 5, 2007

<sup>1</sup> Borer 2005 is something of an exception to this claim. And in general imputation of properties to MST are not intended to hold exceptionlessly for those who contribute to that tradition. Note too that features called *morphological* can be checked without requiring morphology to be overt.

<sup>2</sup> As is usual, each  $F$  in  $Rule_G$  has bounded arity. That is, for some natural number  $k$ , every tuple in  $Dom(F)$  is of length  $\leq k$ .

<sup>3</sup> Rules are partial functions, so we define  $g \circ f$  as that map with domain  $\{a \in Domf | f(a) \in Domg\}$  given by:  $(g \circ f)(a) = g(f(a))$ .

<sup>4</sup> Note that  $P_0$ s denote subsets of  $E^0 = \{\emptyset\}$ , so a  $P_0$  denotes an element of  $\{\emptyset, \{\emptyset\}\} = \{0, 1\}$ , our usual representation of the set of two truth values.

<sup>5</sup> Interpreting  $b$  as an  $n$ -tuple and  $bR$  as  $\{d | (b, d) \in R\}$  the equation in (4) is the general condition for  $F$  a map from  $P_{n+1}$ s to  $P_n$ s.

<sup>6</sup> This definition must be generalized to account for anaphors in a wider range of contexts: *Mary protected John from himself, Multiply 12 by itself, every worker's criticism of himself*, etc.

### References

- Aissen, J.L.(1987) *Tzotzil Clause Structure*. D. Reidel  
 Baker, Mark. 2003. *Lexical Categories*. CUP  
 Bar-Hillel, Y., C. Gaifman, and E. Shamir. 1960. On categorical and phrase structure grammars. Reprinted in *Language and Information* Y. Bar-Hillel Addison-Wesley 1964. pp. 99 – 115.  
 Bobaljik, J.D. 2003. Realizing Germanic inflection: why morphology does not drive syntax. In *Journal of Comparative Germanic Linguistics* 2003.

- Borer, Hagit 2005. *In Name Only*. Oxford
- Bošković, Željko and Howard and Lasnik (eds). 2007. *Minimalist Syntax: The Essential Readings*. Blackwell.
- Brewer, J.W. and M.K. Smith (eds.) (1981) *Emmy Noether: A Tribute to Her Life and Work* New York: Marcel Dekker.
- Bunch, Bryan. 1989. *Reality's Mirror*. John Wiley and Sons. New York.
- Chomsky, Noam. 1957. *Syntactic Structures* Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press
- Chomsky, Noam. 1975. *Reflections on Language*. Pantheon.
- Chomsky, Noam. 1986. *Knowledge of Language* Praeger
- Cole, K.C. 1997. *The Universe and the Teacup* Harcourt Brace & Co.
- Cole, Peter. and Hermon, Gaby. 2008. VP Raising in a VOS Language. *Syntax* **11.2**:144 – 197.
- Cotton, F. Albert. 1990. *Chemical Applications of Group Theory* 3<sup>rd</sup> edition. John Wiley & Sons.
- Darvas, G. 2007. *Symmetry*. Birkhäuser. Basel. Trans D.R. Evans.
- Derbyshire, Desmond. C. (1977). Word order universals and the existence of OVS languages. *Linguistic Inquiry* 8:590 – 599
- Dick, August. 1981. *Emmy Noether 1882 – 1935* Birkhäuser. H.I. Blocher (trans.)
- Gardner, Martin 2005. *The New Ambidextrous Universe* 3<sup>rd</sup> revised edition. Dover
- Gordon, R.G. (2005). *Ethnologue* 15<sup>th</sup> edition. SIL .
- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. 73 – 114. In J. Greenberg (editor) 1963. *Universals of Language*
- Henkin, Leon. 1961. Some remarks on infinitely long formulas. In *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics* pp. 167 – 183. Pergamon Press and Polish Scientific Publishers.
- Higginbotham, James. 1985. On semantics. In *Linguistic Inquiry* **16**:547 – 93.
- Hoffman, Roald. 1995. *The Same and Not the Same* Columbia University Press.
- Hopper, P.J. and E.C. Traugott (1993) *Grammaticalization* CUP.
- Hornstein, Norbert. 1995. *Logical Form: From GB to Minimalism*. Blackwell
- Jakobson, Roman 1963. Implications of language universals for linguistics. In J. Greenberg (ed) *Universals of Language* pp.263 – 279. MIT Press.
- Kandybowicz, J. (2008) *The Grammar of Repetition: Nupe Grammar at the Syntax-Phonology Interface*. Amsterdam: John Benjamins.
- Kayne, R. 1994. *The Antisymmetry of Syntax* MIT Press.
- Keenan, E.L. (1976) Remarkable subjects in Malagasy. In *Subject and Topic* C. Li (ed) Academic Press. Pp. 247 – 301.
- Keenan, E.L. 1981. A Boolean approach to semantics. In J. Groenendijk *et al* (eds.), *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum: 343 – 379.
- Keenan, E.L. 1988. On semantics and the binding theory. In *Explaining Language Universals* J. Hawkins (ed). Blackwell. Pp. 105 – 144.
- Keenan, E.L. 2001. Logical objects. In *Logic, Meaning and Computation* C.A. Anderson & M. Zelény (eds). Kluwer. Pp. 149 – 180.

