

**The Role of Working Memory in
Linearization and Island Effects
in Natural Language**

A Thesis

Presented To

The Graduate School of

Human and Environmental Studies

Kyoto University

In Partial Fulfillment

of the Requirements for the Degree of

Master of Arts

Kirill Vasiltssov

January 2019

Abstract

In this thesis I explore the possibility that natural limitations on working memory explain several unacceptability phenomena together known as *island effects*. The core proposal is that pronunciation (or, technically speaking *Spell-Out*) of a sentence requires linearization of its structure and that this task cannot be carried out without holding some pieces of structure in *buffer memory*. We expect a certain subset of syntactically *possible* structures to be impossible to linearize due to lack of memory resources. As a consequence, we predict these sentences to be judged as unacceptable. The technical analysis of structures that contain islands reveals that their unacceptability is caused by requiring three or more pieces of structure to be held in buffer memory simultaneously. Not all islands are found to be the same with respect to memory resources they require. For example, Adjunct Islands require less memory than Subject Islands. I show that under minimal assumptions about syntax Condition on Extraction Domains (CED) is an incorrect generalization and original island effect data is better explained by several independent alternatives.

Contents

1	Introduction	3
2	Islands	5
2.1	Remarks on acceptability and grammaticality	5
2.2	Islands	7
2.3	Cross-linguistic picture	9
2.4	Summary	13
3	Parsing Theories	14
3.1	Superadditivity	14
3.2	Working memory	18
3.3	Summary	21
4	Modern Structural Approach	22
4.1	Barriers and phases	22
4.2	Linearization and Multiple Spell-Out	28
4.3	Summary	34
5	Island Effects as a Side Effect of Working Memory Limitations	36
5.1	Methodology	38
5.2	Linearization flow	39
5.3	Other islands	54
6	Discussion	55
7	Conclusion	59

Chapter 1 Introduction

In this thesis I explore the possibility that natural limitations on working memory explain several unacceptability phenomena together known as *island effects*. The thesis is written within the generative grammar tradition, but only the technical chapters require familiarity with the main tenets of Government and Binding Theory and Minimalist Program.

The core proposal is that pronunciation (or, technically speaking *Spell-Out*) of a sentence requires linearization of its structure and that this task cannot be carried out without holding some pieces of structure in *buffer memory*. Since our memory resources are naturally limited, we expect a certain subset of syntactically *possible* structures to be impossible to linearize and, as a consequence, be judged as unacceptable. The technical analysis of such structures reveals that their linearization requires holding three or more pieces of structure in buffer memory simultaneously.

In chapter 2 I give a detailed overview of island effects: their history, definition, Subjacency Condition as a classical generative explanation, and the cross-linguistic picture, namely how persistent island effects are in different languages and what might possibly explain this variation.

In chapter 3 I review how experimental psycholinguistic data supports the significance of island effects and what existing parsing theories, i.e. theories that do not rely on constraints expressed in syntactic terms, have to say about island effects. There I also look at how working memory should be best understood nowadays.

In chapter 4 I review modern approaches to island effects within generative grammar. In particular, I focus on two minimalist theories by Nunes & Uriagereka (2000) and Fox & Pesetsky (2005) that try to derive some known constraints on movement (CED and constraints on Object Shift) from the need to linearize syntactic structures.

In chapter 5 I develop a new theory that explains island effects by limited memory resources available to syntax during linearization of structures. I show that most unacceptable cases require three or more pieces of structure to be held in buffer during a certain linearization step, which is more than

normally available to us.

Chapter 6 is a discussion and chapter 7 is a conclusion.

Chapter 2 Islands

2.1 Remarks on acceptability and grammaticality

In syntactic literature it is very common to see two distinctions: **acceptable/unacceptable** and **grammatical/ungrammatical**. The terms acceptable and grammatical are sometimes used interchangeably, but there is a very obvious reason why this is not welcome. Grammaticality is a theoretical notion, while acceptability is an observational one. Grammaticality is a property of sentences, while acceptability is a property of utterances. Of course, since our theories are partially based on observations very often utterances that are judged as acceptable are also grammatical sentences, and utterances that are judged unacceptable are also ungrammatical sentences. By saying that a sentence is grammatical we mean that it is possible to assign an interpretation to it, which in turn means that it conforms to the rules of grammar as specified by a linguist.¹ However, there are utterances that are judged acceptable although it is impossible to assign any interpretation to them at all, and there are utterances that are judged unacceptable although they are perfectly grammatical on analysis. Thus we find all pairs +G/+A, +G/-A, -G/-A, -G/+A. Below are some of the examples that are grammatical but unacceptable:

- (1) a. +G/-A That that that Bill left Mary amused Sam is interesting
is sad.
- b. +G/-A The horse raced past the barn fell.

These are the cases of self-embedding and garden path ambiguity, respectively. On the other hand, the following sentence is judged as perfectly

¹“Interpretation” is used here in the technical sense. In modern generative grammar it is common to assume that Merge is unconstrained and can build infinitely many structures. Not all of these structures, however, match the format that is set by the conceptual-intentional interface, the “meaning” side. Thus, it is possible to assign an interpretation only when the generated structure matches the format. Of course, we still do not know much about the nature of the C-I interface and it is possible that every structure is interpretable. For example, one could say that to have a feeling that an expression is deviant, it must be interpreted first anyway.

acceptable, but on a closer look it doesn't have any meaning at all. This is a good example of a grammatical illusion:

- (2) -G/+A More people visited Rome last year than I did.

