UCLA Working Papers in Linguistics, Theories of Everything
Volume 17, Article 45: 394-410, 2012

Learnability and the autonomy of syntactic categories

Edward P. Stabler

Linguistics might be easier if there were a fixed, finite number of string substitution tests that could distinguish any pair of categories, or if the categories were fixed and finite across grammars of all possible human languages. But the evidence does not support these simple ideas (at least, as we can understand them now), and so linguistics is more of a challenge, and language acquisition is more of a mystery, than might otherwise have been expected. The familiar and persuasive evidence for this assessment is reviewed here, but with particular attention to how these facts are compatible with the existence of significant linguistic universals, and with the fact that languages are learned easily by normal human infants.

Keywords language, grammar, automorphism, learnability

0 Introduction

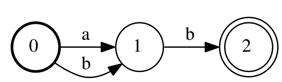
Some linguists have remarked that the conception of linguistic categories and structure in Keenan and Stabler (2003) is very 'intensional' in the sense that structure depends entirely on grammar, and of course grammars vary with speaker and language. This leaves us with the problem of getting from the 'extension' of a language (its spoken expressions), possibly with other relevant evidence, to the right structure. But no oracle has provided us with secure, easily applicable criteria, extensional or otherwise, for telling whether we have the right grammar for any language; we have found no 'royal road' to linguistic truth. But a royal road is essentially what the facts of language learnability seem to require. Learnability considerations are sometimes thought to require that linguistic categories must be pinned down distributionally by finite tests on their pronounced extensions, or else categories given once and for all in the one and only universal grammar. But familiar arguments show that those ideas cannot be right (at least, as we can understand them now). If they were, language would not be the amazing human creation we know it to be. In fact, every reasonable characterization of human language must be 'intensional' in something like the way it is in Keenan and Stabler (2003). But this does not spell doom for significant, restrictive universals, or for language acquisition that depends on them. On the contrary, it draws attention to some basic facts about how language acquisition must work. This paper reviews some advances in learnability, and then some linguistic perspectives, to argue that in spite of apparent differences a coherent perspective is emerging.

1 Learning some simple languages

The Nerode-Myhill theorem shows that the states of a minimal deterministic finite automaton can be regarded as sets of expressions.[1] Given any alphabet Σ, any language $L \subseteq \Sigma^*$ and any sentences $u, v \in \Sigma^*$, let's say $u \equiv_L v$ iff for all $w \in \Sigma^*$, $uw \in L$ iff $vw \in L$. That is, u and v are equivalent iff they can appear in all the same 'good contexts', w. For regular languages, the relevant contexts are the 'good finals', the suffixes of the well formed expressions. If L is regular, the states of minimal deterministic finite automaton for L can be regarded as the sets of 'good finals' $\{\{v\colon\ uv \in L\}\colon\ u \in \Sigma^*\}$. With this perspective, every suffix of every sentence in the language is a member of one or more of these categories. Learnability results for subsets of the regular languages often quite explicitly depend on some such identification, since then it is completely clear how, in any finite sample of the language, the learner is seeing finite approximations to the categories of the minimal automaton that generates the language.

Angluin (1982) provides a learnability result for the infinite subset of the regular languages, the zero-reversible languages, in which finding one good final, v, shared by two different prefixes u_1 and u_2, guarantees that u_1 and u_2 have all their good finals in common. That is, the state reached by accepting u_1 in a minimal deterministic automaton must be the same state reached by accepting u_2. So the minimal automata for these languages are distinguished by being 'deterministic in reverse' too, with exactly one final state, and where every state has at most one entering transition for each $a \in \Sigma$. This property provides a finite basis for category identification, so that learning is possible (and even efficient in a certain sense) from positive data.

When a language is defined with a context free grammar, the situation is more complicated. Each category of a context free grammar derives a set of strings which might occur only in the middle of an expression, and the boundaries of these 'middles' u in their 'good contexts' $\langle v, w\rangle$, where vuw is derivable, may not be indicated in any way. But Clark and Eyraud (2007) show how Angluin's result for zero-reversible regular languages can be generalized to context free languages in which finding one good context, $\langle v, w\rangle$, shared by two different substrings u_1 and u_2, guarantees that u_1 and u_2 have all their contexts in common. These languages are learnable from positive evidence.[2] Clark and Eyraud (2007) point out that the artificial language $\{ab, bb\}$ is zero-reversible but not substitutable, as we can see from the fact that its minimal deterministic automaton is also reverse deterministic:



and from the fact that a, b share some but not all their contexts:

<u>'middles'</u>	<u>contexts</u>
a	$\langle \varepsilon, a\rangle,\ *\langle a, \varepsilon\rangle$
b	$\langle \varepsilon, a\rangle,\ \langle a, \varepsilon\rangle.$

[1]See for example Hopcroft and Ullman (1979:§3.4) or Moll, Arbib, and Kfoury (1988:§8.2).

[2]Clark and Eyraud (2007) first propose a learning function that finds very large and redundant grammars, and then consider the problem of how to get a more compact representation. This is closely related to the natural ideas (discussed below) that the categories should not draw syntactically irrelevant distinctions (Keenan and Stabler 2003:141).

The languages Σ^* and $\{wcw^r\colon\ w \in \{a,b\}^*\}$, on the other hand, are substitutable, where w^r is the reverse of w.

