# Mixed computation: Grammar up and down the Chomsky Hierarchy

Diego Gabriel Krivochen

**Abstract.** Proof-theoretic models of grammar are based on the view that an explicit characterization of a language comes in the form of the recursive enumeration of strings in that language. That recursive enumeration is carried out by a procedure which strongly generates a set of structural descriptions Σ and weakly generates a set of strings S; a grammar is thus a function that pairs an element of Σ with elements of S. Structural descriptions are obtained by means of Context-Free phrase structure rules (e.g., X-bar theory) or via recursive combinatorics (Merge) and structure is assumed to be *uniform*: binary branching trees all the way down. In this work we will analyse natural language constructions for which such a rigid conception of phrase structure results descriptively inadequate, and propose a solution for the problem of phrase structure grammars assigning too much or too little structure to natural language strings: we propose that the grammar can oscillate between levels of computational complexity in local domains, which correspond to elementary trees in a lexicalised Tree Adjoining Grammar.

**Keywords:** Syntax, Derivations, Mixed Computation, Tree Adjoining Grammars, Compositionality

## 1. Introduction

A foundational assumption in generative grammar is the idea that languages are (infinite) sets of (finite) strings (Chomsky, 1956: 114; 2015: 156) whose structural descriptions are enumerated by a recursive procedure (Chomsky & Miller, 1963: 283; Langendoen & Postal, 1984: 18, ff.). The aim of a generative grammar is, then, to enumerate structural descriptions for *all and only* well-formed sentences (Chomsky, 1959: 137). This procedure has the form of a function that relates a structural description to each well-formed sentence in a language. In this sense, a language is a *generated* set insofar as an adequate grammar can recursively enumerate all and only well-formed sequences.

As transformational theory developed towards the 1980s, generative grammar started pursuing the idea that syntactic structure is universally based on a single kind of template far more enthusiastically than ever before, in particular with the extension of the X-bar schema to functional categories in Chomsky (1986) and the raising popularity of Kayne's (1984) highly restrictive format for X-bar trees: well-formed trees are *all and only* single-rooted, binary-branching, and endocentric structures, with no loops or multi-dominance (also Stowell, 1981; Jackendoff, 1977). This means that the structural variety that characterised earlier versions of the generative theory, where flat structures were combined with binary branching (e.g., Ross, 1967; Emonds, 1976), was replaced by a single template that converged in a uniform Context-Free description for the base component of a generative grammar, with some additional power made available through transformations and operations over features (Move and Agree). But this move towards general principles which characterised a single kind of structure (always binary branching and endocentric) was not without its formal issues. It was noted in Pullum (1984) and formally proven in Kornai & Pullum (1990) that infinitely many CFGs strongly equivalent to X-bar theory can be constructed, given the fact that there is no upper boundary on the number of allowed categories (in current frameworks where features can project phrases, the proof carries along since there is no meta-theory that limits the system of arbitrary features). A generative grammar in the transformational tradition consists of a set of context-free rules (which at most can make reference to the last line of a derivation) plus a set of rules that may access the 'derivational history' of a sequence to various effects (e.g., move, copy, delete, and substitute). That architecture entails a set of assumptions about the strong generative capacity of natural language grammars, a point to which we will return briefly. Empirically, the transformational generative view implies *structural uniformity*: the idea that the computational complexity of linguistic dependencies is uniform, and every natural language string can thus be characterized by a unique rule format located at a single point in the Hierarchy. Narita & Fukui (2013: 20) express this view very clearly:

> *considerations of binding, quantifier scope, coordination, and various other phenomena seem to lend support to the* ***universal binary branching*** *hypothesis. However, we do not know why human language is structured that way.* […] *it is likely that theories of labeling and linearization play major roles in this binarity restriction. Moreover, the relevance of third-factor principles of efficient computation has been suggested at times, though arguments are inconclusive* [our highlighting]

Note that the existence of empirical phenomena that are indeed amenable to a binary-branching analysis does *not* preclude the existence of phenomena for which a binary-branching approach is inadequate. However, binary branching as a model of structural uniformity does imply rejecting the possibility that other configurations are available on *a priori* grounds.

Let us consider the consequences of *structural uniformity* for natural language grammars. Recall the distinction between strong and weak generative capacity (Chomsky, 1965: 60): strong generative capacity refers to the structural descriptions enumerated by a grammar, and weak generative capacity to the actual strings it can produce. With this in mind, consider the following quotation, from Stabler (2013: 318):

> *computational consensus was identified by Joshi (1985) in his hypothesis that human languages are both strongly and weakly mildly context sensitive (MCS)*

Stabler proceeds to identify a number of theories that verify this 'computational consensus': from strictly CF grammars (which *may* include HPSG and related theories, although Stabler does not explicitly mention HPSG, LFG, or Dependency Grammars), Tree Adjoining Grammars (TAGs; Joshi, 1985), Combinatory Categorial Grammars (Steedman, 2019) to several kinds of Minimalist Grammars (Stabler, 1997): with phase-constrained extraction, with sidewards movement, with Relativised Minimality constraints…. But, what exactly does this entail for the development of descriptively adequate formal theories of natural language grammars? Stabler's position is clear in this respect:

> *Joshi's original definition of MCS grammars was partly informal, so there are now various precise versions of his claim. One is that human languages are defined by tree adjoining grammars (TAGs) or closely related grammars, and another theoretically weaker (and hence empirically stronger) position is that human language are definable by the more expressive (set local) multi-component TAGs or closely related grammars.*

These 'closely related grammars' refer to the set of formalisms that reside 'between' CF and CS; grammars whose strong generative power is higher than strictly CF but which do not allow for unlimited crossing dependencies, thus not reaching the expressive power of CSG (in the case of TAGs with links, crossing dependencies are limited to two sets of elements, e.g., NPs and VPs; see Joshi, 1985: 221, ff.).

From the perspective of descriptively adequate grammars, it is legitimate to ask whether recognising that a specific construction in a specific language displays, say, restricted CF dependencies (as in the now classic Dutch German and Swiss German examples in Joshi, 1985 and Shieber, 1985) means that *all* constructions (in that language and others) must be assigned a structural description of that computational power. In other words, whether structure uniformity determines that, once an expression in L displays a certain computational complexity, all expressions in L automatically do. What happens with structural uniformity collides with descriptive adequacy?

## 2. Empirical problems

We require of a descriptively adequate theory of grammar that it be able to assign a structural description that represents semantic dependencies between elements in a string. This is essential, since it may be here that one of the crucial differences between natural and formal languages reside. In a formal language such as first order logic, the so-called *unique readability theorem* holds: well-formed formulae are defined in such a way that there is only one way to construct each formula given an alphabet of constants and Boolean connectives and set of formation rules, and only one way to read it (OLP, 2020: 190-192; Epstein, 2011: 8-9). It may be apparent to the reader that natural language does not behave like that: humans use strings of words in ways that do not always obey strict compositionality nor are unambiguous (consider irony, plays on words, jokes…). But we need not consider *how language is used* in order to reject unique readability as a condition on natural language well-formed sentences: it seems to be a property of *how the grammatical system works* rather than *how it is used*. That is: we are facing a problem of *grammar*. Consider the following strings:

1) Wakanda is a big small country
2) Gandalf is an old old man
3) That's fake fake news

To see why (1-3) pose a problem for theories of syntax based on structural uniformity, let us consider what kind of structural description they would assign to our strings. There has been remarkably little attention paid to iteration in generative grammar, although several analyses of the syntax of adjective stacking are available (see Alexiadou, 2014). Adjectives have been claimed to be adjuncts or specifiers to NP/DP (and therefore full phrases; see Jackendoff, 1977; Svenonius, 1994) or heads (Abney, 1987; Cinque, 1994), always with the focus set on the ordering restrictions between different semantic kinds of adjectives as well as their relative position with respect to the noun they modify (pre-nominal vs. post-nominal distribution).