Compare this to the perfectly comprehensible sentence below:

- (3) More people visited Rome last year than visited Venice.

Perhaps it is better to formulate the -G/+A case as assignment of the *wrong* interpretation. The reason is that when a person judges something as acceptable they feel like it does have *some* interpretation. They are confused only if asked to explain what the sentence like the one above actually means. Then we can include another well-known sentence in this category:

- (4) No eye injury is too trivial to ignore.

When people hear this sentence they understand it as meaning that eye injury is never trivial and should not be ignored, while it actually means that no eye injury is so trivial that one cannot ignore it. A structurally equivalent sentence is *No atom bomb is too small to ban*. It actually means that there is no atom bomb so small that it cannot be banned, so all atom bombs can be banned. In these cases people assign the *wrong* interpretation to the sentence.

As we have just seen the degree of acceptability is a poor indicator of grammaticality (and vice versa). The data never comes as grammatical or ungrammatical and it is a linguist's job to determine which one it is. By developing theories we actually "create" +G/+A and -G/-A cases (usually in the form of contrasting *minimal pairs*): they are the points of connect between our theories and observation. Therefore it is incorrect to say that any uncovered acceptable/unacceptable case *per se* makes a theory weaker or that some set of unacceptability phenomena *must* have a principled explanation – doing so is orthogonal to the idea that data is created by a theory.

I chose to mention this because all island data is essentially created data. It takes a person and an argument to declare that a certain set of phenomena is a natural class. Thus, the validity of the declaration depends on the strength of the argument. If it turns out that the argument is not strong enough it

undermines the very idea that such a collective phenomenon even exists. We must evaluate arguments not only by how well they cover all data, but also respect how well they cover what they are constructed to cover. If island data can be split in 2 or 3 sets of phenomena, each of which is elegantly explained, it is better than explaining all phenomena with an ad hoc argument.

2.2 Islands

By *island effects* we usually (broadly) understand a collection of cases where unacceptability of a sentence is attributed to involving a dependency between a *filler* and a *gap* inside a certain class of structures. In the light of the discussion above, consider the following sentences from Chomsky (1964):

- (5) a. Mary saw the boy walk towards the railroad station.
- b. Mary saw the boy who was walking towards the railroad station.
- c. Mary saw the boy walking towards the railroad station.

“Walk towards the railroad station” is a verb complement in (5a). In (5b), on the other hand, “who was walking towards the railroad station” is a relative clause that forms a noun phrase with “the boy.” Both sentences are unambiguous. (5c), on the other hand, is ambiguous, because “walking towards the railroad station” can be interpreted either as a verb complement or as a relative clause. However, Chomsky argues, the question “What did Mary see the boy walking towards?” is unambiguous although its declarative counterpart is not. The only interpretation possible for this question is the one where “walking towards...” is a verb complement. He concludes that since relativization of “the railroad station” is just as unacceptable in (5b) (“**What did Mary see the boy who was walking towards?*”), we are dealing with a syntactic fact about relative clauses – that they are somehow fundamentally different from verb complements. This is supported by the fact that the phenomenon is observed even if we use other lexical items, meaning the cause is not semantic in nature. (5) is a classic example of minimal pairs, where every factor except structural configuration is controlled for.

This fact remained without principled explanation until Ross (1967). Ross identified many other structural configurations that somehow prevent a successful dependency between a gap within them and a wh-word. He dubbed them *islands*.

- (6)
- a. * What do you wonder [CP whether John bought ___]?
 - b. * What did you make [NP the claim that John bought ___]?
 - c. * What do you think [NP the speech about ___] interrupted the TV show?
 - d. * What do you worry [CP if John buys ___]?
 - e. * What did you meet [RC the scientist who invented ___]?
 - f. * What did [CP that John wrote ___] offend the editor?
 - g. * What did John buy [CONJP a shirt and ___]?
 - h. * Which did John borrow [NP ___ book]?

Ross analyzed them in the following way: the sentences are results of “chopping” transformations that remove the resumptive pronoun left by movement of the wh-word and create the gap. If the gap happens to be within an island (relative clause, complex noun phrase, conjunction phrase etc.) then the result of chopping transformation is unacceptable. Chomsky (1973) attempted to find another way to unify these phenomena. For Ross resumption was a result of movement, so it was not movement itself that rendered the sentences unacceptable. Chomsky saw resumption as a *non-movement*. For him, gaps are left by movement and thus the acceptability directly depends on whether movement “crossed” an island boundary. This is much easier to see in contrast with similar, but structurally different sentences which allow movement. Chomsky formulated the constraint as the Subjacency Condition (Chomsky (1977)), slightly reformulated here:

(7) **Subjacency condition**

A rule cannot move a phrase from position Y to position X (or conversely) in ... X ... [α ... [β ... Y ...] ...] ... X ..., where α and β are bounding nodes. Bounding nodes are IP and NP.

