

THE SUBTLE DEPENDENCY BETWEEN COMPETENCE AND PERFORMANCE

CRISTIANO CHESI
ANDREA MORO

Center for Neurocognition, Epistemology and theoretical Syntax (NEtS), IUSS Pavia

1 Introduction to the classic competence vs. performance divide

Since Chomsky's (1957) seminal work, we think of a *grammar* as a formal device (G) providing an explicit description of all productive constraints (e.g. rewriting rules) that restrict the possible geometry of all the significant relations among words (or morphemes) and other (intended) syntagmatic categories. A constituency-based graph (i.e. a syntactic tree) represents the result of the application of the relevant rules and principles and indicates, unambiguously, both first order relations, like precedence and dominance among (certain) nodes, as well as second order relations like ((anti-)symmetric) C-command. The primary goal of G , for a given language L , is to distinguish "grammatical" sequences of words from "ungrammatical" ones, where L is meant to be an infinite set of word sequences each associated with (at least) one relevant structural description. In mentalistic terms, G expresses the *linguistic competence*, that is "speaker/hearer's knowledge of his/her language" (Chomsky 1965:4).

G is *descriptively* adequate with respect to L , if and only if it succeeds in generating nothing but the infinite set of sentences that are accepted by the native speakers of L and if, in addition to the exact definition of the L boundaries, it also suggests how the accepted sentences are understood by an "ideal speaker-hearer" (Chomsky 1965:5). In this sense, G is a theory of L (Chomsky 1965:24) since it describes all possible sentences in a given corpus but also any new possible sentence that is not (yet) in any given finite collection of well-formed expressions.

Notice that any grammatical formalism G must be used "mechanically", that is, given an adequate algorithm (potentially included in G itself), we should always be able to decide if a sentence is well-formed or not without any ambiguity and without any other external knowledge. Moreover, in order to be computable, the algorithm should attain to this goal in a finite amount of time, that is, the computation should terminate in a finite number of steps.

This clear picture might get blurred by the fact that, given a phrase structure grammar G , many different algorithms can perform the generation (and the recognition) task by making use

of the constraints expressed by G : rewriting rules of the context-free kind¹, for instance, are descriptive constraints that can be used by different parsing algorithms in order to recognize² the structure of a well-formed sentence in a finite amount of steps (see Earley or CKY algorithms, Jurafsky and Martin 2009). The compatibility of G with many algorithms might be problematic since it might well be the case that for a given sentence s , different algorithms on G could take more or less time, or, in certain cases, not even terminate (§2). In more precise terms, different algorithms for G can have different complexity, where complexity can be precisely expressed in terms of time (number of steps) and space (memory needed to store working information) as discussed by Barton, Berwick and Ristad (1987).

A problem arises if we assume that the human *linguistic performance*, that is the actual use of the linguistic competence (Chomsky 1965:10) taking into consideration time and space limits, is an approximation of the behavior of a specific algorithm A applied to G . On the one hand, assuming some restriction on A is the only way to justify a divergence from the intuition of the *real native speaker (RS)* and the *ideal speaker (IS)*; on the other, since many algorithms A can be applied to G and many grammars G s can circumscribe similar set of expressions, both including or excluding certain critical sentences for which the actual and the ideal native speaker intuition diverges, we are left with the ambiguity of deciding whether A , G or both are inadequate.

In fact, we can find examples of two different kinds of divergence between RS and IS : on one side, there are sentences that should be well-formed, according to G , but that are considered ill-formed according to RS (e.g. self-embedding, (1), Chomsky 1965:11, Miller & Chomsky 1963), on the other, there are expressions that should be ill-formed but that are processed consistently by real speakers, most of the time being accepted, like (2) (i.e. subject islands violation with pied-piping, Bianchi & Chesi 2014, in press):

- (1) [The man [who the boy [who the students recognized] pointed out] is a friend of mine]
 (2) [Of which masterpiece], is [one reproduction _i] already available?

Since, for the great majority of cases, IS and RS intuitions seem to match, the situation is less disturbing than expected. There are however extreme cases, as usually in science, as (1) and (2), that are richly informative about the competence vs. performance divide and that ask for an explanation. In this paper, we suggest that the tension related to this divide did not disappear within the last 50 years (indeed, it has been increased by the advent of the Minimalist Program), then we want to dig into some crucial evidence, following Chomsky's original intuition, namely that there is no a serious performance model that is not "competence-based" (Chomsky 1965:10). In the end we will conclude that a linguistic theory that subsumes the real speaker performance in contexts like (1) and (2) is explanatory more adequate than a theory that relegates (1) to performance facts and (2) to exceptional cases.