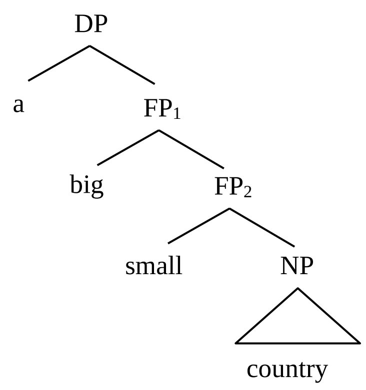
In Cartographic approaches, each adjective must head its own functional projection within the NP/DP; each modifier is placed in a unique head position (Cinque, 2010; Scott, 2002; Bortolotto, 2016) or as specifiers of functional phrases FP (Aljović, 2010). Cartography provides a rich set of functional projections where adjectives of different semantic classes are located. These functional categories are arranged in a strictly binary-branching {Head, XP} fashion, following Kayne (1994, 2018) and Chomsky (2013). An approximation to the functional hierarchy for adjectives is given in (4) (but see Scott, 2002: 114 for a refined version):

4) $[_{DP}\ D^{\circ}\ [Adj_{poss}\ [Adj_{card}\ [Adj_{ord}\ [Adj_{qual}\ [Adj_{size}\ [Adj_{shape}\ [Adj_{color}\ [Adj_{nation}\ [_{NP}\ N^{\circ}]]]]]]]]]]$

Scott (2002: 112) proposes that '*an adjective generated in the hierarchy relates directly to the semantic reading it receives*', and considers only readings of adjective iteration where the adjectives contribute distinct meanings: for instance, in *a good good typist*, he distinguishes the first *good* as pertaining to 'morality' and the second as pertaining to 'manner'; therefore, he assumed a stacked structure of the kind [$_{Subj.CommentP}$ *good* [$_{MannerP}$ *good* [$_{NP}$ *typist*]]]; similarly, in *an old old Etonian*, '*this fragment can only receive the interpretation an old (= in age) old (= former) Etonian*'. Scott's position is representative of the cartographic view, heavily influenced by the Chomsky-Kayne approach to phrase structure, which is committed to structural uniformity. It is, however, surprising that readings like 'a very good typist' or 'a very old Etonian' are not even considered[1].

Going back to our sentences, consider first (1). It only seems to allow for a reading in which one adjective has scope over the other, such that the corresponding interpretation is roughly 'a country that is big for a small country'. We are not concerned with fine-grained issues pertaining to the semantics of adjectives in this paper, but rather with the fact that a PSG would assign (1) a structural description in which, as pointed out above, one adjective c-commands the other, thus having scope over it (Ladusaw, 1980; May, 1985). Using Cinque's (1994) FPs as proxies for whatever non-terminal labels are assumed to be in play, the structural representation would look like (5):

5)

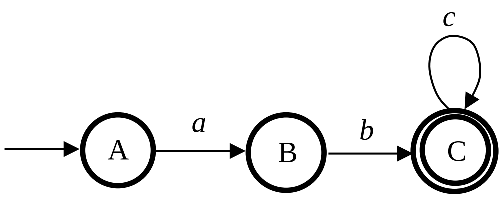



What kind of computational device can generate the structure in (5)? To a certain extent, the answer depends on the status of adjectives in X-bar theoretic terms: heads or specifiers. Recall that FS grammars allow for rules of the form A → *a*B, which means that it is possible to produce the tree in (5) by using only FS rules if adjectives are FP heads, since:

> *a given finite-state language L can be generated either by a psg* [Phrase Structure Grammar] *containing only left-linear rules: Z → aY, Z → a, or by a psg containing only right-linear rules: Z → Ya, Z → a, and a psg containing either only left-linear rules or only right-linear rules will generate a finite state language* (Greibach, 1965: 44)

In other words, the following are equivalent in strong generative power:

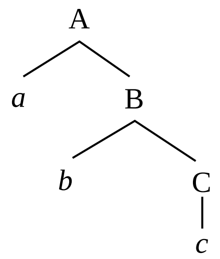
6) A → *a*B
B → *b*C
C → *c*




[1] Cinque (2010: 118) similarly neglects the possibility of intensive reduplication with adjective iteration:

> *If both* [adjectives] *are prenominal, with some degree of cumbersomeness (la vecchia vecchia bicicletta di Gianni 'Gianni's old old bicycle'), the first* vecchia *is* ***necessarily*** *interpreted as POSS-modifying and the second* vecchia *as N-modifying, just as in English.* [our highlighting]

$\delta(q_1, A, a) = (q_2, B)$
$\delta(q_2, B, b) = (q_3, C)$
$\delta(q_3, C, c) = (q_3, C)$




This is an important point: a phrase structure grammar that only allows for production rules involving a terminal and a non-terminal actually generates an FS language (see also Uriagereka, 2012: 53)[2]. If, however, adjectives are taken to be *specifiers* of FP (thus, maximal projections), then the grammar must contain also rules of the form A → BC, with A, B, C nonterminals. Such a grammar is strictly context-free. In either case, however, there is a point to be made: the semantic scope of *big* must be *small country*, in order to get the proper semantic interpretation. That means that, under more or less strong compositional assumptions (consider, e.g., Bach's 1976 rule-by-rule hypothesis[3]; or Jacobson's 2012 *direct compositionality*) the semantic contribution of *big* must apply to the output of a syntactic rule generating an expression *small country*. A *big small country* is *not* a country that is small and big.

The second case contrasts with the first in an interesting respect: the only possible interpretation for this case of adjective iteration is *intensive reduplication*; that is, 'a very old man'. Schmerling (2018: 3) observes that '*the semantic value of an NP with multiple occurrences of an adjective, say* old, *does not increase as the number of instances of* old *increases*.', and attributes this to the finite state syntax of total reduplication in these instances: in *$old_1$ $old_2$ man* it would be inaccurate to say that *$old_1$* takes as its input the semantic value of *$old_2$ man*; rather, *man* is modified by the semantic value of the sequence {*$old_1$ $old_2$*, …*$old_n$*}. The direct consequence of this reasoning is that a structural description

---

[2] A similar argument was used in Reich (1969: 835), who enriches FS devices with a set of Boolean connectives; he replaces rewrite grammars with circuit-like (Hamiltonian) transition graphs which instantiate strictly right-branching or strictly left-branching trees (but no symmetric bifurcation). Reich's conclusion about the 'location' of natural languages in the CH, however, is also strongly committed to structural uniformity:

> *English (and, I suspect, all languages) can, in fact, be described by a finite state device, namely a network of relationships, where each relationship itself a finite state device.* (Reich, 1969: 834)

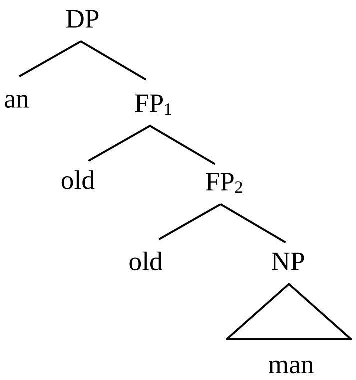
As observed in Shieber (1985), the view that natural languages are regular requires commitment to the idea that they are sets of finite strings, with embedding (and, more extremely, cross-serial dependencies) requiring fixed bounds.

[3] The fundamental tenets of the rule-by-rule hypothesis are the following:

> *a. For every syntactic rule, there is a unique translation rule specifying the translation of the output of the rule as a function of the translation(s) of the input(s).*
>
> *b. All rules apply strictly locally in a derivation, that is, no rule has access to earlier or later stages of a derivation.*
>
> *c. Syntactic rules apply to syntactic structures; translation rules to (already built-up) translations. Neither type of rule has access to the representations of the other type except at the point where a translation rule corresponding to a given syntactic rule is applied.*
>
> *d. The translations of the inputs to a syntactic rule must be 'intact' in the translation of the output (except possibly for changes in the variables, to avoid accidental binding of variables).* (Bach, 1976: 187)

like (5), as in (7) would be empirically inadequate insofar as it would be unable to adequately represent the relations between the adjectives and the semantic representation assigned to (2):

7)




In (7), one instance of *old* takes a constituent *old man* as its complement and takes scope over it, which does not adequately represent the semantics of the NP. This is unavoidable in a system set up as in Cartography or Merge-based Minimalism, given structural uniformity. This is not a completely novel point. The inadequacy of phrase structure representations for strings like (2) has been recognised since at least Chomsky (1963):

> *a constituent-structure grammar necessarily imposes too rich an analysis on sentences because of features inherent in the way P-markers are defined for such sentences.* (Chomsky & Miller, 1963: 297-298)

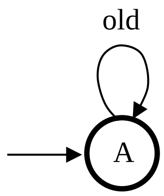
Following up on Chomsky & Miller (1963) and Postal (1964), Lasnik (2011) acknowledges the problem imposing 'too much structure' on structural descriptions for strings if a uniform 'moving up' in the Chomsky Hierarchy is assumed (that is: 'FSGs are inadequate for some substrings, then we proceed to CSGs; these also have limitations for some substrings, thus we go further up…'):

> *In a manner of speaking, what we really want to do is* ***move down*** *the* [Chomsky] *hierarchy. Finite-state Markov processes give flat objects, as they impose no structure. But* ***that is not quite the answer either****.* (Lasnik, 2011: 361. Our highlighting)

While it may be too strong to claim that FS representations 'impose no structure' (since an FS language may be generated by uniformly right- or left-branching phrase structure rules), it seems accurate to want to assign an FS representation to instances of intensive reduplication in English, most of all considering that, although increasing the number of iterations does not modify the semantic value of the adjective beyond perhaps intensifying the degree to which a property holds for an entity (such that *old old old old old* also means 'very old'[4]) it is indeed possible to multiply the occurrences of an adjective beyond two. In this context, what would an adequate structural description for (2) look like? We would represent the adjective iteration in (2) as a loop in an FSA:

[4] The meaning of intensive reduplication is reminiscent of the "rhetorical accent" identified in Newman (1946). The same reading could have been obtained by means of vowel lengthening: *An oooooooold man* ('an o·ld man', in Newman's notation). We thank Susan F. Schmerling for calling this reference to our attention.