So, in (6a), for example, *what* crosses two IPs and violates Subjacency Condition:

- (8) *What do [IP you wonder [whether [IP John bought ___]]]?

In particular, what made the Subjacency Condition a good generalization is that it also predicted that comparatives are also the result of a movement transformation since they are subject to the Subjacency Condition and resumptive pronouns cannot be used in comparatives (9).

- (9) a. Mary isn't different than [what she was five years ago]
 b. Mary isn't different than [what John believes [that Bill claimed [that she was five years ago]]]
 c. *Mary isn't different than [what John believes [Bill's claim [that she was five years ago]]]
 d. *Mary isn't different than [what I wonder [whether she was five years ago]]

2.3 Cross-linguistic picture

After more work on similar phenomena in non-English languages was done, we learned that island effects are persistent cross-linguistically, though with slight variations that eventually had to be explained. It is common to distinguish between *surface* (superficial) differences and *deep* differences. For example, the so-called that-trace effect is observed in English but not in Italian, when a subject occurs post-verbally as in (10) (Rizzi, 1982).

- (10) a. *Who did you say that ___ wrote this book?
 b. Chi_i hai detto che ha scritto questo libro ____i?
 who have.2SG said that has written this book

One example of a surface difference in islandhood is the acceptability of extraction from independently motivated *major subject constructions* (MSC) found in some East Asian languages like Chinese, Korean and Japanese (Han & Kim (2004); Hsu (2008); Ishizuka (2009)). MSC is a construction that contains more than one nominative subject in a clause and matches certain conditions that I will mention below:

- (11) Sinsi-ga yooohuku-ga yogoreteiru
gentleman.NOM clothes.NOM dirty
Lit. ‘The gentleman, the clothes are dirty.’

“Sinsi” is said to be the major subject here. The following are the examples of long-distance extraction that are said to be acceptable:

- (12) a. [[[e_i chuan ____j de] yifu_j] hen piaoliang de] na-ge nühai_i
wear DE dress very pretty DE that.CLF girl
‘That girl is such that the clothes that she is wearing are very pretty.’
- b. [[CP Op_i [[IP ____i [[NP [[CP pro_i ____j kiteiru]
Op pro wearing-is suit.NOM
yooohuku_j]-ga yogoreteiru] [sinsi_i]]
dirty-is gentleman
‘That gentleman is such that the suit that he is wearing is dirty.’

These cases seem to involve a long-distance dependency between a NP and a gap within a relative clause, which is believed to be a strong island. However, some argue that in fact the NP *binds* the *pro* inside the relative clause and thus never crosses the relative clause boundary by movement. What appears to be a case of “double” relativization as in (12) is in fact single relativization (by movement) with the true gap only in the higher clause. This argument is supported by evidence that only a subclass of relative clauses allows such apparent extraction; it is constrained by the same constraints that apply to MSCs. The known restrictions are e.g. sensitivity to argument structure (only *be*-type predicates like *yogoreteiru*) or the restriction that the apparent gap be a subject.²

²While the evidence seems persuasive, we must not use the word “dependency” without explicitly saying what “dependency” means. It remains unclear whether there can exist dependencies of different nature, e.g. a dependency between a moved (internally merged) item and its canonical position, an agreement dependency or binding dependency. Some authors believe that only movement (internal merge) is what can create a dependency. Thus, for example, there is no binding of PRO by an NP antecedent in control configurations - there is only movement of NP from the embedded position and PRO does not exist. This is the idea behind the movement theory of control developed by Hornstein (1999). This is a good example of minimalist theorizing. The examples above are not instances of control; but they do involve a dependency that is supposedly not a result of movement.

The deep differences posit a more serious challenge for the universal theory of islands. One characteristic example is the difference in what counts as a bounding node in English and Italian. Rizzi (1982), on the basis of empirical data, concludes that movement in Italian is subject to the Subjacency Condition, but the bounding nodes must be NP and CP and not NP/IP. This led to the suggestion that this constraint is subject to parametrical variation. The idea that such parameter might exist was further supported by data from other languages, like Spanish (Torrego (1984)):

(13) Absence of WH-island effects in Italian

- a. Il solo incarico che non sapevi a chi
 the only task that not knew.2SG to whom
 avrebbero affidato è poi finito proprio a te.
 have.3PL.COND assigned is then ended up right to you
 ‘The only task that you didn’t know who they would assign
 ___ to was then given right to you.’
- b. Tuo fratello, a cui mi domando che storie
 your brother, to whom myself ask.1SG what stories
 abbiano raccontato, era molto preoccupato.
 have.3PL told, was very worried
 ‘Your brother, who I wonder what stories they told ___, was
 very worried.’

Perhaps the most puzzling example of variation are Scandinavian languages such as Swedish (Engdahl (1980)). It is reported that Subjacency Condition is absent from Swedish since most extractions out of islands that are unacceptable in e.g. English are acceptable in Swedish. If it is true then no surface difference could possibly hint at why such extraction is possible; and even if variation was caused by some parameter it is unclear what kind of parameter it should be. However, further research found that not *all* extractions are acceptable. Kush & Lindahl (2011) show that just changing the main verb in the minimal pair might lead to unacceptability of movement from a relative clause as predicted by Subjacency Condition:

Whatever we may call it (e.g. *binding*), it is crucial to make sure that such a dependency is really plausible from the minimalist point of view.

