"More Languages Than We Might Have Thought. Fewer Languages Than There Might Have Been"

Richard S. Kayne
New York University

March, 2016

1. Introduction.
    In working toward an understanding of the syntactic component of the human language faculty, syntacticians necessarily ask question after question. One prominent question is:
    (1) What properties do all languages have in common?
A related question is:
    (2) How exactly do languages vary in their syntax and what are the principled limits on that variation?

2. Fewer languages than there might have been
    In partial answer to the first question, we can think of the following:
    (3) All languages have negation.[1]
    (4) All languages have demonstratives (such as *this, that*).[2]
The interest of these two properties is enhanced by the observation that other familiar elements are not universally present (at least not in visible form), e.g. definite and indefinite articles, which many languages lack, Russian being one well-known example.
    A second, different type of property that all languages have in common is given by the statement that all languages prohibit sentences such as:[3]
    (5) *Which student were you talking to this student and?
Even in languages in which an interrogative phrase such as *which student* is normally brought to sentence-initial position, as it is in English and in Italian, it is never possible to apply that operation to the part of a coordination that immediately follows *and*.
    A third type of property, different from the first two, that holds of the set of human languages is the following one:
    (6) Every possible language is such that its mirror-image is an impossible language.
Put another way, there are no mirror-image pairs of languages. There is no mirror-image English (or mirror-image Italian), where mirror-image English would be defined as a language identical to English in its vocabulary, but such that every well-formed English sentence would have a counterpart in mirror-image English with the same words in reverse order. For example, corresponding to English:
    (7) The dog was running after the cat.
mirror-image English would have:
    (8) *Cat the after running was dog the.
_____

[1]Cf. Horn (1989, xiii) and Dryer (2005, 454).
[2]Cf. Lyons (1999, xv).
[3]Cf. Ross (1967).

The property given in (6) is clearly valid, as far as anybody can see, even if, for the purpose of evaluating it, we set aside questions of constituent structure. Once we bring in such considerations, however, we can formulate an even stronger principle for which I have used the term 'antisymmetry'.[4] Informally put, antisymmetry states in part that if some constituent structure (tree structure) representation is the correct one for some sentence, or phrase, in some language, then the exact mirror-image of that constituent structure representation cannot be correct for any sentence or any phrase in any language.

Taken together, (3)-(6) have told us that there are many fewer languages than one might have imagined. The set of human languages might have included some lacking negation, or lacking demonstratives, but such languages are not to be found. A language that would allow (5) is imaginable but unattested and is virtually certain never to be attested. One can, finally, readily understand what mirror-image English would look like if it existed, but it doesn't exist and neither does any language exist, as far as we can tell, that is the exact mirror-image of some other possible language. As in the title of this essay, there are many fewer languages than there might have been had the human language faculty been otherwise than it is.

We will of course also want to reach an understanding of why (3)-(6) hold as properties of the human language faculty. The properties given in (3) and (4) can be put into context alongside the wide-ranging cartography work of Cinque (1999) (along with related later work of his), which suggests that there are a considerable number of syntactic elements common to every human language. Future work will ask how best to characterize that set of common elements, as well as how best to distinguish, within that set of common elements, those that can apparently consistently be left silent in some languages from those, like negation and demonstratives, that are always visible (at least sometimes) in every language.

The property concerning coordination given in (5) might be reducible to Rizzi's (1990) relativized minimality principle or to some variant of it. (It might be, more specifically, that the first phrase of the coordination sharply blocks the extraction of the second phrase.) As for (6), an initial answer is that it follows from the Linear Correspondence Axiom (LCA) of Kayne (1994). A further, more ambitious, question asks why the language faculty should follow the strictures of the LCA. The answer suggested in Kayne (2011) rests on the view (not shared by Chomsky[5]) that temporal/linear order is uniformly integrated by the language faculty directly into the merge operation that creates larger constituents out of smaller ones and that lies at the heart of syntax.

3. More currently spoken languages than we might have thought

The second question given toward the beginning of this essay, in (2), was the following:

(9) How exactly do languages vary in their syntax and what are the principled limits on that variation?

