Computational complexity in the brain

Cristiano Chesi & Andrea Moro

NeTS, IUSS Pavia

1. Introduction

How complex is human language? Every chapter in this volume is devoted, in one way or another, to addressing this question. The great majority of the chapters focus on the question of whether languages can be *compared* in terms of their relative degree of complexity. The present chapter, however (along with the one by Trotzke and Zwart) takes a somewhat different approach to the question of grammatical complexity. We ask: 'How might one characterize the degree of complexity of natural language *in general*?' Fortunately there are formal means of measuring the complexity of grammatical systems that go back to the 1950s. We refer, of course, to the Chomsky hierarchy (Chomsky 1956, 1963), which classes grammars, and the automata that recognize them, in formal terms. Our chapter goes beyond that of Trotzke and Zwart, however, in attempting to identify the brain activity that correlates with the formally-arrived at levels of grammar in the hierarchy. We will see that the relationship between the hierarchy and both measurable processing difficulty and detectable brain activity is not a simple one.

The chapter is organized as follows. Section 2 reviews briefly the Chomsky hierarchy. Section 3 investigates the processing (primarily parsing) correlates of the hierarchy and §4 the neurophysiological correlates. Section 5 further probes the modeling by automata of neurolinguistics processes. Section 6 is a brief conclusion.

2. The Chomsky hierarchy

In this section we provide background on the definition of formal grammars, the ranking of languages generated by such grammars as given by the Chomsky hierarchy, and the position of natural languages with respect to the hierarchy. We then relate the generative power of the grammar types on the hierarchy to computational procedures that can be used to calculate complexity costs, through the use of automata.

A formal grammar consists of a finite set of production rules (left-hand side → right-hand side), where each side consists of a sequence of the following symbols: a finite set of nonterminal symbols (indicating that some production rule can yet be applied), a finite set of terminal symbols (indicating that no production rule can be applied), and a start symbol (a distinguished nonterminal symbol). A formal grammar defines (or generates) a formal language, which is a (usually infinite) set of finite-length sequences of symbols (i.e. strings) that may be constructed by applying production rules to another sequence of symbols which initially contains just the start symbol.

The Chomsky hierarchy consists of the following levels:[1]

(1) a. Type 0: unrestricted (or Turing-equivalent) grammars. All rules are of the form $\alpha \rightarrow \beta$, with the only restriction being that α cannot be null.
b. Type 1: context-sensitive grammars (CSGs). All rules are of the form $\alpha A \beta \rightarrow \alpha \gamma \beta$. A must be nonterminal and α, β, and γ can be strings of terminals and nonterminals. The strings α and β may be empty, but γ must be nonempty.
c. Type 2: context-free grammars (CFGs). All rules are of the form $A \rightarrow \gamma$. Only one non-terminal symbol is allowed to the left side of the rewriting rules.
d. Type 3: regular grammars (RGs): All rules are of the form $A \rightarrow xB$ or $A \rightarrow Bx$, where x is a terminal symbol (i.e., a lexical item). Note that any non-terminal symbol must appear on the right edge or to the left edge.

Figure 1 depicts the set inclusions of languages according to the hierarchy. Grammars with fewer restrictions can generate all the languages that are generable by grammars with more restrictions.

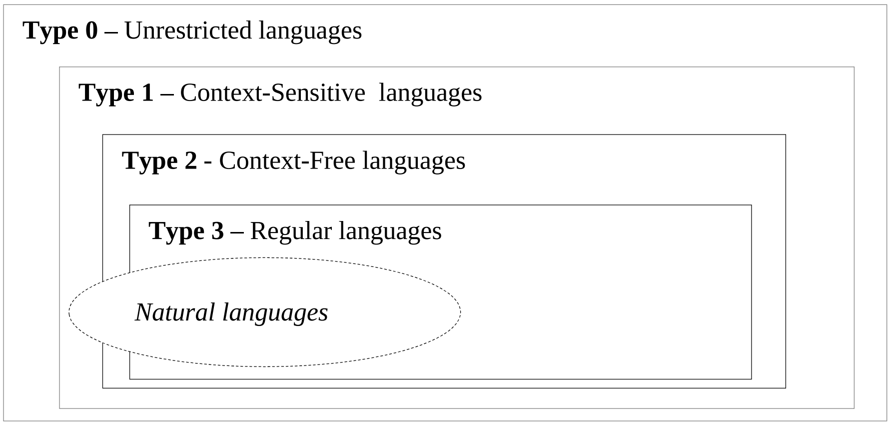


Figure 1: Set inclusions in the Chomsky hierarchy

Note the positioning of 'natural languages'. It has been argued that they require a grammar with a generative power that exceeds that of RGs (Chomsky 1957) and of CFGs (Shieber 1987). The arguments supporting this claim are based on two kinds of recursive structures that are attested in natural languages. Consider the following examples:

(2) a. The rat$_1$ [(that) the cat$_2$ [(that) the dog$_3$ chased$_3$] ate$_2$] died$_1$

---

[1] The remainder of this section greatly oversimplifies the material treated. The interested reader should consult any introduction to mathematical linguistics for more refined detail.

b. Anyone$_1$ [who feels that if$_2$ [so-many$_3$ [more$_4$ [students$_5$ [whom we$_6$ haven't$_6$ actually admitted] are$_5$ sitting in on the course] than$_4$ ones we have] that$_3$ the room had to be changed], then$_2$ probably auditors will have to be excluded], is$_1$ likely to agree that the curriculum needs revision. (Chomsky & Miller 1963:286)

c. Alberto, Bianca … e Xenia sono rispettivamente promosso, bocciata
… e promossa. (Italian)
A.$_{\text{male}}$, B.$_{\text{fem}}$ … and X.$_{\text{fem}}$ are respectively promoted$_{\text{male}}$, rejected$_{\text{fem}}$
… and promoted$_{\text{fem}}$