¹ Context-Free Grammars (CFGs) are Type 2 phrase structure grammars in the Chomsky hierarchy. CFGs use rewriting rules of the form " $A \rightarrow \gamma$ " where " A " is a non-terminal symbol and " γ " represents any sequence of terminal or non-terminal symbols (included the null one) and " \rightarrow " is the rewriting symbol expressing immediate dominance: only one non-terminal symbol is allowed to the left side of the rewriting symbols in CFGs and anything is possible to the right side.

² In psycholinguistics terms, *parsing* and *production* refer to human performance tasks; in recursion and automata theory, the corresponding tasks are usually called *recognition* and *generation* respectively. Here we will use parsing and recognition, as well as production and generation as synonyms, both indicating mechanical procedures.

In order to discuss these points, the paper is organized as follows: first, we will state in a formally clear way the “competence paradox”, that the complex relation between algorithmic and constraining part of *RS* and *IS* (§2); then we will present some relevant data on self-embedding and extraction from subject islands, pointing out to new empirical evidence not yet available in the ’60s (§3). It will be clear in the end (§4) that Minimalism (Chomsky 1995) exacerbates the competence paradox, since constraints on the possible algorithmic procedures are provided that are not easily compatible with any performance task. In the end, we will sketch few alternatives that attempt at bridging the gap between *RS* and *IS*, suggesting configurational triggers for non-local relations (on the line of Moro 2000) and a different orientation of the structure building operations (Chesi 2014).

2 The competence paradox

As long as the infinite set of well-formed sentences is independently and precisely defined (that is, there is at least an ideal speakers that consistently generates and recognizes all well-formed sentences on the one hand and rejects ill-formed ones on the other for a given language) the “*competence problem*” is well stated. Unfortunately, there is no ideal speaker’s competence of L that is distinguishable by $A(G)$, where A is an algorithmic procedure taking as argument the grammar G , and both A and G can be independently tuned in order to generate or recognize L^3 .

Let us exemplify this with a simple, non-trivial (namely recursive) CFG that generates any possible sequence of (non-null) *as* and *bs* in this order (S , as usual, is the starting symbol; rules are numbered for convenience):

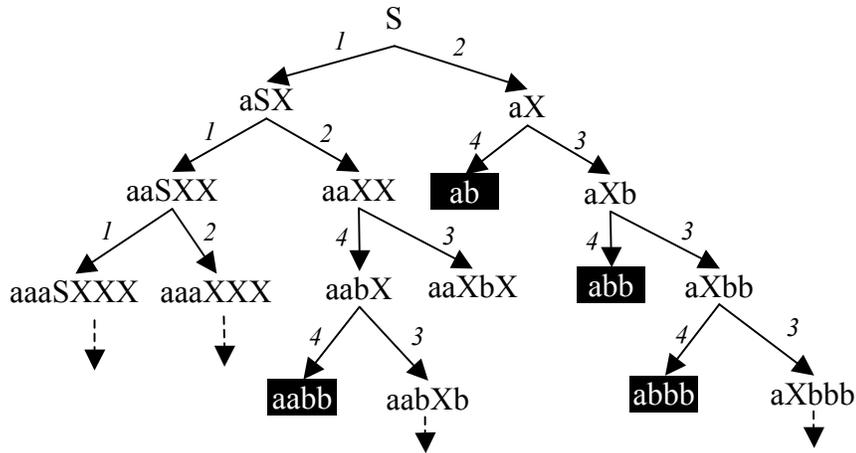
- (1) $G =$
 1. $S \rightarrow aSX$
 2. $S \rightarrow aX$
 3. $X \rightarrow Xb$
 4. $X \rightarrow b$

Rewriting rules “ \rightarrow ” can be read bottom-up (from the right side of the rewriting symbol to its left side) or top-down (in the opposite way, that is the standard sense of the rewriting symbol). This already permits two different kinds of generation algorithms based on distinct problem spaces⁴ that are the result of different order of application of the possible rules; below an example of the problem space generated by a top-down, left-right expansion of non-terminals (left-most symbol is expanded first):

- (2) Partial top-down, left-right problem space expansion (the states of the problem space are labeled with the result of the substitution; the labels on the arrows leading to a certain state indicate the rule applied; black boxes are the generated terminal strings that end the derivation)

³ One might object that A can be included in G . This possibility (exploited in Minimalism) is not logically distinguishable from the one we discuss here, since if there is an algorithmic part in G , independent of the constraints/principles/rules expressed by G , we can always consider this as A , hence unbundle the algorithm from G (e.g. separating structure building operations, A , from lexical features, G , just to use familiar minimalist terms).

⁴ The problem space represents all possible reachable states that can be obtained by applying legal rules.



