Language, Information and Computation(PACLIC 11),1996, 11-20

Configuration vs. information: An informational explanation of command relations

Jean-Yves Morin
University of Montreal
morinjy@iro.umontreal.ca, morinje@ere.umontreal.ca

Abstract

In this paper, we show that configurational notions follow from inherent properties of informational concepts. First, we analyze the hidden complexity of the central family of configurational relations (command) from which all the others (government, proper government, etc.) are derived. We then examine the different types of grammatical information, their sources and their propagation mode. Finally we show how simple and general constraints on these might make syntactic dependencies appear to be governed by configurational relations such as command, but just like the Stars might appear to be turning around the Earth.

1. Introduction

The essential difference between the most widely used syntactic frameworks in linguistics (Principles and Parameters, Minimalist Program) on the one hand and those used in computational linguistics (Generalized Phrase-Structure Grammar, GPSG, Head-Driven Phrase-Structure Grammar, HPSG, Lexical-Functional, LFG, Categorial Unification Grammar, CUG, etc.) on the other hand is that the former crucially make use of *configurational* notions, while the latter use *informational* concepts.

In this paper, we show that configurational notions follow from inherent properties of informational concepts.

First, we analyze the hidden complexity of the central family of configurational relations (command) from which all the others (government, proper government, etc.) are derived.

We then examine the different types of grammatical information, their sources and their propagation mode.

Finally we show how simple and general constraints on these might make syntactic dependencies appear to be governed by configurational relations such as command, but just like the Stars might appear to be turning around the Earth.

2. Command relations

In configurational theories, command relations are part of the central ("hard") kernel. They are ubiquitous and play an essential role.

In all variants of Principles and Parameters (GB, Barriers, etc.), each of the sub-theories or so-called "modules" ((proper) government, binding, Case, control, chains, theta-theory, barriers, etc.) is ultimately defined and "explained" in terms of one or the other notion of COMMAND (c-command, max-command, etc.). In other words, all the important (families of) concepts of this family of theories are defined, directly or indirectly, in configurational COMMAND terms.

This configurational notion of COMMAND remains central even in the Minimalist Program, as Chomsky (1995) himself notes:

> " ... the core relations of phrase structure theory, *dominate* and *c-command*
> [...]
> The core intuition underlying c-command is that
> (36) X c-commands Y if
> (i) every Z that dominates X dominates Y and
> (ii) X and Y are disconnected.
> For categories, we take X and Y to be disconnected if $X \neq Y$ and neither dominates the other."
> (Chomsky 1995 : 418-9. The emphasis is present in the original. I have reformatted the text of the definition to make it more readable, JYM.)

Classes of syntactic phenomena

As for classes of syntactic phenomena, command relations are pervasive in the analysis of short (e.g., selection and subcategorization), as well as medium (e.g., passives, clitics, local agreement) and long (or unbounded dependencies (e.g., questions, relatives, topicalizations, clefts).

Configurational notions in informational theories

In informational theories, such global configurational notions have no direct counterpart. All constraints or principles (e.g., ID, LP, HEAD, SUBCAT, etc. in HPSG, c-constraints, f-constraints and constraint equations in LFG) are essentially local. Global configurational properties of syntactic structures simply follow from the interaction of these local constraints. Pollard & Sag (1994, ch. 6) have shown that even the BINDING THEORY is best represented as a set of (obliqueness) constraints on local values of SUBCAT lists.

Pasting conditions

On the other hand, it might be interesting to define *pasting conditions*, that would allow the evaluation and the importation into an informational theory (like HPSG, for example) of analyses formulated within a configurational framework.

It is quite important then to clarify this fundamental notion of COMMAND. It should not be too surprising that the only systematic study of command relations (Barker & Pullum, 1990) was not conducted within the configurational framework (where it is considered an indisputable axiom, open only to parametrization), but from an informational point of view. Barker & Pullum (1990) made an inventory of existing command relations and proposed an abstract typology of possible command relations. They noticed the problems raised by the irreflexive character of traditional definitions and showed that, once this stipulation is removed, the different notions of command can be characterized by the choice of UPPER BOUND (minimal dominating node with a given set of properties) that they admit. This choice can be purely configurational (e.g., dominating node, branching node, etc.) or make reference to grammatical information (maximal projection, cyclic node, lexically branching node, etc.).

