

Alternative Minimalist Visions of Language

Ray Jackendoff
Brandeis University

Advertisement: Peter Culicover and Ray Jackendoff, *Simpler Syntax*, Oxford University Press, to appear, 2005.

Primary Goal of contemporary linguistic theory: Explanation of how child attains adult competence in language: ability to use rich combinatorial system creatively.

Descriptive Constraint: Theory must account for adult competence in language: As competence proves to be more complex, there is more to explain.

One strategy: Minimize what has to be learned, in two ways:

1. Find ways to minimize complexity in adult grammar, consistent with facts.
2. Pack as much complexity as possible into innate Faculty of Language (Universal Grammar), so that child has less to learn. (This is strategy of Principles & Parameters.)

Evolutionary Constraint, in tension with 2: Minimize what had to evolve to get human ability to attain linguistic competence from apes. In particular, minimize what's special for language (Narrow Faculty of Language, "language instinct"). However, anything that serves other purposes is "free" (e.g. audition, theory of mind, pointing, perhaps imitation: Broad Faculty of Language). Evolutionary Constraint gives "minimalist inquiry" its empirical bite.

Hypothesis of Minimalist Program (Chomsky 1995): The Narrow Language Faculty is "perfect", in the sense that it satisfies the Descriptive Constraint (mapping between sound and meaning) with the absolute minimum of machinery; complexity arises only by virtue of interactions with independent properties of sound and meaning (i.e. Broad Language Faculty). (A recantation?)

Another strategy for satisfying Descriptive Constraint: Find ways to formulate complexity in adult grammar so that more of it can be learned.

Substrategy: Minimize elements of linguistic structure that child cannot infer from overt form: null elements (e.g. PRO), covert syntactic structure (e.g. Deep Structure, LF), and movement that transforms covert syntax into surface form.

Astonished question: How is it possible to do without null elements & movement?
Beginning of answer: Main motivation for these things (which have been taken for granted in syntax since 1960s) is to have a level of syntax that is homomorphic to semantics ("Deep Structure/Logical Form determines/encodes meaning"). All the mismatches are in the relation of covert syntactic form to surface form – "Interface Uniformity."

Alternative: Encode the mismatches directly in the relation of meaning to surface form. E.g. instead of mapping active and passive from single underlying syntactic form, in turn mapped to semantics, view active and passive as alternative ways of mapping directly between meaning and surface syntax.

Question: Isn't this in effect the same thing?

Answer: You have to encode the same mismatches either way. But look at resulting architecture:

Transformational grammar/Principles and Parameters/Minimalist Program:
meaning ← simple mapping → covert syntax ← complex mapping → surface syntax

Direct mapping:
meaning ← complex mapping → surface syntax

So from point of view of minimality in number of components, direct mapping is a priori superior. Feasibility of direct mapping is not a matter of speculation: HPSG and other similar "monostatal" approaches have worked out such approaches in detail (though not particularly with an eye toward Primary Goal or Evolutionary Constraint).

Is direct mapping empirically superior as well? Look at two well-known phenomena.

Passive:

Movement theory of passive commits you to having an NP position to which underlying object moves.

Passive VPs without such a position:

- (1) a. Dick had John followed by the FBI.
- b. The man followed by the FBI is my brother.
- c. My brother heard insults shouted at him by the cops.
- d. Followed day after day by the FBI, John went slowly nuts.

The movement theory can be salvaged by positing a null (or deleted NP) that has undergone movement, but that makes the theory less than minimal: adult must know more about syntax.

Theory of passive in terms of syntactic movement commits you to its being semantically blind.

Passives where lexical semantics makes a difference:

- (2) a. The bed was slept in/on/*under/*beside by John.
- b. The telescope was looked through/*inside by the technician.

The movement theory can be salvaged by appeal to "reanalysis" (*slept in* "counts as" verb), but then "reanalysis" is extra complexity in syntax, and then it can't be semantically blind. Technical difficulties with reanalysis: *They were taken considerable advantage of. She was spoken clearly to.* etc.

Long-distance dependencies:

Movement theory of long-distance dependencies commits you to a position to which a phrase moves. Plausible in (3):

(3) Which man did Fred see (t)?

But what has moved in the following cases?