8) A → *old* A

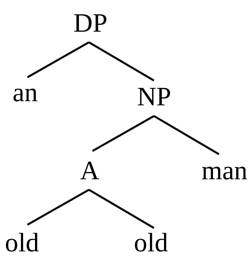



Take (8) to be a representation of a fragment of an FS recogniser: given a string, we ask whether that string belongs to the language defined by a regular grammar. Then, rule (8) means ‘if you are in state A and read input *old*, proceed to state A’. But this representation would be adequate only for the iterative part: whereas it is true that there is no hierarchical structure to be found among the adjectives (which is how we rephrase Lasnik’s ‘no structure’), there is a hierarchical relation between the adjectives and the noun such that the sequence of iterated adjectives modifies the noun. And here is where a pure FS structure results insufficient (as noted by Lasnik): if the sequence of adjectives has the properties of a non-terminal symbol (if it is an adjective phrase AP), then either as an NP specifier or adjunct there is at one derivational line a sequence of non-terminal symbols: AP and NP or N’. In either case, a structure that is beyond FS power. However, locally, a sub-string in (2), namely, the adjective iteration, has no internal hierarchical structure (that is, there is strict *parataxis*) according to any syntactic or semantic test, and is closed under Kleene star. Again, we cannot say that *all sequences of adjectives* have this property: it would be certainly inadequate to assign a ‘flat’ structure to (1). In this sense, Chomsky & Miller’s (1963: 298) proposed (but rejected) solution to the problem of ‘too much structure’ for adjectives in predicative position would also be inadequate as a rule of the grammar:

9) Predicate → $\text{Adj}^n$ and Adj ($n \geq 1$)

The reason for its inadequacy is not (only), as Chomsky & Miller claim, that there are ‘*many difficulties involved in formulating this notion so that descriptive adequacy may be maintained*’, but rather that the combination of (9) with the assumption of structural uniformity creates *empirical* (descriptive) problems. In strictly formal terms pertaining to the expressive power of the grammar, we must note that the intersection of a CF language with a FS language is always a CF language. However, for purposes of grammatical description, this may be taken to mean that the base component of a transformational grammar is *at most* CF, not that the structural description assigned to every single sub-string is strictly CF (since, as we have seen, that results in unnecessary additional structure). What would an alternative structural description look like? We may propose a diagram like (10), for illustrative purposes:

10)



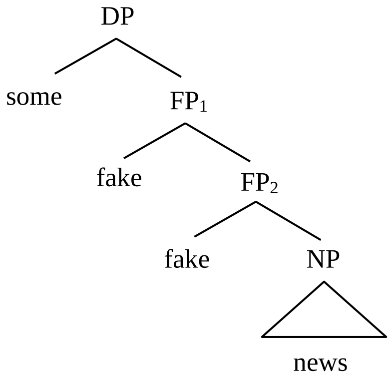

But our point is more than notational or graphical: however it is drawn, an empirically adequate structural representation for (2) simply cannot assign the sequence of reduplicated *old* internal hierarchical structure. What we claim is that *an old old man* is the result of combining a CF structure with a FS one in a grammatically relevant unit (an NP/DP), and that for purposes of grammatical analysis, it must be possible to express the distinction between these two sub-structures. We will come back to this point shortly.

We now turn to (3), *that's fake fake news*. The reason this example is particularly interesting is that in principle there are *two* possible readings, which correspond to two different segmentations:

11) a. [fake [fake news]] (i.e., truthful news, or at least not fake news)
    b. [fake fake [news]] (i.e., very fake news, obviously fake news)

This is a crucial case, since it makes it evident that a descriptively adequate grammar should be able to assign two distinct structural representations to (3) if it is to capture the semantics in any way. Consider (12), a Cinquean approach to adjective stacking:

12)




This representation illustrates an important point: *fake* in $FP_1$ has scope over a constituent *fake news*, whose label is $FP_2$. If the semantic representation is built in tandem with the syntactic structure (Bach, 1976; Jacobson, 2012), then this structure corresponds to segmentation (11a), and therefore with the meaning 'not fake news' (i.e., the semantic value of *fake* applied to the semantic value of *fake news*). However, because of precisely these reasons, (12) is entirely inadequate as a structural description for the reading (11b), which involves intensive reduplication. As argued above, intensive iteration (including lexical reduplication) should receive a finite state treatment. In this respect, the iteration of lexical items from a syntactic (supra-lexical) perspective seems to be computationally simpler than the kind of reduplication studied in morphology (Dolatian & Heinz, 2019) in the sense that no memory seems to be required. Thus, the expressive power of a FSA can indeed (strongly and weakly) generate iterative patterns of adjectives, adverbs, and other categories whose semantic value is intensified in iteration, as in (13)[5]:

13) a. The coffee was *very very* hot
    b. It's been a long *lonely, lonely, lonely, lonely, lonely* time (Led Zeppelin, *Rock and Roll*, 1971)

[5] Susan Schmerling (p.c.) has pointed out to us that intensive iteration as identified here is to be distinguished from the phenomenon illustrated in (i)-(ii) (the examples are Schmerling's):

i) John is older than many of his friends, but he certainly isn't *óld*-old.
ii) I don't want a director's chair; I want a *cháir*-chair.

As Schmerling observes, this kind of iteration is both prosodically and semantically distinct from the cases of iteration analysed here. To begin with, the reduplication pattern in (i)-(ii) is limited to two elements (cf. *'he certainly isn't *old old old*'). Also, the prosodic pattern is different, requiring prominence in the first instance of the reduplicated element. Semantically, in the cases noted by Schmerling we are dealing with a phrase that denotes a *prototypical* instance of the entity or property denoted by the lexical head.

Total reduplication, understood here as the iteration of a complete lexical form (as opposed to *partial reduplication*, where a morpheme is targeted within a lexical item), is amenable to finite-state modelling. In particular, for instance, Dolatian & Heinz (2019) propose a 2-way FSA which scans an input, letter by letter, and produces an output associated to each transition between states. These states allow for loops in such a way that the FSA can produce a reduplicated output. In our case, however, we are not interested in the reduplication of strings of letters *per se*, but rather on the iteration of basic expressions under the same non-terminal (Adj⁀Adj⁀Adj…): in this sense, the number of states needed is in principle only the number of indexed categories in the lexicon. The 'explosion' of states that occurs in FS analyses of partial reduplication in morpho-phonology (Chandlee, 2014; Dolatian & Heinz, 2019) need not happen in syntax if intensive reduplication is adequately restricted. This means, among other things, that there is no need to be able to recognise infinite sequences of iterated categories, since these do not occur in the input: we want the grammar to be able to assign strongly adequate structural descriptions to natural language strings, and in this evaluation, semantics play a crucial role[6]. Claiming that a language that allows for iteration is necessarily context-free is faced with the difficulty of distinguishing instances of iteration in which each occurrence of an expression E has scope over whatever appears at its right (e.g., a putative reading [*lonely* [*lonely* [*lonely* [*lonely* [*lonely* time]]]]]) from instances of iteration in which there is no hierarchical relation between the iterated expressions (e.g., [[*lonely lonely lonely lonely lonely*] time]]).

### 3. Towards a general solution

The picture that emerges from the previous section can be summarised as follows: if the goal of grammatical theory is to assign adequate structural descriptions to natural language strings, and if adequacy is defined in terms of semantic interpretation as well as structural relations (constituency, dependencies, etc.), then the axiom of structural uniformity that is prevalent in generative grammar conspires against the descriptive adequacy of the grammar. Assuming that all syntactic structure is created equal leads to the assignment of too much structure to certain constructions, among which we have identified some cases of adjective stacking and total reduplication (or, rather, iteration) in which semantic interpretation requires that either there be no hierarchical dependency between the iterated elements, or that there be, depending on specific properties of the terminals involved. Identifying what exactly in *fake* in (3) allows for both interpretations but not *old* in (2) is outside the scope of the present paper, but clearly it is a problem that needs to be addressed as part of an explanatory theory of the lexicon-syntax interface.