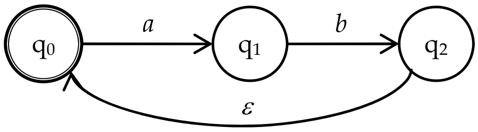
(2a-b) represent cases of nested dependencies of the 'mirror recursion' type XX^R (e.g. *abcddcba*), while (2c) illustrates a case of 'cross-serial dependency' of the XX kind (e.g. *abcdabcd).* Partee et al. (1990) show how sentences containing mirror recursion nested dependencies can not be generated/recognized by RGs and sentences containing cross-serial dependencies can not be generated/recognized by CFGs. Given such data, it appears to be the case that natural language grammars have at least the generative power of context-sensitive grammars.[2]

Our goal now is to relate generative power as ranked by the hierarchy to computational procedures for which complexity costs can be calculated. One way to evaluate complexity, relying on the Chomsky hierarchy, is to use automata that are equivalent in terms of generative power to the grammars in the hierarchy (for discussion, see Hopcroft et al. 2001). Finite state automata (FSAs), for instance, can express any regular language by using states and transitions among states. The diagram below shows how simple recursive rules characteristic of RG, such as (3a), can be subsumed by finite automata like the one in (3b). Another way of putting it is to say that (3a) and (3b) capture the same set of sentences:

(3) a. $S \rightarrow a\ b\ S$

$S \rightarrow \varepsilon$

b.



---

[2] Or perhaps mildly context-sensitive grammars (MCSG) (see Weir 1988 for discussion). MCSG express a class of grammars not included in CFGs, but included in CSGs (the tree-adjoining grammars of Joshi 1985 are an example).

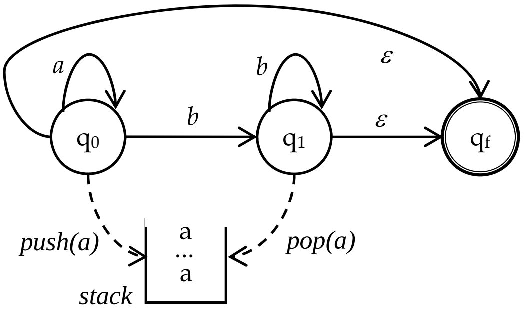
The automaton starts in *q_0* (the initial state, by convention, is labeled *q_0*) and ends in *q_0* (in this special case, the initial state is also the final state; in the diagrams, by convention, we indicate final states using double-circles). The arrows indicate possible transitions from one state to the other and they can be traversed only when the (terminal) symbol (e.g. a lexical item) that labels them is recognized or generated (ε indicates the null symbol; the transition labeled by such symbol can be freely traversed without consuming input tokens). Only if a final state is reached, and every token in the input is consumed, the computation can terminate successfully.

CFGs are more powerful than RGs and hence require more powerful automata. The latter are called push-down automata (PDAs). The example below depicts a sample CFG and a corresponding automaton:

(4) a. $S \rightarrow a\, S\, b$

$S \rightarrow \varepsilon$

b.



Both CF rewriting rules like those (4a) and their corresponding automata (4b) can capture counting recursion of the $a^n b^n$ kind (i.e. *n* sequences of *a* followed by the very same number of *b*). This automaton will start its computation from the initial state *q_0*, then it will 'push' an *a* in the stack any time it reads one from the input. Reading a *b*, the automaton will move to *q_1* and it will start 'popping' an *a* as long as *b*s are in the input. Once *b*s are exhausted, if the stack is empty, the automata will reach *q_f*, the final state, signaling the end of a successful computation. The stack in PDA hence has a 'last in first out' (LIFO) memory structure. That is, last item pushed in the stack will be the first one to be popped out. The computation starts at *q_0*, is the initial state, and ends at *q_f*, the final one.

Now then, where does the issue of 'complexity' enter the picture? Given the above, we can conceptualize complexity in two dimensions: time and space. The time complexity is a function of the number of states traversed. As can be seen in comparing (3b) and (4b), FSAs and PSAs with the same number of states are equivalent in terms of state traversals: given a string of length 4 ("abab" for the FSA, "aabb" for the PDA), both automata need 4 state traversals (the initial states and the states reached using the ε transition are not counted) . However, if we take a 'three dimensional' look at the diagrams, we can see that processing "aabb" using the PDAs is more complex than processing "abab" using the FSAs in terms of space complexity. The space complexity is a function of the additional memory needed to keep track of the dependency. A PDA is thus more complex than a FSA in a very precise way, namely it needs a LIFO memory buffer. So to process a string of length 4 (i.e. *aabb*), in addition to the 4 state traversals, the PDA needs to store 2 items in the memory (the two instances of *a*), using 4 operations (pushing *a* twice and popping *a* twice). For a string of length 6 (i.e. *aaabbb*) the very same PDA needs 6 state traversals, 3 items in the memory buffer, and 6 operations on the memory buffer.

For our purposes, the important thing to keep in mind while reading the following section is that such computational models give us a baseline for comparing precisely the complexity of two different performance tasks (such as processing different levels of hierarchical embedding) in terms of time (the number of computational states to be traversed) and space (the number of items to be stored and retrieved).

3. Computational models of processing complexity

This section takes as its point of departure the idea that the Chomsky hierarchy translates directly into a measure of processing complexity. However, we will see that such is not correct. Many factors affect processing that are independent of the position of a grammar on the hierarchy. We will discuss three examples where complexity as measured by the hierarchy does not directly correspond to complexity as measured by processing tasks. We will then discuss additional factors that may account for this lack of direct correspondence.

