

STUDI

# Linguistic Knowledge and Unconscious Computations

Luigi Rizzi<sup>(α)</sup>

Ricevuto: 24 gennaio 2016; accettato: 27 ottobre 2016

**Abstract** The open-ended character of natural languages calls for the hypothesis that humans are endowed with a recursive procedure generating sentences which are hierarchically organized. Structural relations such as c-command, expressed on hierarchical sentential representations, determine all sorts of formal and interpretive properties of sentences. The relevant computational principles are well beyond the reach of conscious introspection, so that studying such properties requires the formulation of precise formal hypotheses, and empirically testing them. This article illustrates all these aspects of linguistic research through the discussion of non-coreference effects. The article argues in favor of the formal linguistic approach based on hierarchical structures, and against alternatives based on vague notions of “analogical generalization”, and/or exploiting mere linear order. In the final part, the issue of cross-linguistic invariance and variation of non-coreference effects is addressed.

**KEYWORDS:** Linguistic Knowledge; Morphosyntactic Properties; Unconscious Computations; Coreference; Linguistic Representations

**Riassunto** *Conoscenza linguistica e computazioni inconsce* – Il carattere aperto del linguaggio naturale avvalorata l’ipotesi che gli esseri umani siano dotati di una procedura ricorsiva che genera frasi gerarchicamente organizzate. Relazioni strutturali come il c-comando, espresse su rappresentazioni frasali gerarchiche, determinano tutte le proprietà formali e interpretative delle frasi. I principi computazionali rilevanti sono totalmente al di fuori della portata della coscienza introspettiva e così lo studio di tali proprietà richiede la formulazione di precise ipotesi formali e la loro verifica sperimentale. Questo articolo illustra tutti questi aspetti della ricerca linguistica, esaminando gli effetti di non-coreferenza. Si argomenta in favore dell’approccio linguistico formale basato su strutture gerarchiche e contro alternative basate su vaghe nozioni di “generalizzazione analogica” e/o che impiegano il semplice ordine lineare. Nella parte finale si affronta il tema dell’invarianza e della variazione cross-linguistica degli effetti di non-coreferenza.

**PAROLE CHIAVE:** Conoscenza linguistica; Proprietà morfosintattiche; Computazioni inconsce; Coreferenzialità; Rappresentazioni linguistiche



## Introduction

ONE REMARKABLE PROPERTY OF HUMAN language is its unbounded character: speakers are constantly confronted with sentences they

have not encountered in their previous linguistic experience, and still they can easily integrate such new messages, understand them and properly respond in dialogue. In fact, any speaker potentially masters an unbounded

<sup>(α)</sup>Dipartimento di Scienze sociali, Politiche e Cognitive, Università degli Studi di Siena, Complesso San Niccolò, via Roma, 56 - 53100 Siena (I)

E-mail: [luigi.rizzi@unisi.it](mailto:luigi.rizzi@unisi.it) (✉)



number of messages. This property caught the attention of major thinkers in the past: in particular, Descartes saw in it a test capable of discriminating man and machine:<sup>1</sup> no automaton, no matter how sophisticated, would be capable of what any man can do («ainsi que les hommes les plus hébétés peuvent faire»),<sup>2</sup> hear a new sentence and respond appropriately to it, in a way that is neither deterministic nor arbitrary.

Once the unbounded character of language is properly recognized, this has immediate implications for the nature of linguistic knowledge. “Knowing a language” cannot be equivalent to “having memorized a list of messages”, as the list would have to be infinite, and we would not have had enough time (or space in neural circuitry) to memorize it. The “memorized list” model may be essentially correct, at least for a first approximation, for our knowledge of the lexicon: we hear a new word, we figure out what it means, and we add it to our mental lexicon (it should be noted, though, that a list-based model is to be expressed with the proviso that the lexicon is not an unstructured list, but a highly structured system). But the list idea is not appropriate to capture our mastery of sentences. Clearly, the secret of the unbounded character of language is in its combinatorial nature: words can be combined to form sentences, so knowing a language means mastering the combinatorial laws that govern sentence formation. In other words, knowing a language means possessing a generative procedure, capable of computing an unbounded number of messages; acquiring a language means acquiring the lexicon, and mastering the combinatorial procedure to generate new messages.

This is what is sometimes called the “computational view” of linguistic knowledge, an important legacy of the study of language to the broader domain of cognitive neuroscience, as the idea of the “computational mind” proved to be viable in the study of the human cognitive capacities well beyond language, from vision, to reasoning, to motor control: having a cognitive capacity means possessing

a mechanism for computing representations relevant for a specific cognitive domain, with properties which may be in part domain general, and in part task-specific.

