The Complexity of Trees, Universal Grammar and Economy Conditions

Chris Collins, New York University

Abstract: In this squib, I argue that the child faces a severe computational complexity problem in parsing even the simplest of trees: the number of possible trees consistent with UG grows exponentially as a function of the number of lexical items. Economy conditions have the result of drastically decreasing the complexity of the parsing task. I also discuss the relationship between UG, I-language, economy conditions and explanatory adequacy.

Keywords: binary branching, Catalan numbers, Super Catalan numbers, economy conditions, explanatory adequacy

1.   The Complexity of Trees in Acquisition

Imagine a language acquisition scenario. The child who is between 0-4 years old listens to their parents and other children speaking, trying to understand what they are saying. There are many difficulties facing them in trying to understand the sentences of the PLD (primary linguistic data). But surely one difficulty they face is the sheer intractable number of possible trees corresponding to any given string of words. I assume that finding the right tree is a prerequisite to understanding what is being said. Furthermore, I assume that by the age of four the child performs this tree parsing task effortlessly and mostly without deviating from the adult structures. If a tree has n leaves (so the child hears n lexical items), there are a vast number of trees consistent with those lexical items. How does the child arrive at picking out one tree from amongst all the ones possible?

Consider a relatively short sentence with four lexical items (or morphemes). Here are the possible trees consistent with those four lexical items (excluding unary branching, but allowing in other types of branching):

(1)　　a.　　[x y z w]

　　　　b.　　[[x y z] w]

　　　　c.　　[[[x y] z] w]

　　　　d.　　[[x [y z]] w]

　　　　e.　　[x [y z w]]

　　　　f.　　[x [y [z w]]]

　　　　g.　　[x [[y z] w]]

　　　　h.　　[[x y] z w]

　　　　i.　　[[x y] [z w]]

　　　　j.　　[x [y z] w]

　　　　k.　　[x y [z w]]

For a very simple sentence with four lexical items there are eleven possible trees that the child must sort through to find the exact one needed. It is useless to say that meaning guides the choice, because after all the child is trying to understand the meaning that the adult intends to communicate. In other words, the child cannot be guided by meaning, because the child is trying to recover the meaning of the sentence.

So how is the parsing task possible for a young child? This paper will argue that economy conditions limit the total number of trees that the child needs to consider, radically reducing the

computational complexity of the task for the child, and making language acquisition possible. In the conclusion, I will suggest that other economy conditions can be seen in the same light.

I do not discuss any specific theory of natural language parsing, nor do I enter into a specific theory of language acquisition. That is, I am not making any specific claims about how parsing or language acquisition are actually carried out by the child. Rather, I am claiming that economy conditions reduce the total number of possible trees corresponding to a string of words, and therefore contribute to reducing the complexity of the task that the child faces.

In sections 2 and 3, I give a concrete idea of how complex the parsing task is by discussing the so-called Catalan numbers and Super Catalan numbers. In section 4, I discuss binary branching, and how it reduces the computational complexity of the parsing task. In section 5, I show how binary branching is the result of an economy condition. In section 6, I discuss unary branching. In section 7, I discuss the role of economy conditions in explanatory adequacy. Section 8 is the conclusion.

2.   Catalan Numbers

A binary branching tree is defined as follows:

(2)   Each node in a binary branching tree is (a) a terminal or (b) has exactly two daughters.

The number of binary branching trees with n-leaves is given by the Catalan numbers:

(3) $\text{Trees}_{\text{BIN}}(n)$ $=$ $(2n-2)!/n!(n-1)!$

For example, for two leaves, there will be one tree: $(4-2)!/2!(2-1)! = 2!/2! = 1$. For one leaf, there will also be one tree: $(2-2)!/1!(1-1)! = 0!/1!0! = 1$.