- (14) a. Den teorin känner jag ingen som tror på.
 That theory know I nobody that believes in
 ‘That theory, I know nobody who believes in.’
- b. *Den här teori, finns det ingen som tror på.
 ‘That theory, one finds nobody who believes in.’

I will return to the Scandinavian case later during discussion of how island configurations affect sentence processing.

Another challenge that any theory of islands meets is that there are languages where *wh*-words do not move. This means that since there is no extraction from subjects or relative clauses, no island effect is going to be observed. However, evidence suggests otherwise. Japanese and Chinese are among the most studied *wh*-in-situ languages and data shows that certain sentences with a *wh*-word inside a subject or adjunct are unacceptable. This is interesting, because no movement actually crosses the boundaries of islands, but implies that some dependency does exist (e.g. dependency between an interrogative *C* and the *wh*-word).³ When it comes to islands, languages like Japanese or Chinese exhibit the so-called *argument/adjunct asymmetry*. *Wh*-adjuncts such as *why* and *how* cannot appear inside of island structures in *wh*-in-situ languages. Consider the following examples:

- (15) [John-wa [[ADJ kare-no okusan-ga nani-o katta kara]
 John.TOP he.GEN wife.NOM what.ACC bought because
 okotta no]?
 get.angry Q?
 ‘What did John get angry because his wife bought *t*?’
- (16) * [John-wa [[ADJ kare-no okusan-ga naze atarasii doresu-o
 John.TOP he.GEN wife.NOM why new dress.ACC
 katta kara] okotta no]?
 bought because get.angry Q?
 ‘Why did John get angry because his wife bought a new dress *t*?’

I will return to this interesting distinction in chapter 5, where I analyze similar cases in detail.

³This, of course, brings us to the above-mentioned problem of whether there is such a thing as dependency without movement.

2.4 Summary

In this chapter we saw how Subjacency Condition helped us bridge our theories and observations in such a way that a range of *prima facie* unrelated phenomena received a principled explanation. Chomsky's understanding turned out to be more enlightening than Ross's. A closer look at the definition of Subjacency Condition reveals that it is actually not about islands at all – rather about what kind of movement is disallowed. It gives us a very obvious hint – dependencies cannot be too long. From this point of view it is also possible to do metatheoretical work: *if* Subjacency Condition (or its more modern definitions) is correct then natural language might be constrained on a much more abstract, natural level by such things as locality. We also saw that Subjacency Condition doesn't always hold of similar configurations in other languages. This does not mean that locality does not affect those languages – it means that locality might manifest itself in slightly different ways that are not captured by Subjacency Condition. Alternatively, as I mentioned in the introduction, this could mean that even when limited to English data, the success of Subjacency Condition might be accidental. I believe that thinking this way is better, because otherwise researchers of other languages might feel the duty to fit their explanations as close to Subjacency Condition as possible, missing opportunities to formulate locality conditions in other ways. Moreover, in the light of recent findings Subjacency Condition seems to be rather unnatural; it does not make much sense as an independent constraint from a biolinguistic point of view and thus requires re-interpretation in the minimalist framework. In chapter 4, we will see more modern approaches to the problem of locality. But before that, in the next chapter I will briefly cover theories that attempt to derive island effects from the way sentences are processed by speakers.

Chapter 3 Parsing Theories

In this chapter I will review recent experimental studies from psycholinguistics that mainly deal with syntactic phenomena, and, in particular, with island effects. Since many explanations presented require agreement on what working memory is, I will also review how working memory is understood nowadays.

3.1 Superadditivity

Insights from psycholinguistic work are important because if it is true that (at least most) islands effects can be reduced to the complexity of the processing task, we would not have to guess which grammatical explanation is better. Grammatical explanations today are different from grammatical explanations 20 years ago, mainly because the way we evaluate theories has changed. It will change again, but the fact that humans must process linguistic expressions and the way they do it do not change. This is why a successful psycholinguistic explanation of e.g. island effects would be consistent with any grammatical explanation of them, the latter becoming epiphenomenal. Of course, this does not mean that we should aim to eliminate syntax from our theories of natural language. However, eliminating **constraints** formulated in syntactic terms seems to me to be a desirable goal. It would allow us to maintain the view that domain-specific linguistic endowment is extremely minimal—the goal that current minimalist theories aim for.

So-called *processing explanations* seek to explain island effects as arising from constraints on the *parsing system*. Sprouse & Hornstein (2013) argue that it is not often clear what *processing explanations* and *grammatical explanations* refer to: if a processing explanation simply mirrors some constraint on structure-building then it is equivalent to grammatical explanation and not very useful. What we are looking for are *possible* structure-building operations which aren't carried out due to constraints on resources available to the parsing system. For this reason, Sprouse et al. (2012) and Phillips (2013) suggest a term *reductionist explanations*. These explanations “seek to

reduce island effects to one or more components of the sentence-processing system that are motivated by language-independent perceptual or cognitive properties.” The authors see the requirements that any reductionist theory must meet in the following way:

First, reductionist explanations must explicitly specify which mechanisms give rise to the island effect. As mentioned above, these mechanisms must not be structure-building operations or abstract constraints on structure-building operations, otherwise they are equivalent to an explanation in grammatical terms. Second, the mechanisms of the reductionist explanation must be independently motivated; in other words, the mechanisms should be necessary to explain phenomena other than island effects. If the mechanisms are not independently necessary, then the reductionist explanation is not truly reductionist.