Assume that the algorithms we want to test only deal with serial expansions⁵ and that we want to compare how “complex” is generating the string “ab” with these different algorithms, all using the grammar in (1). Below the result of the relevant exploration of the problem spaces generated using 5 different algorithms:

- (3) A1. top-down, left-right, breath-first (short-rule first)⁶:
 goal: *ab*; computation: *S, aX, aSx, ab* (it terminates in 3 steps)
- A2. top-down, left-right, breath-first (long-rule first):
 goal: *ab*; computation: *S, aSx, aX, aaSxx, aaXX, ab* (it terminates in 6 steps)
- A3. top-down, left-right, depth-first (long-rule first):
 goal: *ab*; computation: *aSx, aaSxx, aaaSxxx...* (it never terminates!)
- A4. Bottom-up, left-right, depth-first (greedy, short-rule first)⁷:
 goal: *S*; computation: *ab, Xb, Sxb, X* (solution not found!)
- A5. Bottom-up, right-left, greedy (short-first):
 goal: *S*; computation: *ab, aX, aXb, S* (solution found in 4 steps)

As it is clear, considering the (simpler) generation problem, there are many non-equivalent algorithms all using *G*. Their non-equivalence is demonstrated by the fact that some algorithm does not even terminate (*A3*) while others might conclude the computation without finding a solution (*A4*); their different complexity, on the other hand, is suggested by the fact that *A1* finds the relevant string in 3 steps, *A2* in 6 steps and *A5* in 4 steps (as we said, time complexity is

⁵ In case of multiple possible application of different rules the algorithm expands one rule per time. This is the case of rule 1. and 2., for instance: reading these rules top-down, they both can be applied whenever *S* is found. This produces branching as indicated in (2) (first bifurcation after the root state, marked as “*S*”).

⁶ Breath-first is opposed to depth-first: in the breath-first case, all alternative expansions are evaluated before going deeper; in the depth-first case, one expansion is pursued up to a terminal string is rewritten.

⁷ “Greedy” means exploring first the rules with the maximum number of matches: if “*ab*” is the sequence to match, and there are two rules compatible with the left-prefix, “*X*→*ab*” and “*Y*→*a*”, *X* will take over.

measurable also in terms of number of steps)⁸. This means that if we want to simulate a performance algorithm, we have many different alternatives that produce different (sometimes conflicting) outcomes.

From a generative point of view, any algorithm that explores the problem space completely would simulate the *IS* competence (e.g. A1 or A2), but if we want to explain the possible mismatch between *RS* and *IS*, there are just few possible solutions that we could consider:

(4) *a. Algorithmic Independence*

IS and *RS* use different algorithms to explore the problem space generated from *G*;

b. Algorithmic Equivalence

IS and *RS* use the same algorithm *A*, but *RS* is affected by complexity (e.g. longer derivations are harder to be computed);

c. Algorithmic Equivalence with independent constraints sensitivity

IS and *RS* use the same algorithm, but *RS* is sensitive to some property of the constraints applied and *IS* to others (or none);

In the first case (Algorithm Independence) A_i and A_r (i.e. the algorithms used by *IS* and *RS* respectively) are different and it is possible to expect that for some $s \in L$, s is generable by A_i and not by A_r (or the way around); though logically possible, this solution is highly unexplanatory and requires the formulation of two unrelated procedures; moreover it can not guarantee whether A_i or G are inadequate: in order to account for (2), we should first guarantee that (2) is generable given an appropriate G , then that A_i prevents (2) from being generated while A_r generates it, in the end, giving the suspicion that A_i is simply inadequate. The hypothesis (4).b (Algorithmic Equivalence) is somehow more explanatory since just one algorithm needs to be stated ($A_i = A_r$), but it is a matter of fact that its complexity is the only restriction affecting A_r ; in fact, it is impossible to explain cases in which a sentence s , such that $s \notin L$, is generable (and recognizable) by A_r and not by A_i . Under this circumstance, the language L_r generable by A_r should be fully included in L (namely by the language generable by A_i). Last hypothesis, (4).c (Algorithmic Equivalence with constraints sensitivity), could predict that certain rules chain, like expansions of the same recursive rule in sequence, is penalized by A_r that simply confounds similar rules in self-embedding cases (much on the line of Miller & Chomsky 1963), but how to explain (2)? In the end, all these options could predict a divergence between *IS* and *RS*. Notice however that in all cases, we are left in a fideistic condition in which the adequacy of G (as stated in §1) is guaranteed only by the grammar itself. This leads to the *competence "paradox"*:

(5) *The competence "paradox"*

- i. Given A_i , and A_r , two distinct algorithms (inspired by the *ideal* and the *real* native speakers of L , respectively), a grammar G is adequate with respect to L , if and only if $A_i(G)$ generates nothing but any possible sentence $s \in L$ where the relevant set L is independently generated by *RS* (the real native speaker of L).
- ii. If *RS* and $A_i(G)$ diverge, G is adequate if only if there is A_r such that $A_r(G)$ generates nothing but any possible sentence generable by *RS*.
- iii. If A_r exists, there is no reason to assume A_i . If A_r does not exist $A_i(G)$ (hence G) is not adequate.