Is there an alternative to this computational view? What is sometimes considered an alternative is the enrichment of the list model with the idea that the unbounded character of the system is due to our capacity for analogical generalization: during language acquisition, the child hears a finite number of messages and she memorizes them, so that the initial knowledge is “item-based”; at some later point, the child generalizes their properties to new messages through analogy as in the “neuroconstructivist” approach.<sup>3</sup> The problem with this (very popular) view is that it is neither right nor wrong, if expressed at this level of generality: it is simply vacuous until when we give content to the notions of “analogy” and “analogical generalization” to capture the fact that certain “analogical generalizations” are unerringly made by all speakers (and all language learners), while other conceivable “analogical generalizations” are never made. But once we have properly structured the concept, so much structure emerges that the vague term “analogy” does not seem to be a felicitous terminological choice to address the “projection problem”, the fact that the child projects her finite linguistic experience onto an unbounded set of possible messages.<sup>4</sup>

I would like to illustrate these issues through a simple example: the constraints on referential dependencies between nominal expressions and pronouns. The discussion will aim at making two points. On the one hand, a point that directly bears on the topic of this workshop: the knowledge that speakers unerringly manifest is completely unconscious, there is simply no way to introspectively penetrate the structural computations that we all perform when we evaluate the possible interpretation of a pronoun in context; so the only thing to do to study this mental capacity is to formulate precise structural hypotheses, let them generate predictions, and test the predictions. This is true for this case, as well as for so many other cases of the study of non trivial

mental capacities, in language and other cognitive domains. On the other hand, I will try to show that the knowledge that every speaker has about the possible referential dependencies between nouns and pronouns obeys structural constraints which appear to go well beyond the reach of an unstructured notion of “analogical generalization”. A brief discussion of the invariance and variation observed across languages in the domain of coreference will conclude the paper.

### ■ A constraint on coreference

In certain sentences, a name (or other nominal expression, for instance a definite description) and a pronoun can refer to the same individual, or “corefer”. For instance, in (1), the name *John* and the pronoun *he* can corefer:

(1) John<sub>i</sub> said that he<sub>i</sub> was sick

In other words, the sentence can mean “John said that he, John, was sick”: we may explicitly express coreference through the indexing notation, i.e., by assigning the same index (i) to the two coreferential elements, as in (1). Coreference is not obligatory (i.e., in (1) *he* could refer to *Bill*, who was introduced in the previous discourse in a context like *Nobody understood why Bill had made such a mistake, but then...*), but it is an option in cases like (1).

Coreference is possible in certain environments and impossible in others. Consider (2) for instance:

(2) \* He<sub>i</sub> said that John<sub>i</sub> was sick

The sentence is fine, but *he* can only refer to an individual different from John, e.g., Peter, introduced in previous discourse. Coreference (he, John) is excluded, and this is what the asterisk marking the sentence means: the sentence is well-formed per se, but it is excluded with the interpretation expressed by the indices.

Notice that the contrast (1)-(2) is immediately clear to all the speakers of English, and it

is abstract knowledge, in the sense that it is completely independent from any particular knowledge of the discourse situation, or of the states of affairs in the factual world: I may very well not know anyone named John, and still if I hear somebody utter (1) I will assume that the speaker may mean that a certain guy John said that the same guy John was sick, while the speaker uttering (2) does not intend to convey that meaning. Clearly, we dispose of a procedure allowing us to evaluate coreference, and the procedure discriminates between (1) and (2), completely independently of any knowledge of factual situations.

What is this procedure? Here is the point of immediate relevance for our general theme. We perceive in a crystal clear manner the result of this mental procedure (coreference is possible here and impossible there), but the procedure itself is completely inaccessible to our conscious introspection.

In introductory courses, I always do a little informal experiment in class, first testing students on the contrastive judgment between the Italian equivalents of (1) and (2) (obviously accessible to everyone), and then asking them on what basis they are able to differentiate coreference possibilities in the two examples. Various hypotheses are made in the discussion in class, some rather complex, but one idea that always emerges, and strikes everyone for its simplicity and plausibility, is that speakers may use a linear strategy:

(3) Linear strategy: A noun and a pronoun can corefer when the noun precedes the pronoun in the linear order of the words; if the pronoun precedes the noun, coreference is impossible: i.e.,

a. ok ... N<sub>i</sub> ... Pron<sub>i</sub> ...

b. \* ... Pron<sub>i</sub> ... N<sub>i</sub> ...

This hypothesis has the appeal of simplicity, and a considerable plausibility: after all, it makes perfect sense that we first introduce a referent with a noun, and then we refer back to it through a pronoun. Nevertheless, the linear

strategy is not the one speakers actually use within a complex sentence: there is literally an infinity of sentences (in English, Italian and other languages) in which the pronoun precedes the noun, and coreference is still possible. Here are some examples:

(4) People who know him<sub>i</sub> will say that John<sub>i</sub> is sick

(5) is<sub>i</sub> father said that John<sub>i</sub> was sick

(6) When he<sub>i</sub> is sick, John<sub>i</sub> does not go to work

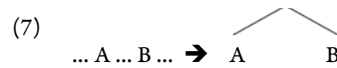
So, (4) can naturally admit the interpretation in which the pronoun *him* refers to John, in (5) *his father* may well mean John's father, with the possessive *his* referring to John, *he* can naturally refer to John in (6), etc. There may be a moment of hesitation when coreference is evaluated in contexts in which the pronoun precedes the noun (contexts of "backward coreference"), but the contrast between (4)-(6) and the sharp impossibility of coreference in (2) is clear.

