12th Int. Cong. Logic Methodology and Philosophy of Science, LMPS'03
Oviedo, Spain, August 7-13, 2003
correspondence to: ekeenan@ucla.edu

Edward L. Keenan · Edward P. Stabler

Linguistic Invariants and Language Variation

Since the publication of Noam Chomsky's field founding *Syntactic Structures* in 1957, generative grammarians have been formulating and studying the grammars of particular languages to extract from them what is general across languages. The idea is that properties which all languages have will give us some insight into the nature of mind. A widely acknowledged problem to which this work has led is how to reconcile the goal of generalization with language specific phenomena and the cross language variation they induce. Good science requires that cross linguistically valid generalizations be based on accurate, precise and thorough descriptions of particular languages. But such work on any given language increasingly leads us to describe language specific phenomena: irregular verbs, exceptions to paradigms, lexically conditioned rules, etc. So this work and cross language generalization seem to pull in opposite directions.

Here we propose an approach in which these two forces are reconciled. Our solution, presented in greater depth in *Bare Grammar* (Keenan and Stabler, 2003), is built on the notion of linguistic invariant. On our approach different languages do have non-trivially different grammars: their grammatical categories are defined internal to the language and may fail to be comparable to ones used for other languages. Their rules, ways of building complex expressions from simpler ones, may also fail to be isomorphic across languages. So languages differ. Nonetheless certain properties and relations may be invariant in all natural language grammars, as we will see below. And it is to these linguistic invariants that we should look for properties of mind.

Our approach contrasts with that of the most widely adopted linguistic theories, where the dominant idea is that *there is only one grammar*, the grammars of particular languages being, somehow, special cases. This has led to a mode of description in which grammars of particular languages are given in a notationally uniform way: the grammatical categories of all languages are drawn from a fixed universal set,[1] as are the rules characterizing complex expressions in terms of their components. It has also led to the postulation of a level of unobservable structure ("LF", suggesting

University of California, Los Angeles, Department of Linguistics

"Logical Form"), where structural properties of observable expressions may be changed in important ways. So this allows that structural generalizations which appear to be false on the basis of observable expressions may be true at LF where structural properties have been modified. We shall be concerned with one such case in this paper.

1. Linguistic Invariants

Consider the minimally complex expressions in (1):

(1) a. Casper coughed
b. Carson sneezed

Different linguistic theories - GB/Minimalism (Hornstein, 1995), HPSG (Pollard and Sag, 1994), LFG (Bresnan, 2001), Relational Grammar and Arc-Pair Grammar (Aissen, 1987) - differ with regard to the structure they attribute to (1a), and of course the notation they use to express that structure. But each of these theories would assign the same structure to (1a) and (1b). And it is this latter type of judgment - Under what conditions do X and Y have the same structure? - that forms the basis of the Bare Grammar (BG) approach.

Consider how we might argue pretheoretically that (1a,b) have the same structure. We agree that replacing 'Caspar' by 'Carson' in (1a) yielding *Carson coughed* does not change structure. And then replacing 'coughed' by 'sneezed' deriving thus (1b) does not change structure. So the intuition is that expressions X and Y have the same structure if each can be derived from the other by a succession of structure preserving transformations.

Here is a more explicit statement, leading up to our definition of *invariant*. We think of a grammar as a way of defining (and semantically interpreting) a class of expressions. Specifically the syntax of a grammar G is primarily a pair $(\mathrm{Lex_G}, \mathrm{Rule_G})$, where, omitting subscripts, Lex is a (normally) finite set of expressions, called *lexical items*, and Rule is a set of functions, called *generating* or *structure building functions*. $\mathrm{L_G}$, the *language generated by* G, is the set of all expressions you can build starting with those in Lex and applying the structure building functions finitely many times.

Lexical items on our view do present some internal structure. Like the expressions in $\mathrm{L_G}$ in general, they are partitioned into classes by grammatical categories. So we represent an expression, and in particular a lexical item, as an ordered pair (s, C) where s is a string over the vocabulary $\mathrm{V_G}$ of G and C is an element of the set $\mathrm{Cat_G}$ of category symbols of G. For any expression $\mathrm{e} = (\mathrm{s}, \mathrm{C})$, $\mathrm{Cat(e)} =_{\mathrm{df}} \mathrm{C}$, its second coordinate. Slightly more formally:

Definition 1. *A* bare grammar *G is a four-tuple,* $\langle \mathrm{V_G}, \mathrm{Cat_G}, \mathrm{Lex_G}, \mathrm{Rule_G} \rangle$, *where* $\mathrm{Lex} \subseteq \mathrm{V} \times \mathrm{Cat}$, *and* Rule *is a set of partial functions from* $(\mathrm{V}^* \times \mathrm{Cat})^+$ *into* $\mathrm{V}^* \times \mathrm{Cat}$. $\mathrm{V}^* \times \mathrm{Cat}$ *is the set of* possible expressions *over* G, *and the* language generated by G, $\mathrm{L_G}$, *is the closure of* Lex *under* Rule.

For any set K we can find a grammar G as above such that K is the set of strings of expressions in L_G. So any universal properties of natural language will have to be given explicitly as axioms (or consequences of other axioms), they do not follow from the mere formalism we use to express the grammar.

Definition 2. *An* automorphism *of a grammar* G *is a bijection* $h : L_G \to L_G$ *which fixes each* F *in* Rule*, that is,* $h(F) = F$. *This just means that* F *maps a tuple* $\langle s_1, \ldots, s_n \rangle$ *to* s_{n+1} *iff F maps* $\langle h(s_1), \ldots, h(s_n) \rangle$ *to* $h(s_{n+1})$.

Fact 1 id_{L_G}, the identity map on L_G, is in Aut_G, the set of automorphisms of G; so is h^{-1} whenever h is, and so is $g \circ h$ whenever g and h are. So Aut_G is a group, as expected.