We have identified the computational complexity of these non-scopal iterated adjective constructions in English as finite state, in the sense that a formal system with no memory and only capable of recognising finite expressions can assign these substrings 'flat' structural descriptions (by virtue of not imposing extra structure in the form of non-terminals nodes) which, we argue, capture their syntactic and semantic properties. However descriptively adequate this approach seems to be, it does clash with the 'consensus' identified by Stabler with respect to the computational complexity of natural language grammars. Given the fact that we are not dealing with a problem of parsing, automatic speech recognition, or translation but of grammar, a *uniform* FS approach to syntax would simply not be even descriptively adequate (as observed by Lasnik). What to do in this context? The proposal in this paper is

[6] *A grammar G is weakly adequate for a string language L if L(G) = L. G is* ***strongly adequate*** *for L if L(G) = L and for each w in L, G assigns an* ***'appropriate' structural description*** *to w* (Joshi, 1985: 208)

that *structural uniformity* must be abandoned if descriptive adequacy is still seen as a goal of grammatical theory. In our view, this entails also abandoning the idea that there is a set of mechanisms for the production and assignment of structure which are completely independent from semantics.

At this point it is legitimate to ask whether, in natural languages, evaluation in terms of levels in the CH should not proceed in terms of *constructions* in specific languages rather than as universal generalisations. In other words: given a well-formed expression, each sub-expression which is also a well-formed expression is assigned the computationally simplest structural description which captures the semantic dependencies between its component parts[7].

The simplicity requirement serves the purpose, in this context, to avoid extra structure: assigning a context-free structural description to an expression for which a finite-state description suffices has consequences at both syntactic and semantic levels, in particular under direct compositionality assumptions: since additional structure is inserted in the form of non-terminal nodes, the possible targets for rules of the grammar multiply. In (14), neither example allows for an intensive reading, and whereas that is fine for *fake fake* (since there is an alternative structure available), *old old* becomes hard to interpret at all:

14) a. Fake is what the fake news was
b. ?Old is what the old man was

The crux of the issue is to define a way in which this may be accomplished.

As a first approximation to the matter, let us assume that the procedure in charge of assigning structural descriptions to natural language expressions contains at least two processes: *chunking* and *substitution/adjunction*. If chunking is sensitive to semantics, then we need to be able to capture the fact that in (2) we have essentially two chunks which display different kinds of structural relations:

15) a. A man
b. old old

The structural dependency assigned to *a* and *man* needs to capture the fact that the quantifier has scope over the noun. Because there is hierarchical structure between the two basic expressions that make up the derived expression (15a), using a term with long provenance in grammar, we will refer to this as *hypotaxis*. But the relation between both instances of *old* is of a different kind: it is strictly *paratactic*. The distinction between *hypotaxis* and *parataxis* is necessary, in this context, to account for relations of (a) modification (b) selection, and (c) iteration. These are theory-independent, so far as we can see.

The relations of *modification* and *selection* involve hierarchy, such that the modifier is hierarchically higher than the modified and the selector, higher than the selected. This interpretation of the rela-

---

[7] A similar desideratum can be found, for instance, in Lowe & Lovestrand's development of an LFG-inspired phrase structure model (2020: 3):

> *Utilize* ***only as much structure as required to model constituency****, avoiding nonbranching dominance chains.*

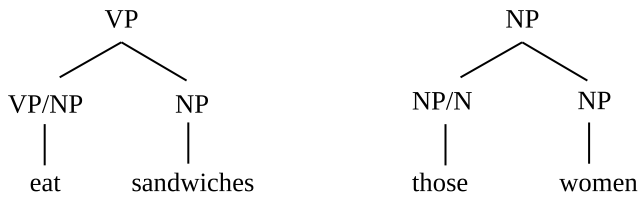
The main difference is that in this paper we attempt to avoid not only non-branching dominance chains, but also branching dominance chains which introduce unnecessary non-terminals.

tions is close to that of Dependency Grammars (Osborne, 2014) and Arc Pair Grammar and its successors (Postal, 2010). In semantic terms, if A is a modifier and B a modified, or A a predicate and B its argument, for ⟦A⟧ the semantic value of A and ⟦B⟧ the semantic value of B, then the relations are defined as follows:

16) a. Modification: ⟦A⟧(⟦B⟧)
b. Selection: ⟦A⟧(⟦B⟧)

The crucial point here is that the structural condition that the predicate be higher that its argument is common to both relations. The difference pertains to the definition of the categories involved: following Dowty (2003: 37), a modifier may be classified as an *adjunct* in a structure (A/A)/A, where A is an indexed category of the language; the format for *argumental* dependencies is (A/B)/B, where A and B are *distinct* indexed categories[8]. The system outlined by Dowty assigns the following analysis trees to modifier-modified and predicate-argument relations[9]:

17)




We can get to the rule from the category definitions by means of so-called *functional abstraction* (Dowty, 2012: 41): if we have an expression of category (A/…B)/B (i.e., a sequence of categories with B as its rightmost element), then removing B will result in an expression of category A/B. In rule format, what we have is

18) a. VP → (VP/NP)/NP
b. NP → (NP/N)/N

Without a rule of *quantifying-in*, a CG is equivalent to a CF PSG (Lewis, 1970: 20), which means that we are within the computational ballpark we are actually interested in, at the level of local syntactic objects. Let us consider the case of *fake fake news* in the sense ‘fake news that are not really fake news’: here, we would say that the first *fake* affects *fake news*. Now we can be more explicit: the semantic value of *fake* applies to the semantic value of *fake news*:

19) ⟦fake⟧(⟦fake news⟧)

And thus they need to be introduced in the derivation sequentially, under (some version of) the direct compositionality hypothesis: syntactic objects are interpreted as they are introduced. We will see how this is operationalised shortly.

The case of iteration is different: here there is no hierarchy between the iterated elements (thus our description of this pattern as *paratactic*). This is a structural scheme that does appear elsewhere in the grammar, as we will see shortly. Furthermore, the iterated elements do not change the category definition of the target of iteration. In other words: if *old* is an expression of category C (AP, NP/NP, etc.), then *old old* is also an expression of category C. The semantic interpretation rule has the same format,

[8] See Joshi & Kulick (1997) for a TAG approach to CG analysis trees.

[9] Presumably, in a system that accepts the existence of both DP languages (e.g., English) and NP languages (e.g., Turkish), the category definition of determiners and demonstratives would be DP/NP, and not NP/N.

of course, but now it is the semantic value of the predicate which gets intensified that has scope over the argument:

20) ⟦old old⟧(⟦man⟧)

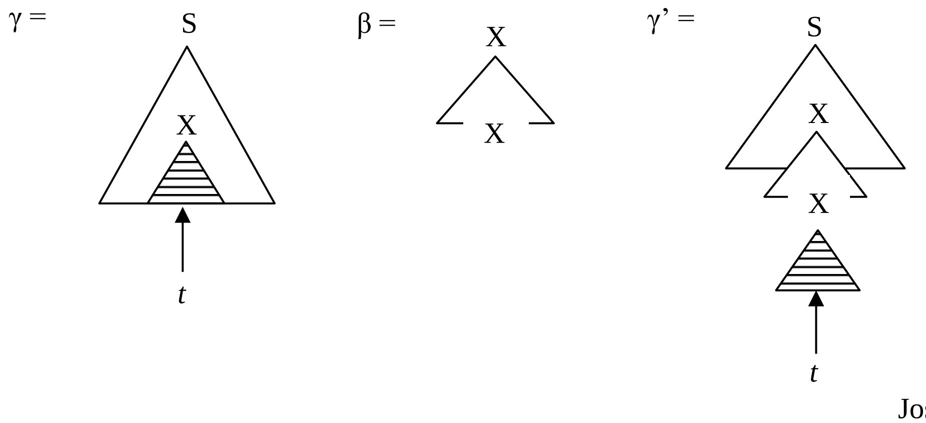
As argued above, the best way to represent the lack of hierarchical dependency between the instances of *old* is to see them as a loop on a single state (i.e., a transition from a state to itself).

Summarising so far: an adequate structural description for a string like *an old old man* combines two kinds of dependencies: iteration and modification. The former imposes no hierarchy on the expressions it contains (and is thus strictly regular), the latter does (and pushes the computational power of the grammar up; in the case under consideration here, to context-free power). If we require of a grammar to assign no more structure than strictly needed to represent semantic dependencies, then there are two distinct sub-structures in *an old old man*, as indicated in (15).