_____

[4]Cf. Kayne (1994).
[5]Cf. Chomsky (2005, 15), who takes linear order to be "restricted to the mapping to the phonetic interface".

This second general question can be approached by linking it to a third:

(10) How many languages are there?

which should itself be broken down into two further subquestions:

(11) How many possible languages are there?

(12) How many languages are currently spoken?

The question in (12) looks more manageable than the one in (10), but it is in fact less easy than it might look. Encyclopedias and other sources often give an answer on the order of 6000 languages currently spoken.[6] But they don't always say how they've done the counting, or justify the criteria they've used.

One criterion often thought of in counting or individuating languages is that of mutual intelligibility, i.e. count two languages as distinct only if they're mutually unintelligible. Although that criterion may be useful for some purposes (even with the complication that mutual intelligibility is in practice not a black and white matter), it is not a sufficiently fine- grained criterion for a syntactician.

Take English. Is there one English or many? If there are many, how many? We all know that British accents differ from American accents, i.e. we all know that the phonology of British English differs from the phonology of American English. What is less widely appreciated is that the syntax of British English differs from the syntax of American English, as we shall now see.

Here is one example:[7]

(13) Are you going to the theater tomorrow? I might (do).

In such cases, British English allows *I might do, He might have done*, etc. in a way that American English does not. To count English as one language would amount to deciding to overlook this rather striking syntactic difference. Let us agree, then, to count British English and American English as two languages (despite substantial mutual intelligibility), each with its own syntax. Of course the syntactic differences between them are smaller than those between either of them and Japanese. But as far as we can see this is a matter of degree, not a matter of kind.

One might wonder where this is leading. What about American English itself? We all know that there are many different accents within the United States. Again, what is less widely appreciated is that there are also, within the United States, many differences in syntax. As an example, consider:

(14) I might could.

which is found in the English of the southern United States.[8] Why would we want to say that the English that allows *I might could* is identical in its syntax to the English that disallows *I might could*? The syntax of the one is simply not identical to the syntax of the other.

In this way, we are led to distinguishing, syntactically speaking, southern American English from, say, northeastern American English, which lacks *I might could*. But northeastern American English is itself not uniform in its syntax, since in northern New England one finds:[9]

––––––––––––––––––––––

[6]For some discussion, cf. Comrie (1987, 2), Comrie et al. (2005, 3), Crystal (1987, 284).
[7]Cf. Algeo (2006, 288).
[8]Cf. Hasty (2014).
[9]Cf. Wood (2014).

(15) Mary is intelligent, but so isn't John.

which is not at all possible in my (New York City) English.

One might at this point think of bringing in the notion 'dialect' and of calling the English that allows (13) or (14) or (15) a dialect of English, a different one in each case. As most linguists recognize, though, the dialect/language distinction is not in essence about phonology or syntax (or semantics), but rather concerns the political/cultural/social importance of the sets of speakers in question and the associated prestige of the language/dialect. We can note in addition that neither American English nor British English is called a dialect of English (at least not by Americans).

Once we do bring in dialects, a revision of the question in (12) is called for. A more satisfactory version is:

(16) How many languages/dialects are currently spoken?

This version makes it clear that as far as syntax is concerned (and similarly for phonology and semantics) what are called dialects must be taken into account.

It would be natural to ask again how far this is taking us. If we look for and find more and more and more syntactic differences within what we call English, how many syntactically distinct subvarieties of English will we end up with?

Before attempting to answer this question, let me point out that, although the syntactic differences so far mentioned can be characterized regionally (British English, southern American English, northeastern American English, northern New England English), that is not always the case, as far as we know. An example of a non-regional syntactic difference would be the one indicated by:[10]

(17) these kind of horses

For a subset of speakers of English it is possible to have singular *kind* in combination with plural *these*, as in (17). For other speakers, (17) is excluded. Yet in this case there is no obvious regional generalization about where such speakers are found.