Definition 3. *For all* $s, t \in L_G$*,* s *is isomorphic to* t*, noted* $s \simeq t$*, iff* $h(s) = t$ *for some* $h \in Aut_G$. *We write* [s] *for* $\{t \in L_G | \ s \simeq t\}$.

We may, when useful, treat [s] as the "structure" of s. In practice we have not found this very useful; $\simeq$, however, is a very useful relation.

Fact 2 For each G, $\simeq$ is an equivalence relation partitioning L_G into blocks $\{[s] | \ s \in L_G\}$.

Now, leading up to our definition of invariant, observe that whenever g is a function from a set A to a set B we can canonically lift g to a map P_g from $\wp(A)$, the power set of A, into $\wp(B)$ by setting $P_g(K) = \{g(x) | \ x \in K\}$. We usually just write $g(K)$ instead of $P_g(K)$. Similarly we can extend g to a map g^* from A^*, the set of finite sequences of elements of A, into B^* by setting $g^*(a_1, \ldots, a_n) = (g(a_1), \ldots, g(a_n))$. Again we usually write g for g^* here.

Definition 4. *The* invariants *of a grammar G are the expressions, properties (sets) of expressions, relations between expressions,... that are fixed, mapped to themselves, by all the automorphisms of G.*

So given a grammar, its (logical) invariants are those linguistic objects (expressions, properties of expressions, relations between expressions, functions from expressions to expressions,...) which cannot be changed without changing structure.

Later we introduce the notion of a *stable automorphism* and define the linguistic invariants of a grammar G to be those linguistic objects fixed by all stable automorphisms. But first let us learn to use the more general notion (and in any event in our initial examples of grammars the automorphisms and the stable automorphisms coincide).

2. Eng, an illustrative grammar for a fragment of English

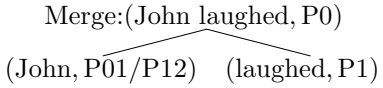
We present a very simple grammar Eng in order to illustrate in a concrete way the notions of grammar and invariant defined above. It has some proper nouns, like *John* and *Bill*, some one place predicate symbols (P1s), like *laughed* and *cried*, some two place predicate symbols (P2s), like *praised* and

criticized. We also have some conjunctions, *and* and *or* which form boolean compounds of expressions in a fairly obvious way. Finally, Eng has a reflexive pronoun *himself* that combines with P2s to form P1s, but does not combine with P1s to form anything. Eng has just two rules: Merge, which combines nominal elements with Pn+1s to form Pn's (we use P0 where many use 'S' for 'sentence'), and Coord which forms boolean compounds with *and* and *or*. Formally, Eng=$\langle V, Cat, Lex, Rule\rangle$, where these are given as follows:

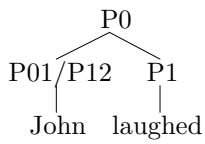
V:	laugh, cry, sneeze, praise, criticize, see,	
	John, Bill, Sam, himself, and, or, both, either	
Cat:	P0, P1, P2, P01/P12, P1/P2, CONJ	
Lex:	P1	laughed, cried, sneezed
	P2	praised, criticized, interviewed
	P01/P12	John, Bill, Sam
	P1/P2	himself
	CONJ	and, or
Rule:	Merge and Coord, defined below.	

Domain	Merge	Value	Conditions
s t A B	$\longmapsto$	s$^\frown$t P0	A = P01/P12, B = P1
s t A B	$\longmapsto$	t$^\frown$s P1	A $\in$ {P1/P2, P01/P12}, B = P2

So the domain of Merge is the set of pairs $\langle(s, A), (t, B)\rangle$, for any s, t in V^* and any A,B in Cat meeting the specified conditions. We summarize the argument that (John laughed, P0) is in L_{Eng} using a Function-Argument (FA) tree in which mother nodes are labeled with the values of generating functions applied to the labels on the daughter nodes:



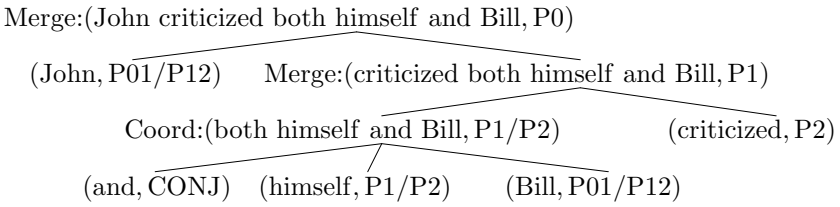
Linguists more often represent this derivation with slightly less explicit "standard" tree like the following:



Letting the set of coordinable categories cC_{Eng} = Cat − {CONJ} and the class of nominal categories nC_{Eng} = {P1/P2, P01/P12}, we define the other generating function Coord as follows:

Domain			Coord	Value	Conditions
and	s	t	$\longmapsto$	both$^\frown$s $^\frown$and$^\frown$t	$C \in cC_{Eng}$
CONJ	C	C		C	
or	s	t	$\longmapsto$	either$^\frown$s $^\frown$or$^\frown$t	$C \in cC_{Eng}$
CONJ	C	C		C	
and	s	t	$\longmapsto$	both$^\frown$s $^\frown$and$^\frown$t	$C \neq C' \in nC_{Eng}$
CONJ	C	C'		P1/P2	
or	s	t	$\longmapsto$	either$^\frown$s $^\frown$or$^\frown$t	$C \neq C' \in nC_{Eng}$
CONJ	C	C'		P1/P2	

This rule is used in the derivation of (John criticized both himself and Bill, P0), as we see in the following FA derivation tree:



3. Some invariants of Eng

E1. At the lowest level, the only expression that is invariant is (himself, P1/P2). The reason is that it has a unique distribution. It is the only lexical item that combines with P2s to form P1s but does not combine with P1s to form P0s.