Locally, then, (2) displays both context-free and finite-state dependencies. But there must be a way to put both chunks together, otherwise, it would be impossible to build a compositional interpretation. We mentioned above that the grammar contains at least two mechanisms, *chunking* and *substitution/adjunction*. It is by means of *chunking* that expressions can be segmented into computationally uniform sub-strings. What we need now is to characterise an operation that can insert a chunk in a designated position within another chunk. This kind of operation is common in syntactic theory, from Chomsky's *generalised transformations* to Joshi's *adjunction*. Martín-Vide et al. (1996) distinguish formal grammars based on *rewriting* (Chomsky, 1959) from grammars based on *adjunction*. Here, then, we explore a version of the latter. The grammar contains two sets of elementary trees: *initial trees* and *auxiliary trees*. *Initial trees* are the target for *adjunction* of *auxiliary trees*, which results in a *derived tree*. Adjunction

> […] *composes an auxiliary tree β with a tree γ. Let γ be a tree with a node labelled X and let β be an auxiliary tree with the root labelled X also. (Note that γ must have, by definition, a node - and only one - labelled X on the frontier)* (Joshi, 1985)

In diagrammatic form,




Joshi (1985: 209)

In particular, our system has the following properties:

21) a. It is *lexicalised*
b. It rejects *structural uniformity*
c. It rejects the autonomy of syntax

The first property, *lexicalisation*, means that local structural units are defined around a single lexical head (Joshi & Schabes, 1991; XTAG group, 2001; Frank, 2013). Crucially, in addition to restricting the size of the elementary units in the grammar, (21a) can be interpreted in a stronger sense, as in Frank's 'Elementary TAG hypothesis':

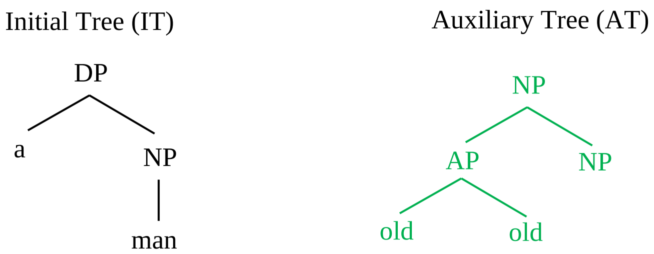
> *Every syntactic dependency is expressed locally within a single elementary tree* (Frank, 2013: 233)

The second, *rejection of structural uniformity*, means that these local units need not be generated by the same system of rules or display the same kind of structural and semantic dependencies: if we consider two local units α and β, α may display only *paratactic* dependencies and β only *hypotactic* dependencies. This represents a departure from at least some versions of TAG, where each elementary tree can be built via Merge (Frank, 2013: 240-241) and therefore the only dependencies possible are established in structurally uniform trees (binary-branching all the way down, endocentric, single-rooted). We do require, however, that dependencies within a single structural unit be computationally uniform.

Lastly, the *rejection of autonomy of syntax* in turn entails two things: on the one hand, that syntactic rules (whichever these are) operate over semantic material (see also McCawley, 1971: 285; Krivochen, 2019), and on the other, that the semantic properties of the lexical head of a local structural unit determines the rules that can apply to that unit.

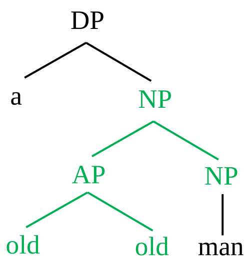
How does this help with our cases? Initially, in a lexicalised TAG the only way to separate elementary trees is to locate a single lexical head since structure is uniform all throughout (in elementary trees and derived trees). However, if we require that structural dependencies be uniform only within elementary trees, then we have a second criterion to define elementary trees: not only as the '*extended projection of a single lexical head*' (Frank, 2013: 239), but also as computationally uniform local units. This is precisely what allows us to get the segmentation in (15): *a man* is the extended projection of the lexical item *man*, and is a local *hypotactic* unit. The iterated adjective *old old* also contains a single lexical head, and only *paratactic* dependencies. The structural description assigned to *an old old man* in our terms, then, should contain the following elementary trees:

22)



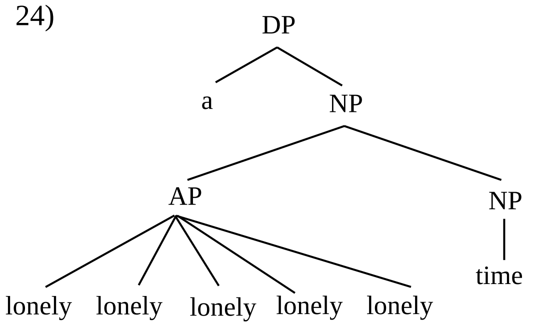

After *adjunction*, the derived tree would be (23):

23)




Of course, the number of iterated elements can be greater than two, as in the line from *Rock and Roll* cited above:

24)




A direct consequence of our proposal is that the two readings for *fake fake news* should receive two distinct structural descriptions, which we take to be a good thing. In a way, the framework advanced in this paper follows rather closely Dowty's (2007: 30) criteria:

> ***Compositional transparency****: the degree to which the compositional semantic interpretation of natural language is readily apparent (obvious, simple, easy to compute) from its syntactic structure.*
>
> ***Syntactic economy****: the degree to which the syntactic structures of natural language are no more complicated than they need to be to produce compositionally the semantic interpretation that they have.*

The specific way in which *syntactic economy* is interpreted here is precisely defined in relation to the CH: in this way, binary branching is not always the 'simplest' possible construal for the analysis of local natural language substrings (cf. Kayne, 1994; Chomsky, 2013). The simplest kind of syntactic construal is a loop on a single state:

25) A → *a*A

This produces a sequence of *a*'s in which there is no hypotaxis by virtue of their being dominated by the same mother label. More concretely:

25') AP → *lonely* AP

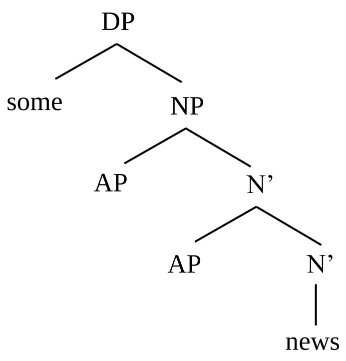
This is the pattern for iteration.

When modification is required, then (as seen above) some additional structure is needed to capture *hypotaxis*, and consequently complexity increases with respect to the *paratactic* case. If we want to capture the non-iterative reading for *fake fake news*, then we need to assign the string a structural description in which *fake* modifies *fake news*. A context-free grammar must then contain, minimally, the following rules:

26) DP → D NP
NP → AP N'
AP → A
N' → AP, N'
N' → N

Which generate the structural description in (27):

27)

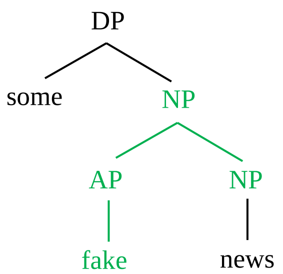



Importantly, the set of rules in (26) is *not* lexicalised, insofar as there are rules that are not lexically anchored (XTAG group, 2001: 5-6). The lexicalised elementary trees for *some fake fake news* would then be as in (28):

28) Initial Tree (IT) Auxiliary Tree (AT)

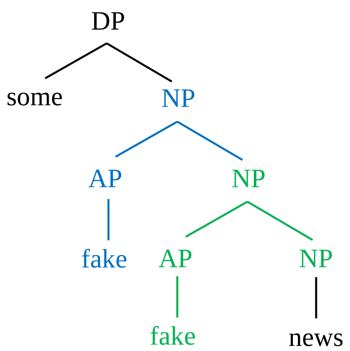

A directly compositional approach, given these elementary trees, delivers the correct reading if adjunction applies step-wise. Let us illustrate the derivation. The interpretation of the initial tree is the extended projection of its lexical 'anchor': *news*. Oversimplifying the issue, let us refer to that as ⟦news⟧. The first application of adjunction produces the derived tree in (29):

29)




At this point, the interpretation is the semantic value of *fake* applied to the semantic value of *news*: ⟦fake⟧(⟦news⟧). But the auxiliary tree gets adjoined once again, yielding (30):

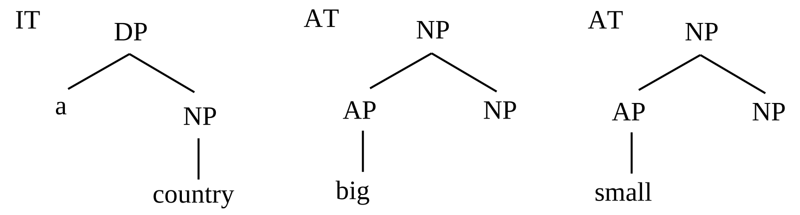
30)



The interpretation at this point is precisely what we have indicated in (19): ⟦fake⟧(⟦fake news⟧).