To apply this way of thinking to islands means to assume that movement across the island boundary is possible but unacceptable because its processing requires more cognitive resources than are available to an individual. Let us take a look at how islands effects are treated in the experimental literature. Sprouse & Hornstein (2013), using the classical minimal pair methodology, designed an experiment using the factorial definition of islands. What does factorial definition mean? From the reductionist point of view, unacceptability is caused by the combination of two factors – the presence of an island configuration (e.g. *whether*-clause) and a long-distance dependency between a matrix and an embedded clause.

- (17) a. Who ___ thinks that John bought a car? NON-ISLAND | MATRIX
 b. What do you think that John bought ___ ? NON-ISLAND | EMBEDDED
 c. Who ___ wonders whether John bought a car? ISLAND | MATRIX
 d. *What do you wonder whether John bought ___ ? ISLAND | EMBEDDED

By designing the experiment this way and providing a baseline in the form of (17a) we can isolate the effects of each of the two factors mentioned above.

The difference between (17a) and (17b) isolates the effect of a long-distance dependency. The difference between (17a) and (17c) isolates the effect of *whether*-clauses.

The simplest prediction that can be made is that each of the factors is associated with additional costs of processing a sentence and the cost of the two factors combined is simply a sum of their independent costs. So, the combined cost is high enough to render a sentence unacceptable, but each cost independently does not cause unacceptability. This prediction can be visually represented with a graph below, where the slope of the line represents the main effect of long-distance dependencies. The lines are expected to be parallel if the combined cost is a sum of the two costs.

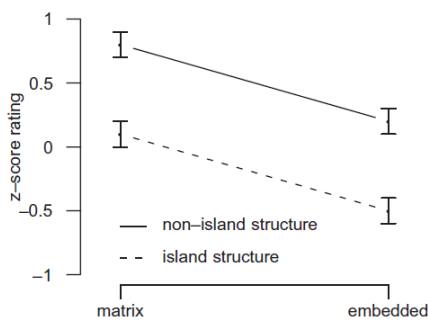


Figure 1: A graphical example of a linearly additive effect with a 2x2 design (Sprouse et. al (2014: 27))

The results of the experiment were not consistent with the simplest prediction. Surprisingly, in all island configurations tested the combined cost turned out to be greater than the sum of parts (see Figure 2). We say in this case that the island effects are *superadditive*. One thing that is clear is that observed superadditivity cannot be explained with the simplest reductionist account. Sprouse believes there are two possible solutions: either this means that there is a real grammatical constraint (like Subjacency Condition or all of its later versions) that functions as yet another factor or we are missing some component of processing (apart from cost) that influences the acceptability outcome. As for the latter, the only strong hypothesis so far was proposed by Kluender & Kutas (1993). Their argument is roughly as

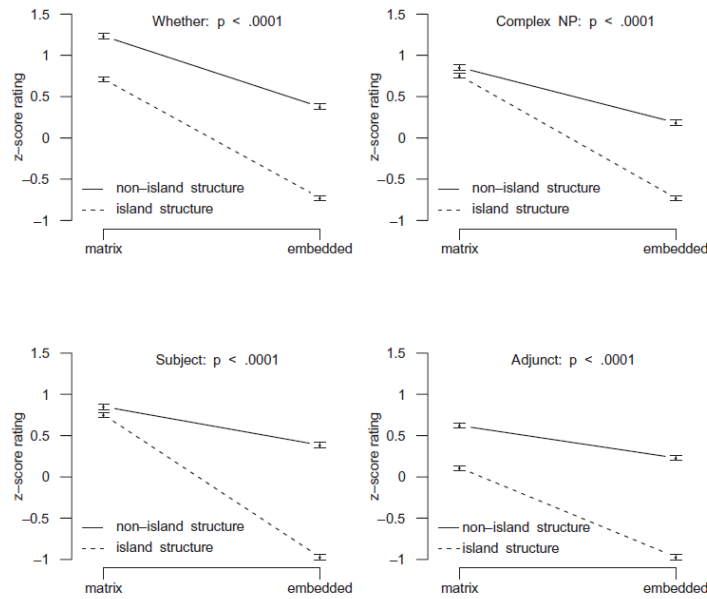


Figure 2: Results of magnitude estimation experiments for four island types from Sprouse et. al (2012: 28)

follows:

1. There are operations for processing long-distance dependencies.
2. There are operations for processing island configurations.
3. There is limited pool of working memory resources.
4. Long-distance dependency operations require working memory resources.
5. Island structure processes require working memory resources.
6. The sum of the resources required by 4 and 5 is greater than the resources available.
7. There is a parsing algorithm that deploys the operations in 1 and 2 simultaneously.

While not uncontroversial this approach provides the reasonable missing factor and makes a very clear prediction: the size of working memory matters. On the other hand grammatical explanations implicitly make the opposite prediction: working memory cannot influence acceptability.