E2. At the level of properties, we find several interesting invariants. First, the property of being a lexical item is invariant. That is, for all automorphisms h of Eng, $h(Lex_{Eng}) = Lex_{Eng}$. Indeed one might think that the property of being a lexical item was invariant in all G, but this is not the case.

E3. For each category C of Eng, the property of being an expression of category C is invariant. That is, for all $h \in Aut_{Eng}$, $h(PH(C)) = PH(C)$, where $PH(C) =_{df} \{s \in L_G |\ s = (t, C)$ for some string $t\}$. This also is not a universal invariant, as we see explicitly later.

E4. A more interesting invariant property in L_{Eng} is: the property of being an anaphor. Informally anaphors are expressions like *himself, both himself and Bill*, etc. which are obligatorily interpreted as referentially dependent in a certain way. (Below we provide a properly semantic, language independent, definition of 'anaphor'.) We can show that the (infinite) set of expressions in L_{Eng} which have this property is fixed by all the automorphisms of Eng.

E5. At the level of relations and functions, the binary relation *is a constituent of* (CON_{Eng}) is invariant, but this is universally invariant in the sense that for all G, CON_{G} is invariant (as explained in the next section). Also invariant, but not universally so, is the three place relation *s is a possible antecedent of an anaphor t in u*. To illustrate the intuition behind this relation consider that in the expressions below *himself* may be understood as referentially dependent as the underlined nominals in the expression, and if there is none it is ungrammatical (indicated by the asterisk):

(2) a. John thought that <u>the duke</u> defended himself well

b. *John thought that Mary defended himself well

c. <u>John</u> protected <u>Bill</u> from himself

E6. And lastly, as an example of an invariant (partial) function on L_{Eng} consider SUBJ_{Eng}, which maps a P0 to its subject it if has one: for any $\text{s} \in \text{L}_{\text{Eng}}$,

$$\text{Domain}(\text{SUBJ}_{\text{Eng}}) = \text{Range}(\text{Merge}) \cap \text{PH}(\text{P0})$$
$$\text{SUBJ}_{\text{Eng}}(\text{s}) = \text{t iff for some u of category P1}, \text{s} = \text{Merge}(\text{t}, \text{u}).$$

So SUBJ_{Eng}(both John and Bill praised Sam, P0) = (both John and Bill, P01/P12). But (Either John laughed or Bill cried, P0) is not mapped to anything by this function, since it is not in the range of Merge.

4. Universal invariants

We referred above to invariants as universal if they are invariant in all G, no matter how implausible G might be considered as a grammar for a natural language. So these are invariants that follow from our definition of a grammar plus that of invariant. But linguistically our interest lies primarily in properties, relations, etc. which are empirically invariant – they hold for all motivated grammars of natural language but admit of formal counterexamples. We shall argue that *is an anaphor* and *is a possible antecedent of* are two such cases. But first, let us list some universal invariants, since they place boundary conditions on empirical invariants and they are very useful in showing that one or another property of a particular grammar G is invariant. In our statements we use 'structural' and 'structurally definable' as synonyms of 'invariant'. We have the following, for all grammars G:

U1. L_{G} is invariant. That is, the property of being grammatical in G is structural.

U2. For any $\text{F} \in \text{Rule}_{\text{G}}$, F is invariant (trivially), as is its domain and range. So the property of being derived by any given $\text{F} \in \text{Rule}_{\text{G}}$ is structural.

U3. If $\mathrm{Lex_G}$ is invariant then for all n, $\mathrm{Lex_n}$ is invariant, where we define the complexity hierarchy $\mathrm{Lex_n}$ by: $\mathrm{Lex_0} = \mathrm{Lex_G}$ and for all n, $\mathrm{Lex_{n+1}} = \mathrm{Lex_n} \cup \{\mathrm{F(t)}|\ \mathrm{F} \in \mathrm{Rule_G},\ \mathrm{t} \in \mathrm{Lex_n^*} \cap \mathrm{Domain(F)}\}$.

Note that $\mathrm{L_G} = \bigcup_{\mathrm{n}} \mathrm{Lex_n}$ and if for all $\mathrm{F} \in \mathrm{Rule_G}, \mathrm{Range(F)} \cap \mathrm{Lex_G} = \emptyset$ then $\mathrm{Lex_G}$ is invariant.

U4. If G is category functional and each Lex(C) is invariant then each PH(C) is invariant, where $\mathrm{Lex(C)} =_{\mathrm{df}} \mathrm{PH(C)} \cap \mathrm{Lex}$ and G is *category functional* iff for all $\mathrm{F} \in \mathrm{Rule}$ and all n-tuples $\mathrm{u, v} \in \mathrm{Domain(F)}$, if $\mathrm{Cat(u_i)} = \mathrm{Cat(v_i)}$ all $1 \leq \mathrm{i} \leq \mathrm{n}$ then $\mathrm{Cat(F(u))} = \mathrm{Cat(F(v))}$.

U5. The set of invariant subsets of $\mathrm{L_G}$ is closed under relative complement and arbitrary intersections and unions, and thus forms a complete atomic boolean algebra (with atoms [s]). So conjunctions, disjunctions, and negations of invariant properties are themselves invariant properties. Comparable claims hold for $\mathrm{R} \subseteq (\mathrm{L_G})^{\mathrm{n}}$, for all n. Equally, cross products of invariant sets are invariant.

So if the property of being a feminine noun is invariant, and the property of being a plural noun is invariant then the property of being a feminine plural noun is invariant, as is that of being a feminine non-plural noun, etc.

