

*Systematicity and Natural Language Syntax*¹

GEOFFREY K. PULLUM AND BARBARA C. SCHOLZ

*Department of ???,
University of California, Santa Cruz*

A lengthy debate in the philosophy of the cognitive sciences has turned on whether the phenomenon known as ‘systematicity’ of language and thought shows that connectionist explanatory aspirations are misguided. We investigate the issue of just which phenomenon ‘systematicity’ is supposed to be. The much-rehearsed examples always suggest that being systematic has something to do with ways in which some parts of expressions in natural languages (and, more conjecturally, some parts of thoughts) can be substituted for others without altering well-formedness. We show that under one construal this yields a grossly weak claim that is not just compatible with a narrow version of associationist psychology but essentially coincides with a formalization of its descriptive power. Under another construal we get a claim (apparently unintended) that requires natural languages to fall within the context-free class, a claim that most linguists regard as too strong. Looking more closely at this proposed reconstruction of systematicity leads us to endorse, with further illustrations, the suggestion of Johnson (2004) that systematicity as a matter of substitutability of co-categorial constituents for one another does not appear to hold of natural languages at all. The appeal of the ill-delineated notion of systematicity may lie in the fact that within certain subclasses of lexical items mutual intersubstitutability does seem to hold, and the explanation for that lies in a limitation on human memory: we simply cannot learn separate privileges of syntactic distribution for all of the huge number of words and phrases that we know.

¹ Our joint work with James Rogers (see Pullum, Rogers and Scholz, forthcoming), partially supported by fellowships from the Radcliffe Institute for Advanced Study at Harvard University, 2005–2006, was important to the development of these ideas; but this paper was prepared without Rogers’ collaboration, so he bears no responsibility for any of its shortcomings or errors. We thank Dunja Jutronic for inviting us to present an earlier version of this paper at the Mental Phenomena conference on philosophy of linguistics in Dubrovnik in September 2006, and we are grateful to Alex Barber, Mark Cain, John Collins, Michael Devitt, Steven Gross, Guy Longworth, Peter Ludlow, Nenad Mišević, Paul Pietroski, and Barry Smith for their comments. None of these people should be assumed to agree with any of what we say in this version.

Key words:

A substantial literature has grown up within the philosophy of the linguistic and cognitive sciences discussing the concept of **systematicity**, mostly in relation to the arguments given by Fodor & Pylyshyn (1988; henceforth F&P) in favor of a “Classical” architecture for models of cognition and language and against “Connectionist” ones. We have no prior commitment to either architecture (if “architecture” is the right term for ideas as broadly and loosely framed as these). It may be that connectionism is, as its opponents would have us believe, an ill-advised associationist insurgency — a revival of bad psychology from the first half of the 20th century that is doomed to repeat history by failing in the long run. Perhaps, on the other hand, they are wrong, and connectionism points to a bright future in which psychology will be led out of its speculation-bound past into a future of rigorous computational modeling. We have no stake in this. What we are concerned with is the notion of systematicity. We want to know what it is.

It must come as a bit of a shock to a reader approaching the literature for the first time to discover that (as a number of authors have remarked; see *inter alia* Niklasson & van Gelder 1994, Cummins 1996, Hadley 1997, and Johnson 2004) the large body of work on systematicity generally operates without benefit of any clear characterization of the crucial notion. For example, Aizawa’s paper ‘Explaining systematicity’ (1997) is not an attempt to explain to the reader what systematicity is, but a discussion of what might be the explanation for the property or phenomenon thus named. But what *is* that property or phenomenon? Hardly anybody says. Instead they mostly rehearse very briefly a couple of utterly trivial examples of the supposed consequences of the systematicity of the language capacity (often the ones given in F&P), and move on quickly.

Here we attempt to identify what systematicity actually is, employing methods and results from formal language theory and model-theoretic syntax in trying to understand it better. We consider first a version that turns out to be way too weak (§1), and then a much more robust interpretation that turns out to be way too strong as a characterization applying to natural language (§2), and then, after a short interlude on thought (§3), we look at a plausible-looking rephrasing by Johnson (2004), and show that its predictions about syntax are even further from being true than Johnson assumed (§4). We look briefly at a relevant contrast between natural and formal languages with respect to systematicity (§5), and we conclude with some remarks on why the core notion animating discussions of systematicity has seemed so attractive to so many (§6).

Robbins (2005) notes that the concept is applied sometimes to human language processing capabilities, sometimes to the capacity to represent propositions, and sometimes to grammar. We agree with Johnson

(2004:113) that it is better to concentrate on issues of what is grammatical rather than allowing distraction to arise from such matters as our abilities to process utterances that are nowhere near being grammatical in our language (*Me Tarzan, you Jane* is processed well enough, but that raises extraneous concerns). So we shall be attempting to locate a clear meaning for the term ‘systematicity’ in the context of a linguistic system, and specifically a syntactic system.

1. *Systematicity as Substring Substitutability*

Systematicity has repeatedly been introduced and illustrated by quoting such remarks as these from the text of F&P (37):

What we mean when we say that linguistic capacities are *systematic* is that the ability to produce/understand some sentences is *intrinsically* connected to the ability to produce/understand certain others . . . You don’t, for example, find native speakers who know how to say . . . that John loves the girl but don’t know how to say . . . that the girl loves John.

Other writers appeal to these illustrations, either repeating them letter for letter, or making trivial changes. Thus Cummins et al. (2001, 168) understand F&P as saying that systematicity entails that “anyone who can understand ‘John loves Mary’ can understand ‘Mary loves John,’” and Aizawa (1997, 119) uses examples including ‘Mary hates John’ and ‘John hates Mary’.

These examples of the supposed consequences of systematicity suggest that we might model it mathematically as a property of *sets of strings of words*. (We will henceforth refer to sets of strings as *string-sets*.) Consider F&P’s “phrase book” illustration: in F&P (p. 37) a phrase book is taken to be a prime example of what the absence of systematicity is like, because

you can learn *any part of a phrase book without learning the rest*. Hence, on the phrase book model, it would be perfectly possible to learn that uttering the form of words ‘Granny’s cat is on Uncle Arthur’s mat’ is the way to say (in English) that Granny’s cat is on Uncle Arthur’s mat, and yet have no idea how to say that it’s raining (or, for that matter, how to say that Uncle Arthur’s cat is on Granny’s mat).

The exposition here is a bit puzzling (whatever the systematicity of our capacity to form utterances in English might be like, a mastery of the form of sentences about cats and mats and Granny doesn’t equip us with a mastery of the form of weather sentences), but the intent is clearly to point out that a phrase book is an *arbitrary* set of well-formed strings of words in a language paired with (synonymous) well-formed word-strings from another language. F&P stresses that “You don’t, for example, find native speakers who know how to say in English that John loves the girl but don’t know how to say in English that the girl loves John.”

What is crucial for them is that “the ability to produce/understand some sentences is *intrinsically* connected to the ability to produce/understand certain others.” For a set of word strings to be systematic, then, would apparently involve being closed under some ‘intrinsically connected to’ relation: all strings that are intrinsically connected (in the right ways) to the ones that are already members must also be members.

We are not suggesting that F&P overtly take systematicity to be a property of stringsets. They say at one point on p. 37 that it is “a property of the mastery of the syntax of a language.” But mastery of syntax manifests itself (under the usual idealizations) in terms of regarding certain strings, and not others, as well formed. The set of all possible products of that mastery will be a stringset of a certain special kind if the mastery is systematic: the stringset will exhibit the systematic intrinsic interconnectedness that is the claimed hallmark of systematicity.

So let us consider a class of sets of strings exhibiting a very general kind of intrinsic interconnectness: a class of sets in which the presence of certain strings guarantees the presence of (typically a large number of) other intrinsically related strings. The class we have in mind consists of all and only those stringsets that satisfy a certain general condition. It will help us in stating it succinctly if we introduce the term *prefix* for a substring that begins at the start of a string and *suffix* for a substring that extends to the end of a string. The condition can then be stated thus:

- [1] Given a string in the set containing some non-empty substring x , any prefix preceding x in some string in the set can be substituted for any other prefix that can precede x in some string in the set, and the result will also be in the set; and likewise any suffix following x in a string can be replaced by any other suffix that can follow x .²

This condition makes explicit in a particular way what might be meant by saying that the presence of some strings is “*intrinsically* connected” to the presence of others. It means the stringset has a certain structural property: it is closed under intersubstitution of prefixes and suffixes that are compatible with particular ‘middles’ of strings. Thus if a string is in the set, certain other strings *must* also be in the set. Consider the set whose members are listed in [2]:

- [2] {*John loves Mary, John loves the girl, Steve loves Mary, Steve loves the girl*}

² More precisely: for any non-empty strings u, v, w, x, y over the relevant vocabulary, if uxv is in the set and so is wxy , then uxy is in the set as well. This is stated in terms of switching endings (changing v to y on the grounds that y can follow x in wxy), but it is easy to show that beginnings can therefore be switched as well: the string uxv must be in the set (because we can switch the ending y to v in wxy); but that yields the same result as if we allowed switching u to w in uxv . So both beginnings and endings can be intersubstituted.