3.2 Working memory

Sprouse & Hornstein (2013) designed another experiment in which the working memory capacity of an individual and the size of the superadditivity effect for this individual are measured. The prediction of the resource capacity theory can be interpreted this way: there must be a strong inverse relationship between the strength of the island effect and working memory capacity for each individual. The results of the experiment also turned out to be incompatible with the predictions: no strong relationship was observed. Hence, at first glance, the results appear to support the grammatical view. But I can think of at least two objections:

1. **Working memory measurement:** In the experiment above working memory was measured by widely accepted methods known as *serial recall* task and *n-back* task. The former requires the participants to recall the series of words in the order that they were presented to the participants. In the latter, a series of random letters are presented to participants one at a time and they are required to press a button if the current letter was also presented n items previously. Although it is argued that these two tasks can account for most individual differences in working memory, it is at least reasonable to doubt whether working memory is exploited in the *same* way when processing actual sentences, where rigid order and linear distance might not matter as much as in the tasks above.
2. **Working memory type:** In fact, some researches hypothesized that there might be some kind of dedicated (encapsulated) working memory for syntactic processing, for example see Caplan & Waters (1999). Analyzing the results of their experiments and previous studies, the authors conclude that “the use of processing resources in assigning syntactic structure and using that structure to determine the meaning of a sentence is separable from a subject’s verbal working memory capacity as measured by standard working memory tasks.”

Another interesting finding regarding processing of dependencies and work-

ing memory comes from Wagers (2013). Ever since the famous paper by Miller (1956) on the limit on number of items that can be maintained in working memory (the “magic number” 7 ± 2), it became common to think of working memory as a kind of a stack with a limited number of slots. Wagers argues that it is also possible to interpret the term “working memory” in a broader way, “as the sum of the many component parts and processes that support the encoding, retention, and recollection of recently encountered information.” He provides support for this broader view of working memory in the form of recent experimental findings. It appears that very few task-relevant encodings can be successfully maintained in memory and directly available during the task, as few as just one. This one piece of information is said to be in the *focus of attention*. There is an observable distinction between information in the focus of attention and other information which is held in working memory, and so we are compelled to view working memory in the broader way. According to Kluender & Gieselmann (2013), studies of visual working memory in monkeys (Zaksas & Pasternak (2006)) and neural imaging of humans (Jha & McCarthy (2000)) “have shown neural activation during both encoding and retrieval phases, but no neural correlates of a maintenance function in between.” Likewise, event-related brain potential (ERP) studies of grammatical but difficult-to-process object filler-gap dependencies yield similar results (Phillips et al. (2005)).

The classical view is that during the processing of e.g. wh-dependencies the dependency is opened upon encounter of a filler (wh-word) and *maintained* in WM until the gap is available. The more recent view is that since all other words of the sentence must also be properly processed they come into the focus of attention, and since the focus is extremely limited the open dependency cannot be actually maintained. So, what matters is not how much can be maintained simultaneously but how hard it is to retrieve some information when needed once it is out of focus. It is also possible that the information out of the focus of attention could be as coarse and minimal as e.g. just a [+wh] feature and not a specific lexical item. As a consequence, if a sentence features more than one wh-word, multiple candidate encodings are available at a gap for retrieval and the system is prone to retrieval of a

wrong item (similarity-based interference). This seems to be a good experimental support for explanations given in grammatical terms, like Relativized Minimality of Rizzi (1990). But what is also important is that while *maintenance* plays a role in working memory measurement tasks like serial recall or n-back, it does not seem to play any role during processing of dependencies in a sentence. Thus arguments like that of Sprouse & Hornstein (2013) against the existence of inverse relationship between WM capacity and the size of the superadditivity effect do not apply in this case.

Similar results are reported by Kluender & Gieselmann (2013) for so-called negative islands. Negative islands are configurations in which a negative element functions as an intervening item between a moved wh-word and its canonical position, much like a wh-word prevents other wh-words from moving in wh-islands.⁴ The classic example would be (18):

- (18) a. How did they clean the windows ___ ?
 b. *How didn't they clean the windows ___ ?

The author argues that there does not seem to be any global constraint that would cover all negative islands, rather such factors as “negation, extraction, and referentiality all carry processing costs evident not only in isolation, but also in combination with each other: effects were cumulative even in grammatical sentences.” Kluender & Gieselmann suggest that degraded acceptability could be a result of similarity-based interference, mentioned above: if what is retrieved is some minimal information like “operator” then since both wh-word and negative item are operators they interfere with each other at the gap.

The fact that superadditivity effect is so consistently observed in experimental studies of island effects can also contribute to solving problems that puzzle proponents of grammatical explanations. One puzzle mentioned in the previous chapter is the apparent absence of island effects in Scandinavian languages. To explain this anomaly in purely grammatical terms requires ad hoc modifications which is not very desirable. Recently, Kush et al. (2018)

⁴Note that movement of argument wh-words such as *what* tends to be acceptable in these cases. But argumenthood is not the only factor, as is clear from sentences like **Which state doesn't he come from ___ ?* (Kuno & Takami (1997: 557)).

found experimentally that superadditivity effect *is* observed in acceptability judgements of island configurations by native speakers. In other words, Scandinavian speakers judge the sentences as acceptable even when a dependency crosses the island boundary, but even so *the way* two factors combine is identical to that of English. This is an impressive finding that implies that while there may not be island *effects* in Scandinavian there are *islands* in some sense.