U6. The is a constituent of relation, CON, is invariant, as are PCON (is a proper constituent of) and ICON (is an immediate constituent of), where for all $\mathrm{s, t} \in \mathrm{L_G}$, we define:

a. sICONt iff for some $\mathrm{u_1}, \ldots, \mathrm{u_n} \in \mathrm{L_G}$ and some $\mathrm{F} \in \mathrm{Rule_G}$, $t = F(u_1, \ldots, u_n)$ and $s = u_i$, some $1 \leq i \leq n$.

b. sPCONt iff for some $n \geq 2$ there is a sequence $v = \langle \mathrm{v_1}, \ldots, \mathrm{v_n} \rangle$ of elements of $\mathrm{L_G}$ with $\mathrm{v_1} = \mathrm{s}, \mathrm{v_n} = \mathrm{t}$ and for each $1 \leq i < n$, $\mathrm{v_i ICON v_{i+1}}$.

c. sCONt iff s = t or sPCONt

U7. The *sister of* relation is invariant, where, s sister of t in u iff some $\mathrm{F(v_1}, \ldots, \mathrm{v_n)}$ is a constituent of u and for some $\mathrm{i} \neq \mathrm{j}$, $\mathrm{s} = \mathrm{v_i}$ and $\mathrm{t} = \mathrm{v_j}$.

U8. CC, *c-commands*, is invariant, where, sCCt in u iff for some constituent v of u, s is a sister of v in u and t is a constituent of v.

U6-U8 define linguistic notions on expressions, not, as is more usual, on derivations or tree-like structures representing derivations. We give the definitions more generally than usual because there are a variety of linguistic phenomena that are not naturally representable with standard trees and in which constituency is not recoverable by merely segmenting the derived string. Examples are reduplication, second position placement of Latin *-que* 'and', and the Dutch crossing verb dependencies (see Keenan and Stabler 2003, Chapter 3).

5. Empirical invariants: Anaphor-Antecedent relations

For illustrative purposes we limit ourselves to the simplest environment in which non-trivial anaphora obtains: that between the two arguments of a binary relation denoting expression (e.g. a transitive verb). Consider the data pattern in English below, where the intended antecedent of the anaphor *himself* is underlined, and constituency is indicated by brackets for later reference:

(3) a. [Every student [criticized himself]]
b. *[Himself [criticized every student]]

A first attempt to describe these data might use left-right order: "X is a possible antecedent of an anaphor Y iff X and Y are co-arguments and X precedes Y". This claim works surprisingly well for quite a range of fairly simple sentences in English. But it is cross linguistically not valid. Languages such as Malagasy (Austronesian; Madagascar) and Tzotzil (Mayan; Mexico) which use Verb+Patient+Agent as a pragmatically neutral order in simple sentences, (4a,b), naturally present anaphors before their antecedents (5a,b)

(4) a. Namono ny akoho Rabe — Malagasy
Killed the chicken Rabe
'Rabe killed the chicken'

b. ʔi-s-poxta Xun li jʔilol-e — Tzotzil (Aissen 1987:90)
Asp-3-care Xun the shaman-clitic
'The shaman treated Xun'

(5) a. Namono tena Rabe — Malagasy
Killed self Rabe
Rabe killed himself

b. ʔi-s-poxta s-ba li Xun-e — Tzotzil
Asp-3-care 3-self art Xun-clitic
'Xun treated himself'

A more comprehensive proposal, accepted by many linguists as valid for natural languages in general, would replace "X precedes Y" with "X c-commands Y". This characterization of the AA (Anaphor-Antecedent) relation is consistent with the Tzotzil and Malagasy data above. But again it seems insufficiently general to account for a quite widespread language type: the verb is peripheral (usually final) and the arguments of the verb carry morphological markings, *case markers*, which identify the arguments. In the verb final case, illustrated below by Korean, the relative order of arguments is often rather free. We give the examples directly with the anaphors, but non-anaphoric nominals may replace them without change.

(6) [Caki-casin-ul [motun haksayng+tul-i piphanhayssta]] Korean
Self-emph-acc all student+pl-nom criticized
'All the students criticized themselves'

(7) [[Sinampal ng babae] ang sarili niya] Tagalog
slap+GF gen woman top self 3poss
'The woman slapped herself'

There is reasonable evidence in these cases that the antecedent of the anaphor does not c-command it; indeed the anaphor seems to asymmetrically c-command its antecedent. But the important structural regularity here concerns the case markers. They cannot be interchanged preserving grammaticality:

(8) *[Caki-casin-i [motun haksayng+tul-ul piphanhayssta]] Korean
Self-emph-nom all student+pl-acc criticized
'All the students criticized themselves'

(9) *[[Sinampal ang babae] ng sarili niya] Tagalog
slap+GF top woman gen self 3poss
'The woman slapped herself'

The c-command relations have not changed, but the case marking has, resulting in ungrammaticality. So case marking plays a structurally important role in these languages, and in our models is provably invariant.

The appropriate generalization for Korean then is: in simple sentences, X is a possible antecedent for an anaphor Y iff X and Y are co-arguments and X is *-i* marked and Y is *-ul* marked.[2] In Tagalog X is *ng* marked and Y is *ang* marked. Based on the Korean data we exhibit a mini-grammar for a verb final case marking language in which case relations determine the distribution of anaphors. We provide a compositional semantic interpretation, including a semantic, language independent, definition of anaphor, thereby establishing that the expressions we call anaphors are indeed interpreted as anaphors. But first let us give the language independent definition of anaphor (for the restricted class of contexts considered).