Various hypotheses have been discussed in literature that propose a more or less straightforward correlation between the automata (and their relative complexity metrics) discussed in the previous section and the degree of difficulty in human sentence processing. For example, it has been proposed that patterns that are not captured by RGs are generally hard to process (Pullum & Gazdar 1982:498). One might assume then that cross-serial dependencies, such as (2c), which are licensed by the more complex CFGs and CSGs, should be even more difficult to parse and generate than nested dependencies such as (2a-b). This is roughly, but not entirely, correct. Bach et al. (1986) compared, in terms of acceptability judgments and accuracy in paraphrase understanding, the linguistic performance on cross-serial and nesting dependencies. They showed that cross-serial patterns (in Dutch) with two levels of dependency are significantly more easily processed than center embeddings with the very same level of dependencies (in German).[3] Comparable results have been obtained with three levels of dependency using artificial grammars, both with humans and with artificial neural network simulations (Christiansen and Chater 1999). In other words, what seems crucial in determining

---

[3] German and Dutch are compared since the equivalent sentence requires a nested dependency [arg_1 [arg_2 verb_2] verb_1] in German and a serial dependency in Dutch [arg_1 arg_2 verb_1 verb_2].

processing complexity is the kind of dependency, with nesting dependencies in general harder than cross-serial dependencies. The position in the Chomsky hierarchy is not the determining factor.

Another way to relate the Chomsky hierarchy to processing complexity involves looking at specific performance tasks. For example, it has been suggested that performance difficulty in parsing is associated with the limitations of the memory stack (Yngve 1960). Given such an idea, the more incomplete (i.e., unexpanded) the phrase structure rules are that the parser needs to store within the stack (to retrieve a complete phrase structure), the more complex the sentence will be. In this vein, Chomsky and Miller (1963) suggested that a stack-based parser might simply get confused while processing self-embedding. The problem is that storing more than once the same context-free rules in the stack might lead the parser to confuse two instances of the very same rule. This problem holds only for CF (or generatively more powerful) rules and crucially not for RG rules, since only in the first case is a new rule stored in the stack while the previous rule is not yet completed. In RGs, top-down expansion of any rule does not need any stack-based storage.

Another problem is that sentences with the same level of embedding (and otherwise equivalent in terms of the Chomsky hierarchy) are perceived as more or less difficult to process depending on the position of the very same lexical items. It seems to be the case that the complexity of certain structures is not related simply to the phrase structure to be expanded or to the necessity of using a memory buffer, but to the kind of non-local dependency to be computed. For instance, consider the sentences in (6) below. The relative clause whose head (*reporter*) is interpreted in the object position within the relative clause, hence an 'object-headed relative clause' (ORC; 6a), takes more time to process than a subject-headed relative clause (SRC; 6b) (King & Just 1991):[4]

(5) a. The reporter$_i$ [who [the senator] attacked _$_i$] admitted the error. (ORC)
b. The reporter$_i$ [who _$_i$ attacked [the senator]] admitted the error. (SRC)

What seems at issue in (5) is the special relation between the filler and the gap (Fodor 1978). In particular, the kind of element that intervenes (*the senator* in 6a) between the filler (*the reporter)* and the gap (the argument position within the relative clause) appears to play a crucial role.

Pursuing the question of the processing costs of the two types of relative clauses, it has been found that pronouns and proper names produce a milder complexity increase in sentence processing than full DPs in ORCs. Syntactic prediction locality theory (SPLT, Gibson 1998) attempts to explain why such is the case. Gibson's idea recasts complexity metrics in terms of *integration* and *memory-load cost*. The first component (integration cost) is associated with new discourse referents to be incorporated in the structure (a full DP

---

[4] This asymmetry has been systematically documented using self-paced reading experiments (King & Just 1991), probe-task paradigms (Warner & Marastos 1978), brain activity analysis (Just et al. 1996), and eye-tracking techniques (Traxler et al. 2002).

introduces a new referent, a pronoun does not). The second component (memory-load) is associated with keeping track of obligatory syntactic requirements, that is, the minimal set of items that could conclude the processed sentence in a well-formed way. For instance, given independently-needed assumptions about the grammar-processor relation, after processing a DP like *the reporter*, the processor expects at least a VP (e.g. *left*) to complete the sentence. Waiting for a VP will have a cost that needs to be added to the cost of other expectations plus the cost of integrating new discourse referents, if any. In (5a), *the senator* introduces a new discourse referent and therefore incurs an integration cost. (5a) incurs a memory-load cost because the processor must wait for the VP until after the relative clause is complete.

PDA appear to be computationally adequate for implementing both integration cost and memory-load cost. As far as the former is concerned, we might assume that the recognition of any new referent (a proper name or a full DP) triggers the insertion of a co-indexed variable in the memory buffer, while a pronoun picks up a referent from it (cf. Schlenker 2005, Bianchi 2009). As for memory-load cost, we might assume that any new sequential DP moves the automaton to a deeper state more removed from the final state. If this were correct, then integration cost and the memory-load would translate directly into space complexity and time complexity respectively. But things, as always, are more complicated than that. Unfortunately, Gibson's hypothesis makes incorrect empirical predictions where intervening proper nouns are involved. Consider (6a-c):

(6) a. *The pictures* that *the boy* took yesterday
b. *The pictures* that *John* took yesterday
c. *The pictures* that *I* took yesterday

(6a) is measurably harder to process than (6b), which is harder than (6c) (Gordon et al. 2001, 2004). To address this problem of incorrect prediction, Warren and Gibson (2005) have refined Gibson's integration cost by defining a hierarchy of referents, with full DPs 'heavier' than proper names, and proper names 'heavier' than pronouns. But doing so explains only the contrast schematized by the paradigm in (6), but not the fact that when the relative clause head is also a proper name (*It is **Bob** that **John** saw yesterday*), the processing difficulty is comparable to that of (6a) (Belletti & Rizzi 2013).

These contrasts suggest that factors other than referentiality are involved in the correct complexity metric for processing. Among these are animacy features (Mak et al. 2002) and agreement features (Belletti et al. 2012). As far as features are concerned, locality constraints (Rizzi 1990, Friedmann et al. 2009) seem to play a fundamental role in accounting for complexity increase in terms of feature intervention. We can rephrase part of Friedmann et al.'s (2009) hypothesis as follows:

(7) Feature intervention: When *X* and *Y* enter a non-local relation and *Z* intervenes within such relation, the complexity of the dependency is proportional to the features shared between *X* and *Z*.