- (4) a. Zero tensed relatives: the man [I love (t)]
- b. for-to relatives: the man [for you to vote for (t)]
- c. Comparatives: The book is longer [than I thought it was (t)]

The movement theory can be salvaged by positing a null or deleted NP (or operator) that has undergone movement, but that makes the theory less than minimal.

Movement theory of long-distance dependencies commits you to its being semantically blind. Long-distance dependencies where lexical semantics makes a difference (Erteschik 1973):

- (5) a. What did Bill say/?grumble that Harry would like (t) for lunch?
- b. The man who Bill said/*grumbled that Harry met (t) ordered a bagel.
- c. The book is longer than you said/*grumbled that it was (t).

No syntactic approaches to this difference are known to me.

Monstrous theories approach these problems in terms of mapping from semantic configurations to syntax, require no commitment to phantom syntactic elements that undergo movement to phantom positions (taken for granted since 1960s).

As known early on, learning structural descriptions of ordered transformations is one of the most severe obstacles to language acquisition (Wexler & Culicover 1980). This was argued to be a strong justification for a sophisticated UG that builds in many general constraints on movement.

A direct mapping theory that also meets Descriptive Constraint, but without movement, has a hope of eluding this need, thereby better satisfying Evolutionary Constraint (see below).

Implementation: Mechanisms for building syntactic structure

Minimalist Program: It's necessary to put units together recursively. Simplest possible way is

Merge: take two constituents and stick them together, give them the label of one of them.

Take A and B and create either [A A B] or [B A B] or [A B A] or [B B A].

What do you start with? A "numeration": set of elements chosen from lexicon. What's in lexicon? Minimally, words and/or morphemes, coded minimally: nonredundantly. All redundancy is squeezed out into rules. Bloomfield, as quoted by Chomsky: "The lexicon is really an appendix of the grammar, a list of basic irregularities."

Assumptions of this approach:

1. Organization of syntactic structure is to be characterized in terms of putting pieces together one after another. (Derivational, a.k.a proof-theoretic)
2. Binary branching is optimal and minimal
3. Lexicon is nonredundant.
4. Strict division between lexicon and grammar (rules/regularities): they are entirely different beasts.
5. Semantics is strictly locally compositional (Fregean): word by word, combined according to structure

Alternatives to these assumptions:

1. Organization of syntactic structure is to be characterized in terms of licensing: checking whether each piece of structure and each relation among pieces of structure online is licensed by a relevant principle. (Constraint-based, a.k.a. representational, model-theoretic) This is the characteristic formalization of all the non-movement theories: Is this different from derivation? (Chomsky usually says it's a notational variant.) To be seen.

A simple case of difference between the two: Constructions that admit non-English terminal elements (Postal 2004)

- (6) a. The space alien said 'k'laatu barrada n'ikto' to Gort.
- b. [leenspeak:] And then, I'm all like, [gesture of exasperation]
- c. The sign @ was invented in 1451.
- d. *Skief* does not rhyme with *nikto*.
- e. *Jean est mangé le pain* is ungrammatical in French.

These cannot be described using derivation starting from numeration (unless lexicon, improbably, includes all manner of non-English items) In constraint-based model, there can be particular contexts that do not constrain constituents to items of English.

2. Fundamental combinatorial device is **Unification** (sort of like Boolean union on feature structures), used in HPSG, LFG, Construction Grammar

Unification of [V, +past] and [V, 3 sing] = [V, +past, 3 sing]

Unification of [_{VP} V NP] and [V, +past] = [_{VP} [V, +past] NP]

Merge is a special case of Unify: Unification of A and [x, y] = [A, y]

Unification of B and [A, y] = [A, B]

Unify cannot be reduced to a special case of Merge.

So Merge is not the conceptually simplest combinatorial operation, as claimed.

3. Multiple branching trees require fewer nodes than binary branching trees. Which is minimal, fewer nodes, or fewer branches per node? Multiple branching recursion is present elsewhere in cognition, so it arguably comes "for free." Example: visual grouping:

```

xxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx 00000 xxxxxx 00000 xxxxxx
00000 xxxxxx 00000 xxxxxx 00000 xxxxxx 00000 xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx
xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx xxxxxx 00000 xxxxxx

```

This is recursive (and could be further embedded in array of arrays). But no justification for binary branching here – only ternary and 5-ary.