6. A semantic definition of 'anaphor'

For each domain E we interpret P2s as binary relations over E, represented as functions from E into $[E \to \{0,1\}]$. Anaphors and ordinary NPs, such *John, most of John's friends*, etc. map P2 denotations into $[E \to \{0,1\}]$. The difference in the two cases concerns what the values of the functions depend on. Compare:

(10) a. Sam criticized most of John's students
b. Sam criticized himself

In (10a) whether the denotation of *criticized most of John's students* holds of Sam is decided just by checking the set of objects that Sam criticized. If that set includes a majority of John's students the whole S is true. We don't need to know who Sam is. If Bill praised exactly the people that Sam criticized then (10a) and *Bill praised most of John's students* must

have the same truth value. In contrast it might be that the individuals Sam criticized are just those that Bill praised but (10b) and *Bill praised himself* have different truth values. Formally,

Definition 5. *Given a domain* E*, a binary relation* R *over* E*, and* $x \in E$*,*

$$xR =_{df} \{y \in E |\ (R(y))(x) = 1\}.$$

So in set notation, $xR = \{y \in E |\ (x, y) \in R\}$.

Let F *map binary relations to properties. Then* F satisfies the Extensions Condition (EC) *iff for all* $a, b \in E$*, all binary relations* R, S *over* E*,*

$$\text{if } aR = bS \text{ then } F(R)(a) = F(S)(b).$$

And F satisfies the Anaphor Condition (AC) *iff for all* $a \in E$*, all binary relations* R, S *over* E*,*

$$\text{if } aR = aS \text{ then } F(R)(a) = F(S)(a).$$

Let D combine with P2s to form P1s. Then D is an anaphor *iff all non-trivial* [3] *interpretations of D satisfy the AC but fail the EC.*[4]

So for example, for E with at least two members, the function SELF from binary relations to sets given by: $SELF(R)(x) = R(x)(x)$ is easily seen to fail the EC but satisfy the AC.

7. Kor, a verb final case marking language

Consider the following language Kor, inspired by Korean:

V: laughed, cried, sneezed, praised, criticized, saw, -nom, -acc, John, Bill, Sam, himself, and, or, nor, both, either, neither

Cat: NP, NP_{refl}, Ka, Kn, KPa, KPn, P0, P1a, P1n, P2, CONJ

Lex:

Kn	-nom
Ka	-acc
P1n	laughed, cried, sneezed
P2	praised, criticized, interviewed
NP	John, Bill, Sam
NP_{refl}	himself
CONJ	and, or, nor

Rule: CM (case mark), PA (predicate-argument) and Coord, as follows.

Domain		CM	Value	Conditions
-nom	t	$\longmapsto$	t$^\frown$-nom	(none)
Kn	NP		KPn	
-acc	t	$\longmapsto$	t$^\frown$-acc	$X \in \{NP, NP_{refl}\}$
Ka	X		KPa	

Domain		PA	Value
s	t	$\longmapsto$	s⌢t
KPn	P1n		S
s	t	$\longmapsto$	s⌢t
KPa	P1a		S
s	t	$\longmapsto$	s⌢t
KPn	P2		P1a
s	t	$\longmapsto$	s⌢t
KPa	P2		P1n

Letting the coordinable, "boolean" categories be

$$\mathrm{cC_{Kor}} =_{\mathrm{df}} \mathrm{Cat} - \{\mathrm{CONJ}, \mathrm{Ka}, \mathrm{Kn}, \mathrm{KPa}, \mathrm{KPn}\}$$

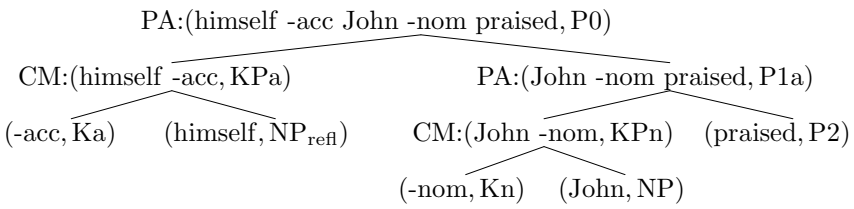
and the nominal categories be

$$\mathrm{nC_{Kor}} =_{\mathrm{df}} \{\mathrm{NP}, \mathrm{NP_{refl}}\},$$

we define a coordination rule as follows:[5]

Domain			Coord	Value	Conditions
and	s	t	$\longmapsto$	both⌢s ⌢and⌢t	$\mathrm{C} \in \mathrm{cC_{Kor}}$
CONJ	C	C		C	
or	s	t	$\longmapsto$	either⌢s ⌢or⌢t	$\mathrm{C} \in \mathrm{cC_{Kor}}$
CONJ	C	C		C	
nor	s	t	$\longmapsto$	neither⌢s ⌢nor⌢t	$\mathrm{C} \in \mathrm{cC_{Kor}}$
CONJ	C	C		C	
and	s	t	$\longmapsto$	both⌢s ⌢and⌢t	$\mathrm{C} \neq \mathrm{C}' \in \mathrm{nC_{Kor}}$
CONJ	C	C'		$\mathrm{NP_{refl}}$	
or	s	t	$\longmapsto$	either⌢s ⌢or⌢t	$\mathrm{C} \neq \mathrm{C}' \in \mathrm{nC_{Kor}}$
CONJ	C	C'		$\mathrm{NP_{refl}}$	
nor	s	t	$\longmapsto$	neither⌢s ⌢or⌢t	$\mathrm{C} \neq \mathrm{C}' \in \mathrm{nC_{Kor}}$
CONJ	C	C'		$\mathrm{NP_{refl}}$	

The following tree represents the argument that (himself-acc John-nom praised, P0)∈L(Kor).



This is the only derivation of this expression, and so, in this expression, (himself,$\mathrm{NP_{refl}}$) c-commands and is not c-commanded by (John-nom,KPn).

8. Some invariants of Kor

K1. The set Lex is invariant. So by U3, $\mathrm{Lex_n}$ is invariant for each n.

K2. The expressions (-nom,Kn) and (-acc,Ka) are both invariants.