This set satisfies [1]. Fixing the word *loves* as the middle part x , any string in the set that can precede *loves* can be replaced by any other string that can precede *loves*, regardless of what follows it; and any string that can follow *loves* can be replaced by any other string that can follow *loves*, regardless of what precedes it. The presence of each string is intrinsically connected to the presence of the others, in the sense that if we removed any of them we would have a set that did not satisfy [1].

In [2] we merely enumerate the list of members of a particular set that satisfies [1]. But infinite sets can satisfy [1], so we need a way of describing such sets with a grammar. What kind of grammar will guarantee satisfaction of [1]? There is a class of grammars that exactly characterizes the class of stringsets satisfying [1].

A simple way to state such grammars (there are other ways) is in terms of a finite list of **bigrams**, i.e., ordered pairs of symbols. (In our examples, the symbols are English words.) Such a list is interpreted as describing the set of all and only those strings that are entirely composed of the bigrams found on the list. One can see it as a production procedure (to form a string, pick some bigrams and paste them together) or as a recognition criterion (given a string, check it from left to right to make sure each pair of adjacent symbols is on the list of bigrams).

In [3] we give a description that would suffice to characterize mastery of the set in [2]. The metalanguage in which the description is stated includes the symbols ‘▶’ and ‘◀’, which represent beginnings and ends of strings permitted in the set: the presence of ‘▶ z ’ in a description means that z is permitted to begin a string, and ‘ z ◀’ means that z is allowed to end a string.

[3] ▶ *John* *girl* ◀ *John loves* *loves Mary* *the girl*
 ▶ *Steve* *Mary* ◀ *Steve loves* *loves the*

A list of bigrams like the one in [3] determines three things: (i) which words may begin a string (in this example, *John* and *Steve*), (ii) which symbols may end a string (here, *girl* and *Mary*), and (iii) for each symbol, which symbols may immediately follow it (x can be followed by y in a string if and only if ‘ xy ’ is on the list). A string is in the set if and only if all of its adjacent symbol pairs are on the bigram list. Thus [3] is a kind of grammar for a stringset, and the form of the grammar determines that if *John loves the girl* and *Steve loves Mary* are recognizable (or producible) members of the defined set, *John loves Mary* must be too (see Figure 1).

Do not be distracted by the fact that the list of bigrams in this illustrative example happens to be longer than the list of members of the described set. We are concerned with the structure of sets here, not size of sets, or economy of description. By adding as few as three bigrams (e.g., ‘*John really*’, ‘*really loves*’, and ‘*really really*’), we get a description of an infinite set of grammatical English strings

Bigram grammars can thus clearly model what is called “productivity” in F&P, something that they claim a connectionist architecture can-

not accomplish. What they say on this topic (pp. 33–35) seems clearly incorrect; but productivity is a separate topic that we will not pursue here.

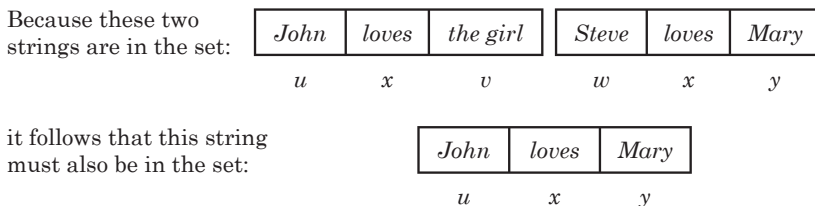


Figure 1: How the presence of two strings determines the presence of a third in a bigram-describable stringset. Any string that can follow *x* can be replaced by any other string that can follow *x*, and the new string thus formed is also in the set.

Bigram grammars certainly captures a kind of intrinsic interconnect- edness that guarantees the presence of certain strings in a set on the basis of structural analogies between it and some other string, which is a characteristic of systematicity that F&P emphasize. But an interesting fact about the bigram-describable stringsets makes it clear that this kind of definition of systematicity is far too weak to serve F&P's purpose of discrediting connectionism. Bigram descriptions are basically just a formalization of a simple version of associationist psychology under which a stimulus or other item can only be associated with the immediately preceding one. Bever, Fodor and Garrett (1967; henceforth BFG) state what they consider to be the general character of associative principles or rules this way:

Associative principles are rules defined over the “terminal” vocabulary of a theory, i.e., over the vocabulary in which behavior is described. Any description of an *n*-tuple of elements between which an association can hold must be a possible description of the actual behavior. [BFG, 583]

That is, a behavior like uttering a sequence of *n* words is to be modeled solely in terms of *n*-tuples of adjacent words; no abstract units like parts of speech or phrase types figure in the description. What we are pointing out that such abstract units are not a prerequisite to explaining systematicity (or even productivity, as we just remarked).

A bigram is a special case of an *n*-tuple where *n* = 2. BFG refers to the celebrated critique of associationism offered by Lashley (1951). Lashley pointed out that the products of typing behaviors cannot be described in terms of associations between adjacent keystroke pairs. Errors like typing ‘*Lalshey*’ for *Lashley* (where the error involves a sequence of three letters) cannot be described in terms of adjacent pairs— i.e., bigram grammars.³ But it is straightforward to generalize bigram

³ It is also the case that we cannot describe the grammatical strings of English in terms of adjacent word pairs. Although there are infinitely many bigram-describable stringsets over the set of English words, the set of all grammatical English strings

description to n -gram description (the relevant generalization of [1] says that the middle substring x has to be at least $n-1$ symbols long).⁴ One might imagine that n -gram descriptions for some $n > 2$ might suffice. BFG want to argue that in the case of natural language, no value of n could suffice.

So if the guaranteed systematic interconnectedness between word strings that we see in n -gram descriptions is the conception of systematicity F&P have in mind, then formalizations of associationist psychology exhibit systematicity. This conclusion is precisely the negation of what F&P argues for. The paper is an attack on connectionism, which is claimed to be nothing more or less than a computer-powered associationist insurgency. The whole point is supposed to be that associationist models of behavior, hence connectionist ones, cannot explain systematicity.

We have applied some elementary formal language theory to the task of making explicit a minimal sense of intrinsic interconnectedness illustrated in F&P's examples. An obvious move for Fodor & Pylyshyn to make would be to argue that systematicity cannot be a structural property of stringsets. Systematicity in F&P's sense may have to do with the character of the objects in the set as well as their interconnectedness. The obvious sort of objects to consider moving to would be those that can represent constituent structure, so we turn to that topic next.

2. *Systematicity as Closure under Subconstituent Substitution*

F&P bring the constituent structure and lexical and phrasal categories of expressions into the picture when they write (F&P, 38):

[4] If you assume that sentences are constructed out of words and phrases, and that many different sequences of words can be phrases of the same type, the very fact that one formula is a sentence of the language will often imply that other formulas must be too: in effect, systematicity follows from the postulation of constituent structure.

Suppose, for example, that it's a fact about English that formulas with the constituent analysis 'NP Vt NP' are well formed; and suppose that 'John' and 'the girl' are NPs and 'loves' is a Vt. It follows from these assumptions that 'John loves the girl,' 'John loves John,'

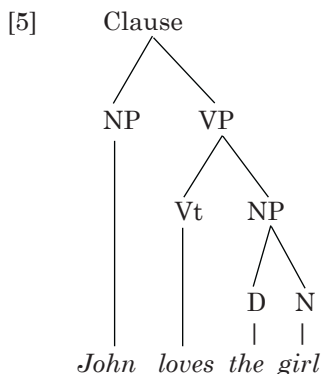
of words cannot be one of them. This is very easy to show. Let $u = \textit{You}$, $v = \textit{yourself}$, $x = \textit{hate}$, $w = \textit{They}$, and $y = \textit{themselves}$. Then since $uxv = \textit{You hate yourself}$ and $wxy = \textit{They hate themselves}$ are grammatical in English, $uxy = \textit{*You hate themselves}$ should also be grammatical if [1] is to be respected. But $\textit{*You hate themselves}$ is not grammatical. Therefore the stringset of English has no bigram description.