Conclusion: From point of view of Evolutionary Constraint, no reason to suppose binary branching. From point of view of Descriptive Constraint, binary branching forces extra complexity. Therefore argument for binary branching based on alleged simplicity is spurious.

Claim that binary branching simplifies binding theory by eliminating linear order from syntactic theory is also spurious: linear order is something you get "for free" in the signal. A priori the theory should use linear order in preference to adding complexity elsewhere.

Claim that binary branching simplifies acquisition is spurious. Acquisition is equally simplified by assuming default n-ary branching (flat structure). Therefore

Uniform binary branching is not a minimal assumption in terms of either the Descriptive or the Evolutionary Constraints.

4. Lexicon is full of redundancy.
 - a. Although Chomsky often asserts that lexicon is nonredundant, no one (to my knowledge) has ever formulated how all redundancy is squeezed out of lexicon into rules. (See below for many problems)
 - b. Best guess is that redundancy is characteristic of brain processes, helps stabilize them & make them more reliable.

Addressing acquisition: What do you have to learn?

Words. There are a lot of them. Need a learning procedure that acquires them – not setting a bunch of parameters.

What else do you have to learn???

Thousands of clichés, titles, etc. (cf. *Wheel of Fortune* corpus, Jackendoff 1997a):

- (7) a. Any friend of yours is a friend of mine.
 - b. all hands on deck
 - c. Can't Buy Me Love
 - d. faster than a speeding bullet
 - e. The Grapes of Wrath
 - f. leave us a message after the tone
- Certainly redundant. Certainly linguistic knowledge (what else could they be?).

Thousands of idioms

- (8) a. kick the bucket
 - b. take NP to task (discontinuous: can't insert under one node)
 - c. son of a gun
 - d. down in the dumps
 - e. fit as a fiddle

Again redundant. How do these get into syntax?

Problem for Merge: If these are single items, they somehow have to constitute a single node – yet they look one hell of a lot like normal syntactic structure, built by composition from smaller units.

Constraint-based approaches: These are lexical items that license structures larger than words, partly redundantly with constraints that license ordinary productive phrase structure.

Noncanonical utterance types, not predicted by X-bar theory

- (9) a. *PP with NP:* Off with his head! Into the trunk with you!
How about a cup of coffee? How about we have a little talk?
- b. *How about X?* What, me worry? Him in an accident? John drunk? (Akmajian 1984)
- c. *NP+acc Pred?* One more beer and I'm leaving. One more step and I shoot. (Culicover 1972)
- d. *NP and S* The Red Sox 4, the Yankees 3
- e. *Scores* The more I read, the less I understand. (McCawley 1988, Culicover and Jackendoff 1999)
- f. *The more ... the more*

These can ONLY be stored as exceptional pieces of syntax, complete with special interpretation – among the original motivations for Construction Grammar.

Other noncanonical pieces of syntax ("syntactic nuts": Culicover 1999)

- (10) Names of geographical features:
 - the Atlantic Ocean, the Hudson River, the Mediterranean Sea
 - the Bay of Biscay, the Gulf of Aqaba, the Sea of Azov
 - Arrowhead Lake, Wissahickon Creek, Laurel Hill, Loon Mountain
 - Lake Michigan, Mount Washington
- These are productive, apply to novel names (*the Bojoric Sea: *Sea Bojoric*). So there are rules (of sorts). In lexicon? In grammar??

(11) Numbers:

three hundred fifty-five billion, fourteen million, one hundred twenty-five thousand, six hundred thirteen

(12) Focus reduplication (Horn 1993, Chomeshi et al. 2004):

You make the tuna salad, and I'll make the SALAD-salad.
Would you like some wine? Would you like a DRINK-drink?
Do you LIKE-her-like her?

Are you guys, um, LIVING-together-living together?

Unlike most reduplication rules, this one can copy phrases as well as words.

(13) N-P-N construction (Williams 1994, Jackendoff 1997a):

dollar for dollar (a dollar-for-dollar exchange; we matched them dollar for dollar)
face to face (a face-to-face confrontation; we stood face to face)
house by house (a house-by-house inspection; we looked house by house for spies)
month after month (of war)
(month-after-month boredom; it went on month after month; month after month of war ruined the economy)
volume (up)on volume of phonology texts

(we accumulated volume on volume of phonology texts)

Productive in some respects, but riddled with special features (Is it a PP? an AP? an NP?)