⁸ It would be interesting to try to demonstrate whether or not an algorithm can uniformly find a solution quicker than other algorithms (see Russel & Norvig 2003:91), but this goes beyond the scope of this paper.

This is a “paradox” since we concluded that there is not theory of competence if there is no theory of performance! Indeed, we think we can make sense out of it: at the beginning of ‘900, the physicists gave birth to quantum theory on the basis of a wave theory of light built on controversial experimental observations (Mehra and Rechenberg, 1982). Taking inspiration from this, we should not relegate systematic RS data to the “performance” rubric, but integrating them, as best as possible, within our theory of *L*. This could lead to less paradoxical conclusions.

3 Rediscovering (and refining) the same old data

Since *Aspects*, the empirical evidence greatly increased in the last decades, both in terms of numbers and quality of sources. In this section we will briefly review the main evidence bearing on the two cases we want to discuss: self-embedding and island domains. In the first case complexity signatures (including brain imaging studies) will be used to show that non-local dependencies are sensitive to feature-matching in a non-trivial way (§3.1); in the second case, we will show that we can deep our understanding using very subtle oppositions by teasing apart variables whose correlation is only evident using robust experimental methods (§3.2).

3.1 New evidence on self-embedding and intervention

Self-embedding is a special case of a nested construction in which two constituents, A and B, with B self-embedded in A, are of the same kind and B falls totally within A (that is, A dominates B and there is some non-null element of A to both the right and to the left side of B, Chomsky 1965:12). The example in (1), repeated below for convenience, is exactly a case of self-embedding, since one relative clause (RC) is fully included in the other, while the subject and the verb of the superordinate RC respectively precedes and follows the whole inner RC:

(1) ??[The man [_{RC} who the boy [_{RC} who the students recognized] pointed out] is ...]

Speculating on the *unacceptability* (which is a matter of performance, as opposed to *ungrammaticality*, which is a matter of competence) of this configuration, Chomsky mentions to “suggestive observations concerning limitations on performance imposed by organization of memory and bounds on memory” (Chomsky 1965:10). He also concludes that “if [the observation that self-embedding contributes even more radically than nesting to unacceptability] is correct, then we have evidence for a conclusion about organization of memory that goes beyond the triviality that it must be finite in size” (Chomsky 1965:14). This is because we should not expect an asymmetry in acceptability between standard nesting and self-embedding if memory size limits would be the only parameter to be tuned to explain RS and IS asymmetry.

As a matter of fact, the presence of similar asymmetries has been deeply studied in RCs and, for instance, the opposition between Subject (SR, (6).a) vs. Object Relative clauses (OR, (6).b) has been widely documented:

- (6) a. The boy [who recognized *the students*] is a friend of mine
 b. The boy [who *the students* recognized] is a friend of mine

The contrast between SRs and ORs, with SRs easier to be generated (and understood) than ORs, is very solid and has been reported using many different approaches: King and Just (1991) started a long tradition of self-paced reading experiments and described a significant slowdown in the critical regions: the RC subject and verb, in ORs, compared to verb and object in the SR condition, revealed longer reading times. These results match with the evidence revealed using standard complexity measurements in eye-tracking experiments (e.g. first-pass and first-pass regressions within the relative clause region⁹, Traxler et al. 2002). The same contrast has been investigated also using brain activity analysis: Just et al. (1996) revealed an increase in the volume of neural tissue activation (number of voxels, i.e. volumetric pixels, produced by an fMRI imaging system¹⁰), mainly in language centers (both on Broca's and Wernicke's areas), when ORs are processed (e.g. "The reporter_i [that *the senator* attacked _i] admitted the error") with respect to SRs condition (e.g. "The reporter_i [that _i attacked *the senator*] admitted the error"); both SR and OR show increased activation with respect to a baseline activation produced by two coordinated active sentences (e.g. "The reporter attacked the senator and admitted the error"). Their interpretation of this fact relies on a left-right processing assumption: the coordinated active sentences are uninterrupted, since the DPs feed the required subject positions in both conjoined active sentences, consistently with canonical SVO order. On the other hand, SRs are more complex than coordinated structures since the matrix sentence (in the example above "The reporter admitted the error") is 'interrupted' by the RC (i.e. "the reporter_i [*that _i attacked the senator*]") leading to embedding, in the sense discussed above. Here the relative clause head ("the reporter" in the example¹¹) feeds the subject position both in the matrix clause and in the relative clause. This does not happen in ORs, where the RC has its own subject, ("*the senator*", in the example) which intervenes between the RC head and the gap in the object position within the ORs (*the reporter_i [that the senator attacked _i]*).