Context free grammars (CFGs) are not appropriate for the definition of human languages, even for ones that may have context free string languages (Chomsky 1956; Stabler 2012). Mildly context sensitive grammars are more expressive than CFGs. For example, multiple context free grammars (MCFGs) which generalize CFGs by allowing a category to derive a pair (or in general a k-tuple) of strings instead of just a single string. The previous learning results can be extended into this class with learners that consider contexts $\langle v,w,z\rangle$ for possibly discontinuous pairs of strings $\langle s,t\rangle$, such that $vswtz$ is well-formed. Let's say that a language is k-substitutable iff whenever two k-tuples share a context, they share all contexts. When $k=1$, we have substitutable context free languages, but with $k>1$, larger classes of languages become learnable (Yoshinaka 2011). These ideas also extend to languages defined by certain 'parallel' MCFGs (PMCFGs), k-dimensional grammars with rules that allow copying (Clark and Yoshinaka 2012).

These recent learnability results – coming from first results on regular languages, then extended to large subsets of context free and context sensitive languages – are exciting partly because they suggest that the some of the evidence linguists use for determining structure might also be used reflexively by language learners. And even if this turns out not to be the whole story, it is still valuable to explore precisely and systematically some of the kinds of substitution tests we find in the linguistics literature – standard fare in introductory linguistics texts. But we can see that the learnability results mentioned here are established with fundamental assumptions that may be troubling:

(Cong) Categories are blocks of a coarsest congruence of L(G) and invariant

(Conc) Complexes are formed by concatenation (or similar functions on tuples[3])

(Subst) If two pronounced (tuples of) sequences have one context in common, they are the same category and have all contexts in common.

First, let's briefly introduce these properties and show how they hold in the classes of artificial languages mentioned above, classes for which we have learnability results.[4]

2 Fundamental properties of some artificial languages

For comparing grammars, it is useful to have a very expressive formalism like 'bare grammars' $G = \langle \Sigma, Cat, Lex, \mathscr{F}\rangle$, where V, Cat are nonempty sets, $Lex \subseteq (\Sigma^* \times Cat)$, and $\mathscr{F} \subseteq [(\Sigma^* \times Cat)^* \rightarrow (\Sigma^* \times Cat)]$. The language L(G) of the grammar is the closure of Lex with respect to the functions in $\mathscr{F}$. For any category C, the phrases of category C

$$Ph(C) = \{\langle s,C\rangle\colon\ \langle s,C\rangle \in L(G)\}.$$

[3] Extensions of the standard string function, concatenation, to tuples of strings are defined by Seki, Matsumura, Fujii, and Kasami (1991). For an analogous logical perspective see Morrill and Valentin (2010).

[4] A careful reader might notice, even before the properties Cong, Conc and Subst are carefully explained, that all of them refer to grammars. This might seem odd since standard measures of learning success depend on the learner's language, not the learner's grammar. But the learnable language classes are defined by grammars with these standard properties, and the learner's hypotheses are grammars with these properties too. As we will see in §3 below, linguists might worry that, with these properties, we have excluded appropriate grammars for human languages.

For the 'start' category (or any category) S, the strings of category S,

$$Str(S) = \{s\colon\ \langle s,S\rangle \in Ph(S)\}.$$

An interpretation maps expressions in $L(G)$, or their derivations, to semantic values. An automorphism of $\langle L(G),\mathscr{F}\rangle$ is a bijection on $L(G)$ that, when applied pointwise to each rule in $\mathscr{F}$, leaves each rule unchanged. We will call a property or relation on expressions that is fixed by the automorphisms structurally invariant, or structural. And two expressions are said to have the same structure iff some automorphism maps one to the other.

Keenan and Stabler (2003:§3.1) point out that CFGs have a straightforward translation into this framework. The CFG on the left below, for example, is represented by the bare grammar on the right:

$S \to ASA$	$\Sigma = \{a,b,c\}$
$S \to BSB$	$Cat = \{A,B,S\}$
$S \to c$	$Lex = \{\langle a,A\rangle,\langle b,B\rangle,\langle c,S\rangle\}$
$A \to a$	$\mathscr{F} = \{f,g\}$
$B \to b$	

where f is defined as follows:

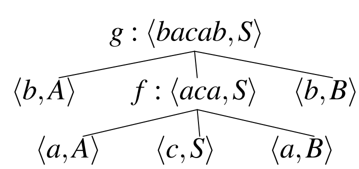
$$\begin{array}{ccc} \text{Domain} & f & \text{Value} \\ \begin{array}{ccc} s & t & u \\ A & S & A \end{array} & \longmapsto & \begin{array}{c} stu \\ S \end{array} \end{array}$$

This notation from Keenan and Stabler (2003) indicates that for any $s,t,u \in \Sigma^*$, f applies to $\langle s,A\rangle$, $\langle t,S\rangle$, and $\langle u,A\rangle$ to produce $\langle stu,S\rangle$. And g is defined similarly:

$$\begin{array}{ccc} \text{Domain} & g & \text{Value} \\ \begin{array}{ccc} s & t & u \\ B & S & B \end{array} & \longmapsto & \begin{array}{c} stu \\ S \end{array} \end{array}$$