So, what should we conclude from these examples? Speakers possess a procedure allowing them to evaluate coreference, and they apply it very efficiently and automatically to the sentences they hear: the judgment on the interpretation of (1), (2) etc. is clear and quickly determined by the speaker. Yet, speakers don't have any conscious introspective access to the procedure: they only "see" the result. Clearly, the students who proposed the linear rule did not have introspective access to the procedure they were using, they simply formulated a hypothesis on the nature of the procedure, based on the data they had access to, their own interpretive judgments on these sentences. The hypothesis turned out to be incorrect, but this is, in fact, the only way to proceed: formulate a precise hypothesis on the nature of the mechanism, submit it to empirical testing, and revise it in accordance with the empirical evidence.

In order to successfully address the problem, we now need to sharpen our assumptions about the structural organization of linguistic representations.

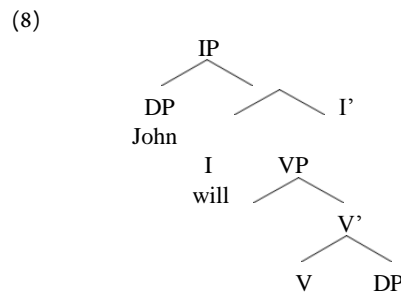
### The hierarchical nature of syntactic representations.

A fundamental property of syntactic structures is that words forming sentences are assembled into hierarchically organized units. Over fifty years of discussion on the fundamental combinatorial device led to the conclusion, within the Minimalist Program,<sup>5</sup> that it has the simplest possible shape, an instruction which says "take two elements and string them together to form a third element". This is the operation Merge:

(7) 

$$\dots A \dots B \dots \rightarrow \begin{array}{c} \diagup \quad \diagdown \\ A \quad B \end{array}$$

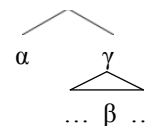
Merge takes two elements, A and B, which can be items drawn from the functional or contentive lexicon, or partial structures already built by previous applications of Merge, and strings them together to form a new structural unit. Repeated applications of Merge give rise to hierarchical structures which can be expressed by familiar syntactic trees like the following:<sup>6</sup>

(8) 

$$\begin{array}{c} \text{IP} \\ \diagup \quad \diagdown \\ \text{DP} \quad \text{I}' \\ \text{John} \quad \begin{array}{c} \diagup \quad \diagdown \\ \text{I} \quad \text{VP} \\ \text{will} \quad \begin{array}{c} \diagup \quad \diagdown \\ \text{V}' \quad \text{VP} \\ \text{V} \quad \text{DP} \end{array} \end{array} \end{array}$$

Similar representations permit the definition of fundamental structural relations which govern the structuring of form and interpretation. One very fundamental relation is c-command,<sup>7</sup> which I will define as follows:

(9)  $\alpha$  c-commands  $\beta$  in the following configuration:



$$\begin{array}{c} \diagup \quad \diagdown \\ \alpha \quad \gamma \\ \quad \quad \diagup \quad \diagdown \\ \quad \quad \dots \beta \dots \end{array}$$

i.e., c-command holds in a configuration in which  $\beta$  is contained in the node  $\gamma$  which has been merged with  $\alpha$ ;  $\gamma$  is also called the “sister node” of  $\alpha$ . So, we can say that  $\alpha$  c-commands  $\beta$  when  $\beta$  is contained in the sister node of  $\alpha$ .

C-command is a formal way to express structural prominence. For instance, as is clear from representations like (9), the subject c-commands the direct object, but not vice-versa: the object is contained in I', the sister node of the subject DP *John*<sub>i</sub>; the object does not c-command the subject because the subject *John* is not contained in the sister node of the object DP *Mary*, which is V.

C-command plays a pervasive role in the principles determining form and meaning of syntactic representations.

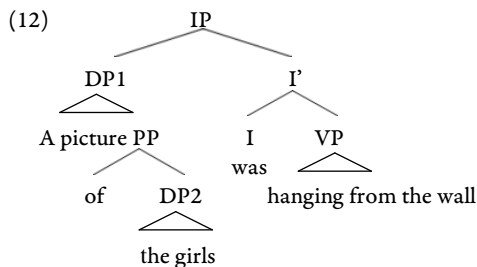
Consider for instance agreement: the subject, but not the object, agrees with the verb in person and number in English:

(10) John has (\*have) met the girls

This follows from the fact that the subject, but not the object, locally c-commands the node I, containing the inflected verb, in representations like (8). Similarly, in

(11) A picture of the girls was (\*were) hanging on the wall

The copular verb agrees with the whole subject DP1 *A picture of the girls*, which locally c-commands it, not with the adjacent DP2 *the girls*, which does not c-command it:<sup>8</sup>



In addition to being operative in all sorts of morphosyntactic processes, such as properties of the case-agreement system, c-

command is involved in interpretive processes. For instance, an anaphoric expression like a reflexive must be locally c-commanded by an antecedent which determines its reference. That is why a subject can bind an anaphoric object, but not vice-versa:

- (13) a *John*<sub>i</sub> criticized himself<sub>i</sub>  
 b \* *Himself*<sub>i</sub> criticized *John*<sub>i</sub>