3.3 Summary

The superadditivity effect is a very strong indicator of the presence of some factor distinct from mere processing costs associated with certain structural configurations. It is not yet known what this factor may be but, as resource capacity theories suggest, it could reasonably be the limit of memory resources available to an individual during the task. This view is controversial since no strong relationship between the WM capacity and superadditivity effect has been confirmed. However, this could be due to the fact that syntax exploits some kind of dedicated WM as argued by Caplan & Waters (1999), not measurable by usual methods like serial recall or n-back. Besides, the traditional methods measure the ability to maintain ordered representations in memory, while processing complexity rather depends on how difficult it is to retrieve information not in the focus of attention anymore. The models of verbal working memory that de-emphasize maintenance in favor of encoding or retrieval appear to make better predictions. Evidence in favor of such models comes from two independent experiments on wh-islands and negative islands that show similar patterns. However, since the explanations rely on similarity-based interference, it is not very clear how they might explain many other types of islands such as Subject Island or Adjunct Island, which are not characterized by such factors as the embedded clause boundary or similar intervening items. In the next chapter, we will see how modern theories of locality explain Subject/Adjunct Islands along with other phenomena like Object Shift.

Chapter 4 Modern Structural Approach

4.1 Barriers and phases

In this chapter I review modern generative approaches to island effects (and more generally, to the problem of locality): barriers and phases. It is worth mentioning that islands are not the only examples of how locality constrains syntax. While the barriers framework certainly aimed to cover all island data, islands were not the main focus of the phases framework. As a result some of the phenomena are explained at a cost of losing coverage that Subjacency Condition already provided. Nevertheless, both these approaches (phases in particular) are superior conceptually, providing much more natural explanations that allow linguistics to become a more interdisciplinary science. Moreover, the drawbacks of these approaches encourage theoretically valuable alternatives, which I will also review below.

Subjacency Condition was no doubt a breakthrough in syntactic theorizing. However, nothing was said about *why* IP and NP should be bounding nodes. What is so special about them? Chomsky (1986a) tried to answer this question and deepen our understanding of locality in syntax in his *barriers* framework. The biggest advantage of the new approach is that Chomsky not only identified nodes which serve as barriers to movement (hence their name *barriers* in the book), but showed how barrierhood of a node is determined by more primitive relations such as being *theta-governed*. The system developed by Chomsky in *Barriers* is very rich and covers such cases of unacceptability as the impossibility of extraction from subjects or adjuncts, unexplained by Subjacency Condition. Moreover, by introducing Empty Category Principle (ECP) into theory, it became possible to explain even the slight difference in acceptability between extraction of adjuncts (less acceptable) and extraction of complements (more acceptable) out of indirect questions. Together with a modified definition of Subjacency (given below), ECP provided great data coverage.

The so-called Subject Islands and Adjunct Islands have traditionally been explained as following from CED (Condition on extraction domain) of Huang

(1982), which bans movement from non-complement phrases. The condition is extremely simple: since neither subjects or adjuncts are complements (being specifiers and adjuncts, respectively), you cannot move anything from them. The nature of CED is rather non-explanatory and barriers were a good chance to unify CED and phenomena previously covered by Subjacency. The definition of a barrier is complex and itself relies on definitions which in turn rely on other definitions (Epstein et al. (1998)):

(19) *Barrier*

Z is a barrier for Y iff

1. Z immediately dominates W, W a blocking category for Y, or
2. Z is a blocking category for Y and Z is not IP.

(20) *Blocking Category*

Z is a blocking category for Y iff

1. Z is not L-marked, and
2. Z dominates Y.

(21) *L-marks*

X L-marks Y iff X is a lexical category that theta-governs Y.

(22) *Theta-Government*

X theta-governs Y iff

1. X is a zero-level category, and
2. X theta-marks Y, and
3. X and Y are sisters.

On the basis of the above definitions, Chomsky redefines Subjacency (as a structural property of a lexical item) as well as the constraint itself which relies on Subjacency in the following way:

(23) *Subjacency*

β is n -subjacent to α iff there are fewer than $n+1$ barriers for β that exclude α .

(24) *Subjacency Condition*

1. In a chain with a link (α_i, α_{i+1}) , α_{i+1} must be 1-subjacent to α_i .
2. 0-subjacency yields a more acceptable structure than 1-subjacency.

The Subjacency Condition formulated this way is superior to that of Chomsky (1977). In the barriers framework, movement is understood as chain-formation between an antecedent (head of a chain) and its canonical position which is a trace (tail of a chain). The constraint applies to representations: a representation violates the Subjacency Condition if there are two or more barriers between an antecedent and a trace in a chain. Let us see how it helps us explain Subject Islands and Adjunct Islands, previously only explained by CED. Consider the following sentences:

(25) **Subject Island**

- a. The man who [IP [NP pictures of ___ are on the table]
- b. The book that [IP [NP reading ___ would be fun]
- c. * Who does [IP [CP ___ that Mary likes ___] surprise you]?