Clearly $L(G) = Lex \cup \{\langle wcw^r,S\rangle\colon\ w \in \{a,b\}^*\}$, and as mentioned above, the string language $Str(S) = \{wcw^r\colon\ w \in \{a,b\}^*\}$ is a substitutable context free language. We can represent derivations with trees (i.e. terms) with lexical items at their leaves and generating functions (and possibly, redundantly, their values) at internal nodes. So with the previous grammar, $\langle aabb,S\rangle$ has the derivation represented by the term on the left or the corresponding tree on the right:

$$g(\langle b,B\rangle, f(\langle a,A\rangle,\langle c,S\rangle,\langle a,A\rangle),\langle b,B\rangle)$$



Multiple context free grammars (MCFGs) generalize CFGs by allowing categories to have more than one string component, and a straightforward extension of bare grammars can represent MCFGs (Keenan and Stabler 2003:§3.3), allowing $\mathscr{F} \subseteq [((\Sigma^*)^* \times Cat)^* \to$

$((\Sigma^*)^* \times Cat)]$. For example, consider the following grammar where T categorizes pairs of strings:

$$\begin{array}{l} V = \{a,b,c,d,e\} \\ Cat = \{A,C,E,T,S\} \\ Lex = \{\langle a,A\rangle, \langle b,d,T\rangle, \langle c,C\rangle, \langle e,E\rangle\} \\ \mathscr{F} = \{h,i\} \end{array}$$

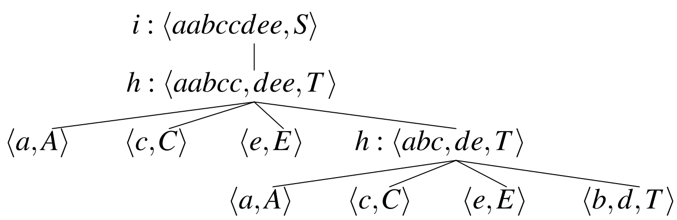
where h is defined as follows:

$$\begin{array}{ccc} \text{Domain} & h & \text{Value} \\ \begin{array}{cccc} s & t & u & v,w \\ A & C & E & T \end{array} & \longmapsto & \begin{array}{c} svt,wu \\ T \end{array} \end{array}$$

and i is

$$\begin{array}{ccc} \text{Domain} & i & \text{Value} \\ \begin{array}{c} s,t \\ T \end{array} & \longmapsto & \begin{array}{c} st \\ S \end{array} \end{array}$$

With this grammar, $Str(S) = \{a^n bc^n de^n |\ n \geq 0\}$, a non-context-free language. The string *aabccdee* has the derivation



Notice that, in the definition of the rules in this example, each string component in the arguments appears at most once (in this case, exactly once) in the value on the right, as MCFGs require.

We can easily relax the string copying requirement to allow multiple copies of argument strings, as in parallel multiple context free grammars (PMCFGs). For example, the following bare grammar represents a PMCFG with a single copying rule, extending our first example with just one rule j:

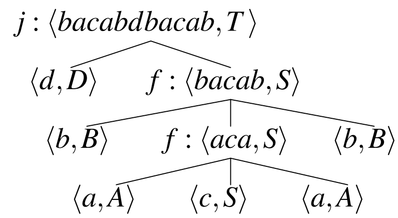
$$\begin{array}{l} V = \{a,b,c,d\} \\ Cat = \{S,T\} \\ Lex = \{\langle a,A\rangle, \langle b,B\rangle, \langle c,S\rangle, \langle d,D\rangle\} \\ \mathscr{F} = \{f,g,j\} \end{array}$$

where rules f,g are unchanged from the first example and j is the rule:

$$\begin{array}{ccc} \text{Domain} & j & \text{Value} \\ \begin{array}{cc} s & t,u \\ D & T \end{array} & \longmapsto & \begin{array}{c} tsu \\ T \end{array} \end{array}$$

With this grammar, $Str(T) = \{wcw^r dwcw^r |\ n \geq 0, w \in \{a,b\}^*\}$. The string *bacabdbacab* has the derivation:

$$j(g(\langle a,A\rangle, f(\langle a,A\rangle, \langle \varepsilon,S\rangle, \langle b,B\rangle), \langle b,B\rangle))$$



It is significant that, as we see in this derivation tree, the copying never needs to compare two independently generated strings, but simply copies one derived string to more than one position.

We have now shown how to represent as bare grammars all of the kinds of grammars mentioned in §1, grammars defining learnable classes of languages: certain regular grammars (i.e. certain CFGs), MCFGs, and PMCFGs. It is easy to see that these formal grammars have the fundamental properties Cong and Conc. With all of them expressed in the bare grammar framework, the following properties are immediate:

Cong: Categorization is a coarsest congruence and invariant. For any bare grammar, the set $\{Ph(C)\colon\ C \in Cat\}$ is a partition of $L(G)$. Let's call that partition the **categorization**. For all the artificial grammars mentioned, the categorization is invariant, that is, every $Ph(C)$ is invariant, because the categories define the domains of the rules, and the rules are fixed by every automorphism. For any sequence u of expressions of $L(G)$, let $|u|$ be the length of u. Then it is easy to see that, for each of our grammars above, the categorization is also a **congruence** with respect to $\mathscr{F}$ in the sense that

$$\forall F \in \mathscr{F},\ \forall u \in \text{Domain}(F),\ \text{if } |u| = |v| \text{ and } \forall 1 \leq i \leq |u|,\ u_i, v_i \text{ have the same category,}$$
$$\text{then } v \in \text{Domain}(F) \text{ and } F(u), F(v) \text{ have the same category.}$$

In general there can be many congruences. For example the trivial partition of singleton sets of elements of $L(G)$ is of course a congruence too. But for all the formal grammars of this section, the categorizations are coarsest congruences. That is, no union of any two distinct $Ph(C)$ and $Ph(D)$ yields a partition of $L(G)$ that is also a congruence.