The case of *small big country* is actually exactly parallel to the scopal interpretation of *fake fake news*, with the only difference that we have *two* auxiliary trees (AT) instead of one:

31)



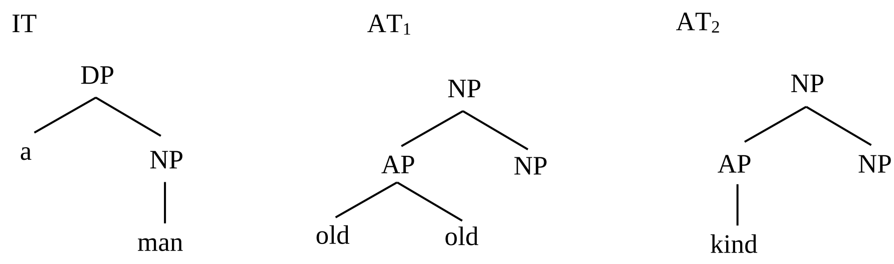

The order in which the auxiliary trees get adjoined gives two distinct semantic interpretations (which correspond to two distinct derivations): *big small country* (a country that is big for a small country) and *small big country* (a country that is small for a big country).

It is possible also to combine these dependencies within a single NP. Consider, for instance, (32):

32) A kind, old old man

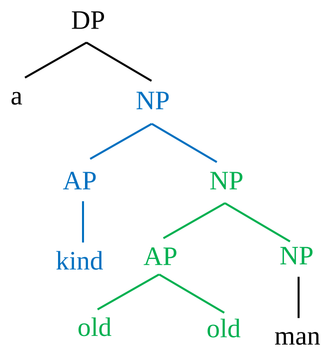
The interpretation of (32) goes along the lines of 'a very old man who was also kind': *kind* has scope over *old old man*, and *old old* has scope over *man*, with neither instance of *old* having scope over the other (as already established). The elementary trees are those in (33):

33)




Adjunction proceeds as follows: first, $AT_1$ is adjoined, yielding *an old old man* (exactly as in (23)). At this point, the semantic interpretation is ⟦old old⟧(⟦man⟧), as above. But then, $AT_2$ is adjoined above $AT_1$. The semantic value of *kind* now has scope over the semantic value of the target of adjunction: ⟦kind⟧(⟦old old⟧(⟦man⟧)). The full derived tree is:

34)




The extra structure introduced by adjunction of *kind* seems to be justified; if we attempt the same test as in (14), we see that *kind* may be clefted, just like *fake* in the non-iterative reading:

35) Kind is what the old old man was

To summarise, a directly compositional lexicalised TAG approach, in addition to the rejection of the axiom of structural uniformity seems to pay off in the empirical analysis of simple cases, yet challenging for many current models of syntactic structure. The strictly CF dependencies that models like Cartography and Minimalism attempt to capture are indeed accounted for, in addition to iterative patterns that may receive a simpler structural analysis, in terms of FS loops.

## 4. Iteration and coordination

The system sketched in the previous section provides adequate structural descriptions for English sentences containing adjectival iteration, for which a CFG results too powerful in the sense that it assigns too much structure which is not justified in semantic terms or syntactic terms. In the present view, iteration is best handled by a set of FS rules. Can we extend the mechanisms proposed above to other iterative phenomena in the grammar of English?

Consider examples (36) and (37)[10]:

36) The Trump Twitter Archive shows the Republican whining about his predecessor's golfing *over and over and over and over and over* again.
--Steve Bonen, 'The problem with Trump's defense of his many golf outings', MSNBC, 13 July 2020

37) She would *wait and wait and wait and wait* / For her steady date
--Robert Byrd, 'Over and Over', (c) Paino Leonted Desert Music O/b/o Leon Rene Family Partnership, 1958

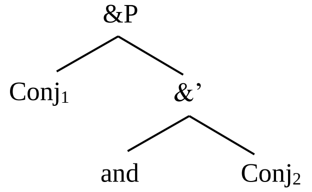
Superficially, sentences like (36) and (37) seem to contain *n*-ary coordinated structures. As observed by De Vos (2005), these structures (which he dubs *repetitive pseudo-coordinations*, ReCo) may yield serial and repetitive readings, but the fact that they are compatible with states (e.g., *wait*) suggests that repetition of an event is not a necessary condition. Even with dynamic verbs, repetitive readings are not guaranteed:

38) Caesar's legions *marched and marched* for days (De Vos' ex. (4))

In (38) there is only one event of marching, which extends for a long period of time. De Vos claims, correctly in our opinion, that a plural subject licenses (but not necessarily coerces) an iterative reading, whereas a singular subject allows for an intensive reading.

What kind of structural description is adequate for examples like (36-38)? In a structurally uniform system like MGG, coordination also complies with the axioms of X-bar theory, thus resulting in structures that always look like (39):

39)



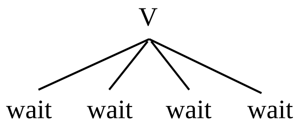

Analyses of this form are defended in Kayne (1994), Zoerner (1995), Progovac (1998), Chomsky (2013), among many others. In some versions of this hypothesis the first conjunct is adjoined to a

[10] We owe examples (36) and (37) to Susan F. Schmerling, as well as much discussion about their analysis.

ConjP formed by the coordinating conjunction and the second conjunct, in others, conjuncts are base-generated as Specifier and Complement of &/Conj, and in others the first conjunct moves from an internal position, in a sort of 'predicate first' analysis (that is: and[X, Y] → [X [and *t* Y]], e.g. Chomsky, 2013). Borsley (2004) provides an excellent overview of arguments against a structure like (39), and observes that in fact such an analysis is *not* assumed widely outside (present-day) MGG. There are two main problematic aspects of a template like (39): (i) structural uniformity (binary-branching and obligatory hypotaxis) and (ii) projection of the coordinating conjunction to phrasal level (which means that if we coordinate NPs, the result will not be an NP, but a ConjP / &P).

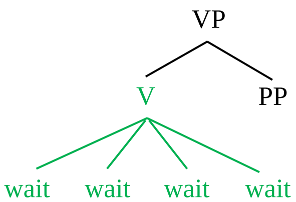
The proposal we want to make with respect to (36-39) is that (a) they are *not* instances of coordination at all (cf. De Vos, 2005: 4); rather, the iterated expression constitutes a terminal for all subsequent intents and purposes, just like *old old old…*. Furthermore, at least in the cases treated here, a very similar semantic effect arises: the semantic value of the iterated expression intensifies. Thus, (36) is not interpreted as literally referring to five events of whining, but rather expresses that whining took place frequently (in an unspecified number); in (37) there are not four events of waiting but only one, which extends for a long time[11]. These cases of iteration, like adjective iteration, are not restricted to a certain number of elements, nor is there anything *inside* the iteration that requires memory storage. Thus, we may assign (36-38) the same *flat* (FS) structure that we assigned to iterated adjectives:

40)




Let us comment on two features of this representation: (i) the lack of any *and* in the structure and (ii) the use of the label V (as opposed to VP). Recall that, in our view, there is no coordination in De Vos' ReCo: the word *and* makes no contribution to the syntax or the semantics of the construction. Just like *and* in cases like *Which dresses is she going to go/up/take and ruin now?* (Ross, 1967), the presence of the word *and* does not mean that we are in the presence of a coordinated structure. Each of the leaves in (40) could be *and wait*, but in this context its grammatical properties would not change. Which takes us to the second point: labelling (40) as V. This entails that a structure with ReCo would require *substitution* (an operation that is at least CF, depending on whether it targets intermediate nodes or not) of V in the frontier of an initial tree by the root of the tree in (40) (obtained by an FS loop).

41)




Note that the PP is *outside* the iterated structure. This makes two predictions: (i) there is no structural position available for an internal argument inside the auxiliary tree (40), and (ii) movement of an NP internal argument in a parallel structure with a transitive verb does not violate the Coordinate Structure

[11] In this case, *wait* may receive rhetorical accent: *wa·it* (Newman, 1946)

Constraint. Note that, if ReCo was an instance of coordination involving Vs or VPs, then NP movement would require chopping the NP from one conjunct. We can rule out the CSC violation with examples like (42):

42) a. Six web series that we can *watch and watch and watch*…https://economictimes.indiatimes.com/magazines/panache/2016-all-oer-again-six-web-series-that-we-can-watch-and-watch-and-watch/articleshow/55838344.cms. Relativisation, A'-movement.
b. It was *said, and said, and said, and said*, and that's why I say it (Tom Mould, *Choctaw Prophecy: A Legacy for the Future*, p. 29). Passivisation, A-movement

Note also, in the case of (42b), that there is only one auxiliary *be*, yet all instances of *say* appear in participial form; in present terms, passive *be* determines that the expressions dominated by V will bear participial morphology; all that V dominates (be it a single word or an iterated word) will be a participle. Finding examples with *wh*-interrogatives is trickier, probably because of the intensive value of iteration: the exchange in (43a) sounds more natural than (43b) (but (43b) is not ungrammatical):

43) a. -What did she say?
-She *said and said and said and said* that I should clean my room

b. -#What did she *say and say and say*?
-She said that I should clean my room

It seems that the finite-state approach to iteration sketched in the previous sections can be extended to ReCo fruitfully.