Similarly, an anaphoric object can be bound by the subject DP, not by a possessive DP contained within the subject. I.e., the following sentences mean “John’s brother criticized John’s brother”, and cannot mean “John’s brother criticized John”, with the possessive acting as the antecedent (whether the possessive is prenominal, as in English, or postnominal, as in Italian):

(14) [[ *John* ]’s brother] criticized himself

(15) [ *Il fratello di [ Gianni ]* ] ha criticato se stesso

“The brother of John has criticized himself”

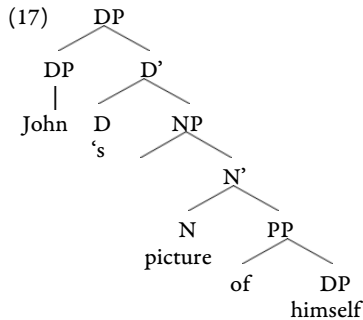
All these properties follow from the fact that anaphor binding requires c-command from the binder: the subject asymmetrically c-commands the object, whence the facts of (13); and a possessive DP does not c-command anything outside the DP that contains it, whence the interpretive properties of (14)-(15). If the anaphor is DP internal, the possessive can bind it because c-command holds:

- (16) [[ *John* ]’s picture of himself]  
 was hanging from the wall

Here too, binding is fine because the nominal specifier (the possessive) asymmetrically c-commands the nominal complement of *picture*:

- (16) [[ *John* ]’s picture of himself]  
 was hanging from the wall

Here too, binding is fine because the nominal specifier (the possessive) asymmetrically c-commands the nominal complement of *picture*:



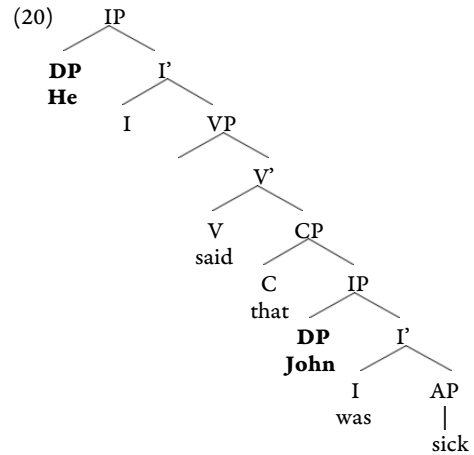
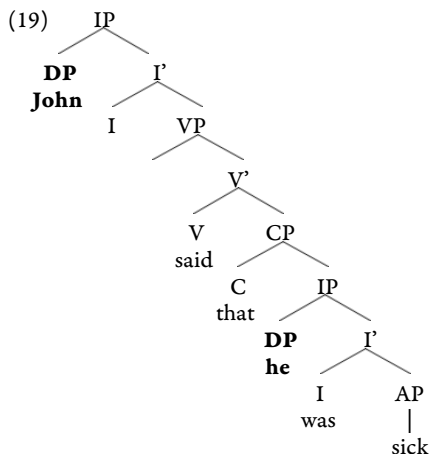
So, here too complement and specifier could not be reversed, because in that case the anaphor would not be properly c-commanded (hence bound):

(18) \* [[Himself]'s picture of John]  
was hanging from the wall

We are now ready to go back to the non-coreference pattern that we started with in section 2, which also crucially depends on c-command.

### Non-coreference

Consider the tree representations of (1) and (2):



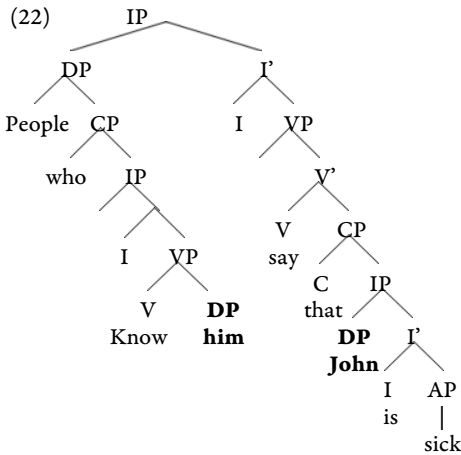
In a classical paper Howard Lasnik<sup>9</sup> observed that the pronominal DP has the name in its domain of c-command in (20), but not in (19), which led him to state the following interpretive principle:

(21) Non-coreference: A pronoun and a name cannot corefer when the pronoun c-commands the name.

So, (21) precludes coreference in (20): there the pronominal subject of the main clause c-commands everything else, including the name subject of the embedded clause. In every other environment, coreference is an option. For instance in (19) the c-domain of the pronoun is limited to the embedded clause, the name subject of the main clause is external to the c-domain of the pronoun, hence (21) does not apply and coreference is possible.

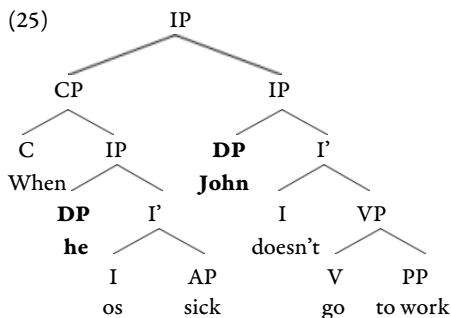
In (19) the name asymmetrically c-commands the pronoun, but coreference is also possible in the environments in which neither element c-commands the other, and the pronoun and the name have disjoint c-domains.