Conc: Complexes are formed by concatenation. In each of the previous grammars G, each rule simply concatenates string components (possibly with repetitions, as in the last example).

Now we are ready to consider why some linguists will be uncomfortable with the fundamental assumptions Cong, Conc, and Subst, and whether their worries are well founded.

3 Linguistic perspectives

3.1 Against Cong: categories may have distinguished subsets

Various of the example grammars in Keenan and Stabler (2003) have categorizations that violate Cong. The example that gets the most attention there is Little Spanish = $\langle \Sigma, Cat, Lex, Rule\rangle$ where

Σ: man, woman, obstetrician, doctor, -a, -o, gentle, intelligent, every, some, very, moderately

Cat: Nm, Nf, A, Am, Af, Amod, NPm, NPf, Agrm, Agrf D, Dm, Df

Lex:

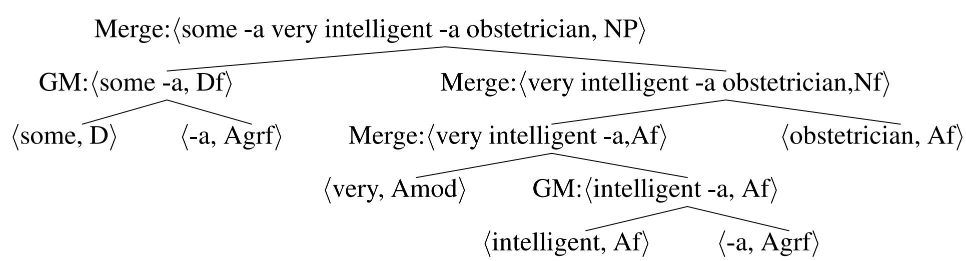
Nm	man, doctor
Nf	woman, obstetrician
A	gentle, intelligent
D	every, some
Agrm	-o
Agrf	-a
Amod	very, moderately

$\mathscr{F}$**:** GM, Merge, where

Domain	GM	Value	Conditions
$\overset{s}{\text{C}}\ \overset{t}{\text{Agrm}}$	$\longmapsto$	$\overset{st}{\text{Cm}}$	$\text{C} \in \{\text{D}, \text{A}\}$
$\overset{s}{\text{C}}\ \overset{t}{\text{Agrf}}$	$\longmapsto$	$\overset{st}{\text{Cf}}$	$\text{C} \in \{\text{D}, \text{A}\}$

Domain	Merge	Value	Conditions
$\overset{s}{\text{Ax}}\ \overset{t}{\text{Nx}}$	$\longmapsto$	$\overset{st}{\text{Nx}}$	$\text{x} \in \{\text{m}, \text{f}\}$
$\overset{s}{\text{Dx}}\ \overset{t}{\text{Nx}}$	$\longmapsto$	$\overset{st}{\text{NPx}}$	$\text{x} \in \{\text{m}, \text{f}\}$
$\overset{s}{\text{Amod}}\ \overset{t}{\text{Ax}}$	$\longmapsto$	$\overset{st}{\text{Ax}}$	$\text{x} \in \{\text{m}, \text{f}\}$

With this grammar, we have derivations like this:



This grammar has automorphisms that exchange masculine and feminine nouns (Nm and Nf), and the masculine and feminine agreement markings (Agrm and Agrf) in the lexicon, projecting these changes through the language to exchange all masculine and feminine expressions. So apparently agreement and other similar dependencies can produce symmetric subcategories, subcategories that we would like to relate, but doing so with an automorphism yields non-invariant categorizations. A similar problem for congruences arises if sometimes a pair of constituents must be identical or in some other very close correspondence at some point in a derivation (ellipsis, X-or-no-X, verbal clefts in Kru languages, X-bu-X questions in certain Chinese dialects, etc.) In such cases, we cannot generally replace one member of the pair by something else of the same category without replacing the other member of the pair, and so the categorization cannot be a congruence.

3.2 *Against Conc: Expressions not formed by concatenation*

For the description of human languages, assumption Conc is potentially problematic too. When two strings s,t are concatenated to produce st, obviously their edges do not overlap or change each other. But it might seem that human languages are not like that. There are various kinds of string mergers, reversals, and fusions across brackets. For example, traditional grammars tell us that when P combines with [D NP], we do not always get [P [D NP]] or [[D NP] P], the possibilities that concatenation allows. Sometimes, for example, we get [P+D NP], where P+D is a single, fused element.[5] In Chomskian syntax, we find many similar types of complexes. [T [V DP]] can becomes [V+T [_ DP]], and so on. Changes at the edges of constituents are also produced by morphophonological processes: fusion (Halle and Marantz 1993; Embick 2010), m-merger (Matushansky 2006), span instantiation (Svenonius 2012; Williams 2003), and other similar schemes have been proposed to form portmanteau morphemes for agreement and tense, for case and number, or other elements.