Pretheoretically case markers are grammatical formatives, so the fact that they are provably invariants in Kor supports that our formal notion of invariant identifies expressions independently judged to be grammatical in nature. So no automorphism can interchange (-nom,Kn) and (-acc,Ka).

K3. The expression (himself,$\mathrm{NP_{refl}}$) is invariant, but (Bill,NP) is not.

K4. For all $\mathrm{C} \in$ Cat, the set PH(C) of expressions of that category is invariant.

K5. The *co-argument* relation is invariant, defined by: s co-argument t in u iff for some v of category P2, either PA(s,PA(t,v)) or PA(t,PA(s,v)) is a constituent of u.

9. Semantic interpretation for Kor

This section provides L(Kor) with a compositional semantics which shows that sentences with reflexives are interpreted correctly in all cases. Those willing to take our word for this can move directly to the next section. We assume a modest familiarity with a model theoretic semantics and boolean lattices.

Definition 6. *Given a non-empty universe* E*, we let* $\mathrm{R_0} =_{\mathrm{df}} \{0,1\}$*, regarded as the boolean lattice 2 where the* $\leq$ *relation coincides with the numerical one. In general* $\mathrm{R_{n+1}}$ *is* $[E \rightarrow R_n]$*, regarded as a boolean lattice with* $\leq$ *understood pointwise:* $\mathrm{f} \leq \mathrm{g}$ *iff for* $\mathrm{x} \in \mathrm{E}$, $\mathrm{f(x)} \leq \mathrm{g(x)}$.

Type 1 is the set of functions from n+1-ary relations to n-ary ones, for all n:

$$\{\mathrm{f} \in [\bigcup \mathrm{R_{n+1}} \rightarrow \bigcup \mathrm{R_n}] | \text{ for all n, all } \mathrm{r} \in \mathrm{R_{n+1}},\ \mathrm{f(r)} \in \mathrm{R_n}\}.$$

Definition 7. *A model for* L(Kor) *is a pair* $\mathrm{M} = \langle \mathrm{E}, \mathrm{m} \rangle$*,* E *a non-empty domain and* m *a function mapping elements* $\langle \mathrm{v}, \mathrm{C} \rangle$ *of* Lex *into* $\mathrm{Den_E(C)}$*, the set of possible denotations of expressions of category* C *in* M*, defined as follows. Note in particular the definition of NOM(f); its value at properties determines its value at relations.*

$\mathrm{Den_E(NP_{refl})}$ = {f ∈ Type 1| if nontrivial, f satisfies AC and fails EC}
$\mathrm{Den_E(P0)}$ = $\mathrm{R_0}$
$\mathrm{Den_E(P1n)}$ = $\mathrm{R_1}$
$\mathrm{Den_E(P2)}$ = $\mathrm{R_2}$
$\mathrm{Den_E(NP)}$ = Type1
$\mathrm{Den_E(KPa)}$ = Type1
$\mathrm{Den_E(P1a)}$ = [Type1 → $\mathrm{R_1}$]
$\mathrm{Den_E(CONJ)}$ = $\{\wedge_{\mathrm{C}}, \vee_{\mathrm{C}}\}$, *where* $\wedge_C$ *is the greatest lower bound operator in* $\mathrm{Den_E(C)}$ *and* $\vee_C$ *is the least upper bound operator*
$\mathrm{Den_E(KPn)}$ = {NOM(f)| f ∈ Type1}, *where for any* f ∈ Type1, *NOM is the function with domain* $\mathrm{R_1 \cup R_2}$ *such that for* $\mathrm{P \in R_1}$, $\mathrm{NOM(f)(P) = f(P)}$ *and for* $\mathrm{R \in R_2, h \in}$ Type1, $\mathrm{NOM(f)(R)(h) = f(h(R))}$

1. m *at elements of* Lex *satisfies the following conditions:*

a. for all $\mathrm{s \in Lex(NP)}$, $\mathrm{m(s) \in \{I_b|\ b \in E\}}$, *where for all* $\mathrm{R \in R_{n+1}}$, $\mathrm{I_b(R) = R(b)}$

b. $\mathrm{m(}$-*acc*$\mathrm{, Ka)}$, *noted* ACC, *is the identity map on Type 1.*

c. $\mathrm{m(}$-*nom*$\mathrm{, Kn) = NOM}$, *defined above*

d. $\mathrm{m(himself, NP_{refl}) = SELF}$, *that map from* R2 *to* R1 *defined earlier*

e. for all $\mathrm{x, y \in Den_E(C)}$, C *boolean,*

$$\mathrm{m(and, CONJ) = \wedge_C} \qquad \textit{and} \qquad \mathrm{m(or, CONJ) = \vee_C}$$

2. m *extends to a function* $\mathrm{m^*}$ *on* L(Kor), *called an interpretation of* L(Kor) *relative to* M, *by:*

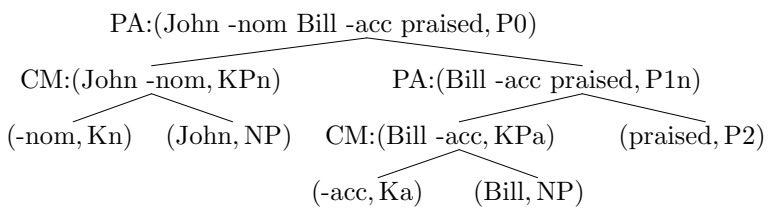
a. $\mathrm{m^*(CM(s,t)) = m(s)(m^*(t))}$

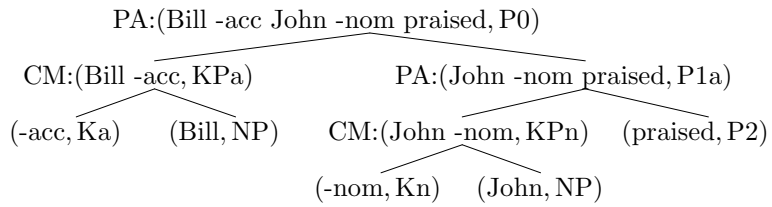
b. $\mathrm{m^*(PA(s,t)) = }\begin{cases} m^*(s)(m^*(t)) & \text{if Cat(s) = KPn and Cat(t) = P1n} \\ m^*(t)(m^*(s)) & \text{otherwise} \end{cases}$

c. $\mathrm{m^*(Coord(s,t,u)) = m(s)(m^*(t), m^*(u))}$

Using these definitions one computes that (11a,b) are logically equivalent (always interpreted the same): for all models $\mathcal{M} = \mathrm{(E, m)}$, $\mathrm{m^*}$(11a) = $\mathrm{m^*}$(11b).