How is meaning related to meaning of preposition, if at all?)

All these little patterns have to be learned. They are rules. But they don't follow from any standard notion of UG. Presumably every language has lots of these sorts of things.

Possible objection from Minimalists: These are "merely" "peripheral" aspects of grammar. The problem of language acquisition and the goal of "perfection" apply only to "core" grammar (e.g. argument structure, passive, raising, long-distance dependencies, basic cases of binding). So these sorts of phenomena are irrelevant.

First response: Such an approach explicitly abandons the Descriptive Constraint: the theory is no longer responsible for the structure of the adult grammar (or postpones it indefinitely).

Second response: If you have a learning procedure that can acquire words and all these "peripheral" grammatical phenomena, can't the same learning procedure acquire "core" grammar as well (Culicover 1999)? Without an account of the learning of the "periphery", you can't tell. Research strategy of Minimalist Program is to idealize away from "periphery" – not to mention its acquisition – so it will never investigate this question.

Third response: Similar peculiarities are found in indisputably "core" areas of grammar

VP constructions, where V does not license complements (Goldberg 1995, Jackendoff 1990, 1997b)

(14) a. He sang/drank/slept/laughed his head off.

(V his head off = 'V excessively')

b. Bill belched/lurched/joked/laughed his way out of the restaurant.

(V his way PP = 'go PP while/by V-ing')

c. Sara slept/drank/sang/laughed the whole afternoon away.

(V NP away = 'spend NP amount of time V-ing')

d. The trolley squealed/rumbled around the corner.

(V PP = 'go PP, motion inducing V-ing sound)

e. Bill drank the pub dry.

(V NP AP = 'make NP AP by V-ing')

These are noncanonical distributions of argument structure, in that direct object, AP and PP arguments are not determined by the verb. Rather, these are to be thought of as idioms with VP structure in which the verb is an argument. No account of these known to me in GB or MP. But argument structure is "core grammar".

"Signatures" of long-distance dependency constructions (Ginzburg and Sag 2000)

(15) a. Wh-questions: wh-phrase at front, allows preceding preposition (*From whom...?*), requires inversion in main clauses *except* with *how come*, which prohibits it.

b. Initial indirect questions: wh-phrase at front (*I wonder who to see*), allows preceding preposition (*I wonder to whom to speak*), disallows *for*-subject (**I wonder who for Bill to see*)

c. Wh-exclamatives: *what (a) or how at front* (**Who we saw!*), does not allow preceding preposition (**From what a gentleman we received a letter!*), no inversion

d. Tensed relative clauses

i. wh-phrase at the front, allows preceding preposition (*the man [to whom you spoke]*)

ii. *that* at the front, no preceding preposition (*the man [that you spoke to]*)

iii. zero at the front (*the man [you spoke to]*); normally subject cannot be the gap

(**the man [spoke to you]*)

e. Initial relative clauses

i. wh-phrase at the front, must be preceded by preposition (*the man [to whom to speak]; *the man [who to speak to]*)

ii. *for*-NP subject at the front, no wh-phrase (*the man for you to hire; *the man with whom for you to talk*)

iii. bare infinitive at the front (*the man to see*)

f. Free relatives

wh-phrase or *wh-ever* phrase at the front, no preceding preposition (*what you saw; whatever you saw; *to whomever you speak*)

g. So topicalization

so-phrase at front, inversion in main and subordinate clauses (*so many men did he see that ...; I can assure you that so many men did he see that ...*)

h. The more constituents

the more/the A-er phrase at front, must be in parataxis with another such clause (or a comparative clause) (*the more you read, the less you understand; you understand less, the more you read*)

These particulars are not predictable from a general rule that says to move things to the front. They have to be learned. Not easy to write rules that come up with these configurations after fronting. The attempt was abandoned with the onset of Principles and Parameters (i.e. in disregard of Descriptive Constraint).