3.3 *Against Subst: no simple substitutability*

The learnability results mentioned above were for languages with the property that if two sequences (or tuples of sequences) of pronounced elements share a common context, they share all contexts. Obviously, English is not substitutable in this sense, as shown for example by the following observations of strings that share some contexts but not others, from Keenan and Stabler (2003:135ff):

	'middles'	contexts
(1)	Sue laughed	$\langle\varepsilon,\varepsilon\rangle$, $\langle$the boy who kissed,$\varepsilon\rangle$
(2)	it rains in Spain	$\langle\varepsilon,\varepsilon\rangle$, *$\langle$the boy who kissed,$\varepsilon\rangle$
(3)	and	$\langle$I saw John,Bill$\rangle$, $\langle$I saw both John,Bill$\rangle$
(4)	or	$\langle$I saw John,Bill$\rangle$, *$\langle$I saw both John,Bill$\rangle$

In the former pair of cases (1,2), we fail to respect the traditional constituency; many substrings are non-constituents. The latter pair of cases (3,4) raises a different issue, failing to respect the dependency between *both* and *and*.

4 A reconciliation

There are various ways to reconcile recent directions in learnability theory with mainstream linguistic proposals. I outline one perspective which I think looks most promising, and then conclude by proposing a new candidate axiom for the theory of human language, a fundamental grammatical principle that this perspective assumes. Anticipating: (i) we must of course generalize Subst, (ii) Conc is not as restrictive as it seems, and it could be rejected over morphemes but still preserved over phonological features, and (iii) I propose Cong as a law of language.

[5] Various kinds of P+D contraction and fusion are found in French (Zwicky 1987), Italian (Napoli and Nevis 1987), German (Waldmüller 2008), and other languages (Svenonius 2012; Williams 2003).

4.1 Categories are not fixed and universal across languages

The situation with categorizations would perhaps be easier to assess if there were a finite, universal, fixed set of syntactic categories, but this claim is difficult to understand when the categories are never defined in cross-linguistically applicable ways (Keenan and Stabler 1994; Stabler and Keenan 2007). For example, the claim that there are $\pm N$ and $\pm V$ categories is not a substantial claim that can be assessed without a theory of the properties of those categories. When 'subcategorizing' features, agreement, phrase level, and so on are factored in, of course we expect to have more than four distinctions among syntactic elements, and it is not unclear what those distinctions will need to be, or why we should assume, in advance, that there is a linguistically principled finite bound on the set of distinctions. One could adopt the methodology of beginning with some basic, 'core' distinctions (Chomsky 1981:§1), but then it is a mistake to confuse that very reasonable methodological strategy with strong empirical claims about what is really needed to get an adequate theory.

Some 'cartographic' approaches to syntax have aimed to identify particular structural positions which have similar roles across languages – a clausal template, or partial order of projections along the spine, or something similar (Cinque 1999; Rizzi 2004; Cinque and Rizzi 2008: among many others). Even on these accounts, languages vary in what can move into these positions (overtly or covertly), how these positions interact or not with agreement and other phenomena, and so the syntactic properties of the phrases in these positions vary significantly even though they may have some common semantic and syntactic features. What kind of arguments could support the claim that there are finitely many fixed positions with significant universal semantic and syntactic features? Various different kinds of arguments could in principle be relevant. It could be that by exploring one language after another, we will find a finite system of positions, perhaps with some finite range of syntactic variation in each position, that is descriptively and explanatorily adequate for all well-studied human languages, or successful enough to suggest that any remaining difficulties are bound to be managed without fundamental changes in our assumptions. But we will not be in that situation anytime soon. Some researchers seem to think that another kind of support for the hypothesis of a fixed, finite clause structure for all languages comes from learnability. But, on the one hand, a very large but finite number of options does not necessarily support a feasible learning theory, and on the other hand, many infinite classes of artificial languages are known to be rather easily learnable (with appropriate senses of 'learnable'). Consequently, it is hard to imagine a persuasive argument coming from considerations of this sort. Keenan and Stabler (2003) point out that studies of artificial languages show that languages can be similar in important respects, and learnable, without being limited to a finite number in principle, and without having the same grammatical structure, and without having any substantial parts of their grammars in common. It is difficult to find any reason to assume that human languages would have to be more similar than these sorts of examples, especially when, by 'language' we include the identification of properties of lexical items and other aspects of language that may be 'peripheral' to the interests of some linguists.

4.2 *Syntax is not sensitive to phonology, but invariants are*

One other idea in the informal linguistics literature is that syntactic rules cannot refer to phonological properties (Pullum and Zwicky 1988; Zwicky and Pullum 1986; Katz and Bever 1976). One sense in which this may be right is that our rules of syntax could be category functional (Keenan and Stabler 2003:p.153, Axiom 4). But it might also seem that the question of how many different expressions there are, how many different nouns, for example, is not a structural or syntactic matter. Similarly non-structural, it seems, is the question of whether any expressions of one category are pronounced the same way as expressions of another category. But the notion of structure in Keenan and Stabler (2003) leads to a different perspective on these matters. For example, we have, almost immediately from our definitions:

Theorem 1. *If no lexical item is also derived, the lexicon is a structural invariant (Keenan and Stabler 2003:23), and so the number of lexical items is too.*

Theorem 2. *For any invariant categories C,D, the number of elements with the same pronunciation, $|\{s\colon (s,C),(s,D) \in L(G)\}|$, is a structural invariant. So, in particular, if any structural category has just one expression, that expression is a structural invariant.*

The point here is unlike the previous debates; we are not concerned with whether these properties should be regarded as 'syntactic' in the sense of any earlier authors. The point here is that pronunciation matters for what we are calling structural invariants. Keenan and Stabler (2003:4) point out that, with invariants defined as the fixed points of the grammatical automorphisms, morphological forms themselves, that is, particular categorized, pronounced expressions, can be structural invariants.