As a second instance of what seems to be a syntactic difference across varieties of English that is not characterizable in regional terms, consider:

(18) a woman that's husband is quite wealthy

which has been attested in various parts of the English-speaking world.[11] The standard version of (18) is:

(19) a woman whose husband is quite wealthy

in which the relative clause is introduced by *whose*. Examples like (18), in which the relative clause is introduced by *that's*, are not at all possible in my English. So I was astonished many years ago to hear one example like it produced by an old friend of mine who had grown up in New York City only a few miles from where I did.

It would be easy to draw up a longer and longer list of syntactic differences within English, such as the ones mentioned so far, that split the set of English speakers into overlapping or non-overlapping subsets. Although useful and instructive for both descriptive and theoretical purposes, drawing up even a partial list of that sort would be a long-range enterprise that far exceeds the bounds of this essay.

─────────────────────────

[10]The existence of which is noted in the *Oxford English Dictionary* under *kind (of)*.
[11]Cf. Seppänen (1999) and Herrmann (2005).

A quicker answer to (16) can be reached using the following thought experiment. Take any two speakers of English. How long would it take an English-speaking syntactician to discover a clear difference in syntax between those two speakers? Assume, as my experience working with syntax leads me to believe true, that there is, for every pair of speakers of English, an answer to this question, i.e. that there is a finite length of time, whether five minutes or five hours or five days or five years, within which I could find a clear syntactic difference between the English of the one and the English of the other. If so, then no two speakers of English have exactly the same syntax. (By parity of reasoning, the same is almost certain to be true of all other languages/dialects.)

The English language, then, has a current population of speakers whose syntax (and phonology and semantics) is sufficiently similar to justify, against the background of political/social/cultural considerations, the convenience of a single term 'English'. I have tried to show that this can be so even if, as seems extremely likely, there are no pairs of speakers of English whose syntax is identical across the board.

Extrapolating from English to the rest of the world, the number of syntactically distinct languages currently spoken may well be at least (given bilingualism) as great as the number of people currently alive (setting aside babies up to a certain age).

4 More possible languages than we might have thought

The question in (11) should also be revised to integrate dialects. Revising it yields:

(20) How many possible languages/dialects are there?

Included in this set, in addition to those currently spoken, are those languages or dialects that were spoken in the past, those that will be spoken in the future and even those possible languages that may for one reason or another never be spoken. To try to answer (20), we can proceed as follows.

We already know of a great many syntactic differences found across languages and dialects. A small number have been discussed earlier. To approach (20), let us take some arbitrary pair of syntactic differences and ask whether or not they are independent of one another. (In technical terms, this amounts to asking whether or not the two differences in question can be traced back to a single difference in the value of a single more abstract syntactic parameter.)

To see how this kind of reasoning works, consider the following example. English and Italian differ in that Italian allows sentences with no visible subject in a way that English does not:

(21) Parla troppo.
(22) *Talks too much.

English and Italian also differ in that in sentences with an auxiliary and a past participle Italian has an alternation between auxiliary 'be' and auxiliary 'have' that English does not. Thus Italian allows:

(23) Gianni è arrivato ieri.

whereas English would require 'have':

(24) John has/*is arrived yesterday.

It seems virtually certain that these two Italian-English differences are independent of one another. Evidence comes in part from other languages, e.g. from the fact that Spanish is like Italian with respect to (21) but like English with respect to (24), while French is like Italian in (23) but like English in (22). (Put more technically again, there is

almost certainly no single parameter that underlies both the subject difference and the auxiliary difference at issue.)

The key question is now:

(25) How many such independent differences are there, cross-linguistically?

Put another way:

(26) How large is the set of syntactic differences such that each is independent of all the others?

More technically:

(27) How many independent syntactic parameters does the language faculty allow for?

There is no simple way to answer this question, but it seems to me reasonable to think, just taking into account those syntactic differences that we are already aware of, that the answer to (25)-(27) may well be on the order of 100.

If so, and if each parameter has two possible values that can, by assumption, fluctuate independently of the values of all the other parameters, then the number of possible combinations of parameter values, and hence the number of possible syntactically distinguishable languages, will be 2 to the 100th power, which is approximately 10 to the 30th power, i.e. the number that would be written out as 1 followed by 30 zeros. In other words, if the assumed answer to (25)-(27) is 100, or even close to that, then the total number of syntactically distinct languages made available in principle by the language faculty is 'astronomical'.