We could adapt (7) to a model of processing cost by modifying Gibson's idea of integration cost in the following way:

(8) Feature-based integration cost (FIC, revision of Gibson 1998's model of integration cost):

> Any operation in the memory buffer has a cost that is proportional to the number of features the item that must be stored or retrieved shares with other items already present in the memory buffer.

We assume that in ORC like (6a), the head of the relative clause, *the pictures*, which is the filler to be stored in memory, is expressed in terms of features as [+D +N] (i.e. determined, nominal entity), as is *the boy*. Therefore two features are shared and two operations (storage and retrieval from memory) are needed for each of them. We combine the two factors in one naïve complexity function C_{FIC} as follows:

(9) $C_{FIC} = O^F$ (naïve Feature-based Integration Cost function)

O is the number of operations accessing memory (two in this case, namely storage and retrieval) and F is the number of shared features (also two in this case).

Note that that this rather simple complexity function automatically explains the contrasts in (6a-c). Assuming that what distinguishes a common noun from a proper noun is a sub-specification of one single nominal feature (i.e. N vs. N_{proper}, as in Belletti & Rizzi 2013), we could consider this overlapping as half of the cost of a full feature overlapping (i.e. 0.5). Hence in (6a), the FIC for *the boy* will be $C_{FIC} = 4$ (i.e. 2^2), while in (6b), *John* will have a $C_{FIC} = 2.8$ (i.e., an approximation of $2^{1.5}$, since +D is fully shared but N and N_{proper} are counted as half sharing). Finally, in (6c) *I* will cost $C_{FIC} = 2$ (since only +D is shared, assuming that a pronoun lacks the N feature (Friedman et al. 2009). Hence, (8) not only explains the processing asymmetry revealed in (6a-c), but also the fact that when the head of the relative clause and the filler are both proper nouns, (*it is **Bob** that **John** saw yesterday*) we get a C_{FIC} for *John* of 4, exactly as in (6a).

To summarize, we have seen that patterns that are captured by more powerful grammars in the Chomsky hierarchy are not always more complex to process than patterns that require less powerful grammars. Independently of the hierarchy, the constructions discussed in this section require the specification of at least two components for a correct complexity function characterization: a hierarchical component and a component that encodes features involved in non-local dependencies of the filler-gap type. We turn now to the question of complexity as revealed by operations in the brain.

4. Syntax in the brain: Autonomy, hierarchy and locality

Research on the biological foundations of language has seen a dramatic development since the turn of the century (see Cappa 2012, Friederici 2011, Kandel et al. 2012 for reviews). In particular, the possibility of exploiting neuroimaging techniques has offered interesting opportunities for deepening our understanding of the relationship between syntax and the brain. In this section, we discuss whether the main components of complexity discussed in Section 3, hierarchy (time complexity, i.e. level of structural embedding) and locality (space complexity, i.e. memory) are distinguishable, not only at the computational level, but also at the neurological level. The following two subsections develop the grammar-brain relationship. Section 4.1 reviews evidence that syntactic computation activates a dedicated network and takes on the question of where in the brain syntax is activated. Section 4.2 discusses the complexity signatures of hierarchical syntactic processing and non-local dependency formation.

4.1 Where to look for complexity: Syntactic networks in the brain

The classical assumption is that the language centers in the brain are Broca's and Wernicke's areas. However, imaging techniques such as PET and fMRI have shown that the story is not quite that simple.[5] From one side, additional areas are involved in language processing (some of them deeply brain-internally), on the other, both Broca's and Wernicke's area are finer subdivided according to their specialization revealed in distinct functions.

To begin, isolating syntax from other linguistic components by pointing to a difference in neural activity is not an easy task, since processing does not treat syntax as an isolated entity. Moro et al. (2001) tackled this problem by inventing a lexicon consisting of fake nouns, verbs, adjectives, etc., whose entries were phonotactically compatible with Italian. Real Italian functional morphemes (determiners, inflections, etc.) were combined with these fake content words, thereby creating pseudo-Italian sentences. The experimental stimuli consisted in introducing selective errors at different levels in 'quasi-Italian' sentences of this kind (for ease of exposition we replace them with 'quasi-English' sentences):

(10)	a.	The gulk ganfles the brals	(grammatical sentence)
	b.	The gulk ganfrzhrld the brals	(phonological error)

---

[5] A PET scan is an imaging technique that exploits (fludeoxy)glucose uptake to reveal neural metabolic activity by tracing regional concentration of these molecules. PET has a good spatial resolution, but a poor temporal resolution. An fMRI is an imaging technique that allows us to visualize neural activity by detecting associated changes in blood flow by using blood-oxygen-level-dependent contrast detectors. fMRI has an excellent spatial resolution and a decent temporal resolution.

c. The gulks ganfles the brals (morphosyntactic error)
d. Gulk the ganfles brals the (syntactic error)

The subjects were prompted with baseline stimuli (well-formed pseudo-sentences like 10a) and their brain activity was compared during the processing of sentence with errors of various types (10b-d). The subjects were asked to read silently the visually presented sentences and, for the three error types under analysis, to provide an acceptability judgment. Brain activity was recorded using a PET (Positron Emission Tomography) scan.

Interestingly, the results show that syntactic anomalies activate specific brain regions, in particular Broca's areas BA44 and BA45 (see Figure 2) and the Left Nucleus Caudatus (LNC, part of the basal nuclei in the inner brain).