However, even if (properly modified) command relations may be consistent, they do not follow from anything else. It might seem strange that in theories where stress is emphatically put on explanation rather than description, there is precisely no explanation for the central role of exactly this type of configurational relation.

Indeed, of all the possible configurational relations between two nodes x and y in a tree T, why should it be such extended avuncular relations (command) which are central to syntax, rather than, say, maternal relations (immediate dominance), ancestor relations (dominance), sister relations (strict locality), elderliness relations (linear precedence), generation depth, distance to common ancestor or to root, community of depth or position, or any other relation in the huge universe of possible configurational relations?

3. Hidden complexity

Even though standard definitions of command (like the one by Chomsky above) look fairly simple, they are actually quite complex, both from a logical and a computational viewpoint.

Some sources for this complexity are the following :

(i) The irreflexive character of the relation (noted by Barker & Pullum 1990).

This irreflexive character entails that the domain of a (classical) command relation is NEVER a constituent, but always as set of disjoint trees, since the root of these trees cannot be part of the relation (i.e., it cannot command itself).

(ii) The use of (proper) dominance (rather than immediate dominance) as a primitive.

If (proper) dominance can easily be defined as the (irreflexive) transitive closure of immediate dominance, taking it as a primitive and deriving the simpler notion of immediate dominance from it is unnatural and computationally costly (Rogers 1994). It is usually abandoned in actual implementations, even the most faithful ones (Fong 1991).

(iii) The mixture of configurational (e.g., branching node) and informational (e.g., lexically branching node, maximal projection) notions.

As we shall see below, it is the (grammatical) informational properties that are important. Superficial configurational properties follow from deeper informational properties.

(iv) The necessity of tree-walking predicates (e.g., testing whether a branch actually leads to a lexically filled node, whether a maximal projection intervenes between two nodes, etc.).

Notice that checking intervening nodes between x and y in terms of dominance involves comparing all nodes that x dominates and all nodes that dominate y, a computationally complex process. Recall that visual inspection, which easily solves the problem in a particular case (and is the usual way in which it is "demonstrated" that a given configurational relation holds) cannot be used in the general case (just as in the traveling salesman problem, where a visual map of a specific problem would turn the general problem into a much simpler specific one). It is the general problem which is interesting (and intractable).

(v) The low filtering value of configurational relations.

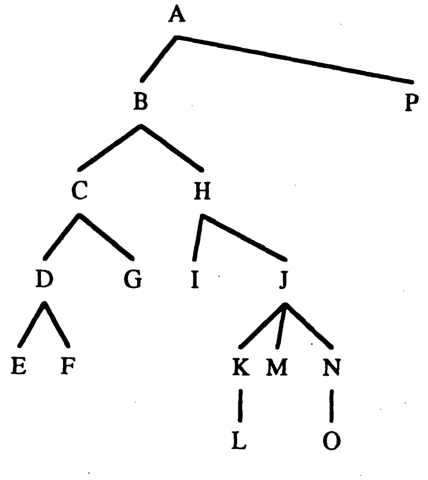
For any branching tree (tree with a branching root), ALL nodes in the tree, except the root are in the classical (branching node) c-command relation, both as c-commanders and c-commanded. This is easily shown. Any node N, except the root, is immediately dominated by one and only one node M, which is either branching (a) or not (b).

(a) If M is branching, N c-commands all the nodes that M dominates except those that N itself dominates.
(b) If M is not branching, c-command is delegated to it. N c-commands all the nodes that M itself c-commands.

For the passive relation (c-commanded), in all branching subtrees S_k of tree T, each of the nodes N_i immediately dominated by M (the root of S_k) and, hence, all the nodes N_i dominates, are c-commanded by each of the other nodes N_j immediately dominated by M. In a non-branching subtree, there is no internal c-command relation (but all the nodes are still c-commanded in the first branching subtree above). This obviously covers all the nodes in the tree. (Only in a degenerate, purely non branching tree, is there no c-command relation at all. But, if there is a c-command relation in a (sub)tree, all nodes, except the root, are part of it.) Because of the transitivity of dominance, most nodes appear many times in the relation, both as c-commanders and c-commanded.