⁴ The proposal of Wickelgren (1969), adopted by Rumelhart & McClelland (1986) and discussed at length by Pinker and Prince in the same issue of *Cognition* where F&P appears (see Pinker & Prince 1988, 89ff), sets $k = 3$, and thus employs trigram descriptions. Rumelhart & McClelland call trigrams (over a vocabulary of phonological segments) 'Wickelphones'.

‘the girl loves the girl,’ and ‘the girl loves John’ must all be sentences. It follows too that anybody who has mastered the grammar of English must have linguistic capacities that are systematic in respect of these sentences; he *can’t but* assume that all of them are sentences if he assumes that any of them are.

Again it is suggested that systematicity is a property guaranteeing that if certain sentences belong to the language then necessarily certain others must too; but in this case the members are English expressions that have been analyzed into their phrasal and lexical constituent parts.

F&P (uncontroversially) assume in [4] that there are English sentences with the phrasal constituents ‘NP Vt NP’ (Noun Phrase + Verb (transitive) + Noun Phrase), in that order, and that both *John* and *the girl* are NPs. If ‘NP Vt NP’ is an expression in the set, then the words *the girl* can replace *John* in the first NP as in the second, and *John* is in the set of English expressions, i.e., it is as grammatically permissible in the second NP as in the first. A standard representation of the constituent structure of *John loves the girl* in the form of a tree diagram would look like this:



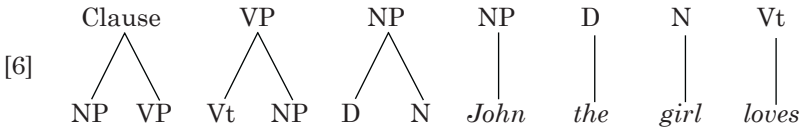
As characterized in [4], the systematicity in question appears to involve the interchangeability of the NPs represented in the tree diagram in [5].

Cummins (1996, 594) interprets F&P as claiming somewhat more: that a set of expressions that can be processed is systematic when it is closed under substitution of “systematic variants”, where “systematic variation is understood in terms of permuting constituents or (more strongly) substituting constituents of the same grammatical category.” Permuting the NPs in the constituent structure in [5] yields only *The girl loves John*. Substituting NPs from other sentences produces a much larger systematic set of expressions.⁵

⁵ Johnson (2004, 114) says “it is not clear that this [latter] version is any stronger than the other” because “there are only finitely many primitives in our language, and sentences can be arbitrarily finitely long” (p. 114); but it seems to us that he is wrong about this. At least, there are certainly sets of trees that are closed under permutation but not under substitution. One example is the set (considered by Rogers

We simply cannot tell from F&P what is intended with regard to the distinction just drawn; the paper is far too inexplicit. But it does seem reasonable to interpret F&P in [4] as equating systematicity with closure under substitution. This is because the paper contains the remark that a speaker with systematic mastery “*can’t but assume that all are sentences if he assumes that any of them are*” (p. 38). That implies that constituent substitution holds not just for certain cases, but globally. To say that a competent speaker cannot do anything else but take *every* substitution by a similarly-labeled constituent to be grammatically permissible is to specify that the whole set of structures that are well formed for the speaker is closed under like-labeled subtree substitution. This is a strong condition on the set of possible grammatical English expressions, or the possible grammatical expressions of any natural language. And we can see just how strong by appealing to certain relevant results in formal language theory that emerged from the fields of computer science and logic in the late 1960s.

A **tree** in the linguist’s sense is a directed, ordered, acyclic, singly-rooted, node-labeled graph meeting two further conditions. (We will be very informal here, not troubling to observe the technical distinction between tree and a picture of a tree, so we can say that the first condition is that any two arbitrary nodes in a tree are related by the downward lines indicating dominance iff they are not related by left-to-right order on the page, and the second is the downward lines in a picture like [5] never cross.) A **local tree** is a tree consisting of just a root node and at least one child (immediately dominated node). Thus the trees in [6] are local trees:



More complex trees can be constructed from local trees by fitting them together, superimposing root node labels with child node labels. It should be obvious that the tree shown in [5] is entirely put together from the local trees in [6]. A set of trees is a **strictly local tree-set** iff it is the set of *all and only those completed trees that can be constructed* out of some finite set of local trees, where ‘completed’ means that the bottom line is a string of words like ‘*the*’ and ‘*loves*’ rather than category labels like ‘NP’. If we form the set of all and only those completed trees that can be constructed out of the local trees in [6], we get four trees, representing the structures of the four sentences mentioned in the quote in [4]. The

1998, 60) of all binary finite trees where exactly one node in each tree is labeled *B* and all other nodes are labeled *A*. It is closed under permutation (any two subtrees with root label *A* may be interchanged and the result will still be in the set), but not under substitution by *A*-rooted subtrees out of other trees in the set (because that might result in bringing in an *A*-rooted subtree containing an extra *B*, creating a tree not in the set).

claim about constituent structure in [4] entails that the set of all possible expressions in any natural language, as modeled by sets of trees, is a strictly local tree-set. (Notice that the interesting tree-set referred to in footnote 5 is *not* a local tree-set.)

Define the **string yield** of a tree as the string formed by reading off the words at the bottom from left to right, and the string yield of a tree-set as the set of all the string yields of its trees. It was proved by Thatcher (1967) that if we the following holds:

[7] The string yield of a strictly local tree-set is a context-free string-set.

A stringset is context-free (henceforth CF) iff it can be generated by a context-free grammar (CFG), the kind of grammar that is called type 2 in Chomsky's early study of generative grammars (1959). The local tree-sets, in fact, are exactly the ones that CF phrase structure grammars produce as parse trees.

So there is a demonstrable consequence — doubtless, unintended and unappreciated — of the strong interpretation of F&P's passage [4]: that *the set of all possible grammatical word sequences of any given natural language is CF*.

This is still a highly controversial topic (see Pullum & Gazdar 1982 for a review of the earlier literature, and Pullum & Rawlins, in press, for a recent discussion); but it is widely believed by linguists that not every natural language is CF. In particular, Dutch is thought to have a non-local tree-set (Bresnan et al. 1982), and for a closely related language, the Zurich-area dialect of Swiss German, Shieber (1985) has given a compelling argument to the effect that it does not even have a CF stringset, which entails that *no* tree-set describing it can possibly be a local set. Similar claims have also been made for the stringsets of various other languages. The received view among linguists today is that it is just false that the set of all grammatical structures of the expressions of a natural language can always be modeled as a strictly local tree-set.

Thus F&P's second sense of systematicity is much too strong, if taken as a claim about the grammatical structures of any arbitrary natural language. And we do not see any way of weakening it that does not evacuate it of content.

For example, take any three out of the four sentences cited by F&P in [4]; consider, for example, the tree structures corresponding to (i) *John loves John*, (ii) *John loves the girl*, and (iii) *The girl loves the girl*. Each of these trees has two NP nodes. Some of these are substitutable by some of the others: you can substitute *the girl* for *John* if it precedes *loves* but not if it follows, and you can substitute *John* for *the girl* anywhere. If a speaker could understand just these three sentences with these structures, but not (iv) *The girl loves John*. Is that enough to make the speaker's capacity systematic or not? F&P is completely explicit on this: a speaker "*can't but* assume that all of them are sentences

if he assumes that any of them are”, so assuming that only three of the four were sentences would be the hallmark of having a *non*-systematic linguistic capacity, since (i)–(iii) without (iv) constitute an arbitrary set. But let’s suppose someone thought this was just a case of F&P being hyperbolic in [4], and really it was enough for the capacity to support partial substitutability, here and there. The problem is that under such a weakening, a capacity to understand almost *any* random set of sentences will count as systematic: the condition is so weak that it will always be satisfied unless all nodes in all trees in the set are labeled differently from all others.

Weakening substitution closure to subsets fails to distinguish the situation the property that is supposed to be explained by the classical architecture from the non-systematic property illustrated by a phrase book. We see all sorts of partial or local substitutability evidenced in the Croatian phrase book that we bought before traveling to the conference in Dubrovnik where we presented this paper; but it is a phrase book, exactly the archetype of what F&P insists is *not* a case of systematicity. Learning the sentences in it together with the partial similarities that hold between some of them, would surely not endow us with anything like a systematic grasp of Croatian. The weakening just discussed would imply that it did.