(11) a. (John-nom Bill-acc praised, P0)
b. (Bill-acc John-nom praised, P0)





The logical equivalence of these sentences relies on the interpretation of (-nom,Kn). When the nominative KP looks at a P2, in effect, it knows to wait until the next KP denotation comes along. So the interpretation of bound morphology here is critical. Moreover the same reasoning shows that the result of replacing (Bill,NP) by (himself,$\mathrm{NP}_{\mathrm{refl}}$) in (11a,b) are also logically equivalent:

$$\mathrm{m}^*(\text{John-nom himself-acc criticized}, \mathrm{P0}) = \mathrm{m}^*(\text{himself-acc John-nom criticized}, \mathrm{P0})$$

Thus the interpretation of *himself* as an anaphor does not depend on it being c-commanded by its antecedent. We note that these sentences, like (11a,b), have isomorphic derivation trees (standard or FA). But the expressions are not isomorphic in L(Kor) since automorphisms can't map KPn's to KPa's, P1n's to P1a's, etc.

10. Two further invariants of Kor

Now we are in a position to state invariants that involve semantic notions.

K6. The property of being an anaphor is invariant, where the expressions interpreted as anaphors following Definition 5 are precisely those in PH(P1/P2).

K7. The Anaphor-Antecedent relation is invariant in Kor, where we define:

s AA t in u iff t is an anaphor and s co-argument t in u

(AA is invariant because it is defined as a boolean compound of invariants).

11. Concluding remarks on Kor

It is unproblematic that anaphors asymmetrically c-command their antecedents. The interpretation of case markers guarantees the right semantic interpretation (sentence internally) independent of c-command. We also note that a compositional interpretation of L(Eng) is even easier than of L(Kor), and that *himself* in Eng denotes SELF, just as *himself* in Kor does. So our claims about anaphors are claims about expressions with the same denotation.

Morphology is structural, independent of c-command relations within the clause. The case markers, (-nom,Kn) and (-acc,Ka), are invariant even

though the KPs they build do not have fixed structural positions. Specifically a KPa does not always combine with a P2 to form a P1; it also combines with P1s to form P0s.

Our formulation of Kor abstracts away from the conditioned variants of the case markers: *-i/-ka* for -nom and *-ul/-lul* for -acc. This seems reasonable when our concern is syntax and semantics, as these differences in form are phonologically conditioned.

Still, an interesting option arises when we do distinguish two categories of NP in Lex, say NPc and NPv (according as the string coordinate ends in a consonant or a vowel). So Lex would contain (John,NPc) and (Joe,NPv) of different categories, but ones that had the same distribution except for the choice of case marker: *-i, -ul* in the first case, *-ka, -lul* in the second. And we would then find that if the cardinalities of the lexical NPv's and NPc's were the same (permitting a bijection between them) we could design an automorphism that would map all NPv's to NPc's and conversely. It would also interchange (-i,Kn) with (-ka,Kn) and (-ul,Ka) and (-lul,Ka). The resulting grammar would be one in which not all PH(C) were invariant.

12. Categorial symmetry and stable automorphisms

The case of conditioned variants noted above for Korean has much more extensive and systematic manifestations in other grammatical subsystems. In BG for example we present a grammar, Span (Spanish), illustrating basic adjective and determiner agreement with masculine (m) and feminine (f) nouns. The Lexicon arbitrarily distinguishes Nm's and Nf's, and when adjectives and determiners combine with them they get marked with an *-o* or an *-a*, of category Agr(m) and Agr(f) respectively. The m/f distinction is inherited by NPs built from the Nm's and Nf's, and then the P1s show predicate agreement with them.

And analogous to the Korean case, if we design the grammar so that the number of lexical Nm's and Nf's is the same then we can find an automorphism of Span which interchanges PH(Nm) and PH(Nf), as well as the derived masculine and feminine adjectives, NPs and P1s. So again not all PH(C) are invariant in Span. However the automorphisms that can effect this category swapping are unstable in that slight additions to the Lexicon rule out their existence. Thus if we add just one new feminine noun, say (poet,Nf) making no other changes then no automorphism changes category and all PH(C) are invariant since then the lexical Nm's and the lexical Nf's would have different cardinalities, so there could be no bijection between them.

The possibility of category changing automorphisms above reveals a categorial symmetry present, in principle, in natural language. Noun classes partition a subset of the expressions in such a way that the blocks of the partition can be structurally interchanged. This possibility is "unstable" in the sense that many "minor" changes in the language, ones we agree are insignificant, such as adding new lexical items, result in languages in which these blocks cannot be interchanged.

Ignoring this accidental possibility would be, we feel, a mistake. A grammar with unequal numbers of lexical Nm's and Nf's could always be extended by adding new lexical items to one in which the numbers evened out again, permitting category changing automorphisms. And the ability to add new content words freely is a basic property of a NL. More generally various types of allomorphy present a similar phenomenon. In English we might distinguish classes of Nouns according to how their plural is formed: with /z/ as in *dog/dogs*, with /s/ as in *cat/cats*, with /əz/ in *judge/judges*, /f/→/vz/ as in *leaf/leaves*, *-on→-a*, as in *phenomenon/phenomena*, no change as in *sheep→sheep*, etc.