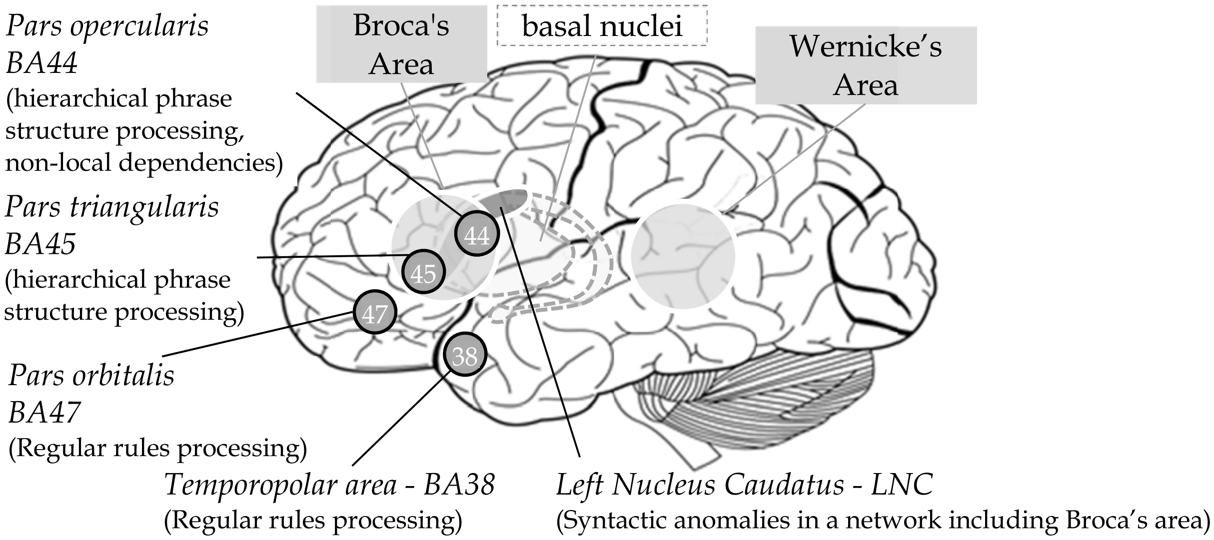


Figure 2: A very schematic summary of brain locations (and their possible functional roles)

These results are compatible with other studies that have attempted to isolate syntactic processing networks. For instance Embick et al. (2000), also using a violation task, found a selective activation of Broca's area in grammatical vs. orthographic errors. This latter experiment, however, did not exclude the lexical semantic contribution and thus provided a weaker result, in that it was possible that syntactic errors also activated semantic networks.

Before concluding that BA44/45 and LNC are the (only) places to look for syntactic complexity, we should take into consideration a few more studies that greatly refine our understanding of BA44/45 and LNC. For example, Musso et al. (2003) provided evidence that the distinction between possible (i.e. Universal Grammar-compliant) and impossible (i.e. Universal Grammar-unrealistic) rules is not just an artificial taxonomy resulting from cultural or conventional constructs, but is instead the reflex of a genuine neuropsychological structure. In order to test this hypothesis, an fMRI experiment was designed to compare syntactic rules sensitive to hierarchical constituent structure to impossible rules targeting a fixed position in the sentence. German speakers from strictly monolingual communities in the pre-unification German Democratic Republic participated in the experiment. Subjects were asked to learn 'artificial grammars' of a pseudo-Italian language, that is, those composed of sentences containing Italian lexical items, but generated either by real Italian rules or by fake (impossible) rules.[6] In the experiment in which real Italian rules were used, the experimental subject would be expected to invert the grammatical subject and the object positions to create a passive (11a). In another, he or she would be expected to place the complementizer *che* between the matrix and subordinate clause. In another experiment, impossible rules were used. For instance, to correctly complete the task the participant would be expected to place the negation marker after the third word (12a) or to force agreement between the first and the last word in the sentence (12b). Note that all rules apply to the baseline sentence *Paolo mangia la pera* ('Paolo eats the pear').

(11) Possible rules

a. **La pera** è mangiata **da Paolo** (passivization)

the pear is eaten by Paolo

b. Pia dice **che** Paolo mangia la pera (complementation)

Pia says that Paolo eats the pear

(12) Impossible rules

a. Paolo mangia la **no** pera (negation)

Paolo eats the no pear.

b. Un**a** bambino mangia una per**a** (agreement)

a$_{fem.sing}$ *child eats a pear*$_{fem.sing}$

Interestingly, it was only the processing of possible rules that correlated with stronger activation of part of Broca's area (pars triangularis, BA45), although all subjects rapidly acquired the same capability to manipulate both possible (11a-b) and impossible (12a-b) rules. Hence we have robust evidence in favour of the neurobiological reality that only

---

[6] See Moro (2008) for a discussion of the characterization of "possible" and "impossible" rules.

hierarchical (syntactic) rules activate a specific brain region and that human language is not purely a matter of cultural or arbitrary convention.[7]

Other studies bear out the idea that BA44/45 is involved in hierarchical and/or non-local processing of linguistic stimuli, for example, those involved in counting dependencies of the a^nb^n kind, which require, as we have seen, the generative power at least of CFGs or PDAs (see Fitch & Friederici 2012). However, patterns that do not activate BA44/45 include strings that conform to the pattern $(ab)^n$, namely sequences of any length of couples of items like the syllables *pa* and *do*: *pa-do-pa-do-pa-do…*). This pattern selectively activates brain regions BA38 (the temporopolar area) and BA47 (part of the inferior frontal gyrus, also known as pars orbitalis).

The hypothesis that only the activation of BA44/45 is selective to hierarchy and/or non-local dependencies is also consistent with the fact that a selective activation of the posterior part of Broca's area has been revealed during tasks comparing syntactic vs. purely lexical (semantic) conditions. On this point, Dapretto & Bookheimer (1999) showed, in a fMRI study, that BA44 is more active when syntactic relations are involved (13a-b), as opposed to purely lexical ones (14a-b):

(13) a. The policeman arrested the thief.
a'. The thief was arrested by the policeman.

b. The teacher was outsmarted by the student.
b'. The teacher outsmarted the student.

(14) a. The lawyer questioned the witness.
a'. The attorney questioned the witness.

b. The man was attacked by the Doberman.
b'. The man was attacked by the pit bull.

Earlier, using the PET technique, Stromswold et al. (1996) and Caplan et al. (1998) reported selective activation of BA44 when subjects made plausibility judgments about center-embedded relative clauses (15a) compared to right-branching relative clauses (15b):

(15) a. The juice that the child spilled stained the rug.
b. The child spilled and the juice that stained the rug.