*4.2 Symmetric coordination and flat structures*

Consider the distinction between *symmetric* and *asymmetric* coordination (Schmerling, 1975). As observed by Schmerling, there are cases of conjunction in which the following equivalence holds:

44) $p \wedge q \equiv q \wedge p$

These are instances of *symmetric coordination*, and can be illustrated by the pair in (45)[12]:

45) a. Paris is the capital of France, and Rome is the capital of Italy.
b. Rome is the capital of Italy, and Paris is the capital of France.

This is the conjunction available in propositional logic, but as is frequently pointed out in introductory textbooks in logic, $\wedge$ and the word *and* are not equivalent, since natural language *and* is not always commutative (Hedman, 2004: 1-2; Allwood et al., 1977: 33, ff.). Thus, commutativity is *not* a property of the coordination in (46):

46) a. Harry stood up and objected to the proposal.
b. Harry objected to the proposal and stood up.

In (46) there is a temporal and causal ordering of the events, such that the situation described by (46a) is not the same as that described by (46b). In the case of (46) the permutations, while semantically different, are both grammatical. However, it is easy to come up with examples in which not even this weaker condition holds:

47) a. Smile and the world smiles with you

[12] Examples (45), (46), and (47) are taken from Schmerling (1975: 211)

b. *The world smiles with you and smile

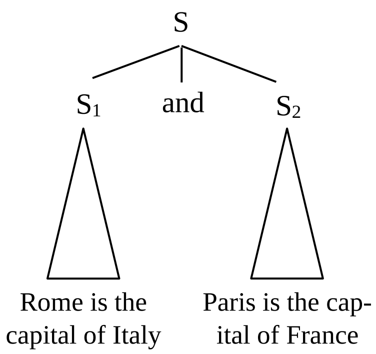
Coordinations of this sort are usually referred to as *asymmetric*. This much seems evident. Less evident, however, is the question of whether *symmetric* and *asymmetric* coordination should receive different structural analyses. From the perspective of MGG the answer is clearly 'no', even though it is not clear how to account for the distributional and semantic differences between them. Similarly, McCawley (1998: Chapter 9) assumes a flat structure for all coordinations (although his system is flexible enough to accommodate various segmentations for the same string with different interpretations, as in [John and [Peter and Bill]] and [John and Peter [and Bill]]). Our view is that the semantic differences between *symmetric* and *asymmetric* conjunction need to be accounted for in the structural descriptions assigned to them. In particular, we will focus on the structure of symmetric conjunction.

As before, one of the crucial issues is *iteration*. Consider (48):

48) Rome is the capital of Italy, Paris is the capital of France, and Berlin is the capital of Germany

The treatment of coordinated structures with more than two conjuncts is not straightforward: Ross (1967), McCawley (1998) and others would assign (48) a ternary branching structure; such a structure would be unavailable in X-bar theory or Merge-based formalisms (Kayne, 1994; Johannessen, 1998). Borsley (2004) provides a series of arguments (based on the distribution of *both* and the availability of *gapping*) that in a coordination like *Tom ate a hamburger, and Alice drank a Martini and Jane a beer* neither the first and the second nor the second and the third conjuncts form a constituent; his arguments extend to all cases of conjunction of *n* terms, for *n* > 2. The argument we would like to put forth here is that symmetric coordination of *n* terms behaves just like iteration, in the sense that it needs to be assigned a 'flat' structure in which there is no hierarchy between the conjuncts (thus, *parataxis*); asymmetric coordination does require there to be hierarchy between the conjuncts, in the form of a *hypotactic* structural description. Symmetric conjunction defines an *unordered* set of terms (NPs, Ss…) which presents the property of *commutativity*; asymmetric conjunction defines an *ordered* set of terms, which do not commute. The present proposal, then, answers the question posed in Abeillé (2003: 6) of whether the structure of coordinations is hierarchical or flat in a novel way: it depends on the semantic relations between the conjuncts. Thus, if we consider the semantic interpretation assigned to (45), under direct compositionality, the equivalence between $p \wedge q$ and $q \wedge p$ only holds if the terms of the coordination are introduced all at the same time (i.e., by the application of a single rule). The structure would then look like (49):

49)




Note that there is no hierarchical relation between $S_1$ and $S_2$, the structure is *paratactic*. In this manner, more conjuncts can be added to the symmetric structure (as in (48)) simply by adding more branches. Importantly, (49) differs from the structure assigned to iteration in that the latter may be modelled as a finite-state loop, but the former requires context-free power, since we have a non-terminal dominating a string of non-terminals and a terminal: the rule format of CF grammars in Greibach-normal form. However, it also means that an X-bar template (like (39)) would again assign too much structure.

## 5. Some conclusions

In this paper we examined some empirical consequences of assuming structural uniformity in natural language syntax; its counterpart in formal language theoretic terms takes the form 'natural languages are…' (FS/CF/mildly-CS, depending on the author). Structural uniformity in syntax is procrustean, assigning too much structure in some cases and too little in others; we illustrated the problem of 'too much structure' with English attributive adjective iteration. We argued that a descriptively adequate theory of the grammar for natural languages must take into consideration how syntactic configuration compositionally specifies semantic interpretations; an adequate theory of natural language syntax cannot be independent of semantics. In proposing that syntactic structure is not uniform, we are also forced to make explicit the kinds of dependencies that we find within local domains; here formal language theory proves an invaluable tool. An exploration of the consequences of our view, which we have called 'mixed computation', leads to interesting proposals concerning the nature of local domains in syntax: cycles are defined as chunks of structure with a single lexical anchor and uniform dependencies which allow to be targeted by composition operations (*adjunction* and *substitution*) and rules of semantic interpretation. *Grosso modo*, change the dependencies, and you change the cycle; change the lexical anchor, and you change the cycle (by virtue of changing the elementary tree). The dependencies include hypotaxis and parataxis, as indicated above, but also within hypotaxis we need to be able to refine the system: the descriptive adequacy of center embedding is not the same as tail recursion, or crossing dependencies. Let us close by briefly considering the case of sentential complementation. The same propositional content may adopt a variety of forms in different languages, for example:

50) a. Jan Piet Marie zag helpen zwemmen (Dutch German)
*Jan Piet Marie saw help swim*
'Jan saw Piet help Mary swim'

b. Merve Ömer'in Esra'nın yüzmesine yardım ettiğini gördü (Turkish)
*Merve Ömer Esra swim help give saw*
'Merve saw Ömer help Esra swim'

c. John saw Peter help Mary swim (English)

If we consider the relations between NPs and the VPs to which they correspond, we obtain the following abstract format for (51a-c):

51) a. $NP_1$ $NP_2$ $NP_3$ $VP_1$ $VP_2$ $VP_3$ (crossing dependencies between two sets of elements)
b. $NP_1$ $NP_2$ $NP_3$ $VP_3$ $VP_2$ $VP_1$ (center embedding)
c. $NP_1$ $VP_1$ $NP_2$ $VP_2$ $NP_3$ $VP_3$ (tail recursion)

The differences in structure between the sentential complementation patterns of different languages suggest that, again, a universal template may not be the best aid for grammatical analysis. After all, if the English example can be generated using only *substitution*, why assume anything more complex? If English and Turkish are both well within CF territory, do we really need to have CS machinery available in the grammatical descriptions for these languages? Would it not be the equivalent of using a formal cannon to kill a computational mosquito? We do not aim at settling the question here, but merely want to pose it as a fundamentally empirical question in what pertains to the relation between FLT and grammar.

**References**


Abeillé, A. (2003) A lexicalist and construction-based approach to coordinations. In S. Müller (ed.), *Proceedings of the HPSG03 Conference* (pp. 5-25). Stanford: CSLI.

Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. PhD thesis, MIT.

Alexiadou, A. (2014). *The syntax of adjectives*. In A. Carnie, Y. Sato & D. Siddiqi. (eds.), *Routledge Handbook of Syntax* (pp. 89-107). London: Routledge.