This number can be approximated by the following function, which shows clearly that the number of trees grows as an exponential function of the number of leaves:

(4) $\text{Trees}_{\text{BIN}}(n) = (\text{approx.})\ 4^{(n-1)}/(n-1)^{1.5}\sqrt{\pi}$

The first few Catalan numbers are given in (5). For example, for 10 leaves, there will be 4,862 binary branching trees. For 14 leaves, there are 742,900 binary branching trees.

(5) leaves=n	$\text{Trees}_{\text{BIN}}(n)$
1	1
2	1
3	2
4	5
5	14
6	42
7	132
8	429
9	1,430
10	4,862
11	16,796
12	58,786

13	208,012
14	742,900
15	2,674,440
16	9,694,845
17	35,357,670
18	129,644,790
19	477,638,700
20	1,767,263,190

3.   Super Catalan Numbers

The Super Catalan numbers correspond to the number of trees of n leaves (excluding unary, but allowing other types of branching, e.g., binary, ternary, etc.). I have not found a simple formula that gives the Super Catalan numbers, rather they are defined recursively in terms of preceding Super Catalan numbers. I will write the function for Super Catalan numbers as Trees(n). On the relevance of Catalan and Super-Catalan numbers to coordination, see Wagner (2010: 193).

The first few Super Catalan numbers are given below (from http://oeis.org/A001003). In other words, for a tree of 10 leaves there are 103,049 trees. It is easy to see that the Super Catalan numbers are much larger than the Catalan numbers.

(6)	leaves=n	Trees(n)
	1	1
	2	1
	3	3

4	11
5	45
6	197
7	903
8	4,279
9	20,793
10	103,049
11	518,859
12	264,6723
13	13,648,869
14	71,039,373
15	372,693,519
16	1,968,801,519
17	10,463,578,353
18	55,909,013,009
19	300,159,426,963
20	1,618,362,158,587

4. Binary Branching

Consider again the toy problem introduced in (1). There we saw that the total number of bracketings for four morphemes is eleven, which is a Super Catalan number (see (6) above). But let's restrict ourselves to binary branching structures, crossing out the non-binary trees. There are only five trees left:

(7)   a.   ~~[x y z w]~~

    b.   ~~[[x y z] w]~~

    c.   [[[x y] z] w]

    d.   [[x [y z]] w]

    e.   ~~[x [y z w]]~~

    f.   [x [y [z w]]]

    g.   [x [[y z] w]]

    h.   ~~[[x y] z w]~~

    i.   [[x y] [z w]]

    j.   ~~[x [y z] w]~~

    k.   ~~[x y [z w]]~~

Therefore, a child who has binary branching Merge, which excludes unary branching and other kinds branching (ternary branching), would have many fewer possible trees to exclude when trying to arrive at the correct parse for a sentence. In other words, binary branching Merge allows the child to ignore large number of possible analyses of the data (PLD), focusing only on a small subset of possibilities.

Haegeman (1994: 143) describes the relationship between binary branching and language acquisition as follows (see also Kayne 1984): "A child equipped with a UG that implements only binary branching will have fewer decisions to make when assigning syntactic structure to the data he is exposed to than a child equipped with a less constrained UG which allows ternary or four-way branching. It is easy to see that the more elements are involved the more

choices are available. The unconstrained theory will consistently offer more choices than the binary branching theory and hence will make the child's task of deciding on the structure harder."

For a tree with n leaves, the number of binary branching trees is given by the Catalan numbers. The number of trees of any branching (excluding unary branching) is given by the Super Catalan numbers. The Super Catalan numbers increase far more rapidly than the Catalan numbers. Even for a relatively small number of leaves, such as 6, the ratio of the Super Catalan number to the Catalan number is $197/41 = 4.69$. Therefore, Merge being restricted to binary branching allows the child to reduce the total number of possible trees by a factor of 4.69 in the case of a short sentence with six lexical items.