3. *A Short Interlude for Thought*

An interesting question naturally arises at this point concerning systematicity as it applies to the syntax of thought. We will make one brief point on this topic.

It is assumed in F&P, quite uncontroversially, that an analysis of the systematicity of natural language syntax will model the structure of expressions with (at the very least) something like ordered trees. It seems reasonable to us to assume that thoughts differ from natural language expressions in one respect: they do not have a precedence dimension. That is, we take it that when speakers of English, Turkish, Irish, Malagasy, Hixkaryana, and Xavante entertain the thought that a storm has damaged the house, they are all thinking exactly the same thought, even though the typical expression of that thought in the six languages would have six different precedence orders of subject, verb, and object (SVO, SOV, VSO, VOS, OVS, and OSV, respectively).

This suggests that an appropriate theoretical device for modeling the syntactic form of thoughts might be **unordered** trees: directed, acyclic, singly-rooted, node-labeled graphs that have a dominance order (down from the root node, which dominates all nodes in the tree) but with no a left-to-right precedence sequence defined on the node set.

Now, a strictly local unordered tree-set would be a set of trees that was the closure under substitution of a set of unordered local trees (where an unordered local tree is simply a root node label paired with a multiset, rather than a sequence, of child node labels). It is only the

presence of a linear precedence order (without tangling of branches) that prevents the tree-sets for languages like Dutch and Swiss German from being strictly local, and prevents the stringset of Swiss German from being CF. If no precedence relation is defined on the structure of thoughts, then it does not follow that we have a CF stringset (the notion of a stringset is not even defined for thoughts). And that would mean that it *could* be the case that the set of all thinkable thoughts *does* form an unordered tree-set that has the systematicity property. Thoughts would exhibit part-whole structural relations, but there would be no preceding or following of parts of thoughts by other parts.

Or on the other hand it might not. Matthews (1994) asks whether just because he can think that the object *a* is the sole member of the singleton set $\{a\}$ he can therefore think that the singleton set $\{a\}$ is the sole member of the object *a*. Presumably not. Likewise, we can ask whether being able to think that your mother gave birth to you implies you can think that you gave birth to your mother. We can ask whether being able to think that you would like to eat an ice cream this afternoon implies also being able to think that an ice cream would like to eat you this afternoon. We can ask whether being able to think about \$0 being shared between 37 people implies also being able to think about \$37 being shared among 0 people.

But these are observations about what semantic content is thinkable. It is not easy to connect them to a claim about the syntax of thoughts. In fact it is not clear to us what an objection to a claim like ‘Thoughts have the syntactic form of unordered trees’ could possibly look like, let alone the claim that the entire set of all thinkable thoughts constitutes a local set of unordered trees.

We note also that under the hypothesis explored by Carruthers (2002), it could be the case that parts of our thinking are systematic and other parts not. Carruthers claims “that natural language is the medium of non-domain-specific thought and inference” (p. 665), in the sense of being causally implicated in cognition, and constitutive of it, though he recognizes that “much propositional thinking also takes place independently of natural language” (p. 664). Thus if natural languages did *not* have the property of systematicity (in some agreed sense), the parts of our thinking that are constituted by tokening of natural language sentences would not be systematic, even if perhaps there is also a non-linguistic kind of propositional thought that does have that property.

We will leave this topic in its present highly speculative state, and return to our own central concern, which has to do with the syntax of natural languages.

4. Syntactic Systematicity and Categorizing Constituents

So far we have seen a version of syntactic systematicity that is far too weak because it is compatible associationist psychology, and a version

that is far too strong because it imposes an untenable upper bound on the structural complexity of natural languages. But we have not found anything systematicity could be that will do the work F&P want it to do. We consider now a slightly relaxed conception of systematicity that Johnson (2004) arrives at by adapting remarks of Cummins (1996). We quote Johnson's definition (2004, 114) in [8]:

[8] A language L is systematic if and only if (S) holds for all A :

(S) A is a constituent of L only if for all B of the same linguistic kind as A , and for all things C , C can compose with A (in a certain way) to form a sentence if and only if C can compose with B (in that same way) to form a sentence.

What makes a language systematic, in other words, is that the only constituents permitted in it are those whose category-mates (constituents "of the same linguistic kind") all compose in the same way with exactly the same other linguistic material. The lack of specificity about the meaning of "compose with" leaves open the possibility that composition of expressions might not involve mere concatenation of constituents. That could be very useful in providing an answer to the problem of Dutch, Swiss German, etc., which cannot be syntactically described in terms of combination of linearly continuous local trees by substitution in ordinary trees (where branches are not allowed to tangle), but can be described by free combination of local trees if one constituent is allowed to combine with another by 'wrapping' (or crossing of branches), so that a VP can combine with an NP in a way that puts the head V before the NP and the rest after it (see Pullum 1984 and Ojeda 1988 for discussion of head-wrapping in such cases). The idea brings up various technical issues about syntactic structure that it would not be appropriate to explore here, but we will tentatively assume from now on that Johnson's formulation might provide the necessary weakening of the subtree-substitution version of systematicity to avoid the difficulty noted in section 2.

It would be easy to make the mistake of thinking that constituents could be defined as belonging to the same category (and thus defined as category-mates) by [9]:

[9] A constituent A belongs to the same category as a constituent B iff B can always be substituted for A in a grammatical expression without destroying grammaticality.⁶

But [9] is not a coherent proposal for assigning constituents to categories. This becomes apparent the moment we realize that 'belongs to the same category as' has to be an equivalence relation (reflexive, symmetric, and transitive), but 'can be substituted for' on the set of constituents need not be.

⁶ We raise this point mainly because a parallel mistake appears to have been made in the history of American structuralist linguistics; see Pullum (1972) on a case involving phonology.

To see this in terms of a concrete example, think of using some binary relation between individuals as the basis for a rule determining which pairs of people sit at the same table at a large alumni banquet. Some relations can be coherently and successfully employed in this way. If we use the ‘classmate’ relation, we get one table for each class year; the tables correspond to the equivalence classes in the partition that the classmate relation determines on the set of alumni. But a proposal to use the ‘admires’ relation is incoherent, if defined on a typical set of a few hundred normal human beings. Why? Because admiration is by no means always mutual. Suppose *a* admires *b* but *b* despises *a*. What does the rule ‘*x* and *y* should be at the same table iff *x* admires *y*’ say about who should be at which table? No seating arrangement is admissible: on the variable assignment $x = a$ and $y = b$, the rule says that *x* and *y* must be at the same table, but on the one where $x = b$ and $y = a$, the rule requires that they mustn’t. So there is no course of action that complies with the rule. (The same is true if the ‘admires’ relation on the set is not transitive: if *a* admires *b*, and *b* admires *c*, but *a* despises *c*, no table assignment complies with the rule.)

What holds for the ‘admires’ relation on a typical set of people also holds for the relation ‘can be substituted for’ on the set of English constituents: it is not an equivalence relation.

We can show this by showing that it is not symmetric. Consider the adjectives *fond* and *proud*. It appears always to be possible to change any instance of *fond* in a grammatical sentence to *proud*, preserving grammaticality. For example, taking the first sentence in the *Wall Street Journal* corpus⁷ containing *fond*, namely [10a], and substituting *proud*, we get [10b]:

[10]a. *Brezhnev, for instance, was fond of saying that “the Soviet Union is on the seacoast of the Universe.”*

b. *Brezhnev, for instance, was proud of saying that “the Soviet Union is on the seacoast of the Universe.”*

This is grammatical. And such substitutions seem always to preserve grammaticality:

- | | |
|---|--|
| [11]a. i. <i>He is fond of his children.</i> | ii. <i>He is proud of his children.</i> |
| b. i. <i>She grew ever fonder of him.</i> | ii. <i>She grew ever prouder of him.</i> |
| c. i. <i>That’s one of my fondest memories.</i> | ii. <i>That’s one of my proudest memories.</i> |

However, grammaticality is not always preserved if we replace tokens of *proud* by tokens of *fond*. Again, taking the first occurrence of *proud* in the *Wall Street Journal* corpus, namely [12a], and substituting *fond*, we get [12b]:

⁷ The *Wall Street Journal* corpus is a convenient collection of some 44 million words of newspaper articles from 1987, 1988, and 1989 that was made available on an inexpensive CD ROM by the Association for Computational Linguistics in 1993. It is often used for testing purposes in work on natural language engineering. It is used here purely to make our point very concretely.