The debate on such asymmetry source has been reactivated recently: Friedmann et al. (2009) propose that the greater processing complexity of OR with respect to SR might be due to purely grammatical facts. Their solution goes in the direction that formal constraints, like the locality principle (Rizzi 1990), which are part of our competence, if properly interpreted, can lead to an explanation of performance effects. In the specific case, when a non-local relation is established between X (the head of a non-local dependency) and Y (the tail of such dependency) and a structural intervener Z is present in between, processing difficulty is assumed to be proportional to the number and kind of relevant features shared between the moved item and the intervener. In their work, three main cases are discussed: the *identity* case, in which X and Z share the crucial features triggering the dependency (e.g. [_{+A} X], [_{+A} Z]); the *inclusion* case, in which Z has only a subset of the features of X needed to trigger the dependency (e.g. [_{+A +B} X] and [_{+A} Z]) and the *disjunction* case, in which X and Z are fully distinguishable and the relevant features on X triggering the dependency are not present in Z (e.g. [_{+A} X], [_{+B} Z]). The *identity* case leads to full

⁹ "First-pass" is the sum of all the fixations beginning with the reader's first fixation of a region until the reader's gaze leaves such region (i.e. the relative clause). "First-pass regressions" counts how many times the reader's gaze crosses the left edge of a relevant region (e.g. the verb segment within the RC) after a first-pass fixation.

¹⁰ Functional Magnetic Resonance Imaging (fMRI) is an imaging technique that allows us to visualize neural activity in vivo by detecting associated changes in blood flow by using blood-oxygen-level-dependent (BOLD) contrast detectors. fMRI has an excellent spatial resolution and a decent temporal resolution.

¹¹ We have no space here to discuss different interpretations of the relation between the head and the gap in restrictive RC. However both raising and matching analyses require a non-local relation between the head and the gap to be established (see Bianchi 2002 for a review). See also Friederici (2012) for a comprehensive review on the neurological evidence on RCs processing.

ungrammaticality (as in the classic *wh*- islands effects (e.g. “[_{+wh} *what*]_X do you wonder [_{+wh} *who*]_Z read _X”), while *inclusion* and *disjunction* are generally accepted with *inclusion* (e.g. “[_{+wh} *which book*]_X do you wonder [_{+wh} *who*]_Z read _X”) harder to be processed than *disjunction* (e.g. “[_{+wh} *what*]_X do you wonder [_{+NP} *John*]_Z read _X”). This assumption seems to be borne out since, exactly in the case of restrictive relative clause, ORs (and not SRs) create an *inclusion* context, because of the presence of the overt subject.

The theory is fine grained enough also to account for subtle differences in the featural make-up of the constituents involved in the dependency: studying object cleft sentences, by means of self-paced reading experiments, Warren and Gibson (2005) revealed a scale of complexity dependent on the DPs occupying the head and the subject of the cleft sentence, with DP_S type varying from definite descriptions to proper names or pronouns. The rough results of their experiments assessing all possible DP-type combinations are reported below (Warren p.c.):

(7) Sentence	reading time (ms) at avoided (SD)
a. It was <i>the banker</i> that <i>the lawyer</i>	avoided _ at the party 365 (19)
b. It was <i>the banker</i> that <i>Dan</i>	avoided _ at the party 319 (12)
c. It was <i>the banker</i> that <i>we</i>	avoided _ at the party 306 (14)
d. It was <i>Patricia</i> that <i>the lawyer</i>	avoided _ at the party 348 (18)
e. It was <i>Patricia</i> that <i>Dan</i>	avoided _ at the party 347 (21)
f. It was <i>Patricia</i> that <i>we</i>	avoided _ at the party 291 (14)
g. It was <i>you</i> that <i>the lawyer</i>	avoided _ at the party 348 (18)
h. It was <i>you</i> that <i>Dan</i>	avoided _ at the party 311 (15)
i. It was <i>you</i> that <i>we</i>	avoided _ at the party 291 (13)

Warren and Gibson explain the fact that the full DPs matching condition (7).a was harder than (7).b (full DP - proper noun mismatch condition) or (7).i (pro matching condition) under a memory load account, since inserting a definite DP has an integration cost that is greater than the cost of integrating a proper noun or a pronoun in the structure; this is because proper nouns and pronouns are lower on the referentiality hierarchy scale (Warren & Gibson 2005). Their account¹² however falls short in explaining why the proper noun matching condition (7).e registers reading times on the critical verbal region that are comparable with the definite description matching condition (7).a. Due to the accessibility hierarchy scale, proper nouns should have been uniformly more accessible (hence faster to process) than definite DPs.

On the other hand, embracing a feature-based perspective, where definite descriptions constituents decorated by [+D N] features, proper nouns by [+D N_{proper}] features (with N and N_{proper} distinguishing the “lexical restriction”, on the line of Belletti and Rizzi 2013) and finally pro with just [+D] features (because of the missing lexical restriction), we would expect definite description and proper noun matching conditions to be harder than the pro matching condition as

¹² Here we simplified a bit their points in order to keep the discussion focused on few relevant contrasts. Moreover, we do not have space to discuss other similarity-based accounts as well as “usage-based” models. See Belletti and Rizzi (2013) for a review.

well as other mismatch condition since in the number of feature shared is greater in (7).a and (7).e than in the other cases¹³.