This happens, for instance, in (4), whose structural representation is expressed with some simplification by the following tree:

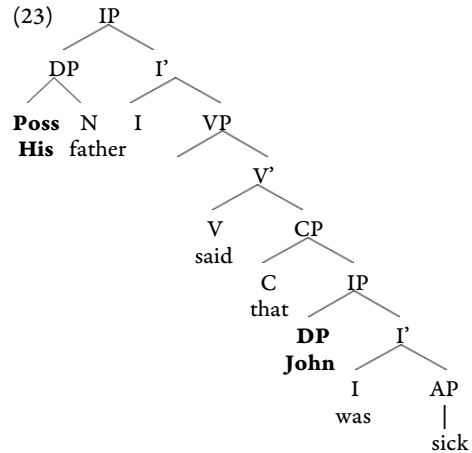


Here neither *him* nor *John* contain the other element in their c-domain, hence (21) does not apply and coreference is a viable option.

When the pronominal element is part of a larger phrase, as the possessive *his* in (5), it can corefer with a following name because the possessive does not c-command the name, as is clear from the following tree representation (the domain of c-command of the possessive is the DP *his father*, which does not include the name *John*):



Notice that possessives are not exempted from the non-coreference constraint: if the possessive c-commands the name, coreference is barred. So *his* cannot refer to *John* in the following, i.e., the sentence cannot mean “John’s picture of his (John’s) father is hanging from the wall”:



(24) \* *His<sub>i</sub> picture of John<sub>i</sub>'s father was hanging from the wall*

Finally, in (6), the pronominal subject of the preposed adverbial clause *When he is sick* does not c-command the main clause and its content, hence the name *John* and the pronoun *he* can corefer, as the non-coreference principle does not apply here:

In conclusion, a complex array of interpretive facts involving coreference follows from a simple principle such as Lasnik’s constraint, applying on hierarchical structural representations.

The subtle distinctions that we have observed in the complex non-coreference pattern raise an interesting challenge for an unstructured notion of “analogical generalization”. The language learner will have access to a sample of sentences containing pronouns; she will figure out from context that in some cases coreference with a nearby noun is intended, and in other cases it is not. E.g., in connection with (1):

(26) John said that he was sick, and that’s why he couldn’t come. He’s always looking for excuses! (coreference between *John* and *he* is likely to be intended)

(27) Nobody had understood why Bill had not showed

up, but then John said that he was sick (coreference between *John* and *he* is likely not to be intended)

On the basis of this evidence, why doesn't the learner simply conclude, by analogical generalization, that coreference is *always* an option? But no one draws that conclusion, as everyone systematically excludes coreference in (2).<sup>10</sup>

Even assuming that the child has some way of inferring that with structures like (2) coreference is never intended, hence that some constraint must be assumed, why wouldn't she assume a linear constraint and extend it by analogy to all cases of backward anaphora, thus implicitly assuming the linear rule (3) that is typically explicitly proposed in our toy experiment? But no language learner does that, as speakers readily recognize the possibility of backward coreference in contexts like (4), (5), (6). It seems clear that, in order to reach empirical adequacy, an analogy-based approach should build c-command into the notion of analogical generalization (something like "in evaluating coreference, analogize only constructions involving identical c-command configurations between nouns and pronouns"); but this move would *de facto* assume the highly structured configurational notions that the analogical approach is intended to avoid. In order to adequately capture what speakers know and do, reference to the structured hierarchical notion of c-command just seems unavoidable.

Then, the question arises of the "further explanation" of constraint (21): should it be considered a primitive principle of the human language faculty, or could it be derived as a theorem from deeper and more general principles? Obviously, the desirable option is that a successful path of "further explanation" may be identifiable. Here different approaches have been proposed. Chomsky originally suggested that (21) may follow from principle C, a component of the binding theory, the module expressing configurational constraints on the possible referential

dependencies between linguistic expressions; Reinhart<sup>11</sup> proposed that the non-coreference effect may have its roots in the computation at the interface between syntax and pragmatics, and be amenable to principles of optimization akin to Grice's maxims. And other approaches have been proposed.<sup>12</sup> I will not try to address the issue of the "further explanation" here. The relevant point is that all these approaches (definitely including Reinhart's interface approach) crucially make reference to the hierarchical configuration of c-command, which just seems unavoidable if the approach aims at meeting empirical adequacy, and capture the selective effects we have reviewed.