- [12]a. *Analogic was proud when all 100 electronic instruments worked in its first shipment to Toshiba seven years ago.*
 b. **Analogic was fond when all 100 electronic instruments worked in its first shipment to Toshiba seven years ago.*

This is not grammatical. Grammaticality will be quite generally destroyed with examples having *proud* in predicative function with no complement preposition phrase:

- [13]a. i. *He is very proud.* ii. **He is very fond.*
 b. i. *She is a proud woman.* ii. **She is a fond woman.*

Substitution fails to preserve grammaticality here because the set of acceptable syntactic contexts for *proud* properly includes the set of contexts for *fond*. That leads to failures of symmetry. Hence grammaticality-preserving substitutability is not an equivalence relation.

Other examples of symmetry failure abound. The adjectives *probable* and *likely* provide another example. *Probable* can always be replaced by *likely*:

- [14]a. i. *That it will rain is probable.* ii. *That it will rain is likely.*
 b. i. *It is probable that it will rain.* ii. *It is likely that it will rain.*

But it is not true that *probable* can always be substituted for *likely*:

- [15] i. *It is likely to rain.* ii. **It is probable to rain.*

The upshot is that we cannot base any syntactic category system for English on the relation ‘substitutable for without loss of grammaticality’ between constituents: that relation is not an equivalence relation on the set of constituents in English, so it does not provide a coherent criterion for categorizing constituents.

Notice, though, that the ‘substitutable for’ relation could turn out to be symmetric in some language. That might even be the case in some natural language (we do not know), and it generally will be if we pick an invented logical language, because they tend to be designed in a way that guarantees we get lucky. In a typical logical language, whenever some monadic predicate *F* can be substituted for a monadic predicate *G* in any context, that always means that substitution of *F* for *G* and substitution of *G* for *F* will always preserve grammaticality; and the same holds for dyadic predicates, and names, and quantifiers, and so on. It’s just that when we come to consider languages like English, that is not how things are.

Now, we can get a coherent categorization principle for English if instead of using [9], we use instead a related principle based on the symmetric relation ‘is substitutable for and substitutable by’:

- [16]A constituent *A* belongs to the same category as a constituent *B* iff, in all grammatical expressions, *B* can be substituted for *A*, or vice versa, without destroying grammaticality.

According to [16], *x* and *y* are in *different* categories iff there is any context in which grammaticality-preserving intersubstitution *fails*. Under that principle, we get categories whose members are intersubstitutable

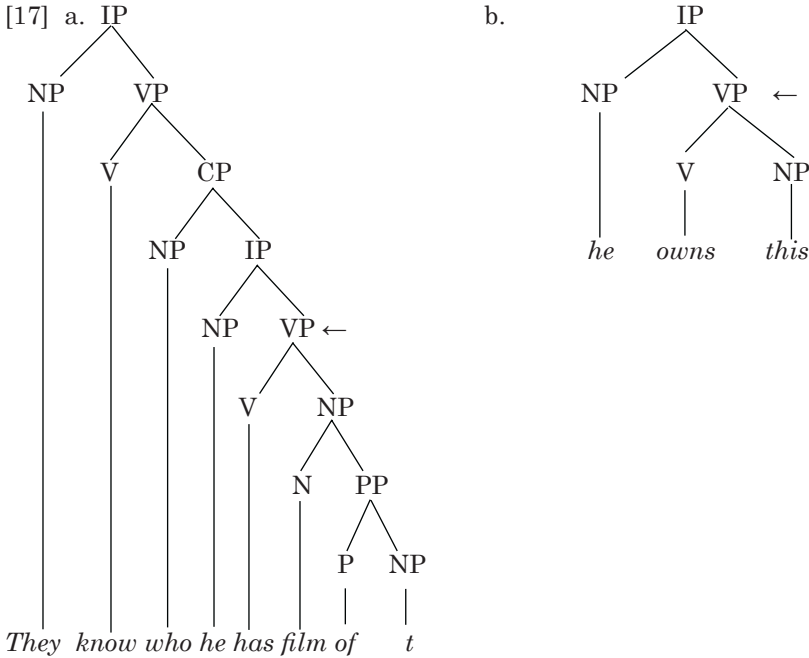
in all contexts.⁸ Any case of an *A* that cannot be replaced by some category-mate *B* in some context would be a counterexample to [16], and thus a counterexample to Johnson's (S) in [8], which demands that (if the language is to be systematic) every constituent must have the same combinatory syntactic behaviors as all of its category-mates.

Johnson argues against the idea that systematicity, defined as in [8] in terms of intersubstitutability of category-mates, is a property of natural language syntax (or at least, of English syntax). In effect he argues that [16] does not yield a system of syntactic categories that is anywhere near plausible: categories needed for syntactic description of English are not a set of equivalence classes determined by the intersubstitutability relation. He further argues that stipulating a category system that satisfies intersubstitutability results in a trivialized conception of systematicity. He presents a litany of examples to illustrate a dilemma: on the one hand, if anything like a standard system of syntactic categories for English is assumed, English is clearly not systematic in the sense of [8]; and on the other hand, if (S) in [8] is stipulated to hold, then English will have an arbitrary and extremely fine-grained set of categories that no syntactician could be a realist about. In what follows, we investigate this tension a bit further.

Note first that [16] is supposed to apply to phrasal categories as well as lexical ones. Curiously, F&P do not appear to have noticed that it cannot possibly be thought to hold of the set of derived structures of English sentences under the standard conception of how constituents are categorized in transformational-generative grammars.

Take the structure of a sentence like *They know who he has film of* as described under contemporary treatments of syntax. It could be represented roughly as in [17a]. Comparing the arrow-marked VP node with the one marked in [17b] should make the problem clear.

⁸ This yields a coherent and solvable constraint satisfaction problem—a kind of graph-coloring problem, in fact.



Systematicity as intersubstitutability fails in [17] given a theory of categories of the sort assumed in any current theory of syntax: the substitution of the arrow-marked VP of [17b] in [17a] produces the ungrammatical [18a], and substituting the other way would yield [18b], also ungrammatical.

- [18]a. **They know who he owns this.*
- b. **He has film of.*

Under standard transformational-generative grammar of almost any variety or vintage, this particular kind of substitution will fail to preserve grammaticality. The reason is that under a transformationalist analysis the VP under consideration in [17a] has had an NP ‘moved’ out of it by *wh*-movement (the ‘*t*’ marks the spot from which it was moved). A VP that has had a constituent moved out of it cannot be substituted for one that hasn’t.

To preserve systematicity, we would need to modify the syntactic category system to distinguish the two arrow-marked labels in [17], for example, by distinguishing ‘VP containing an NP trace’ (in [17a]) from ‘VP’ (in [17b]).⁹

⁹ Gazdar (1981), of course, actually proposed this policy, using ‘VP/NP’ as the label of a VP containing an NP trace, and he exhibited some interesting consequences for the description of unbounded dependencies and coordinate structures. The framework known as GPSG (generalized phrase structure grammar) emerged out of that work (see Gazdar et al., 1985). GPSG is a framework for syntactic theory that (in effect) insists that the set of trees described be closed under co-categorical subtree substitution. It follows, as noted above, that GPSG cannot describe Swiss German.

It may be that Fodor and Pylyshyn's failure to notice problems of the sort illustrated by [17] is due to the extraordinarily simple nature of the examples they used. But in fact even the systematicity of F&P's 'NP Vt NP' clause pattern is highly fragile. It does not even survive substituting pronouns for the NPs. The examples in [19] all have 'NP Vt NP' constituent phrase structure; but English has gender agreement in its reflexive pronouns, so we get (what Fodor and Pylyshyn would have to regard as) a failure of systematicity with respect to NPs.