In a constraint-based (non-movement) theory, these can be learned as idiosyncratic configurations associated with surface forms – i.e. syntactic idioms with particular constructional meanings. Generalizations about long-distance dependencies are not a consequence of movement, but a consequence of relating the signature to a gap within the clause.

Conclusion: There is an unbreakable continuity between “core” and “peripheral” phenomena, and between “core” generalizations and complete lexical idiosyncrasy. A theory that posits a principled difference between them is missing a deep and important fact about language (not to mention abandoning the Descriptive Constraint).

A derivational movement-based theory does not lend itself to expressing this insight. A constraint-based theory does. Therefore:

Derivational and constraint-based theories are not notational variants, and constraint-based theories are more adequate for expressing insights about the texture of linguistic structure. Moreover:

Virtually all basic properties of implementation in MP are either formally non-minimal, empirically inadequate, or methodologically unsound.

Learning and innateness – satisfying the Evolutionary Constraint

Another objection from Minimalists: A constraint-based theory requires a proliferation of rules in order to meet the Descriptive Constraint (not entirely addressed in contemporary derivational theories). How does it address acquisition, so as to comply with Evolutionary Constraint (i.e. reduce innate components of Narrow Language Faculty)?

Difference between a word and a rule: Both are pieces of structure stored in memory, but a rule has variables as part of its structure, which must be satisfied by unification with something else.

(16) The smooth transition from idiosyncrasy to maximal generality:

- VP idiom – no variables: [vp [v kick] [np [the] [n bucket]]]
- VP idioms with variable: [vp [v take] [NP [p to] [vp task]]]
[vp [v [np pro's head] [n off]]]
- VP structure with more variables: [vp V (NP) (PP)]
[vp V ...]
- Head parameter for VP: [vp V ...]
- X-bar theory: [vp ... X ...]

“Core” principles of phrase structure are general schemata; idiosyncratic rules and fully specified items are specializations. There can also be idiosyncratic rules that are not specializations of more general schemata (e.g. N-P-N).

What does this formulation of rules say about learning? Tomassello 2003, Culicover and Nowak 2003 (also Braine in the 1970s):

- Learn particular constructions holistically
- When multiple items share a part, create a new item (i.e. a rule) that consists of the constant part plus a variable corresponding to parts that differ from item to item.
- More and more general schemata arise by recursive application of 2. Much easier in a theory without movement!

How is this different from plain analogical learning?

- Variables acquired by generalizing over stored items.
- UG plays a role, providing the most general schemata, towards which child's generalizations will gravitate, all else being equal. (Jackendoff 2002)

Some aspects of UG:

- Basic organization of conceptual structure, growing directly out of primate cognition (hence part of Broad Faculty of Language).
 - The notion of words being used symbolically to communicate intentionally about perceived world – THE evolutionary breakthrough (Deacon 1997). The rest is refinement.
 - Use of Unification plus variables in stored structures to permit productivity, recursion
 - Basic principles of phrase structure
 - X-bar theory (16e)
 - Other common alternatives such as conjunction schema
 - Basic default principles of syntax-semantics interface
 - Semantic heads map to syntactic heads, semantic arguments to syntactic arguments, semantic modifiers to syntactic adjuncts (default counterpart of Interface Uniformity)
 - Agent First
 - Topic First, Focus Last
 - Basic principles of morphological agreement and case-marking
 - Basic principles of long-distance dependencies
 - A Grammatical Function tier (a la LFG, RG), to deal with distribution of NP arguments alone; responsible for passive, raising, structural case, agreement, pro-drop
- Not to mention UG aspects of phonology and morphology.

This is not “perfect” by any means. But it appears relatively minimal, given need to satisfy Descriptive Constraint. Unlike MP, this conception of grammar allows for proliferation of learned rules, under a potentially realistic learning regimen. Learning rules is mostly an extension of learning words. Narrow Language Faculty as a “toolkit” that results in tendencies toward language universals. And, unlike MP, it is nicely compatible with incremental evolution of the language faculty.

My prediction: If any approach to language is eventually going to satisfy the Primary Goal of linguistic theory, satisfy the Descriptive and Evolutionary Constraints, and make meaningful contact with cognitive neuroscience and evolutionary biology/psychology, it will be an approach growing out of constraint- and construction-based minimalism, not out of the Minimalist Program.