- Keenan, E.L. 2004. Further excursions in natural logic: the mid-point theorems. in *Logics for Linguistic Structures* Fritz Hamm and Stephan Kepser (eds). OUP pp. 87 – 104.
- [K&S] Keenan, E.L. & E.P. Stabler. 2003. *Bare Grammar: Lectures on Linguistic Invariants* CSLI. See Stabler's Home Page for corrections.
- Klein, Felix. 1872. Comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society* vol II. Trans. M.W. Haskell. 1892-93:215 – 249.
- Kobelev, Gregory (2006) *Generating Copies: An Investigation into Structural Identity in Language and Grammar*. PhD. Diss. Dept. of Linguistics, UCLA
- Lee, Thomas Hun Tak. 1986. Acquisition of quantificational scope in Mandarin Chinese. *Papers and Reports on Child Language Development* no 25. Stanford University.
- Lewis, David. 1970. General semantics. In *Synthese* 22:18 – 67. reprinted in D. Davidson and G. Harman (eds) *Semantics of Natural Language* 1972:169 – 218.
- Montague, Richard. 1973 The proper treatment of quantification in ordinary English. Reprinted in *Formal Philosophy* Richmond Thomason (ed), Yale Univ. Press. 1974:247 – 270
- Radford, Andrew. 1997. *Syntax: A Minimalist Introduction*. CUP
- Rotman, J.J. 1999. *An Introduction to the Theory of Groups*. Springer, Graduate Texts in Mathematics. 2<sup>nd</sup> edition. Springer.
- Schachter, Paul. (Ed). 1984. *Studies in the Structure of Toba Batak*. UCLA Occasional Papers 5.
- Sternberg, S. 1995. *Group Theory and Physics* CUP.
- Stewart, Ian and Martin Golubitsky. 1999. *Fearful Symmetry*. Blackwell
- Strocchi, Franco. (2005) *Symmetry Breaking* Lecture Notes in Physics 732. Springer
- Tarski, Alfred. 1986. What are logical notions? (J. Corcoran, editor, from a 1966 lecture) in *History and Philosophy of Logic* 7:143 – 154.
- Van Fraassen, Bas C. 1990. *Laws and Symmetry* OUP.
- Weyl, Hermann 1952. *Symmetry* Princeton University Press
- Wigner, Eugene P. (1979) *Symmetries and Reflections: Scientific Essays*. Ox Bow Press. Woodbridge, Conn.
- Zee, A. 1986. *Fearful Symmetry*. Macmillan. New York.

\*\*

The Duality Corollary invites a deeper, less speculative comparison with MST: *eliminating redundancy can be overridden*. Among the virtues of axiom systems mathematicians include independence (non-redundancy) – no axiom should follow from the others. But

### **Symmetry trumps redundancy.**

For example, common axiomatizations for Boolean algebra contain the two distributivity laws:

$$(8) \quad \text{a. } (x \wedge (y \vee z)) = (x \wedge y) \vee (x \wedge z) \qquad \text{b. } (x \vee (y \wedge z)) = (x \vee y) \wedge (x \vee z).$$

In (8a) meet  $\wedge$  distributes over join  $\vee$  and in (8b) join distributes over meet. If (8a,b) are both removed then neither is entailed by the remaining axioms. But either one is eliminable and provable from the remaining ones, so including both is redundant. The reason for the redundant inclusion is symmetry: There is no basis for choosing among (8a) and (8b) – each is derivable from the other by duality. Choosing just one as an axiom would imply that it was basic and the other “derived”, creating an asymmetry where there is none. So here symmetry conflicts with redundancy and symmetry wins.

Lest the reader think Boolean algebra is atypical, here is a more fundamental example. A *group* is set  $G$  with an associative binary relation  $\bullet$  satisfying two additional axioms:

$$(9) \quad \text{a. Identities} \quad \text{There is an } e \in G \text{ such that for all } x \in G, e \bullet x = x \text{ and } x \bullet e = x \\ \text{b. Inverses} \quad \text{For all } x \in G \text{ there is a } y \in G \text{ such that } y \bullet x = e \text{ and } x \bullet y = e$$

Now the second conjuncts of (9a) and (9b) can be simultaneously eliminated and proven from the remaining axioms. But doing that implicates that having a left identity element and a left inverse is more basic than having a right identity and a right inverse – their existence being a “mere” theorem, not axiomatic. Again these implicatures introduce an unwarranted asymmetry. We can in fact keep both right hand conjuncts in (9a,b) and eliminate the two left hand ones, deriving them as theorems. Again symmetry trumps redundancy.