In conclusion, we observe that experimental evidence suggested a gradable description of the RS performance and that this might be explained in a promising way looking at formal constraints that characterize also IS competence, that is, the sensitivity to locality constraints expressed in terms of relevant features. This seems to confirm Chomsky's intuition that "memory" (where "memory" is naively consider any kind of formal device that would make possible a non-local dependency) is sensitive to precise aspects of the phrase structure. Indeed, here we realized that "bracketing", as in embedding, plus "labeling of bracketing" (Chomsky 1965:10) might not be enough to explain these facts; accessing a deeper feature structure, as well as taking position on the directionality of the algorithm used might be necessary.

3.2 New evidence on Subject Islands

On the other side of our grammatical boundaries, there are cases that create the converse problem of self-embeddings: while in §3.1 we dealt with sentences that should be generable by IS but that are hard (or impossible) to be processed by RS, here we will focus on sentences that can be correctly processed by RS but that are usually supposed not to be generable by IS. This is the case of subextraction from certain subject islands (example (2) in §1 repeated below):

(2) [Of which masterpiece]_i is [one reproduction __i] already available?

Notice that the subject constituent in the preverbal position is a (simple) case of left-branching construction in the sense of Chomsky (1965:12): Chomsky notices that left-recursion reduces acceptability; even though he focused on genitive constructions like "[[[[John]'s brother]'s father]'s uncle]" (1965:13), also right-recursion on the left-branch of the tree (i.e. ROs formation on preverbal subjects) creates self-embedding (this might suggest that the cases discussed in §3.1 and the one we will discuss here might not be so unrelated).

Our argument here hinges on the fact that the idealization of subject islands, considered as absolute domain from which subextraction is always banned (Huang 1982), has been seriously challenged both from a grammatical and from a psycholinguistic point of view recently. From the grammatical perspective, Chomsky himself (contradicting *Barriers* assumptions, Chomsky 1986:31¹⁴) observes that certain subjects allows for subextraction (Chomsky 2008:147):

¹³ Belletti and Rizzi 2013 assume that the head of the cleft is marked with a +R feature that distinguishes the cleft head from the subject. This permits to consider these contrasts under the *inclusion* case discussed above. See Chesi and Moro (2014) and Chesi (2014) for precise complexity metrics on these data.

¹⁴ Chomsky reported an elusive contrast in subextraction with respect to preposition stranding. On extraction from adjuncts, preposition stranding seems to be favored (Adriana Belletti's original observation):

- (i) a. *he is the person *to whom* [they left [before speaking _]]
- b. ?he is the person *who* [they left [before speaking to _]] (before meeting _)

While on subject subextraction the opposite seems to be true:

- (ii) a. ?he is the person *of whom* [pictures _ are on the table]
- b. *he is the person *who* [pictures of _ are on the table]

Chomsky decided to "put these matters aside here, merely noting a potential problem" (Chomsky 1986:32); see Bianchi and Chesi (in press) for a proposal of explanation of the contrast in (ii).

- (8) a. *[Of which car]_i did [the (driver, picture) __i] cause a scandal?
 b. [Of which car]_i was [the (driver, picture) __i] awarded a prize?

Unless we assume that between the PP and the gap there is no a real dependency¹⁵, we should enforce our grammatical power in order to include selectively (4).b and keep out (4).a. Bianchi and Chesi (2014) reviewed different proposals aiming at explaining the intricacies within the exceptional cases of subject island subextraction, eventually concluding that:

- i. Those are real case of “sub-extraction” and on the dependency formation constraints active in (4) must be explained (pace Longobardi 1991¹⁶ and Cinque 1980¹⁷);
- ii. The oppositions between base vs. derived subject position (Takahashi 1994) as well as internal vs. external argument (Chomsky 2008), per se, are not sufficient to explain the contrast revealed in (4)¹⁸;
- iii. The sole possible explanation is considering the interpretation of the subject (on the line of Diesing 1992), distinguishing presuppositional subjects (i.e. semantic categorial structures, in the sense of Ladusaw 1994, in which the subject is interpreted outside the predicative nucleus) from non-presuppositional ones (i.e.thetic structures, Ladusaw 1994, in which the subject must be interpreted within the predicative nucleus); from the first, extraction is banned; from the second, that are interpreted in the “reconstruction” (base-generated) site, extraction is possible.