In general, for a tree with n-leaves, adopting binary branching allows a computational savings by a factor of $\text{Trees}(n)/\text{Trees}_{\text{BIN}}(n)$. A calculation of this ratio for a small range of n is given below:

(8)	leaves	Trees(n)	$\text{Trees}_{\text{BIN}}(n)$	Ratio
	1	1	1	1
	2	1	1	1
	3	3	2	1.5
	4	11	5	2.2
	5	45	14	3.21
	6	197	42	4.69
	7	903	132	6.84
	8	4,279	429	9.97

9	20,793	1,430	14.5
10	103,049	4,862	21.19
11	518,859	16,796	30.89
12	2,646,723	58,786	45.02
13	13,648,869	208,012	65.62
14	71,039,373	742,900	95.62

The following generalization holds:

(9)   There are many more (non-unary) unrestricted trees than binary branching trees for a given
        number n of leaves, specified by the ratio: $\text{Trees}(n)/\text{Trees}_{BIN}(n)$.

5.   Binary Branching and Economy

Following Collins (1997: 75-78), I assume that binary branching is the result of economy
conditions constraining unrestricted Merge:

(10)   "We will see that it is possible to restrict the possible structures allowed by unrestricted
        Merge by appeal to very general principles, including economy of derivation. The basic
        idea is that phrase structure is binary because binary Merge is the smallest operation that
        will ensure that some structure actually gets built. In other words, binary branching will
        be seen to be a result of Minimality. If this idea can be worked out, adopting unrestricted
        Merge will allow us to perceive a deep relationship between movement and phrase
        structure: that both are subject to Minimality." (pg. 76)

9

More concretely, Collins (1997: 77) compares two derivations:

(11)   Merge(A,B,C) = {A,B,C}

(12)   Merge(A,B) = {A,B}

Merge(C, {A,B}) = {A,B,C}

Collins notes (pg. 77): "At the point in the derivation where the operation Merge(A,B,C) is possible, the operation Merge(A,B) (which involves the merging of fewer elements) is also possible. Therefore, we can derive binary branching from Minimality and unrestricted Merge."

Crucially, this account relies on a notion of economy where the size of particular operations is minimized, not necessarily the total length of the derivation (global economy). Even though (11) has only one operation, (12) wins out, because Merge(A,B) is a smaller operation than Merge(A,B,C).

Merge (the fundamental operation of UG) constrained by economy conditions (third factor conditions not stipulated as part of UG, see Chomsky 2005) has effectively made the task of parsing a tree much simpler for a child facing the PLD. It is important to note that I have not made any specific claims about how parsing is done. Rather, I am making observations about the nature of the task that faces the child in acquiring a language. If Merge (a UG operation applying as part of a syntactic derivation) is constrained by economy, yielding binary branching, then the total number of possible trees is much smaller than if Merge is not subject to economy conditions (and hence does not yield binary branching trees).

A very similar economy-based explanation of binary branching is given in Rizzi (2013: footnote 10): "One could ask the question of why Merge should be binary, rather than n-ary, which would permit more than two elements to be strung together at once. One possible answer is that binary Merge is the computationally most parsimonious combinatory principle capable of yielding an unbounded set of structures, in the following sense: it requires just two temporary buffers of operative memory in which to store the elements to be strung together.... Unary Merge would be insufficient to yield an infinity of structures with distinct elements, and n-ary Merge would require more memory resources. From this perspective, binary Merge is optimal."

Rizzi's explanation emphasizes the amount of memory resources, while Collins' explanation emphasizes the size of the Merge operation. But both economy-based explanations have the effect of forcing binary branching.

6.   Unary Branching

If we allow unary branching nodes, the number of trees (for a given number n of lexical items) is unlimited. For any tree constructed, we can just take the root and add a unary branching node dominating the root, without changing the number of leaves. We can drastically reduce the number of possible trees with the following assumption:

(13)   There is no unary branching.

Given this principle, the child would never even consider unary branching trees, taking an unlimited number of possible trees off the table. I will assume that (13) is part of UG, specified innately as a property of UG (not reducible to third factor economy principles).