### Invariance and variation

Comparative considerations become relevant at this point. If non-coreference effects follow from general principles of syntactic organization and interpretation at the interfaces, one expects to observe little or no cross-linguistic variation, under the assumption that such principles are shared by our species. More precisely, one may expect variation, but not of an unconstrained kind: if general shared principles are involved, one may expect cross-linguistic studies to show a limited variation within a strongly invariant architecture. In fact, non coreference effects have been shown to hold in historically and typologically very different languages. A small sample:<sup>13</sup>

(28)

a (Italian)

\*        \_\_\_ pensa che Gianni vincerà  
          \_\_\_ thinks that Gianni will win

b (M. Hebrew)

\* hu ma'amin she-John yenaceax  
  he thinks    that John    will win

c (Thai)

\* khaw khit waa coon chálaát  
  he thinks    that John is smart

d (Vietnamese)



\* no' tin John se thang  
he believes John will win

e (Gungbe)

\* e vedo do Kofi na wa  
he believes that Kofi will come

f (Mohawk)

\* wa-hi-hrori-'tsi Sak ruwa-nuhwe'-s  
 fact-1sS/MsO-tell-punc that Sak FsS/MsO-like-hab  
 I told him that she likes Sak

Nevertheless, the recent literature reports that non-coreference effects of the simple kind considered so far are not, strictly speaking, universal. Some languages seem to allow coreference in a configuration in which the pronoun c-commands the name in particular structural configurations.

Davies<sup>14</sup> gives a comprehensive analysis of one such case, St'át'imcets (also known as Lillooet), an American Indian language member of the Salish family, spoken in the southwestern interior of British Columbia, Canada. The language manifests a certain freedom in word order, with a tendency to have predicate initial clausal structures and VOS order in transitive structures. Davis shows that for some speakers of the language, a pronominal subject of a main clause can be coreferential with a name or definite description in an embedded clause:

(29)

Tsút=tu7 [kw=s=cuz' nas ts'úqwaz'-am s=Mary natew].

say=then [DET=NOM=going.to go fish-MID NOM=Mary tomorrow]

"Mary said *she* was going fishing tomorrow".

More literally: *pro* said **Mary** was going fishing tomorrow.)<sup>15</sup>

(30)

Skenkín [lh=w=as nmatq xát'em ti7 ku=qelhmémen' sqaycw áta7 tsítcw-s=a].

slow [COMP=IMPF=3CNJ walk uphill that DET=old man to house-3POSS=EXIS]

"That old man walks slowly uphill to *his* house".

(More literally: *pro* is slow when **that old man** walks uphill to **his** house.)

In other words, for some speakers of this language, the equivalent of "She said that Mary was going to finish tomorrow" allows the reading "she = Mary" (while other speakers reject this reading, generally judged impossible by speakers of English).

What conclusion should be drawn from this element of variation for the general theory of non-coreference effects? The rational approach, here and elsewhere, is to study an exceptional case to a very general pattern with great care, in order to determine the exact scope and the fine structural properties of the "exception". This is what Davies does, and the conclusion he reaches is that the observed variation is not "wild", but highly constrained. The language offers clear evidence for a configurational organization, and a sensitivity of non-coreference effects from structural properties, such as c-command, e.g.:

1. Apparent violations of principle C in St'át'imcets are limited to cases in which the pronoun and the name are in two distinct clauses; if they are in the same clause, familiar non-coreference effects are found, much as in English and the languages in (28).<sup>16</sup> In other words, the St'át'imcets equivalent of (31)a allows coreference, while the equivalent of (31)b does not, much as its English counterpart:

(31)

a She said Mary was going fishing tomorrow (\* in English, ok in St'át'imcets)

b She smokes in Mary's house (\* in English, \* in St'át'imcets)

2. Violations of non-coreference are found for pronoun-name configuration but not for demonstrative – name (i.e., the equivalent of English "That one said that John was sick" disallows coreference between That one and John, and much as the English equivalent does.)

3. The language manifests Strong Crossover effects, which are traditionally seen as

cases of principle C violation. I.e., in English, while (33)a allows the bound reading of the pronoun (for which  $x$ ,  $x$  said that Mary likes  $x$ ), (33)b does not:

- (32) a Who<sub>i</sub> \_\_\_<sub>i</sub> said that Mary likes him<sub>i</sub>?  
 b \* Who<sub>i</sub> did he<sub>i</sub> say that Mary likes \_\_\_<sub>i</sub>

Davies shows that such strong cross-over effects hold in general in St'át'imcets as well, with no observed variation across speakers. This is illustrated by the following pair:

- (33)  
 a. *Swat* ku=kw7ikwłacw *e* [kw=s=cuz' melyíh *pro* kalál]?  
 who DET=dream [DET=NOM=going.to marry soon]  
 "Who *e* dreamed [*s/he* was going to get married soon]?"  
 b. \**Swat* ku=s-kw7ikwłacw-s *pro* [kw=s=cuz' melyíh *e* kalál]?  
 who DET=NOM-dream-3POSS [DET=NOM=going.to marry soon]  
 \* "Who did *s/he* dream [*e* was going to get married soon]?"<sup>17</sup>

Once the empirical scope of the exceptional behavior is identified, Davies adopts the approach to non coreference presented in Safir,<sup>18</sup> and argues in detail for a formal parametrisation of one of the principles of Safir's theory (within the guidelines of Chomsky's binding theoretic approach) to capture the observed cross-linguistic difference.