Invariants depending on pronunciation and ambiguity (structural and lexical) are directly relevant to central linguistic interests, as we see for example from learnability results like those mentioned earlier; pronunciation obviously matters for the definition of learnable classes of languages. For example, it is easy to see that while the regular grammar,

$$S \to aB \quad B \to bB \quad B \to \varepsilon$$

is zero-reversible, the grammar that results from replacing *b* by *a* in the second rule,

$$S \to aB \quad B \to aB \quad B \to \varepsilon$$

is not. The derivations have the same shape, differing only in the lexical item in one rule, but that is significant. And similarly, as noted above, the context free grammar

$$S \to ab$$

defines a substitutable language, but

$$S \to ab \quad S \to bb$$

does not. (Trivial existence proofs like this are important because they signal the possibility of the infinitely many more difficult but fundamentally similar cases.) Differences like this, differences in how things are pronounced, can and do affect whether a language is learnable.

4.3 Cong: Categorization is a coarsest congruence and invariant

The arguments against Cong in §3.1 are not persuasive. What we see there is that it is possible to define Little Spanish in such a way that a match is required between different categories, matching the gender implicit in {Nf, Af, Df}, or in {Nm, Am, Dm}. But grammars that require repeated matching operations like this are missing generalizations, separating things that are really unified. Similarly for X-or-no-X constructions and the other examples mentioned. The grammars are improved when brought into accord with Cong, and so we will propose that Cong should always be respected.

4.3.1 Little Spanish′

The simple language for Spanish-like NPs in §3.1 above has these properties:

- It has 12 categories, with masculine and feminine varieties of the categories N, A, Agr, D, and NP.
- There is no category N, but instead a (lexically-sensitive) automorphism exchanging Nm and Nf.
- For $C \in$ {Agrm,Agrf,Nm,Nf,Dm,Df}, the phrases $Ph(C)$ are not invariant
- A rule of gender marking combines Nm with Agrm, Nf with Agrf.
- A rule called merge combines Af with Nf, Am with Nm, Df with Nf, Dm with Dm.

But it is possible to capture these agreement relations more simply.[6] Instead of assigning nouns to different categories depending on their gender, and then matching those categories with agreement categories of the same gender, we can formulate an essentially similar grammar with these nicer properties:

- It has not 12 but 5 simple categories: N, A, Amod, D, and NP.
- Masculine and feminine nouns both have category N. They have the same structure, in the sense that they are interchanged by our automorphisms (without the lexical sensitivity of the earlier grammar).
- For all $C \in Cat$, the set of phrases of category C is invariant.
- No separate gender marking rule is required. Instead, lower agreement affixes are attached to selecting heads.
- The rules of the grammar do not need special cases and conditions for masculine and feminine instances.

One way to achieve this is with the following alternative Little Spanish′=$\langle \Sigma, Cat, Lex, \mathscr{F} \rangle$:

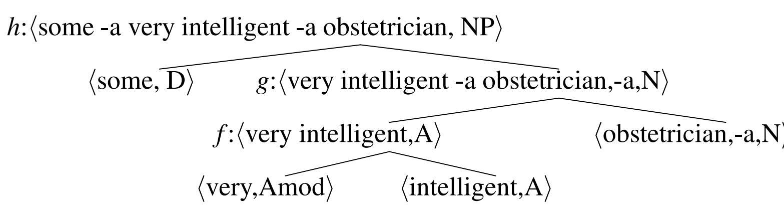
[6] The idea for this reformulation came from Hilda Koopman's suggestion to me (p.c.) that grammars like the one in §3.1 are more complex than necessary because they ignore the evidence that surface-discontinuous agreement is often (perhaps always) underlyingly local.

Σ:	man, woman, obstetrician, doctor, -a, -o, gentle, intelligent, every, some, very, moderately	
Cat:	N, A, Amod, NP, D	
Lex:	N	⟨man,-o⟩, ⟨doctor,-o⟩, ⟨woman,-a⟩, ⟨obstetrician,-a⟩
	A	gentle, intelligent
	D	every, some
	Amod	very, moderately
$\mathscr{F}$:	f,g,h	

where rules f,g,h are the following mappings, respectively:

$$\begin{array}{ccc} \begin{array}{cc} s & t \\ \text{Amod} & \text{A} \end{array} & \longmapsto & \begin{array}{c} st \\ \text{A} \end{array} \\[1em] \begin{array}{cc} s & t,u \\ \text{A} & \text{N} \end{array} & \longmapsto & \begin{array}{c} sut,u \\ \text{N} \end{array} \\[1em] \begin{array}{cc} s & t,u \\ \text{D} & \text{N} \end{array} & \longmapsto & \begin{array}{c} sut \\ \text{NP} \end{array} \end{array}$$

With this grammar, we have derivations like this:



This grammar captures the agreement regularities more simply and elegantly than the one in §3.1. Gender is marked just once, by the lexical entries. Parsimony weighs in favor of this alternative too, if fleshed out versions of these agreement marking rules have a form that is similar to other rules needed in Spanish and other languages. And here, the rules are category functional and string functional, the lexicon is invariant, the sets Ph(C) are invariant, and the sets Ph(C) are a congruence:

Theorem 3. *In Little Spanish′,* $\forall C \in Cat$, $Ph(C)$ *is invariant.*

Proof. This follows immediately from the fact that the $Ph(C)$ are domains and ranges of the generating functions, which must be fixed by the automorphisms. □

Theorem 4. *In Little Spanish′,* $\{Ph(C) : C \in Cat\}$ *is a coarsest congruence.*

Proof. Since each rule is total on the respective domain of pairs $Ph(C) \times Ph(D)$ (where C,D differ in each rule), and since each rule is category functional, we see immediately that the categorization is a congruence.