The conclusion that semantic interpretative options must be considered during the derivation does not weaken the main assumption on competence but simply suggests that also interpretative constraints must be taken into consideration by the algorithmic part of the grammatical knowledge. In fact, assuming standard brick over brick (that is merge-based) phrase structure building we are left in an unsatisfactory condition: we should decide whether a certain subject will be interpreted in the base position or not (or, even worse, undo an already occurred movement operation forcing “reconstruction” in the base position) on the basis of part of the predicative structure that is not merged yet within the structure.

In the end, we would briefly consider a “processing-based” perspective, from which an island violation is simply a matter of performance. Kluender (2004), among many others, suggests that the unacceptability of subextraction (also) from subjects depends mainly on the special status of

¹⁵ As discussed in Jurka (2010:151), the PP in (8) might be base-generated in a hanging topic position (paraphrasing (8).b: “of which car was it the case that the driver of this car was awarded a prize?”); it goes without saying that this solution would not explain the contrast between (4).a and (4).b since also (4).a paraphrase would lead to a grammatical sentence: “of which car was it the case that the driver of this car caused a scandal?”.

¹⁶ Following Longobardi (1991), the presence of a definite article could suggest that the wh- phrase is not really extracted but directly generated in a topic-like position leading to a ‘possessive’ interpretation within the subject (where the possessor is implicit). This solution however does not explain the contrast discussed in Bianchi & Chesi (2014): partitive subjects permit the resumptive “possessor strategy” (iii).a but not for extraction (iii).b:

- (iii) a. A proposito di Maria, ci ha telefonato uno dei figli
 speaking of Maria us has phoned one of_i the sons (*Speaking of Maria, one of her sons called us*)
 b. *Maria, di cui ci ha telefonato uno dei figli
 Maria of whom us has phoned one of_i the sons (*Mary, one of whose sons called us*)

¹⁷ Following Cinque (1990), we should expect a DP, not PP, to be able to exploit the proposed null resumptive pronoun strategy in case like (4).b, that according to Cinque (1990) are just apparent cases of extraction.

¹⁸ As for the base vs. derived subject position, the example (4).b suggests that also the subject in a derived position can permit subextraction; as for the internal vs. external argument, the relevant counter example is provided by Jiménez Fernández (2009):

- (iv) *De qué cantante* te parece que [algunas fotos _] han escandalizado a la audiencia
 of which singer you seem that some pictures have shocked to the audience

the pre-verbal subject in SV languages: first, as attested by the preference for subject over object *wh*- dependency formation (see §3.1), the subject constituent introduces a “memory load” whenever a non-local dependency spanning over it must be established; second, the subject also induces an “integration cost” in terms of new referents to be computed (Gibson 1998, §3.1). Despite the fact that these considerations cannot help, by themselves, in accounting for the contrasts in (8) (using Kluender’s terminology, both “memory load” and “integration cost” are constant in (8)), this suggests (again) that RS performance can be explained in terms of “memory usage” looking at phrase structure “from-left-to-right”.

4 A matter of degrees: reconciling competence and performance

While distinguishing between grammaticality and acceptability Chomsky (1965:11) observes that both concepts are gradable, in the sense they can be assessed on scales. Even though grammaticality and acceptability scales do not coincide, being grammaticality just one of the many aspects contributing to acceptability (the other beings semantic plausibility, pragmatic adequacy and intonation for instance), it is purely an experimental exercise to keep such external factors constant (for instance removing semantic, pragmatic and intonational biases) while looking at subtle structural variables correlations, if any, with respect to RS performance. This is how recently fine grained distinctions have been attested using grammaticality judgments (Jurka 2010, Bianchi and Chesi 2014 a.o.). This permitted to consider seriously systematic evidence where, in the past, we simply found performance-related noise.

The emerging picture is, from many perspectives, even simpler and psycholinguistically more reasonable. For instance, assuming that our algorithmic derivation proceeds Top-Down (Chesi 2014) both self-embedding and subject island “violations” can be accommodated. On one side, we obtain a left-right (that is real-time based) derivation (this is because external merge is regulated by selection features on the already processed lexical items); this fulfills the “processing-based” assumption for which the computation of constituents must be ordered roughly from left to right (§3.1, Chesi 2014). On the other side, we can easily decide if a certain selection requirement introduced by a constituent (e.g. the possessor requirement associated, in generation, to a given noun phrase within the subject pre-verbal position) must be readily processed (this is marked by the presence of the stranded preposition) or must wait to be interpreted within the predicative nucleus (i.e. “reconstruction” takes place, hence deriving extractability and the problematic preference for pied-piping, Bianchi and Chesi in press).