Syntacticians need not despair (nor need anyone else despair). Syntacticians do not need to (and could not possibly) study each and every one of these 10^{30} languages individually.[12] We do. on the other hand, need to reach an understanding (and are in the process of doing so) of what the set of parameters is, of what the set of possible syntactic differences looks like, of what the limits are on those differences, and of why those limits hold. Simultaneously, we need to reach an understanding (and are in the process of doing so) of what properties all possible human languages have in common, and of why they have in common those properties and not others.[13]

References:

Algeo, J. (2006) *British or American English? A handbook of word and grammar patterns*, Cambridge University Press, Cambridge.

Chomsky, N. (2005) "Three Factors in Language Design," *Linguistic Inquiry*, 36, 1-22.

Cinque, G. (1999) *Adverbs and Functional Heads. A Cross-Linguistic Perspective*, Oxford University Press, New York.

Comrie, B., M.S. Dryer, D. Gil and M. Haspelmath (2005) "Introduction," in M. Haspelmath, M.S. Dryer, D. Gil and B. Comrie (eds.) *The World Atlas of Language Structures,* 1-8.

Comrie, B. (1987) "Introduction," in B. Comrie (ed.) *The World's Major Languages,* Oxford University Press, New York, 1-29.

―――――――――――――――――

[12]For important methodological discussion of the study of 'exotic' languages, see Davis et al. (2014).

[13]For some relevant discussion, see Chomsky (2005).

Crystal, D. (1987) *The Cambridge Encyclopedia of Language*, Cambridge University Press, Cambridge.

Davis, H., C. Gillon and L. Matthewson (2014) "How to Investigate Linguistic Diversity: Lessons from the Pacific Northwest," *Language* 90, e180-e226.

Dryer, M.S. (2005) "Negative Morphemes," in M. Haspelmath, M.S. Dryer, D. Gil and B. Comrie (eds.) *The World Atlas of Language Structures,* 454-457.

Hasty, J. D. (2014) "We might should be thinking this way. Theory and Practice in the Study of Syntactic Variation," in R. Zanuttini and L. R. Horn (eds.) *Micro-Syntactic Variation in North American English*, Oxford University Press, New York, 269-293.

Herrmann, T. (2005) "Relative Clauses in English Dialects of the British Isles," in B. Kortmann, T. Herrmann, L. Pietsch, and S. Wagner (eds.) *A Comparative Grammar of British English Dialects. Agreement, Gender, Relative Clauses*, Mouton de Gruyter, Berlin, 21-123.

Horn, L.R. (1989) *A natural history of negation*, The University of Chicago Press, Chicago.

Kayne, R.S. (1994) *The Antisymmetry of Syntax*, The MIT Press, Cambridge, Mass.

Kayne, R.S. (2011) "Why Are There No Directionality Parameters?" in M. Byram Washburn, K. McKinney-Bock, E. Varis, A. Sawyer and B. Tomaszewicz (eds.) *Proceedings of the 28th West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project, Somerville, MA, 1-23 (also in 2013 *Studies in Chinese Linguistics*, 34, 3-37; and in 2014 T. Biberauer and M. Sheehan (eds.) *Theoretical Approaches to Disharmonic Word Order*, Oxford University Press, 219-244).

Lyons, C. (1999) *Definiteness,* Cambridge University Press, Cambridge.

Rizzi, L. (1990) *Relativized Minimality*, MIT Press, Cambridge, Mass.

Ross, J.R. (1967) *Constraints on Variables in Syntax*, Doctoral dissertation, M.I.T.

Seppänen, J. (1999) "Dialectal variation in English relativisation", *Lingua*, 109, 15-34.

Wood, J. (2014) "Affirmative Semantics with Negative Morphosyntax. Negative Exclamatives and the New England *So AUXn't NP/DP* Construction," in R. Zanuttini and L. R. Horn (eds.) *Micro-Syntactic Variation in North American English*, Oxford University Press, New York, 71-114.