7. Explanatory Adequacy

The above discussion focusses on the role of economy conditions in constraining the hypothesis space of a child encountering the PLD. A closely related issue is the role economy conditions play in explanatory adequacy. Chomsky (1986: 53) defines explanatory adequacy as follows:

(14) "A theory of UG meets the conditions of explanatory adequacy to the extent that it provides descriptively adequate grammars under the boundary conditions set by experience." (pg. 53):

The definition of descriptive adequacy is as follows:

(15) "Continuing to think of a grammar as a theory of a language, we may say that a grammar is *descriptively adequate* for a particular language to the extent that it correctly describes this language." (pg. 53)

But neither UG nor particular I-languages incorporate economy conditions. Chomsky (1986) defines UG and I-language as follows:

(16) a. "UG may be regarded as a characterization of the genetically determined language faculty." (pg. 3)

b.   "The I-language, then, is *some* element of the mind of the person who knows the language, acquired by the learner, and used by the speaker-hearer." (pg. 22)

According to these definitions, economy conditions (and structural consequences of economy conditions, such as binary branching) are not part of UG or particular I-languages. The economy conditions are not genetically specified, and they are independent of particular I-languages. For the example at hand, unrestricted Merge is a component of UG and of particular I-languages, but binary branching Merge is the result of economy conditions applying to unrestricted Merge. The situation is illustrated as follows:

(17)   a.   UG: unrestricted Merge

        b.   I-Language: unrestricted Merge

        c.   I-Language + economy: binary branching Merge

To maintain the concepts of explanatory and descriptive adequacy in light of the introduction of economy conditions, the following change is needed:

(18)   A grammar is *descriptively adequate* for a particular language to the extent that <u>in combination with economy conditions</u>, it correctly describes this language.

Chomsky (1986: 77) discusses how attributing principle P to the initial state S_0 is a step toward explanatory adequacy:

(19) "Such general principles as the principle of cyclic application of rules, the island constraints, the subjacency condition, conditions on representations, and so forth serve to restrict the class of permissible rules because it is no longer necessary to incorporate within the rule itself the conditions on its application; in effect, these conditions are factored out of many rules and attributed to the initial state S_0 Formulation of such principles, then, is a step toward explanatory adequacy…"

The reason why attributing P to S_0 is a step toward explanatory adequacy is that the resulting I-languages will now have P independent of the PLD ("boundary conditions set by experience"). In this paper, we are claiming that binary branching is "factored out" of Merge and attributed to economy conditions, not the initial state S_0. If a principle P is an economy condition, then it will also be independent of the PLD, which would also be a step toward explanatory adequacy as defined in (14).

Discussions of explanatory adequacy are usually focused on the relationship between the 'initial state' and the 'steady state' (Chomsky 1986: 25):

(20) "The language faculty is a distinct system of the mind/brain, with an initial state S_0 common to the species…and apparently unique to it in essential respects. Given appropriate experience, this faculty passes from the state S_0 to some relatively stable state S_S, which then undergoes only peripheral modification….The attained state incorporates an I-language (it is the state of having or knowing a particular I-language)."

A general claim of this paper is that economy conditions constrain all I-languages, including stable-state I-languages (mentioned in (20)) as well as the I-languages of new-born infants trying to make sense out of the PLD:

(21)   Economy conditions constrain all I-languages.

8.   Conclusion

The child, faced with primary linguistic data (PLD) which includes speech from the adults and other children surrounding them, must be able to recover the correct tree for a string of words encountered in the PLD. But parsing an individual sentence imposes an apparently computationally insurmountable problem. Each sentence with n morphemes has an exponentially large number of possible parses (the Super Catalan numbers).

The computational complexity of the problem is reduced considerably by the interaction of UG (Merge) with economy conditions. In the example presented in this paper, economy conditions cut down the hypothesis space for a sentence of length n words by the ratio $Trees(n)/Trees_{BIN}(n)$.