Going into the details of Davis's analysis is beyond the scope of the present article. Let me just observe that the important point to be retained from his approach for our theme is that non-coreference effects, like so many other linguistic phenomena, show elements of invariance and elements of variation. Here, as elsewhere, the observed variation never is "wild" and unconstrained: rather, languages typically allow "local" points of variation in an otherwise invariant structural architecture, based on configurational hier-

archies and relations such as c-command (as in the classical "principles and parameters" approach).<sup>19</sup> Some properties appear to be strictly universal (as the strong cross-over effects),<sup>20</sup> others appear to allow a limited range of variation within a tightly constrained and otherwise uniform system.<sup>21</sup>

## Conclusion

Humans constantly produce and understand sentences they have not heard before. This simple and crucial property of normal language use calls for a computational approach: knowing a language amounts to having mastered a generative device capable of computing new structures. Studying even the most elementary formal and interpretive properties of linguistic representations, we quickly realize that the structural organization of such representations is hierarchical, with relations like c-command playing a crucial role in all sorts of syntactic processes, in morpho-syntactic properties like agreement, in properties at the interface with semantics and pragmatics such as the determination of referential dependencies of pronouns and other nominal expressions. Generic notions of analogy and analogical generalization don't even begin to capture the fine properties and role of such hierarchical principles, which appear to be intimately related to the functioning of the fundamental combinatorial operations (recursive merge, in minimalist models).

Mental computations extend well beyond what is accessible to consciousness and introspection. We have conscious access to the representations which are computed, and this allows us to produce and understand new structures, and to express metalinguistic judgments of well-formedness and interpretation. But we have no introspective access to the underlying computational mechanisms. So, the only way to address and conduct the scientific study of our cognitive capacities for language is to proceed as we would in the study of any other natural object: formulate precise hypotheses, and submit them to em-

irical verification through techniques as diverse as possible.

We have only used a set of metalinguistic judgments in different languages in our illustrative example, but there is no reason to put any limitation to the kinds of linguistic evidence that we may want to consider: so, the study of language acquisition (as in the references mentioned above), language pathology, psycholinguistic experimentation, brain imaging studies can all be brought to bear on precise models of linguistic knowledge. If mental computations for language were accessible to consciousness, studying our combinatorial capacities would be straightforward; as linguistic computations are well beyond the reach of introspection, studying the properties of the system is a complex undertaking which requires detailed formal hypotheses and structured techniques of empirical testing, much as in the scientific study of any non-trivial aspect of the natural world.

### Acknowledgments

Thanks are due to Giuseppe Varnier for helpful comments on a preliminary version of this article.

### Notes

<sup>1</sup> See R. DESCARTES, *Discours de la méthode* (1637), Union générale d'éditions, Paris 1951; for a discussion see N. CHOMSKY, *Cartesian Linguistics. A Chapter in the History of Rationalist Thought*, Harper & Row, San Francisco 1966.

<sup>2</sup> R. DESCARTES, *Discours de la méthode*, cit., p. 86.

<sup>3</sup> See, e.g., M. TOMASELLO, *Do Young Children Have Adult Syntactic Competence?*, in: «Cognition», vol. LXXIV, n. 3, 2000, pp. 209-253.

<sup>4</sup> On some problematic aspects of the “item-based” neo-constructivist approach to the acquisition of syntax see J. FRANCK, S. MILLOTTE, R. LASSOTTA, *Early Word Order Representations: Novel Arguments against Old Contradictions*, in: «Language Acquisition», vol. XVIII, n. 2, 2011, pp. 121-135.

<sup>5</sup> See N. CHOMSKY, *The Minimalist Program*, MIT Press, Cambridge (MA) 1995.

<sup>6</sup> Without addressing technicalities, let us just say that DP, or “determiner phrase” is the label of a nominal expression, IP, or “inflectional phrase” is the label of a clause, and VP, or “verb phrase” is the label of the verbal predicate; for the sake of clarity I use here the familiar X-bar notation; nothing substantially changes, for the points at issue, if Bare Phrase Structure (see N. CHOMSKY, *The Minimalist Program*, cit.) is used, a more parsimonious notation adopted in much minimalist research. We omit many details here, and we don't discuss the algorithm through which the elementary structures created by Merge receive structural labels, a procedure which we leave at the intuitive level: a verb and a nominal expression merged together form a verbal projection, a verb phrase, etc. (see L. RIZZI, *Cartography, Criteria, and Labeling*, in: U. SHLONSKY (ed.), *Beyond Functional Sequence. The Cartography of Syntactic Structures*, Oxford University Press, Oxford/New York 2015, pp. 314-338. For details on the labeling algorithm see N. CHOMSKY, *Problems of Projection*, in: «Lingua», vol. CXXX, 2013, pp. 33-49, special issue *Core Ideas and Results in Syntax*).

<sup>7</sup> See T. REINHART, *The Syntactic Domain of Anaphora*, Ph.D. Dissertation, MIT, Cambridge, MA 1976; T. REINHART, *Anaphora and Semantic Interpretation*, University of Chicago Press, Chicago 1983; T. REINHART, *Coreference and Bound Anaphora: A Restatement of the Anaphora Questions*, in: «Linguistics and Philosophy», vol. VI, n. 1, 1983, pp. 47-88.