- | | |
|----------------------------------|--|
| [19]a. <i>John respects him.</i> | e. <i>John can't help himself.</i> |
| b. * <i>Him respects John.</i> | f. * <i>John can't help herself.</i> |
| c. * <i>John respects he.</i> | g. * <i>The girl can't help himself.</i> |
| d. <i>He respects John.</i> | h. <i>The girl can't help herself.</i> |

To preserve systematicity in the face of these and similar examples, the 'NP' constituent type has to be refined into a multitude of categories, enough to allow for (in English) at least number, person, gender, case, pronominal status, and reflexive form (which already yields over two hundred NP categories). And then it also has to be broken up according to presence of internal traces: in *Which car do you have the key to?* the underlined sequence is a 3rd person singular non-human non-genitive NP containing an NP trace; in *To which car do you have the key?* the underlined sequence contains a PP trace; and so on). Most of the distinctions drawn to make all these subcategories turn out to cross-classify, so that categories have to be intersected repeatedly, making them more and more specific. Take English adjectives, for example:

- some take complements (as in *happy with that*) and some don't;
- of those that do, some take PPs (*fond of it*), some take non-finite clauses (*bound to be of use*), some take finite clauses (*aware it happened*), some take more than one of these (*glad of it, glad to be of use, glad it happened*), and so on;
- some have obligatory complements but most have optional complements;
- some are optionally usable in attributive modifier function (before a noun), some can only be used attributively, and some can never be used attributively;
- some are optionally usable in postpositive complement function (after the head noun in an NP, as in *anyone intelligent*), some can only be used postpositively (*trouble aplenty*), and some can never be used postpositively;
- some are optionally usable in predicative complement function (in a VP, as in *feel sad*), some can only be used predicatively, and some can never be used predicatively;

and so on (see Huddleston & Pullum 2002, chapter 6, for details). All these properties are relevant to whether one adjective can be substi-

But our point here is not about GPSG; we are pointing out that a transformational grammarian would have to do exactly what GPSG does, and break up the VP category into VP proper, VP/NP, VP/PP, etc.

tuted for another. More and more categories must be set up for what are lumped together as adjectives in the dictionary as we try to ensure the identical syntactic behavior among category-mates that is required by complete intersubstitutability. And for an entire set of expressions to be systematic, such intersubstitutability must hold in *every* category of constituent therein. As Johnson says, whether a language is systematic depends on the theory of categories assumed.

But even Johnson does not give a full sense of how radically implausible the categories would have to be in order for English to be systematic. Take the case of the nouns *ape* and *baboon*. At first it seems that if these are not in the same equivalence class for substitution, no two words could be: both are regular, non-genitive, count nouns denoting natural kinds (African primates, in fact). No standard syntactic properties seem to distinguish them in any way. Yet intersubstitution is not possible. At first it seems to be:

- [20]a. *The ape/baboon ate the banana.*
 b. *This is a large ape/baboon.*

But here it fails:

- [21]a. *This is an ape.*
 b. **This is an baboon.*

So even *ape* and *baboon* are not intersubstitutable. The key factor, of course, is that the English indefinite article has a syntactic quirk: it has two different phonological and orthographical manifestations, and the condition for their selection depends on the phonological segment that begins the following word (the spelling is ‘*an*’ before letters pronounced as vowels phonetically, but ‘*a*’ before letters pronounced as consonants). If the systematicity of English is to be preserved, *ape* and *baboon* must be members of different categories, in order to cover the fact that substitution can fail when they occur in indefinite NPs.

Again, it would be a mistake to think that this is just a fact about words, to be dealt with by fixing up the dictionary to give different part of speech designations for *ape* and *baboon*. That will not work. We are concerned not with a property of these nouns themselves, but also a property of the phrases built from them. If we add an attributive adjective to build up what Huddleston & Pullum (2002) call a Nominal (bracketed in [22]), we find that *intelligent ape* (in [22ai]) and *clever ape* (in [22aii]) must belong to different subcategories of Nominal:

- [22]a.i. *It was an [intelligent ape].* ii. **It was an [clever ape].*
 b.i. *It was an [intelligent baboon].* ii. **It was an [clever baboon].*

There are indefinitely many Nominals in English (they can be recursively formed, to yield phrases such as *clever, affectionate, playful, energetic, linguistically competent ape*). Thus they cannot just be listed. The entire Nominal category has to be bifurcated. And in fact other categories do as well: what has to be syntactically guaranteed here is that in a member of the category ‘vowel-initial Nominal’ the leftmost word of

the leftmost subconstituent is a vowel-initial word. That would mean the category of adverb phrases has to be bifurcated as well (compare [*an* [[*extremely clever*] *baboon*]] with *[*an* [[[*really*] *extremely clever*] *baboon*]], and so on). The massive reconstruction of the category system that is called for should not be underestimated.

The temptation is to wonder whether we could somehow treat *a* and *an* as a single item: a little class of word forms that for syntactic purposes we would treat as just one word. But remember, the proposal under consideration is that the relation of mutual substitutability in all contexts, as given in [16], should be the criterion for *deciding* category membership. If we can stipulate on non-distributional grounds that the difference between where *a* is found and where *an* is found will be ignored, we have completely given up [16].

Johnson says that by proceeding to more and more fine-grained constituents we make systematicity “a triviality” or a tautology” (p. 126). In one way this is true: we have let systematicity determine the categorization of constituents, and thus it is no surprise that, given the resultant categorization, systematicity holds. The notion is trivial in the technical sense that the question ‘Is set *X* systematic?’ has the same answer for all *X*: the answer is always ‘Yes’, because we make it so by using [16] as our analytical criterion.

But in another it is false that systematicity is a trivial property, either for natural languages or invented languages. As regards natural languages, there will be some set of categories into which English constituents can be classified that is minimal among category sets having the property of full intracategorical mutual substitutability, and it will be a non-trivial fact about English that some particular set suffices. How large such a set of categories might be, linguists have no idea as yet: it is of currently unknown extent, probably thousands or tens of thousands, possibly much larger. Seeking a set of categories with the relevant property would be a substantial research enterprise, and finding a minimal one would be an empirical discovery about English (albeit a discovery that would be irrelevant to syntactic description as ordinarily done, whether formal or informal).

And with regard to invented formal languages, we pointed out above that they are specifically designed to have a classification of constituents into equivalence classes characterized by intracategorical substitutability: in the propositional calculus all sentential variables can substitute for each other, and all binary logical connectives can intersubstitute; in the predicate calculus all relation symbols of a given arity can freely replace each other, and all the quantifiers can replace each other, and so on. But it is not a triviality that precisely when we design formal languages for use in elucidating logic or computation, we abstract away from the complexities of natural language: we design these languages to be syntactically systematic relative to a conveniently small set of categories—categories that match up perfectly with what we want in the semantics. It is an important and useful feature of the design of these

languages that they are systematic in this respect, and by no means is it a tautology that they are.

The implications for the agenda of F&P are radical. If the extent to which the members of particular lexical or phrasal categories can be intersubstituted is heavily circumscribed, even fragmentary, can we conclude from the phenomena of language that the connectionist architecture is dead in the water, so that the classical architecture wins out and should be embraced by every investigator who is accessible to reason? As far as we can see, the answer has to be no. What the classical architecture predicts is not at all similar to an irregular landscape of fragmentary and highly circumscribed partial intersubstitutability of subexpressions. The classical architecture — closely allied, as F&P stresses, to the realm of Turing machines and deductive logic — predicts the full and perfectly regular intersubstitutability that we do not find.

5. *Structure, Universals, and Logic*

There are two conceptions of syntactic categories in 20th-century American linguistics that contrast sharply. One stems from a research program in American anthropological linguistics (much influenced by a desire to describe Amerindian languages without seeing them through Latin-tinted spectacles) that culminates in the work of Harris (1951). Harris worked out in minute detail a set of procedures for syntactic analysis that, if applied to some language with full rigor, would lead to the discovery of a set of categories for the language that guarantees intracategorical intersubstitutability. The categories for any given language will not necessarily have anything to do with the categories for any other; not even in the case of something as simple and apparently transferable as the category normally called ‘feminine noun’: the category called that in French is not the same as the one in German that goes by the same name, because the French category contrasts only with ‘masculine noun’, while the German one contrasts with two others, ‘masculine noun’ and ‘neuter noun’. Such structural contrasts are crucial: in this tradition the whole classification depends on them.

Categories, for linguists such as Harris, are identified by references to distributions—syntactic contexts in which constituents appear. A French noun is feminine because it appears with *la* rather than *le*, with *cette* rather than *ce*, and so on. German does not have these words; instead, we find similar semantic functions being served by *der*, *die*, and *das*. All the contexts are different, and so is the number of gender distinctions drawn. And the syntactic categories are defined entirely by reference to equivalence of syntactic contexts, i.e., to mutual substitutability. We therefore get a guarantee that the kind of intersubstitutability referenced in [8] will hold.

The other conception of syntactic categories derives from an earlier tradition with a much longer history. It is reflected clearly in the tradi-

tional grammars of English that began to be produced in the late 16th century and survives largely unchanged in most contemporary pedagogical grammars. It takes linguistic categories to have a cross-linguistic basis in something independent of the distributions of the particular forms in any particular language—something inhering in meaning, logic, or the structure of thoughts. This older tradition was revived in generative grammar as developed by Harris’s student Chomsky. Chomsky (1965) comments approvingly that the “universal grammar” of traditional language study (e.g., by the 17thC Cartesians, whom he feels that linguists like Harris had unwisely dismissed) “advanced the position that certain fixed syntactic categories (Noun, Verb, etc.) can be found in the syntactic representations of the sentences of any language, and that these provide the general underlying syntactic structure of each language” (p. 28).