Economy conditions have the following properties:

(22)   a.   They are not part of UG.

        b.   They are independent of any particular I-language, hence universal.

        c.   They are not learned.

        d.   The language faculty is constrained by them at birth.

e. They did not evolve.

Even the I-language of a newborn child will be constrained by economy conditions. The child can only generate binary branching trees, without actually having learned that trees are binary branching. So, the economy conditions will immediately lead the child to ignore vast numbers of alternative possibilities that are otherwise consistent with the PLD.

The roles of UG and the economy conditions on this account are different. UG (Merge) specifies the kind of mental representations (hierarchical structures) that are available in the mind of the child. Without UG, the whole language acquisition task would not even get off the ground. But the number of representations consistent with a string of lexical items is vast. Economy conditions work to narrow those down.

It may be that all other economy conditions can be seen in the same light. A preliminary list of economy conditions is given below (for formalization and further references, see Collins and Stabler 2016):

(23) a. Minimal search for Agree (Chomsky 2000: 122)

b. Minimal search for labeling (see Collins 2002, Chomsky 2014)

c. Minimality (binary branching, see section 4).

d. No Tampering Condition (NTC) (see Chomsky 2005)

e. Inclusiveness (Chomsky 2000: 113)

f. Minimal Spell-Out of occurrences at PF/SM-Interface.

g. Phase Impenetrability Condition (PIC) (Chomsky 2000: 108)

Economy conditions may also constraint Spell-Out, the mapping of syntactic structure to linear order. Putting aside movement, if X and Y are sister {X Y}, then they should be contiguous X-Y (X immediately precedes Y) or Y-X (Y immediately precedes X). In other words, the mapping of syntactic structures to representations at the SM-interface (PF-interface) is as straightforward and simple as possible (see Kayne 1994 for a much stronger constraint) with no strange cases allowed (e.g., X and Y are sisters, but Y is spelled-out internal to X or vice versa). These comments suggest the following:

(24)   Spell-Out obeys economy conditions.

If the economy conditions in (23) (and perhaps many others) constrain the narrow syntax and the mapping to the interfaces, they drastically reduce the number of alternative structural hypotheses for any particular sentence, just as we have seen with binary branching above. They severely limit the space of hypotheses that the child must navigate during language acquisition. Since they are economy conditions, they are unlearned, and therefore available to the child immediately. In the ideal case, economy conditions would restrict the number of possible hypotheses for a particular sentence to one, or to a very small number.

Furthermore, as argued above, formulating economy conditions constitutes a major step toward a theory achieving explanatory adequacy.

References:

Catalan number. *Wikipedia, The Free Encyclopedia.*

*(*https://en.wikipedia.org/w/index.php?title=Catalan_number&oldid=1057260992)

Chomsky, Noam. 1986. Knowledge of Language. New York: Praeger.

Chomsky, N. 2000. Minimalist inquiries. In Roger Martin, David Michaels & Juan Uriagereka

(eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155.

Cambridge, MA: MIT Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36:1–22.

Chomsky, Noam. 2013. Problems of Projection. *Lingua* 130, 33-49.

Collins, Chris. 1997. *Local Economy*. Cambridge: MIT Press.

Collins, Chris. 2002. Eliminating Labels. In Samuel Epstein and Daniel Seely (eds*.), Derivation

and Explanation in the Minimalist Program*, 43-64. Oxford: Blackwell.

Collins, Chris and Edward Stabler. 2016. A Formalization of Minimalist Syntax. *Syntax* 19, 43-

78.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory, Second Edition.*

Oxford: Blackwell.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris Publications.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge: MIT Press.

Rizzi, Luigi. 2013. Introduction: Core Computational Principles in Natural Language Syntax.

Lingua 130, 1-13.

Super Catalan/Little Schroeder Numbers. YouTube.

(https://www.youtube.com/watch?v=JNk8s8m-QKs)

Wagner, Michael. 2010. Prosody and Recursion in Coordinate Structures and Beyond. *Natural Language and Linguistic Theory* 28, 183-237.