---

[7] Similar results have been obtained in other experiments, such as those reported in Tettamanti et al. (2002; 2008), which exploited both pseudosentences and non-linguistic symbolic strings within an autonomous learning environment.

To summarize, in this section we pointed out that Broca's Area (BA44 and BA45) is selectively activated in the processing of hierarchical structures. Non-local dependencies (passive contractions and object relative clauses), also involve a special role for BA44.

In the next subsection we will try to relate this selective activation to signatures of complexity during performance, as a preliminary to further discussions of computational complexity in §5.

4.2 The complexity effects of hierarchical syntactic processing and non-local dependency formation

Let us now turn in more detail to syntactic hierarchies and non-local dependencies such syntactic movement. It appears that as far as these are concerned, an increase of complexity in sentence processing is expressed by an increase of activity in the relevant brain areas. It also appears that working memory is involved in non-local dependencies of the filler-gap type (Wager et al. 2003). This means that when constituents become hierarchically deeper and memory load increases (because of increased distance between the filler and the gap and/or because of intervening items), the areas specifically activated would be likely to show stronger metabolic activity or longer metabolic activity and that adjacent areas should be recruited for the more demanding tasks.

All these hypothesis seem to be at least partly borne out. For example, in a pioneering study, Just et al. (1996) noted an increase in the volume of neural tissue activation (number of voxels, i.e. volumetric pixels produced by an fMRI imaging system), mainly in Wernicke and Broca's areas, that was proportional to sentence complexity. The sentences they used are given in (16) below:

(16) a. The reporter attacked the senator and admitted the error.
b. The reporter that attacked the senator admitted the error.
c. The reporter that the senator attacked admitted the error.

In (16a), the two conjoined active sentences are on the lower side of the complexity scale. This is because in left-to-right processing, each sentence is uninterrupted and the first DP feeds the required subject positions in both phrases, consistent with canonical SVO order. On the other hand, as discussed in §3, (16b) is more complex than (16a) In (16b), the matrix sentence *The reporter admitted the error* is 'interrupted' by the SRC (*the reporter*$_i$ *[that* _$_i$ *attacked the senator*]). Here too *the reporter* feeds the subject position both in the matrix clause and in the relative clause.[8] This does not happen in (16c), where the relative clause has its own subject, namely, *the senator,* which intervenes between the head and the gap

---

[8] We have no space to discuss here different interpretations of the relation between the head and the gap in the restrictive relative clause. 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).

in the object position within the ORC ((*the reporter*$_i$ *[that the senator attacked _*$_i$*]*)). As discussed in §3, this involves greater processing complexity than with (16a) and (16b) (see Friedmann et al. 2009 for a review). In other words, the study of Just et al. (1996) tells us that behavioral complexity revealed in psycholinguistic experiments correlates with an activation increase of the areas specifically involved in language processing, and this increase is proportional to hierarchical depth and to the filler-gap distance.[9]

The role of the distance between the filler and the gap must be tested separately from the intervention effects discussed above. For example, looking at scrambling in German, Friederici et al. (2006) noted that the greater the distance between the scrambled constituents and their base (canonical) thematic position, the stronger the activation of BA44.[10] Example (17c) below, where both the indirect object and the direct object are scrambled over the subject, is significantly more complex (i.e., there was stronger activation of BA44) than (17b), where only the indirect object is scrambled across the subject; and (17b) was significantly more complex than (17a) (all arguments are in their base position):

(17) a. Heute hat der Opa dem Jungen den Lutscher geschenkt. (S IO DO)

Today has [the grandfather]$_{\text{NOM}}$ [to the boy]$_{\text{DAT}}$ [the lollipop]$_{\text{ACC}}$ given.

b. Heute hat [dem Jungen]$_i$ der Opa _$_i$ den Lutscher geschenkt. (IO$_i$ S _$_i$ DO)

Today has [to the boy]$_{\text{DAT}}$ [the grandfather]$_{\text{NOM}}$ _ [the lollipop]$_{\text{ACC}}$ given.

c. Heute hat [dem Jungen]$_i$ [den Lutscher]$_j$ der Opa _$_i$ _$_j$ geschenkt.(IO$_i$ DO$_j$ S _$_i$_$_j$)

Today has [to the boy]$_{\text{DAT}}$ [the lollipop]$_{\text{ACC}}$ [the grandfather]$_{\text{NOM}}$ given.

Today, the grandfather has given the lollipop to the boy.

According to Grodzinsky (2000), BA45 also seems to play an important role during processing of certain kinds of non-local dependencies. The evidence is based both on

---

[9] This study however does not help us to understand which component of the complexity function, as we presented it in §3, is involved, because the distance between the filler and the gap (that is, Gibson's 1998 memory load cost) and the nature of the item that intervenes between the filler and the gap (that is, what we call Feature-based Integration Cost 8) are non-identical in (16b) and (16c).

[10] Ben-shachar et al. (2004), testing topicalization vs. dative-shift in Hebrew, reported a more articulated activation of BA44/45 and BA6/9. We do not have space to discuss this matter here. It is simply worth stressing that scrambling and topicalization might involve different neural pattern activation. This is surely true for shifting operations like dative-shift (Larson 1988, Ben-shachar et al. 2004).

aphasic patients with selective lesions and fMRI experiments (Santi and Grodzinsky 2008).[11]

Makuuchi et al. (2009), in a 2-way factorial design study, compared the filler-gap distance (short vs. long) component with the level of embedding (see sentences 18a-b, where the first involves embedding and the second does not):

(18) Peter wusste, dass (Peter knew that) …
  a. [Maria$_1$, [[die Hans]$_2$, [[der gut]$_3$ aussah$_3$] liebte$_2$] Johann geküsst hatte$_1$].
    Maria who Hans who was good looking loved Johann kissed. [literal]
    Maria who loved Hans who was good looking kissed Johann.

  b. Achim$_1$ den großen Mann gestern am späten Abend gesehen hatte$_1$.
    Achim the tall man yesterday at late night saw. [literal]
    Achim saw the tall man yesterday late at night.