Notice that also non-feature based movement can be reconciled with a top-down movement perspective: according to Moro (2000) movement is not triggered by any specific feature (see Chesi 2014 for the critical problems a feature-based movement approach causes from an algorithmic perspective) but by a “symmetric” geometrical configuration that does not permit to determine univocally the relevant (asymmetric) C-command relations between elements. From a reverse (top-down) perspective, we expect that symmetric structures must be recreated during the course of the derivation (e.g. to assign thematic roles) and that certain “asymmetric” configurations should trigger movement. We interpret “asymmetry” as the mismatch between the expected (i.e. selected) features and the actual features introduced in the derivation (following Chesi 2014, unexpected features force movement) and we assume that a “symmetric” configuration is created, when the pending expectation (i.e. selection) is fulfilled (that is, when the relevant holding constituent is “reconstructed” in an appropriate thematic position).

We conclude by mentioning that the idea behind a Top-Down approach is not new at all: Yngve (1960) suggested a similar derivational constraint to explain, from a competence perspective, cases like center-embedding. Chomsky (1965:197) liquidates this approach by saying that he sees “no plausibility at all to the assumption that the speaker must uniformly select sentence type, then determine subcategories, etc. finally, at the last stage, deciding what he is going to talk about; or that the hearer should invariably make all higher-level decisions before doing any lower-level analysis”. Here we do not want to compare our approach with Yngve’s one (the two are clearly different in many respects), but we want to stress that any theory of grammar that is mechanical and derives the correct (graded) oppositions without any uncertainty is totally plausible and cannot be excluded a priori. In this sense, we see no difference in a top-down or bottom-to-top approach since they are both “competence-based” assumptions.

References

- Barton, G. Edward, Robert C. Berwick, & Eric S. Ristad. 1987. *Computational complexity and natural language*. Cambridge, MA: MIT Press.
- Belletti, Adriana & Luigi Rizzi. 2013. Intervention in grammar and processing. In Ivano Caponigro & Carlo Cecchetto (eds.), *From grammar to meaning: The spontaneous logicity of language*, 293-311. Cambridge: Cambridge University Press.
- Bianchi, Valentina. 2002. Headed relative clauses in generative syntax. *Glott International Part I* 6.7:197–204, Part II 6.8: 1-13
- Bianchi, Valentina & Cristiano Chesi. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45(4): 525-569.
- Bianchi, Valentina & Cristiano Chesi. in press. On a PP/DP asymmetry in extraction. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, Strategies and Beyond*. Amsterdam: John Benjamins.
- Chesi, Cristiano. 2014. On directionality of phrase structure building. *Journal of Psycholinguistic Research* 44(1): 65-89.
- Chesi, Cristiano & Andrea Moro. 2014. Computational complexity in the brain. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring linguistic complexity*. Oxford: Oxford University Press.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133-166. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1980. On extraction from NP in Italian. *Journal of Italian Linguistics* 1/2:47-99.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: The MIT Press.
- Friederici, Angela D. 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences* 16(5): 262-268.
- Friedmann, Naama, Adrian Belletti & Luigi Rizzi. 2009. Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua* 119: 67-88.
- Gibson, Ed. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68: 1-76.

- Huang, James. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. thesis, MIT.
- Jiménez Fernández, Ángel. 2009. On the composite nature of subject islands: a phase-based approach. *SKY Journal of Linguistics* 22: 91-138.
- Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing*. Upper Saddle River, New Jersey: Prentice-Hall.
- Jurka, Johannes. 2010. *The importance of being a complement*. Ph.D. thesis, University of Maryland.
- Just, Marcel A., Patricia A. Carpenter, Timothy A. Keller, William Eddy & Keith R. Thulborn. 1996. Brain activation modulated by sentence comprehension. *Science* 274(5284): 114-116.
- King, Jonathan & Marcel A. Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*. 30(5): 580-602.
- Kluender, Robert. 2004. Are subject islands subject to a processing account? In *Proceedings of WCCFL 23*, 101-123.
- Ladusaw, William. 1994. Thetic and categorial, stage and individual, weak and strong. In *Proceedings of SALT IV*, 220-229.
- Longobardi, Giuseppe. 2000. "Postverbal" subjects and the Mapping Hypothesis. *Linguistic Inquiry* 31: 691-702.
- Mehra, Jagdish & Helmut Rechenberg. 1982. *The historical development of quantum theory*. New York: Springer Verlag.
- Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology*, vol. 2, 419-491. New York: Wiley.
- Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT press.
- Russell, J. Stuart & Peter Norvig. 2003. *Artificial intelligence: a modern approach* (2nd ed.). Upper Saddle River, New Jersey: Prentice-Hall.
- Takahashi, Daiko. 1994. *Minimality of movement*. Ph.D. dissertation, University of Connecticut.
- Traxler, Matthew J., Robin K. Morris & Rachel E. Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47(1): 69-90.
- Wanner, Eric & Michael Maratsos. 1978. An ATN approach to comprehension. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 119-161. Cambridge, MA: MIT Press.
- Warren, Tessa & Edward Gibson. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20: 751-767.
- Yngve, Victor H. 1960. A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society*, 444-466.