Take for example the following tree configuration (from Guilfoyle et al. 1992: 382), in which implicit branching has been expanded and all grammatical information has been removed, keeping only configurational relations.

(1)(a) [A [B [C [D E F] G] [H I [J [K L] M [N O]]]] P]
(b)

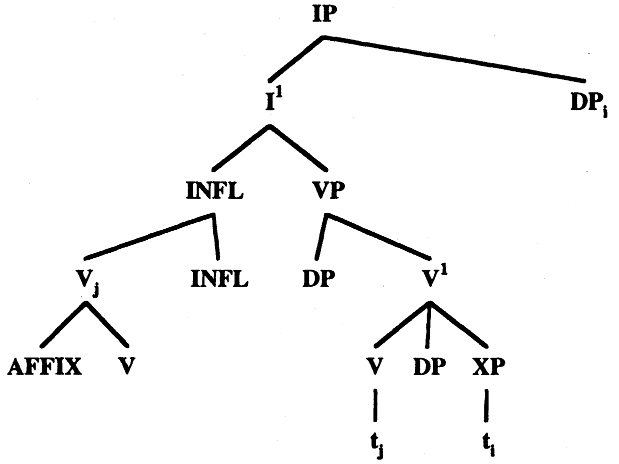


The classical c-command relation (UPPER BOUND = branching node) for this tree is : {<B,P>, <C,H>, <C,I>, <C,J>, <C,K>, <C,L>, <C,M>, <C,N>, <C,O>, <D,G>, <E,F>, <F,E>, <G,D>, <G,E>, <G,F>, <H,C>, <H,D>, <H,G>, <H,E>, <H,F>, <I,J>, <I,K>, <I,M>, <I,N>, <I,L>, <I,O>, <J,I>, <K,M>, <K,N>, <K,O>, <L,M>, <L,N>, <L,O>, <M,K>, <M,L>, <M,N>, <M,O>, <N,K>, <N,M>, <N,L>, <O,K>, <O,M>, <O,L>, <P,B>, <P,C>, <P,H>, <P,D>, <P,G>, <P,I>, <P,J>, <P,E>, <P,F>, <P,K>, <P,M>, <P,N>, <P,L>, <P,O>}.

So, for the 15 relevant nodes (the root is excluded by the irreflexivity of the classical definition), there are 57 c-command pairs, while there are only 48 pairs in the general dominance relation and 15 pairs in the immediate dominance relation. The filtering value of c-command is very poor.

As soon as grammatical information is reintroduced, one can see that it is this information which is important for selecting the relevant pairs and not the configurational information alone.

(2)(a) $[_{IP}\, [_{I^1}\, [_{INFL}\, [_{V_j}$ AFFIX V] INFL] $[_{VP}$ DP $[_{V^1}\, [_{V}\, t_j\,]$ DP $[_{XP}\ t_i\,]]]]$ $DP_i\,]$



This tree is intended to represent the S-structure of a sentence in the circumstancial voice in Malagasy (or Tagalog):

(3) An-(s)asa-n' ny zazavavy ny lamba ny savony.
AFF-wash-INFL the girl the clothes the soap.
The soap is what the girl is washing the clothes with.

DP_i is the (surface) subject, the DP immediately dominated by VP is the agent (underlying subject), the DP immediately dominated by V^1 is the patient (underlying object) and XP_i is any kind of (local, temporal, instrumental, causal, etc.) PP. Once this grammatical information is given (recall that indices are also part of grammatical information in Principles and Parameter Theory), it is easy to check that only a handful of the 57 c-command pairs are meaningful:

(a) V_j and $[_V\, t_j\,]$,

(b) DP_i, and $[_{XP}\, t_i\,]$,

(c) $[_V\, t_j\,]$ and both the DP and the XP under V^1,

(d) DP under VP and $[_V\, t_j\,]$.