The observed activation patterns suggest a contrast between BA44, which is sensitive to structural embedding, and the dorsal portion of both BA44 and BA45 (i.e., the Inferior Frontal Gyrus), which is involved in memory-demanding tasks requiring movement. Using an imaging technique that allows us to draw probabilistic connections between areas by looking at movement of water molecules within the brain tissues (diffusion tensor imaging), Makuuchi et al. (2009) also identified a strong connection between the dorsal portion of the two regions BA44 and BA45. The increase of activation is present both for nesting and non-local dependencies. In other words, these experiments were only partly able to disentangle the role of memory from the role of hierarchy.

To conclude, in this section we showed how specific brain regions that are active during hierarchical syntactic processing and non-local dependency formation become more active when hierarchy gets deeper, as in relative clause embedding, and when dependencies require extra working memory, as is the case with longer dependencies containing constituents that structurally intervene between the filler and the gap. We can now go back to our computational complexity model to draw some conclusion from what we learned.

5. Possible and impossible rules and automata

This section concludes our discussion by relating the computational complexity findings of §3 to the neurophysiological evidence we presented in §4. What we discovered is that the distinction between possible and impossible rules, as investigated by Musso et al. (2003), lends itself to modeling in terms of FSAs and PDAs. (19-20) repeat their crucial sentences, while (21-22) formalize their results in terms of automata:

---

[11] Santi & Groszinsky (2010) found that an activation on BA45 during syntactic movement processing, but not when just pronominal binding was involved.

(19) Possible rules

a. **La pera** è mangiata **da Paolo** (passivization)

the pear is eaten by Paolo

b. Pia dice **che** Paolo mangia la pera (complementation)

Pia says that Paolo eats the pea

(20) Impossible rules

a. Paolo mangia la **no** pera (negation)

Paolo eats the no pear.

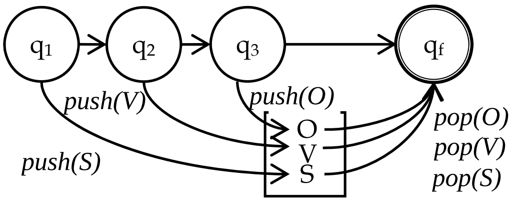
b. Un**a** bambino mangia una per**a** (agreement)

$a_{fem.sing}$ *child eats a* $pear_{fem.sing}$

(21) Possible rules

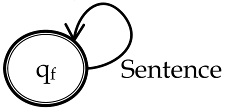
a. Given a sentence, passivize it by inverting the subject and the object:

(*non-recursive, hierarchical*)



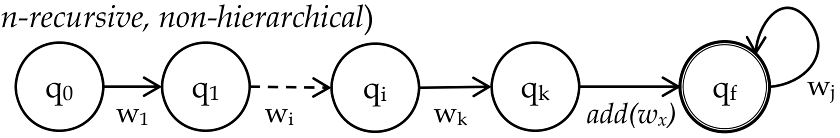
b. Expand a sentence with another sentence by complementation

(*recursive, hierarchical*)



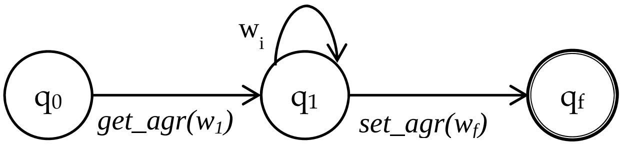
(22) Impossible rules

a. Insert a word w_x at k^{th} position (requires $k+1$ states);
(*non-recursive, non-hierarchical*)



b. The first, w_1, and the last element, w_f, in the string should agree

(*recursive, non-hierarchical*)



We observe that only passivization, which involves a non-local dependency, requires a device that is as powerful as PDA. Complementation simply requires FSA power. This is an important point, since the activation of the areas (BA44 and BA45) discussed in Musso et al. (2003) seems sensitive to hierarchy, but not to recursion or to memory demand. [12]

6. Conclusion

In this chapter we investigated the relationship between complexity, as defined by the Chomsky hierarchy and its associated hierarchy of automata, and the relative complexity of operations in the brain. What we found is that two main components are needed in order to correctly characterize a complexity function that is computationally sound (§2), psycholinguistically tenable (§3) and neurophysiologically testable (§4). These components are the hierarchical embedding (time complexity) and the memory demand expressed in terms of intervening features within a filler-gap dependency (space complexity). We hypothesize that our results, which are based on studies of a few familiar languages, will generalize to human language in general. Whether this is or is not the case, as well as the question of whether all languages avail themselves of the same neurologically-instantiated mechanisms of the same degree of complexity, are matters for future research.

---

[12] Because of Makuuchi et al.'s (2009) experiment results we should expect a stronger activation of the dorsal portion of BA44 and BA45 proportional to the complexity of the non-local depencency, §4.2. Since no difference in pattern activation is revealed by Musso et al. (2003) in (21a) vs. (21b), we conclude that the feature-based integration cost, discussed in (8), is rather low in the passive construction task exemplified by (21a).

REFERENCES

Bach, Emmon, Colin Brown & William Marslen-Wilson (1986). Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Language and Cognitive Processes*, 1(4): 249–262.

Belletti, Adriana & Luigi Rizzi (2013). Intervention in grammar and processing. In Ivano Caponigro & Carlo Cecchetto (eds.) *From Grammar to Meaning: The Spontaneous Logicality of Language*. pp. 293-311. Cambridge: Cambridge University Press.

Belletti, Adriana, Naama Friedmann, Dominique Brunato & Luigi Rizzi (2012) Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua*, 122(10):1053–1069.

Ben-Shachar, Michal, Dafna Palti & Yosef Grodzinsky (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *NeuroImage* 21: 1320–1336.

Bianchi, Valentina (2002) Headed relative Clauses in Generative Syntax. Part I. Glot International 6(7):197–204

Bianchi, Valentina (2009). A note on backward anaphora. *Rivista di Grammatica Generativa*, 34: 3-34

Caplan, David, Nathaniel Alpert & Gloria S. Waters (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, 10(4): 541-552.