That the congruence is coarsest follows from the fact that taking the union of any pair of distinct $Ph(C), Ph(D)$ will produce a strict superset of the domain of at least one of the rules, so substitution within this larger set will not preserve grammaticality. □

For language learners, this last result immediately gives us the nice consequence:

Corollary 5. *For any $a,b,c \in L(Little\ Spanish')$, if a occurs in b and c has the same category as a, then substituting c for any (i.e. one or more) occurrences of a in b produces a well-formed expression that has the same category as b.*

4.3.2 Little X-or-no-X

The English X-or-no-X and other apparent copying constructions have received some attention in the literature (Manaster-Ramer 1986; Pullum and Rawlins 2007). We construct a little grammar that gets the X-or-no-X phrases in examples like,

Linguistics test or no linguistics test, I will enjoy myself
Long day trip or no long day trip, they should be home by now

Let Little X-or-no-X = $\langle \Sigma, Cat, Lex, \mathscr{F} \rangle$ where,

Σ:	long,short,dull,exciting,day,trip,linguistics,test	
Cat:	N, A, NP, X	
Lex:	N	day,trip,linguistics,test
	A	long,dull,short,exciting
$\mathscr{F}$:	f,g,h,i	

where f,g,h,i are defined as the following mappings, respectively:

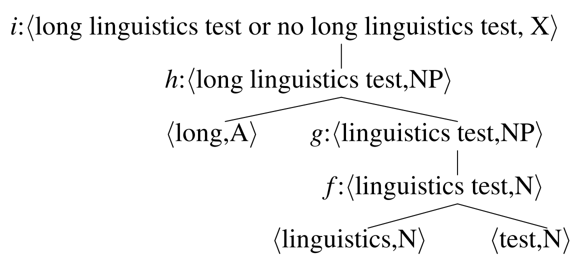
$$\begin{array}{c}s \quad t\\ \mathrm{N} \quad \mathrm{N}\end{array} \longmapsto \begin{array}{c}st\\ \mathrm{N}\end{array}$$

$$\begin{array}{c}s\\ \mathrm{N}\end{array} \longmapsto \begin{array}{c}s\\ \mathrm{NP}\end{array}$$

$$\begin{array}{c}s \quad t\\ \mathrm{A} \quad \mathrm{NP}\end{array} \longmapsto \begin{array}{c}st\\ \mathrm{NP}\end{array}$$

$$\begin{array}{c}s\\ \mathrm{NP}\end{array} \longmapsto \begin{array}{c}s \text{ or no } s\\ \mathrm{X}\end{array}$$

So then we have derivations like this:



The important thing here is that there are no two distinct constituents NP that are copies of each other, and so the simple substitution required by the congruence can work.[7] When a string is repeated, but not a copy, then of course any repeated instance can be replaced, independent of the others, preserving grammaticality. We can have copying even when the

[7] Pullum and Rawlins (2007) note that at least some English speakers allow expletives to be inserted, as in *linguistics test or no damn linguistics test*, or even *long linguistics test or no long bloody linguistics test*. This could plausibly be done by a late adjunction step, after the copy is produced. As in the simpler example displayed here, this would not require two constituents corresponding to the two pronounced copies.

categorization is invariant, and a coarsest congruence. As we saw in the previous example, here again these nice properties are trivially established. The form of the grammar guarantees a simple correspondence between categories and rules.

4.4 Conc: fusion is concatenation (if you choose the right basic elements)

It is not difficult to model head-movement-like relations in a Conc-respecting grammar, even ones that 'fuse' the head complexes (Michaelis 1998; Stabler 2001; Kobele 2002). In fact, since most or all morphology and phonology is finite state, mildly context sensitive formalisms are powerful enough to swallow them entirely, as has been noted before (Kobele 2011; Graf 2011). This would mean that the phonological features of instantiated lexical forms would become the alphabet, replacing the morphemes or words that are used in simple examples like the ones in this paper. The possibility of doing this, in principle, does not mean that it is the best thing to do, but it provides tools for studying how an integration could be achieved without losing the generalizations captured by either the syntax or the morphophonology.

4.5 Beyond substitutability

In human languages, it is a fact that the simple substitutability criteria mentioned earlier – sharing a single context – is not enough to show that two subsequences are (the pronounced parts of) constituents of the same category, or that they are constituents at all. Linguists use other sorts of evidence, and language learners certainly do too. Significant developments in learnability theory for human language are likely to come from better ways of assessing constituency and similarity, especially now that we have positive results reaching into parts of the Chomsky hierarchy where human languages seem to be. We see this trend already in the research briefly reviewed in §1. We see there that when constituents can have discontinuous parts, or copies, an inference from substitutability is still possible in certain settings. Ongoing research aims to find ways of relaxing the strong substitutability requirements on which those preliminary results are based.

5 Laws of language

Keenan and Stabler (2003:§4) propose a number of restrictions on categories and generating functions, but stop short of requiring:

(Cong) Categories are blocks of a coarsest congruence of L(G) and invariant.