Notice that the symmetrical relation, between the verb and the DP under VP, which is also meaningful (this DP is the first semantic argument of the verb), is not part of the c-command relation, neither is the relation between V_j and this NP. Therefore, while many meaningless relations are not filtered out by c-command (e.g., between DP_i and all the other nodes in the tree, between DP and XP inside V^1, etc.), some meaningful ones are not covered (e.g., between V_j or $[_V\, t_j\,]$ and the NP immediately under VP).

(vi) The multiplicity of distinct notions of UPPER BOUND necessary to account for different phenomena.

Here is a sample of some types of nodes that have been proposed in the literature: branching, lexically branching, t-dominating, cyclic, barrier, enumerated types (S, NP, ...), governing category, maximal projection. Once again, this constitutes a clue that it is grammatical information (and not configuration) which is relevant.

(vii) The opacity of the different configurational notions.

Take the following definitions of GOVERNMENT (from a review of the literature on the Empty Category Principle (Hornstein & Weinberg 1995)).

> "A governs B iff for all X, X a maximal projection, X dominates A iff X dominates B."
> (Hornstein & Weinberg 1995: 246, attributed to Aoun & Sportiche 1981)
>
> "*Government* X governs Y iff X m-commands Y and there is no W, W a barrier for Y, such that W excludes X.
>
> *Exclusion* X excludes Y if no segment of X dominates Y.
>
> *m-commands* X m-commands Y iff X does not dominate Y and every W (W maximal) that dominates X dominates Y."
> (Hornstein & Weinberg 1995: 267, attributed to Chomsky 1986)
>
> "A governs B iff all maximal projections dominating B also dominate A and for B = Y^{max} if A governs B then A governs the head of B (i.e. Y^0)."
> (Hornstein & Weinberg 1995: 273, attributed to Aoun, Hornstein, Lightfoot and Weinberg 1987)

The notions of *segment* and *exclusion*, are introduced precisely in order to cope with problems in the classical definition of government. According to Chomsky 1986: 9, in adjunction structures of the form $...\delta...[_{\gamma}\ \alpha\ [_{\gamma}\ ...\beta...]]$, the two γ's are ***segments*** of γ and α is NOT $dominated_2$ by γ, since it is not $dominated_1$ by all its segments. (The indices are not present in Chomsky 1986 and have been introduced here to distinguish the classical relation $dominate_1$ from the new one $dominate_2$.)

There are logical as well as conceptual problems with these definitions. On the logical side, there are the scope of the *iff* connectors, the use of conditionals in the *definiens* and the equivocation on the two meanings of *dominate* in the definition of *exclusion*).

On the conceptual side, the first and third definitions allow A and B to be in a government relationship even when they are in a dominance relationship. A can dominate B or B dominate A, provided neither of them is a maximal projection). So, according to these definitions, in a structure $[_{N^{max}}\ [_{N^1}\ N^0\ ]]$, N^1 and N^0 govern each other. The second definition, while explicitly excluding that X dominates Y, thereby excluding N^1 as an m-commander and, consequently, as a governor for N^0, would still admit N^0 as an m-commander and, consequently, as a governor for N^1.

Apart from that, the only thing that is clear in these definitions is that the relationship between government and command is completely opaque.

Barker & Pullum (1990) have proposed that the core notion of government can be represented as the intersection of a command relation using one type of UPPER BOUND (branching node, i.e., C-COMMAND) and the INVERSE of a command relation using another kind of UPPER BOUND (maximal projection, i.e., MAX-COMMAND). In other words, for x to govern y, x would have to c-command y and y would have to max-command x. Different notions of government would then be related to the choice of UPPER BOUND for the relation and its inverse.

As for the different types of government (proper government, antecedent-government, head-government, θ-government, etc.), they could all be defined by different sets of grammatical (i.e., informational and not configurational) constraints on the grammatical type of the governor and/or the governed category.

Even if it were possible, at least in principle, to give a precise content to configurational notions (essentially by enriching them with informational content), the main question would remain. Why should these particular configurational relations (i.e., extended avuncular relations) be especially relevant to syntax?