Aljović, N. (2010). Syntactic positions of attributive adjectives. In P. Cabredo Hofherr & O. Matushansky (eds.), *Adjectives: Formal analyses on syntax and semantics* (pp. 29-52). Amsterdam: John Benjamins.

Allwood, J., Andersson, L-G. & Dahl, Ö. (1977). *Logic in Linguistics.* Oxford: OUP.

Bach, E. (1976). An Extension of Classical Transformational Grammar. In *Problems of Linguistic Metatheory* (pp. 183-224). Michigan State University.

Bortolotto, L. (2016). *The Syntax of Relational Adjectives in Romance: a Cartographic Approach*. PhD thesis, University of Venice.

Chandlee, J. (2014) Strictly Local Phonological Processes. PhD thesis, University of Delaware.

Chomsky, N. (1959). On Certain Formal Properties of Grammars. *Information and Control* 2, 137-167.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1986). *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, N. (2013). Problems of Projection. *Lingua* 130, 33-49.

Chomsky, N. (2015). *The Minimalist Program* [2nd Edition]. Cambridge, Mass.: MIT Press.

Chomsky, N. & Miller, G. (1963). Introduction to the Formal Analysis of Natural Languages. In R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology* (pp. 269–321). New York: Wiley & Sons.

Cinque, G. (1994). On the Evidence for Partial N-Movement in the Romance DP. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini (eds.), *Paths towards Universal Grammar* (pp. 85‑110). Washington, D.C.: Georgetown University Press.

Cinque, G. (2010). *The Syntax of Adjectives. A Comparative Study*. Cambridge, Mass.: MIT Press.

De Vos, M. (2005). *The syntax of pseudo-coordination in English and Afrikaans*. PhD thesis, Leiden University.

Dolatian, H. & Heinz, J. (2019). Learning reduplication with 2-way finite-state transducers. *Proceedings of the 14th International Conference on Grammatical Inference*. 67-80.

Dowty, D. (2003). The Dual Analysis of Adjuncts/Complements in Categorial Grammar. In E. Lang, C. Maienborn, & C. Fabricius-Hansen (eds.), *Modifying Adjuncts* (pp. 33–66). Berlin: Mouton de Gruyter.

Dowty, D. (2012). Compositionality as an empirical problem. In C. Baker & P. Jacobson (eds.), *Direct Compositionality* (pp. 23-101). Oxford: OUP.

Emonds, J. (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.

Epstein, R. (2011). *Classical mathematical logic*. Princeton: Princeton University Press.

Frank, R. (2013). Tree adjoining grammar. In M. den Dikken (ed.) *The Cambridge Handbook of Generative Syntax* (pp. 226-261). Cambridge: CUP.

Fukui, N. & Narita, H. (2014). Merge, labelling, and projection. In A. Carnie, Y. Sato & D. Siddiqi (eds.) *The Routledge Handbook of Syntax* (pp. 3-23). London: Routledge.

Greibach, S. (1965). A New Normal-Form Theorem for Context-Free Phrase Structure Grammars. *Journal of the ACM* 12(1), 42-52.

Hedman, S. (2004). *A first course in logic*. Oxford: OUP.

Jacobson, P. (2012). Direct Compositionality. In W. Hinzen, E. Machery & M. Werning (eds.) *The Oxford Handbook of Compositionality* (pp. 109-129). Oxford: OUP.

Jackendoff, R. (1977). *X-bar syntax: A study of phrase structure*. Cambridge, Mass.: MIT Press.

Johannessen, J. (1998). *Coordination*. Oxford: OUP.

Joshi, A. K. (1985). Tree adjoining grammars. In D. Dowty, L. Karttunen & A. Zwicky (eds.) *Natural Language Parsing* (pp. 206-250). Cambridge, Mass.: CUP.

Joshi, A. K. & Kulick, S. (1997). Partial Proof Trees as Building Blocks for a Categorial Grammar. *Linguistics and Philosophy* 20(6). 637-667.

Joshi, A. K. & Schabes, Y. (1991). Tree-Adjoining Grammars and Lexicalized Grammars. Technical Reports (CIS). Paper 445. http://repository.upenn.edu/cis_reports/445.

Kayne, R. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris.

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

Kayne, R. (2018). The place of linear order in the language faculty. Talk delivered at the University of Venice, June 16th, 2018. https://as.nyu.edu/content/dam/nyu-as/linguistics/documents/Kayne%200118%20Venice%20The%20Place%20of%20Linear%20Order%20in%20the%20Language%20Faculty.pdf

Kornai, A. & Pullum, G. K. (1990). The X-bar theory of phrase structure. *Language* 66. 24-50.

Krivochen, D. G. (2019). On trans-derivational operations: Generative Semantics and Tree Adjoining Grammar. *Language Sciences* 74. 47-76.

Ladusaw, W. (1980). *Polarity sensitivity as inherent scope relations*. Bloomington: Indiana University Linguistics Club.

Langendoen, D. T. & P. M. Postal. (1984). *The Vastness of Natural Languages*. Oxford: Blackwell.

Lasnik, H. (2011). What Kind of Computing Device is the Human Language Faculty? In A-M. Di Sciullo & C. Boeckx (eds.) *The Biolinguistic Enterprise* (pp. 354-365). Oxford: OUP.

Lowe, J. & Lovestrand, J. (2020). Minimal phrase structure: a new formalized theory of phrase structure. *Journal of Language Modelling* 8(1), 1–52.

Newman, S. (1946). On the stress system of English. *Word* 2. 171–187.

Martin-Vide, C., Paun, G. & Salomaa, A. (1996). Characterizations of RE Languages by Means of Insertion Grammars. Technical Report. Turku Centre for Computer Science.

May, R. (1985). *Logical Form: Its Structure and Derivation*. Cambridge, Mass.: MIT Press.

McCawley, J. D. (1968). Concerning the base component of a transformational grammar. *Foundations of Language* 4, 243-269.

McCawley, J. D. (1971). Interpretative semantics meets Frankenstein. *Foundations of Language* 7, 285-296.

McCawley, J. D. (1998). *The Syntactic Phenomena of English*. Chicago: University of Chicago Press.

Open Logic Project. 2020. *The Open Logic Text*. https://openlogicproject.org/

Osborne, T. (2014). Dependency grammar. In A. Carnie, Y. Sato & D. Siddiqi. (eds.) *Routledge Handbook of Syntax* (pp. 604-626). London: Routledge.

Postal, P. M. (1964). *Constituent Structure*. Bloomington, Indiana: University of Bloomington.

Postal, P. M. (2010). *Edge-Based Clausal Syntax*. Cambridge, Mass.: MIT Press.

Progovac, L. (1998). Structure for coordination. *Glot International* 3(7), 3-6.

Pullum, G. K. (1984). Chomsky on The Enterprise. *Natural Language & Linguistic Theory* 2(3), 349-355.

Reich, P. (1969). The finiteness of natural languages. *Language* 45(4), 831-843.

Ross, J. R. (1967). *Constraints on Variables in Syntax*. PhD Thesis, MIT.

Schmerling, S. (1975). Asymmetric Conjunction and rules of Conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 211-231. New York: Academic Press.

Schmerling, S. (2018). Rhetorical meaning. *Linguistic Frontiers* 1(1). 1-8.

Scott, G-J. (2002). Stacked adjectival modification and the structure of nominal phrases. In G. Cinque (ed.) *Functional structure in DP and IP*. (pp. 91-120). Oxford: OUP.

Shieber, S. (1985). Evidence against the Context-Freeness of Natural Language. *Linguistics and Philosophy* 8(3), 333-343.

Stabler, E. (1997). Derivational Minimalism. In C. Retoré (ed.), *Logical Aspects of Computational Linguistics* (pp. 68-95). New York: Springer.

Stabler, E. (2013). The epicenter of linguistic behavior. In M. Sanz, I. Laka & M. Tanenhaus, (eds.), *Language Down the Garden Path*. Oxford: OUP. 316-323

Steedman, M. (2019). Combinatory Categorial Grammar. In A. Kertész, E. Moravcsik & C. Rákosi (eds.), *Current Approaches to Syntax A Comparative Handbook* (pp. 389–420). Berlin: Mouton de Gruyter.

Stowell, T. (1981). *Origins of Phrase Structure*. PhD Thesis, MIT.

Svenonius, P. (1994). On the structural location of the attributive adjective. *Proceedings of the West Coast Conference on Formal Linguistics* 12, 439–454.

Uriagereka, J. (2012). *Spell-Out and the Minimalist Program*. Oxford: OUP.

XTAG group. (2001). A lexicalized TAG for English. Technical report, University of Pennsylvania. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1020&context=ircs_reports

Zoerner, E. (1995). *Coordination: The syntax of &P*. PhD thesis. University of California, Irvine.