We will treat agreement and allomorphy by distinguishing among automorphisms according as they remain stable under such changes. Informally, an automorphism is stable if it remains an automorphism after the addition of new expressions isomorphic to old ones. "New" means not inducing new derivations of expressions in the original language (thanks to Greg Kobele for this formulation, and thanks to Philippe Schlenker for forcing us to treat allomorphy):

Definition 8. *For* $G = \langle V, Cat, Lex, Rule\rangle$ *and* $S \subseteq_{finite} V \times Cat$,

a. $G[S] =_{df} \langle V, Cat, Lex \cup S, Rule\rangle$. *Write* $G[s]$ *or* G_s *for* $G[\{s\}]$, $s \in V \times Cat$. *So* Gs *results from adding* s *to* Lex_G *with no changes in* Cat *or* Rule.

b. G *is free for* s *in* $V \times Cat$ *iff*

i. for all $t \in L(G_s)$, *if* $t \in L_G$ *then* $\neg$(sCONt), *and*

ii. For some $h \in Aut_{G_s}$ *and some* $t \in Lex_G$, h *interchanges* s *and* t *and fixes all other elements of* Lex_{G_s}.

iii. G *is free for* S *iff for all* $s \in S$, G *is free for* s *and* G_s *is free for* $S - \{s\}$. *(Note that all* G *are free for* $\emptyset$.*)*

So (b.i) blocks adding as new lexical items expressions that are already in L_G.

Definition 9. $h \in Aut_G$ *is* stable *iff* h *extends to an* $h' \in Aut_{G[S]}$, *all finite* S *for which* G *is free.*

An expression, a property of expressions,... over G is a linguistic invariant *iff it is fixed by all stable automorphisms.*

Of course all logical invariants of a grammar are linguistic invariants since an object fixed by all automorphisms is a fortiori fixed by all stable automorphisms. But the converse may fail. In Kor enriched with the phonologically conditioned case markers PH(NPv) is a linguistic invariant but not a logical one. Equally each case marker (-i, Kn), (-lul,Ka), etc. is a linguistic invariant (but not a logical one). And in Span PH(Nm) is a linguistic invariant but not an logical invariant, as is each agreement marker (-o,Agr(m)), (-a,Agr(f)).

13. Conclusion

We have provided a way of establishing invariants of natural languages while countenancing that different languages may have quite different grammars. Our specific claims, that *is an anaphor* or *is a possible antecedent of* are invariant in all natural languages, are empirical, not mathematical, and further empirical research could show them false.

In addition our approach has led us to formulate several conceptually new generalizations about natural language. Here are two, of somewhat different sorts:

Stable Categories In adequate natural language grammars G, each PH(C) is a linguistic invariant

Thesis Grammatical Formatives are linguistically invariant lexical items.

The Thesis above offers a characterization of those expressions linguists variously call "function words" or "grammatical formatives". To our knowledge this is the first non-stipulative characterization of these objects. In contrast, Stable Categories is offered as an axiom of a theory of language structure. It provides a principled account of how the expressions of a language may be partitioned into grammatical categories. They are sets of expressions fixed by all stable automorphisms.

Notes

[1]Advocates of this approach intend more than the claim that we use the same notation for grammatical categories in different languages but it is quite unclear what this "more" is.

[2]In more detail, an expression is -nom marked iff it is suffixed with *-i* if it is consonant final and with *-ka* if it is vowel final. It is -acc marked iff it is suffixed with *-ul* if consonant final and *-lul* if vowel final. In addition either argument (but not both) can have their -nom/-acc suffixes replaced with a topic marker *-un/-nun* preserving the pattern of antecedence. Then a more accurate statement of the AA relation would be: "...X is -nom marked and Y is -acc marked or topic marked, or X is -nom marked or topic marked and Y is -acc marked". The important point remains: the relevant factor governing the distribution of anaphor and antecedent in simple sentences concerns their morphological marking, not their left-right order or c-command relations.

[3]It is assumed here that the universe E of interpretation always has at least two elements. The non-triviality condition is intended for cases like at least two of the ten students besides himself, which requires for non-triviality that the E contain exactly ten students.

[4]The definition of EC and AC and hence of anaphor generalizes directly to maps from n+1-ary relations to n-ary ones just by interpreting a and b as n-tuples rather than "1-tuples".

[5]In head initial languages (Verb initial, or SVO as in English) framing coordinations follow the English pattern (*both X and Y, either X or Y, neither X nor Y*), though the more typical case is where the conjunctive morphemes are the same, as in French: *et Jean et Marie, ou Jean ou Marie, ni Jean ni Marie*. A case can be made that in verb final languages the order is *X and Y and, X or Y or*, etc. though in our examples from Korean we did not find such framing expressions, only infix coordinators. We include the framing construction to avoid semantic ambiguities with iterated coordinations. We are not really studying either coordination or ambiguity here, but we include coordination so that many categories of expression will have infinitely many members, forcing us to avoid non-general definitions by listing cases.

References

Aissen, Judith. 1987. *Tzotzil Clause Structure*. Reidel, Dordrecth.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell, Oxford.

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton, The Hague.

Hornstein, Norbert. 1995. *Logical Form: From GB to Minimalism*. Basil Blackwell, Oxford.

Keenan, Edward L. and Edward P. Stabler. 2003. *Bare Grammar*. CSLI Publications, Stanford, California.

Pollard, Carl and Ivan Sag. 1994. *Head-driven Phrase Structure Grammar*. The University of Chicago Press, Chicago.

Index