Cappa, Stefano (2012) Imaging semantics and syntax. *Neuroimage*, 61: 427-431.

Chomsky, Noam (1956). Three models for the description of language. IRE *Transactions on Information Theory* (2): 113–124. doi:10.1109/TIT.1956.1056813

Chomsky, Noam (1957). *Syntactic structures*. Berlin: Mouton, The Hauge.

Chomsky, Noam, Geroge A. Miller (1963). Introduction to the formal analysis of natural languages. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), Handbook of Mathematical Psychology, vol. 2. pp. 269–321. New York: Wiley.

Christiansen, Morten H. & Nick Chater (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2): 157-205.

Dapretto, Mirella & Susan Y. Bookheimer (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron*, 24(2): 427-432.

Embick, David, Alec Marantz, Yasushi Miyashita, Wayne O'Neil & Kuniyoshi L. Sakai (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences*, 97(11): 6150-6154.

Fitch, W. Tecumseh & Angela D. Friederici (2012) Artificial grammar learning meets formal language theory: an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598): 1933-1955.

Friederici, Angela D. (2011). The brain basis of language processing: from structure to function. Physiological reviews, 91(4), 1357-1392.

Friederici Aangela D., Christian J. Fiebach, Matthias Schlesewsky, Ina D. Bornkessel & D. Yves von Cramon (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex* 16:1709–1717.

Fodor, Janet D., (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*. 9: 427–473.

Friedmann, Naama, Adrian Belletti & Luigi Rizzi (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua*. 119: 67–88.

Gibson, Edward (1998). Linguistic complexity: locality of syntactic dependencies, *Cognition,* 68: 1–76.

Gordon, Peter C., Randall Hendrick & Marcus Johnson (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 27: 1411-1423.

Gordon, Peter C., Randall Hendrick & Marcus Johnson (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language.* 51: 97-114.

Grodzinsky, Yosef (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and brain sciences*, 23(01): 1-21.

Hopcroft, John E., Rajeev Motwani & Jeffrey D. Ullman (2001). *Introduction to Automata theory, languages and computation*. 2nd edition. Boston: Addison-Wesley.

Joshi, Aravind K. (1985). How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky, (eds.). *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives*. pp. 206–250. New York, NY: Cambridge University Press.

Just, Marcel A., Patricia A. Carpenter, Timothy A. Keller, William F. Eddy & Keith R. Thulborn (1996). Brain activation modulated by sentence comprehension. *Science*, 274(5284), 114-116.

Kandel, Eric R., James H. Schwartz, & Thomas M. Jessell, Steven A. Siegelbaum, A. J. Hudspeth (2012) *Principles of Neural Science*, Fifth Edition. McGrawHill Medical, New York.

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.

Larson, Richard K. (1988). On the double object construction. *Linguistic Inquiry*, 19: 335–391.

Mak, W. M., Vonk, W., Schriefers, H. (2002). The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language*, 47(1) 50-68.

Makuuchi, Michiru, Jörg Bahlmann, Alfred Anwander & Angela D. Friederici (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences*, 106(20): 8362-8367.

Moro, Andrea (2008). *The boundaries of Babel: The brain and the enigma of impossible languages*. MIT Press.

Moro, Andrea, Marco Tettamanti, Daniela Perani, Caterina Donati, Stefano Cappa & F. Fazio (2001). Syntax and the brain: disentangling grammar by selective anomalies. *NeuroImage*, 13: 110-118.

Musso, Mariacristina, Andrea Moro, Volkmar Glauche, Michel Rijntjes, Jürgen Reichenbach, Christian Büchel & Cornelius Weiller (2003) Broca's area and the language instinct. *Nature neuroscience*, 6: 774-781.

Pullum, Geoffrey K. & Gerald Gazdar (1982). Natural languages and context-free languages. *Linguistics and Philosophy*, 4: 471–504.

Rizzi, Luigi (1990). *Relativized minimality*. Cambridge, MA: MIT press.

Santi, Andrea & Yosef Grodzinsky (2007). Working memory and syntax interact in Broca's area. *Neuroimage*, 37: 8–17.

Santi, Andrea & Yosef Grodzinsky (2010). fMRI adaptation dissociates syntactic complexity dimensions. Neuroimage, 51(4): 1285-1293.

Schlenker, Philippe (2005). Non-Redundancy: Towards a Semantic Reinterpretation of Binding Theory. *Natural Language Semantics*, 13(1): 1-92

Shieber, S. M. (1987). Evidence against the context-freeness of natural language. In Walter J. Savitch, Emmon Bach, William Marsh, Gila Safran-Naveh (eds.) *The Formal Complexity of Natural Language*. Studies in Linguistics and Philosophy, (33): 320-334

Stromswald, Karin, Caplan, David, Alpert, Nathaniel & Rauch, Scott (2006). Localization of Syntactic Comprehension by Positron Emission Tomography. *Brain and language*, 52: 452–473

Tettamanti, Marco, Irene Rotondi, Daniela Perani, Giuseppe Scotti, Ferruccio Fazio, Stefano Cappa & Andrea Moro (2008) Syntax without language: neurobiological evidence for cross-domain syntactic computations. *Cortex*, 45(7): 825-38.

Tettamanti, Marco, Hatem Alkadhi, Andrea Moro, Daniela Perani, Spyros Kollias & Dorothea Weniger (2002) Neural correlates for the acquisition of natural language syntax. *NeuroImage*, 17:700-709.

Traxler, Matthew J., Robin K. Morris & Rachel E. Seely (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory & Language*, 47(1), 69–90.

Wager, Tor D., Edward E. Smith (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 3(4): 255–274.

Wanner, Eric & Michael Maratsos (1978). An ATN approach to comprehension. In Halle, M., Bresnan, J. & Miller, G. A. (eds.) *Linguistic Theory and Psychological Reality*. pp. 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.

Weir, David J. (1988). *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Ph-D. thesis, University of Pennsylvania.

Yngve, Victor H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104: 444–466.