As Johnson points out, by reference to any set of categories that is of roughly the traditional sort—including “verbs (transitive and intransitive), quantifiers, connectives, adjectives, nouns, and singular terms” (2004, 115)—systematicity in the sense of [8] simply does not hold. The categories posited by traditional grammar are based on arity of predicates, argument structure of propositions, function-argument application, singularity vs. multiplicity, etc. These are quite reasonable linguistic notions, and Johnson refers to the traditional categories as “kinds”, suggesting that they are analogous to the natural kinds of biology; but the cross-linguistically applicable categories based on these notions do not coincide with the categories arrived at by distributional classification; and it is the latter that leads to intersubstitutability as demanded by [8]. If Johnson had based his remarks on a set of categories from contemporary generative linguistics (CP, IP, DP, VP, V, D, N, Tense, Agr, etc.), it would not have affected the general drift of his conclusions.

Putting things this way runs counter to the stereotyped standard history of 20th-century linguistics. The standard story about Harris and the linguists and linguistic anthropologists whose methods he formalized (Boas, Bloomfield, Hockett, and others) is that they took languages as they found them, stayed close to the empirical ground, analyzed raw data in ad hoc ways without ethnocentric preconceptions, and assumed “that languages could differ from each other without limit in unpredictable ways” (Joos 1966, 96). Above all, they eschewed the application of methods that assumed natural languages were like the invented languages of formal logic. The standard story about Chomsky, on the other hand, is that he reintroduced aprioristic ideas from 17thC philosophical grammar, cherished the content the grammatical traditions from earlier centuries that the anthropological linguists had shunned, melded them with techniques from formal logic, insisted on universal properties of languages and grammars, and developed a mathematical conception of grammar that treated natural languages just the way logicians and computer scientists treat formal languages.

We are saying that this standard view has it backwards in at least one respect. Chomsky's requirement that there be a basis for the syntactic category system that is independent of any particular language, and a universal framework that holds constant across all natural languages, guarantees that the category system will be very much *unlike* that of formal languages. The sense of 'verb' (for example) that permits us to identify verbs across all human languages does *not* make all verbs within a language intersubstitutable. The same holds for anaphoric items like pronouns, qualifying words like adjectives, and so on. The universal category system, when applied to a particular language, will not be a system of categories within each of which all members are mutually substitutable. Rather, it is Harris's methods of distributional analysis that will guarantee, for any language to which they are applied, that a set of categories in which the full intersubstitutability that is characteristic of formal languages will result. That is simply a property of the kind of system of categories upon which distributional analysis insists.¹⁰

In F&P, despite the remark that "the systematicity argument for combinatorial structure in thought exactly recapitulates the traditional Structuralist argument for constituent structure in sentences" (F&P, 37), the two different conceptions of syntactic categories are not distinguished. There is in fact a tradeoff between (i) preserving systematicity of natural languages by analyzing them in terms of equivalence classes for intersubstitution and (ii) ensuring breadth of scope of generalizations over categories. Bloomfield (1933) was well aware of this: though he acknowledges that "The categories of a language, especially those which affect morphology (*book* : *books*, *he* : *she*), are so pervasive that anyone who reflects upon his language at all, is sure to notice them" (p. 270), he also points out that "Form-classes are not mutually exclusive, but cross each other and overlap and are included one within the other" (p. 269). This complex cross-classification and intersection of syntactic and morphological categories is the basis for our discussion above.

Fodor and Pylyshyn seem to have made the first of Bloomfield's observations without the second: the fact that categories are pervasive and evident does not fix an answer to the question of whether systematicity (under the intersubstitutability characterization) holds or not.

¹⁰ We note in passing that Johnson (2004, 120) misunderstands what Harris's methods mean for the issue of systematicity. He attributes to Harris the strange view "that natural language lacks systematicity so much that *A* and *B* are distinct words if and only if there is a *C* such that *C(A)* is grammatical but *C(B)* is not." Harris maintains no such thing. In one direction, Johnson's remark is trivial (naturally, if two phonetic forms have different distributions they must be different), and in the other it is just wrong: Harris would regard *scurrilous* and *scabrous* as different words even if they had identical distributions, because they have different phonetic forms. What it would entail if they exhibited no distributional differences would be that they would be assigned to all the same distributionally determined categories, not that they would be regarded as the same word.

Gross (1979) wrote at the conclusion of a major project to write a transformational grammar of French that of the lexemes his team had studied, they had not found any two that had the same range of syntactic behaviors. Since he was using the device of transformations to capture generalizations about syntax, he regarded this burgeoning evidence of idiosyncratic categories as an indication of “The failure of transformational grammar.” It does not need to be seen thus. The discoveries made by his team concerned the multitude of distinct syntactic behavior patterns for central members of the basic vocabulary of the language, particularly the verbs. There will be, indeed, a very large number of syntactically differentiable kinds of verb. An upper bound on the number is given by the size of the power set of the set of all **strict subcategorization** frames for verbs (see Chomsky 1965). There are, *inter alia*:

- [23]a. verbs taking no complement (like *elapse*),
 b. verbs taking an NP direct object (like *possess*),
 c. verbs taking two NPs (like *hand*),
 d. verbs taking an NP and a *to*-PP (like *donate*),
 e. verbs taking an NP and a *with*-PP (as in *supply someone with something*)

and so on through dozens of other subclasses determined by co-occurrence with complements. A given verb can belong to more than one subclass; for instance, *give* participates in the syntactic behavior of [i] (*Give if you can*), [ii] (*Give what you can*), [iii] (*Give me that*), and [iv] (*Give this to her*). A verb will only be freely substitutable for another verb if it belongs to all and only the same subcategories. So each verb is inter-substitutable only with those verbs that belong to all and only the same strict subcategorization subclasses.

But then there is more variety on top of all this: some verbs that take an NP direct object participate in the truth-condition-preserving active/passive alternation (as does *possess*); others occur only in the active form (like *have* when it means ‘possess’); and a few occur only in the passive form (*born*; *rumored*). Some have raising syntax (like *tend*: note *There tend to be multiple apertures*) and some have control syntax (like *try*: note **There try to be multiple apertures*). And so on. It seems to us it is small wonder if a study of a sample of common verbs in a natural language found that no two belonged to exactly the same strict subcategorization subclasses and participated in exactly the same syntactic alternations (Gross’s transformations).

Consider what this means for a simple generalization like that in English the verb begins the verb phrase. It holds for verbs of all the dozens or hundreds of different verb subclasses without exception. It cannot be stated in a unitary way on categories refined enough to ensure inter-substitutability. Instead, it must be replaced by a lengthy list stating the must-be-first requirement for each of the different intersubstitution-closed microcategories containing items that used to be called verbs.

The syntactic theorist has to be responsive to two competing desiderata: that categories should be expansive enough to allow for the statement of broad syntactic generalizations (like those about constituent order), and that they should be narrow enough to allow for the statement of the detailed matters (like complement selection) that make constituents mutually interchangeable. The tension is normally addressed by means of feature decompositions of syntactic categories; for example, GPSG treats *elapse* as a V[SUBCAT 1], *possess* as a V[SUBCAT 2], and so on, taking such complex symbols to have the common element V and the separable elements [SUBCAT 1], [SUBCAT 2], etc. Fodor and Pylyshyn cannot claim that natural languages have systematicity relative to the broader categories like V, because it simply isn't true. They can at best claim that natural languages have systematicity relative to some huge set of highly refined categories that has never yet actually been constructed or even sketched (not even by Harris).

But now consider how Fodor and Pylyshyn are to apply the same reasoning to thought and inference, which is very much at the heart of their project. Whether thought has intersubstitutable categories depends on whether constituents of thought turn out to be classified into equivalence classes for substitution; and it is *not* a tautology that they must. Yet when Fodor and Pylyshyn give examples of the nature of thought and inference, as with the discussion of inferring *P* from *P&Q* to which they return repeatedly in their paper, they make use of invented formal languages — logical calculi — that have been designed to have a small, convenient, and semantically motivated set of categories within which intersubstitutability holds.

It is hardly surprising that Fodor and Pylyshyn conclude that thought and inference are systematic given that they have, in effect, stipulated it through their choice of formalization in representing their examples of thought and inference.