If configurational relations can be shown to be (more or less accidental) consequences of deeper (informational) factors, this would make the logical structure of syntactic theories much clearer.

Moreover, from a computational viewpoint, if configurational relations can be shown to follow from simpler notions, rather than being directly invoked all the time, this could greatly reduce computational complexity.

4. Dimensions of grammatical information

Grammatical information, the fabric grammatical objects are made of, can be considered from a variety of viewpoints or dimensions.

On the dimension of *origin* of grammatical information, we can distinguish γ-information (coming from grammar) and λ-information (coming from the lexicon).

On the dimension of *type*, we can distinguish *inherited* information (appearing on a category by virtue of its being a start category, or being a member of the right-hand side of an ID-constraint) and *synthesized* information (appearing on a category by virtue of its being a lexical category -and coming from its lexical entry-, or being the left-hand side category of an ID-constraint).

On the dimension of *propagation*, we can distinguish *original* information (directly specified on a category, by grammar rule(s) or lexical entry) and *propagated* information (propagated, through propagation constraints, from category to category in a local tree).

5. Configuration vs information

In information-based theories, in order for two syntactic objects in a given structure to be related they must either

(a) be part of the same local tree (only in that case, can they be in the scope of a language particular or universal grammatical constraint) or

(b) share a common ancestor, in a local tree, toward which the relevant constraints can be propagated and ultimately checked.

Case (a) covers all types of local dependencies : ID-constraints, LP-constraints, etc. Obviously, non root nodes in a local tree are all in a c-command relation (and its inverse), but this is just a consequence of strict locality, a much narrower concept. Moreover, the root node is also relevant, although it is excluded in classical c-command.

Case (b) covers long distance dependencies and reduces them in fact to (propagated) local dependencies. The apparent configurational constraints on distance are just consequences of constraints on the information load of intervening categories.

6. A different brand of minimalism

We can even go a step further.

Let us adopt a minimality discipline with respect to configurations. In other words:

(a) configuration changing operations are never used and

(b) grammatical information is never encoded into configurational relations.

Although (a) is a basic postulate of all informational (non derivational) theories, (b) is less often explicitly adopted. This type of approach is in sharp contrast with Chomsky's Minimalist Program (1995) where just about everything is encoded into tree configurations and tree operations. (The Minimalist Program, highly reminiscent of Generative Semantics, as many have already noted, amounts to a reduction of syntactic theory to Q-systems (Colmerauer 1970), not a very welcome move, on computational grounds, despite the pervasive computational metaphors.)

In an information-based framework, the (configurational) minimality discipline is embodied in the following constraints.

UNIVERSAL PROJECTION (UP)

All lexical level categories are directly projected into a phrasal sign.

There is no intermediate level of projection between lexical categories (MINOR or MAJOR) and phrasal projections. That is, all projections are maximal. In other words all lexical signs are useful. They contribute some grammatical information. There is no configurational difference between lexical signs, only informational differences. The difference between MAJOR and MINOR lexical signs is that the former are projected onto an autonomous phrasal sign (e.g., N onto NP, V onto VP), while the information brought by the latter has to be unified with an independently projected phrasal sign (e.g., Det into NP, Classifier into NP or VP).
In other words, we assume that there is

- a *lexical* level, corresponding to objects imported from the lexicon,

- a *phrasal* level, where lexical and non lexical information is packaged and, eventually,
- a *clausal* level, where designated arguments (or even non-arguments) are adjoined to (normally unsaturated) phrases.

CHAPTALIZATION (CHAP)

The grammatical information on a phrasal sign is blended.

The SOURCE (lexicon or grammar), the TYPE (original or propagated) and the DIRECTION (inherited or synthesized) of grammatical information are indistinguishable on a phrasal sign.

FIXATION (FIX)

Grammatical information is packaged into phrasal signs and frozen.

Only grammatical information explicitly typed as global is accessible for further projection.

INTEGRITY CONSTRAINT (IC)

Grammatical information on phrasal signs must be COHERENT and RESOLVED.

Functions must be well-defined and operations (partially) solved.