On closer examination, though, it seems the reasons for rejecting Cong were unsound. Cong requires a simple, desirable correspondence between the categorization and the generating functions, and it licenses a simple and powerful substitution principle defined over the constituents of the language. The language learner does not hear or see the constituents directly, of course, but can hear or see their pronounced parts, and plausibly gets evidence of constituency from other sources too (fragments, prosody, etc.). With any hypothesized constituency, categorization must be identified using some kind of substitution-based reasoning, assessing the similarity of contexts of pronounced parts of constituents.

Acknowledgements

I have learned so much from Ed Keenan and Greg Kobele about these matters that they should get the credit for anything I have gotten right here. Thanks also to Alex Clark, Thomas Graf, Hilda Koopman, and Dick Oehrle for stimulating discussions.

References

Angluin, Dana. 1982. Inference of reversible languages. *Journal of the Association for Computing Machinery* 29:741–765.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* IT-2:113–124.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, Guglielmo, and Luigi Rizzi. 2008. The cartography of syntactic structures. Technical report, Studies in Linguistics, STiL 2, Università di Siena.

Clark, Alexander, and Rémi Eyraud. 2007. Polynomial identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research* 8:1725–1745.

Clark, Alexander, and Ryo Yoshinaka. 2012. Beyond semilinearity: Distributional learning of parallel multiple context-free grammars. In *Proceedings of the 11th International Conference on Grammatical Inference, ICGI 2012*.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, Massachusetts: MIT Press.

Graf, Thomas. 2011. Closure properties of minimalist derivation tree languages. In *Logical Aspects of Computational Linguistics, LACL'11*.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. MIT Press.

Hopcroft, John E., and Jeffrey D. Ullman. 1979. *Introduction to automata theory, languages and computation*. Reading, Massachusetts: Addison-Wesley.

Katz, Jerrold J., and Thomas G. Bever. 1976. The fall and rise of empiricism. In *An integrated theory of linguistic ability*, ed. Thomas G. Bever, Jerrold J. Katz, and D. Terence Landendoen. NY: Crowell.

Keenan, Edward L., and Edward P. Stabler. 1994. There is more than one language. In *Langues et Grammaire-1*, ed. Léa Nash and Georges Tsoulas, 217–235. Université Paris-8.

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

Kobele, Gregory M. 2002. Formalizing mirror theory. *Grammars* 5:177–221.

Kobele, Gregory M. 2011. Minimalist tree languages are closed under intersection with recognizable tree languages. In *Logical Aspects of Computational Linguistics, LACL'11*.

Manaster-Ramer, Alexis. 1986. Copying in natural languages, context freeness, and queue grammars. In *Proceedings of the 1986 Meeting of the Association for Computational Linguistics*.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In *Proceedings, Logical Aspects of Computational Linguistics, LACL'98*, 179–198. NY: Springer.

Moll, R.N., M.A. Arbib, and A.J. Kfoury. 1988. *An introduction to formal language theory*. NY: Springer-Verlag.

Morrill, Glyn, and Oriol Valentin. 2010. On the calculus of displacement. In *The 10th International Conference on Tree Adjoining Grammars and Related Formalisms, TAG+10*.

Napoli, Donna Jo, and Joel Nevis. 1987. Inflected prepositions in Italian. *Phonology* 4:195–209.

Pullum, Geoffrey K., and K. Rawlins. 2007. Argument or no argument? *Linguistics and Philosophy* 30:277–287.

Pullum, Geoffrey K., and Arnold M. Zwicky. 1988. The syntax-phonology interface. In *Linguistics: The Cambridge survey*, ed. Frederick J. Newmeyer, 255–280. NY: Cambridge University Press.

Rizzi, Luigi. 2004. Locality and left periphery. In *Structures and beyond: Cartography of syntactic structures, volume 3*, ed. Adriana Belletti, 104–131. NY: Oxford.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science* 88:191–229.

Stabler, Edward P. 2001. Recognizing head movement. In *Logical aspects of computational linguistics*, ed. Philippe de Groote, Glyn Morrill, and Christian Retoré, Lecture Notes in Artificial Intelligence, No. 2099, 254–260. NY: Springer.

Stabler, Edward P. 2012. Two models of minimalist, incremental syntactic analysis. *Forthcoming* .

Stabler, Edward P., and Edward L. Keenan. 2007. Universals across languages. Workshop on Model Theoretic Syntax, ESSLLI'07.

Svenonius, Peter. 2012. Spanning. Manuscript, http://ling.auf.net/lingbuzz.

Waldmüller, Estela. 2008. Contracted preposition-determiner forms in German: Semantics and pragmatics. Doctoral Dissertation, Universitat Pompeu Fabra, Barcelona.

Williams, Edwin. 2003. *Representation theory*. Cambridge, Massachusetts: MIT Press.

Yoshinaka, Ryo. 2011. Efficient learning of multiple context-free languages with multidimensional substitutability from positive data. *Theoretical Computer Science* 412:1821–1831.

Zwicky, Arnold. 1987. French prepositions: No peeking. *Phonology* 4:211–291.

Zwicky, Arnold, and Geoffrey Pullum. 1986. The principle of phonology-free syntax. *Ohio State University Working Papers* 32:63–91.

Affiliation

Edward P. Stabler
UCLA
stabler@ucla.edu