Conclusion: The Seductiveness of Systematicity

Why has it proved so tempting to so many to think that cognitive and linguistic competence requires some kind of intersubstitutability of constituents? It appears to be related to the observation we cited above from Bloomfield: that lexical and morphological categories “are so pervasive that anyone who reflects upon his language at all, is sure to notice them” (1933, 270). But just because it is obvious (given a moment's reflection) that some words fall into classes within which some items can be substituted for each other, it does not follow that it is a deep syntactic universal, or even a syntactic or semantic property, that requires an explanation in terms of cognitive architecture. Rather, it seems very likely that this property can be explained by limitations on human memory.

Human beings have the capacity to remember the partially idiosyncratic behaviors of many irregular words and idioms, but many people's vocabularies run into the tens or even hundreds of thousands. Many

of the items, though, are quite rare. As we progress from frequent and familiar items to rarer and less familiar ones, our memory limits dictate that there have to be items that share all aspects of syntactic behavior. We know many inflectionally peculiar nouns like *child* and *goose* and *tooth* and *mouse*, and syntactically peculiar nouns like *sake* and *dint* and *self* and *one*, but we could not possibly remember special syntactic privileges for every noun we ever encountered; eventually, there have to be classes of nouns like *pontification*, *transsubstantiation*, *intertranslation*, etc., that share exactly the same behaviors and thus can always be grammatically substituted for each other. We do not have the memory capacity for it to be otherwise. Something similar is true for verbs (see the extensive discussion in Pinker & Prince 1988), and for adjectives too.

Sets of grammatically idiosyncratic forms exhibiting partially overlapping subregularities are a familiar feature of the most frequently-occurring items in the vocabulary, but out in the long tail of the frequency distribution, where the rarer words are, there has to be a degree of regularity and predictability — some clusters of items sufficiently unfamiliar that all their syntactic behavior can be inferred on the basis of general facts about whole equivalence classes of words.

This is the grain of truth in the notion of systematicity as intra-categorical intersubstitutability, and it is a truth about the structure of the lexicon, on which to some extent (under the insight formalized by X-bar theory) certain truths about phrases depend. But it leaves wide scope for explanatory theories to account for the ways in which natural languages are systematic, and for other details of what languages are like. The division of linguistic rules into separate phonological, morphological, and syntactic systems might be explained by considerations of the design of speech control mechanisms; the existence of hierarchical phrase structure might be explained by reference to phylogenetic evolution (Simon 1962; Sampson 2005, 141ff); some aspects of syntax like the apparent broad tendency for languages to have nouns and verbs might be explained by reference to the function of propositional communication; the morphological irregularities and exceptions in many natural languages might be due to the exigencies of intergenerational cultural transmission; and the explanation for why at least some syntactic categories really do have thousands of members sharing identical syntactic properties and distributions might have to do with brute memory limitations.

The lexical systematicity that natural languages exhibit may well be explained not by some special capacity — some Turing-machine-style cognitive architecture with which we are endowed — but rather by an incapacity. Human languages have large sets of intersubstitutable words and phrases because we humans simply lack the capacity to remember separate idiosyncratic distributions for all of them.

References

- Aizawa, Kenneth (1997). "Explaining systematicity." *Mind & Language*, 12, 115–136.
- Bever, Thomas G., Jerry A. Fodor & Merrill Garrett (1967). "A formal limitation of associationism." In Theodore R. Dixon, & David L. Horton, *Verbal Behavior and General Behavior Theory* (Englewood Cliffs: Prentice Hall), 582–585.
- Bloomfield, Leonard (1933). *Language* (New York: Holt).
- Bresnan, Joan; Ronald M. Kaplan; P. Stanley Peters; and Annie Zaenen (1982). "Cross-serial dependencies in Dutch." *Linguistic Inquiry*, 13, 613–635.
- Carruthers, Peter (2002). "The cognitive functions of language." *Behavioral and Brain Sciences*, 25, 657–674 (peer commentary and bibliography 674–725).
- Chomsky, Noam (1959). "On some formal properties of grammars." *Information & Control*, 2, 137–167.
- Chomsky, Noam (1965). *Aspect of the Theory of Syntax* (Cambridge, MA: MIT Press).
- Cummins, Robert (1996). "Systematicity." *Journal of Philosophy*, 93, 591–614.
- Fodor, Jerry A. and Zenon Pylyshyn (1988). "Connectionism and cognitive architecture: a critical analysis." *Cognition*, 28, 3–71. Reprinted in Macdonald & Macdonald (1995), 90–163. Page references are to the original.
- Gazdar, Gerald (1981). "Unbounded dependencies and coordinate structure." *Linguistic Inquiry*, 12, 155–184.
- Gazdar, Gerald; Ewan Klein; Geoffrey K. Pullum; and Ivan A. Sag (1985). *Generalized Phrase Structure Grammar* (Oxford: Basil Blackwell).
- Gross, Maurice (1979). "On the failure of generative grammar." *Language*, 55, 859–885.
- Hadley, Robert (1997). "Cognition, systematicity, and nomic necessity." *Mind & Language*, 12, 137–153.
- Harris, Zellig S. (1951). *Methods in Structural Linguistics* (Chicago: University of Chicago Press). (Later reissued in paperback as *Structural Linguistics*.)
- Huddleston, Rodney and Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language* (Cambridge: Cambridge University Press).
- Johnson, Kent (2004). "On the systematicity of language and thought." *Journal of Philosophy*, 101, 111–139.
- Joos, Martin (1966). *Readings in Linguistics I: The Development of Descriptive Linguistics in America 1925–56*. Second edition (Chicago, Illinois: University of Chicago Press).
- Lashley, Karl S. (1951). "The problem of serial order in behavior." In L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior* (New York: Wiley), 112–136.
- Matthews, Robert (1994). "Three-concept Monte: explanation, implementation, and systematicity." *Synthese*, 101, 347–363.
- Macdonald, Cynthia and Graham Macdonald (eds). (1995). *Connectionism: Debates on Psychological Explanation* (Oxford: Basil Blackwell).
- Niklasson, Lars and Tim van Gelder (1994). "On being systematically connectionist." *Mind & Language*, 9, 288–302.

- Ojeda, Almerindo (1988). "A linear precedence account of cross-serial dependencies." *Linguistics and Philosophy*, 11, 457–492.
- Pinker, Steven and Alan Prince (1988). "On language and connectionism: analysis of a parallel distributive processing model of language acquisition." *Cognition*, 28, 73–193.
- Pullum, Geoffrey K. (1972). "Free variation isn't symmetric, either." *International Journal of American Linguistics*, 38, 268–270.
- Pullum, Geoffrey K. (1984). "Syntactic and semantic parsability." *COLING 84: Proceedings of the 10th International Conference on Computational Linguistics* (Menlo Park, CA: Association for Computational Linguistics), 112–122
- Pullum, Geoffrey K. and Gerald Gazdar (1982). "Natural languages and context-free languages." *Linguistics and Philosophy*, 4, 471–504.
- Pullum, Geoffrey K. & Kyle Rawlins (in press). "Argument or no argument?" To appear in *Linguistics and Philosophy*.
- Pullum, Geoffrey K., James Rogers and Barbara C. Scholz (in preparation). *Model-Theoretic Syntax*.
- Robbins, Philip (2005). "Systematicity." In Keith Brown (ed.), *Encyclopedia of Language and Linguistics (2nd edition)*, Vol. 12 (New York: Elsevier), 440–443.
- Rogers, James (1998). *A Descriptive Approach to Language-Theoretic Complexity* (Stanford, CA: CSLI Publications).
- Rumelhart, David & J. L. McClelland (1986). "On learning the past tenses of English verbs." In J. L. McClelland, D. L. Rumelhart and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models* (Cambridge, MA: MIT Press).
- Sampson, Geoffrey (2005). *The 'Language Instinct' Debate* (London: Continuum).
- Shieber, Stuart (1985). "Evidence against the context-freeness of human language." *Linguistics and Philosophy*, 8, 333–343.
- Simon, Herbert A. (1962). "The architecture of complexity." *Proceedings of the American Philosophical Society*, 106, 467–482. Reprinted in Herbert A. Simon, *The Sciences of the Artificial* (Cambridge, MA: MIT Press, 1969).
- Thatcher, James W. (1967). "Characterizing derivation trees of context-free grammars through a generalization of finite automata theory." *Journal of Computer and System Sciences*, 1, 317–322.
- Wickelgren, W. A. (1969). "Context-sensitive encoding, associative memory, and serial order in (speech) behavior." *Psychological Review*, 76, 1–15.