Assuming these constraints, which are just fairly natural consequences of a framework like HPSG, we can explain why syntactic dependencies appear to be governed by avuncular command constraints.

For two signs to share constraints, they must either

(a) be part of the same package, (immediate constituents of the same phrasal sign), on which this constraint is fixed and resolved, or

(b) one of them must be a lexical sign, with specific (but precompiled, not computed on line) constraints (e.g. V_{SUBCAT}, X_{WH}).

In the case of subcategorization, for instance, SUBCAT originates in a lexical sign. As soon as a phrasal sign is projected (UP), this SUBCAT information must be (partially) resolved with the immediate phrasal daughters (FIX, CHAP and IC). Only unresolved constraints can be transmitted (and only if the phrasal type allows this, e.g. VP).

Fillers (wh, clitics, etc.) and gaps (SUBCAT or ADJUNCTS values, etc.) can only be as far removed from the common ancestor as there are possible intervening projections where they can remain unresolved.

For fillers, this is severely limited. For example, unresolved pronominal clitics (like those found in Romance languages) must be resolved as soon as a projection with a SUBCAT list is hit (i.e., a VP). For wh words, which are lexical signs, the wh feature is projected into a phrasal sign (UP), no matter if the wh word is the head, a specifier or a marker. The information on this phrasal sign is blended (CHAP) and fixed (FIX). Only the wh information can permeate and propagate farther (as it is explicitly typed as global). It must be resolved as soon as a (verbal) projection with a SUBCAT list is hit. The fact that, in some languages, intervening linear cascades of NPs and/or PPs allow wh constraint propagation (i.e., pied-piping) is simply a consequence of the relative openness of these categories with respect to global grammatical information. For gaps it is less restrictive, especially in languages allowing unresolved VP's (i.e., external subjects). Unresolved SUBCAT (or ADJUNCTS) information is propagated until it can (or must) be solved.

This explains why fillers appear to be somewhat closer to the common ancestor than gaps (i.e., filler information cannot remain unresolved after a SUBCAT projection is hit, while SUBCAT and ADJUNCTS information can). It also explains why different notions of UPPER BOUND seem to be needed (i.e., different types of fillers have different domains where their constraints have to be resolved, and different languages may vary with respect to the permeability of their categories).

Thus, the apparent avuncular position of fillers with respect to gaps simply follows from general properties of grammatical information and the way it is propagated and checked.

7. Conclusion

In this paper, we have examined some of the problems raised by configurational notions. And we have shown that the apparent configurational asymmetries (avuncular relations) in natural language syntax simply follow from the informational properties of grammatical objects and constraints.

References

Aoun, J., N. Hornstein, D. Lightfoot and A. Weinberg. 1987. Two types of locality. *Linguistic Inquiry* 16: 623-637.

Aoun, J. & D. Sportiche. 1981. On the formal theory of government. *The Linguistic Review* 2: 211-236.

Barker, Chris, & G. Pullum. 1990. A theory of command relations, *Linguistics and Philosophy* 13, 1 : 1-34.

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

Chomsky, N. 1995. Bare phrase structure. In Gert Webelbuth, ed., *Government and Binding Theory and the Minimalist Program*, 383-439. Oxford: Blackwell.

Colmerauer, Alain. 1970. Les systèmes-Q ou un formalisme pour analyser et synthétiser des phrases sur ordinateur, PI-43, Département d'informatique, Université de Montréal. Reprinted in *TAL* 33, 1-2, 1992: 105-148.

Fong, Sandiway. 1991. *Computational Properties of Principle-Based Grammatical Theories*. Doctoral dissertation, MIT.

Guilfoyle, E., H. Hung and L. Travis. 1992. Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language and Linguistic Theory* 10: 375-414.

Hornstein, N. & A. Weinberg. 1995. The Empty Category Principle. In Gert Webelbuth, ed., *Government and Binding Theory and the Minimalist Program*, 241-296. Oxford: Blackwell.

Rogers, James. 1994. *Studies in the Logic of Trees with Applications to Grammar Formalisms*. Doctoral dissertation, University of Delaware.