<sup>8</sup> If one assumes, as in many minimalist analyses, that agreement is checked by establishing an “agree” relation between the inflectional node and the subject DP in its thematic position, the agree relation reaches the subject DP1 and not DP2 because of locality. For Relativized Minimality see L. RIZZI, *Relativized Minimality*, MIT Press, Cambridge (MA) 1990; for minimal search see N. CHOMSKY, *Minimalist Inquiries: The Framework*, in: R. MARTIN, D. MICHAELS, J. URIAGEREKA (eds.), *Step by Step. Essays in Minimalist Syntax in Honor of Howard Lasnik*, MIT Press, Cambridge (MA) 2000, pp. 89-155. For evidence that agreement is checked both via the agree relation and in the local specifier-head relation discussed in the text see M.T. GUASTI, L. RIZZI, *On The Distinction between T and Agr: Evidence from Acquisition*, in: G. CINQUE (ed.), *The Structure of DP and IP. The Cartography of Syntactic Structure*, vol. I, Oxford University Press, New York/Oxford, pp. 167-194; J. FRANCK, G. LASSI, U. FRAUENFELDER, L.

RIZZI, *Agreement and Movement – A Syntactic Analysis of Attraction*, in: «Cognition», vol. CI, n. 1, 2006, pp. 173-216.

<sup>9</sup> See H. LASNIK, *Remarks on Coreference*, in: «Linguistic Analysis», vol. II, n. 1, 1976, pp. 1-22.

<sup>10</sup> Not only adults, but also children round age 3 or earlier, as shown in experimental studies such as S. CRAIN, C. MCKEE, *Acquisition of Structural Restrictions on Anaphora*, in: S. BERMAN, J. CHOE, J. MCDONOUGH (eds.), *Proceedings of the NELS XVI, GSLA*, University of Massachusetts, Amher 1986, pp. 94-110; M.T. GUASTI, G. CHERCHIA, *Backward vs. Forward Anaphora in Child Grammar*, in: «Language Acquisition », vol. VIII, n. 2, 1999-2000, pp. 129-170. See also S. CRAIN, R. THORNTON, *Investigation in Universal Grammar*, MIT Press, Cambridge (MA) 1998.

<sup>11</sup> See T. REINHART, *Anaphora and Semantic Interpretation*, cit.; T. REINHART, *Coreference and Bound Anaphora: A Restatement of the Anaphora Questions*, cit.

<sup>12</sup> See e.g. P. SCHLENKER, *Non Redundancy. Towards a Semantic Reinterpretation of Binding Theory*, in: «Natural Language Semantics», vol. XIII, n. 1, 2005, pp. 1-92.

<sup>13</sup> Notice that Italian is a Null Subject (or pro-drop) Language, i.e., a language allowing phonetically null pronominal subjects. With respect to the non-coreference effect, the null pronominal subject of Italian behaves exactly as the overt pronominal subject of English.

<sup>14</sup> See H. DAVIES, *Cross-linguistic Variation in Anaphoric Dependencies: Evidence from the Pacific Northwest*, in: «Natural Language and Linguist Theory», vol. XXVII, n. 1, 2009, 1-43.

<sup>15</sup> *Ivi*, p. 20.

<sup>16</sup> See L. MATTHEWSON, H. DAVIS, D. GARDINER, *Coreference in Northern Interior Salish*, in: W. Shipley

(ed.), *28th International Conference on Salish and Neighboring Languages*, University of Washington, Seattle 1993, pp. 217-232.

<sup>17</sup> H. DAVIES, *Cross-linguistic Variation in Anaphoric Dependencies: Evidence from the Pacific Northwest*, cit., p. 56. .

<sup>18</sup> See K SAFIR, *The Syntax of (In)dependence*, MIT Press, Cambridge (MA) 2004.

<sup>19</sup> See N. CHOMSKY, *Lectures on Government and Binding*, Foris Publications, Dordrecht 1981.

<sup>20</sup> See P. POSTAL, *Skeptical Linguistic Essays*, Oxford University Press, Oxford 2003.

<sup>21</sup> The special parametric property permitting “upside down” coreference in cases like (29) may well be linked to other properties of the language. Davies speculates that it may be connected to another special property of pronouns in the language: their ability to be introduced in discourse contexts without a prior discourse referent. See also Vincent Homer (V. HOMER, *Backward Control in Samoan*, in: S. CHUNG, D. FINER, I. PAUL, E. POTSDAM (eds.), *Proceedings of the 16<sup>th</sup> Meeting of the Austronesian Formal Linguistics Association (AFLA)*, University of California, Santa Cruz 2009, pp. 45-59), who proposes to link analogous “upside-down” coreference options in Samoan to “backward control”, another “upside down” case of referential dependency (see M. POLINSKY, E. POTSDAM, *Backward Control*, in: «Linguistic Inquiry», vol. XXXIII, n. 2, 2002, pp. 245-282). Whatever other property may be linked to “upside-down” coreference, learnability considerations suggest that the special parametric value permitting it must be a marked property, assumed to hold by the language learner only in the presence of evidence directly justifying the choice of this option, which is otherwise not entertained by the learner.