Since the subject XP *pictures of who* is not L-marked by any lexical head it is a blocking category (and a barrier) for a moving phrase. IP inherits barrierhood from the subject since it immediately dominates it. There are two barriers between the antecedent and the trace. Therefore, the sentence is ungrammatical.

(26) **Adjunct Island**

- a. * Whom did [IP they leave [before speaking to ___]]?
- b. * To whom did [IP they leave [before speaking ___]]?

Here, adjunct XP *before speaking to whom* is not L-marked and IP, again, inherits barrierhood from VP. There are two barriers between the antecedent and the trace. Therefore, the sentence is ungrammatical.

Barriers proved to be an empirically strong theory that is capable of explaining islandhood but what can we say about its contribution to our overall understanding of the faculty of language? If we look at all definitions that the definition of a barrier requires it seems that they are clearly too much to

regard as primitive relations that human syntax relies on. It is not clear how these relations can be learned or, if they are a part of our linguistic endowment then why *these* relations and not others and how they fit in the larger (neuro)biological picture. For example, in their famous book on syntactic relations, Epstein et al. (1998) argue at length that it is neither desirable to posit so many artificial syntactic relations from the minimalist point of view nor necessary since many relations can be derived from the most basic operations such as Merge, which, in their opinion, gives birth to the only primitive relation c-command.

The problem of locality (and with it, the problem of islands) received another treatment in the relatively new *phase* framework. Originally, phases were conceived to solve a problem of movement in existential sentences. To illustrate the problem, first consider the following sentences:

- (27) a. There seems to be a man in the garden.
 b. * There seems a man_i to be t_i in the garden.

Here, *a man* can't satisfy feature-checking requirements (EPP) of the infinitival T by movement. Chomsky (1995) argues that Merge-over-Move principle (MOM) explains this by assuming lexical insertion operations (such as insertion of *there*) to be less expensive than movement operations. However, MOM couldn't explain cases like the following:

- (28) There was a rumor that a man_i was t_i in the room.

In this example, although *there* was available for insertion, *a man* moved. If MOM is true we would wrongly predict that this sentence is ungrammatical. This is the problem that the phase framework addressed. Chomsky (2000) introduces the concept of lexical subarrays such that (at least) a CP clause has its own lexical subarray. Now (28) can be explained by saying that there is no expletive in the lexical subarray of the embedded clause and that is why *a man* has to move. The reason is that all feature requirements have to be satisfied within a phase because it is *transferred* to the interpretive components before the next phase begins, cyclically. This is the conceptual argument based on the idea that syntax is economical and uses the minimal necessary computational resources.

The original motivation for dividing building of a sentence into phases is different from what Subjacency aimed to explain. But can the locality facts be explained from this new angle? Chomsky (2000) proposed that Phase Impenetrability Condition (PIC) explains most of the cases in question:

(29) *Phase Impenetrability Condition*

In phase α with head H, the domain of H is not accessible to operations outside α , but only H and its edge.

Informally, this means that once a piece of structure that counts as a phase is built, nothing can be moved from within a complement of a phase head. This does not apply to whatever is in the specifier position of the phase head, its “edge.” Thus, for example, the embedded CP in the sentence below is a phase.

(30) * What do you wonder [CP whether John bought ___]?

The sentence is ungrammatical because by the time *what* is able to move (after the matrix C head is merged), the lower CP is already built and does not allow extraction out of its phase complement. But, as is permitted by PIC, *what* could have moved to the lower Spec/CP, which functions as the phase edge, before moving further. Why isn't this option possible here? One possible answer is that lower C does not attract the wh-word because one is already there (*whether*). However, if we adopt the idea that features do not mutually delete, but are simply “checked,” then nothing should prevent *what* from being attracted to Spec/CP. Another possible answer is that all sorts of items can be attracted to the edge by some other feature, like a generalized EPP, and a C head lacks it for some reason. This is very weak reasoning, since the generalized EPP itself seems to exist just to support the argument.

In a careful review of the phase framework by Boeckx & Grohmann (2007), they argue that in fact, “phases as currently conceived of are poorly understood, and that we would do well to explore alternative strategies to deal with the syntactic phenomena that phases have been designed to address.” They identify numerous problems that the phase theory faces. Some of the most important ones are the following:

1. As Ceplova (2001) argues, since all phase heads allow the material with unsatisfied uninterpretable features to move to its specifier (edge), *nothing* should be an island. This is especially problematic under the view that Merge is “free,” meaning that it does not require a purpose to apply but just does so. Accidentally, some of the structures built conform to what would otherwise be explained as movement to the required position, and thus are successfully interpreted. So, whether generalized EPP exists or not, an edge is always available for movement and PIC alone cannot explain the restrictions in question.
2. Unlike in the barriers framework, phase heads (at least C and ν) are identified as “rigid” bounding nodes and not dynamically. But this choice remains unjustified due to many existing counterarguments (see the original paper for details).
3. What happens to the transferred chunks of the representation? They need to be *kept* somewhere and then *recombined* in the right order (since, for example, we obviously cannot start pronouncing all complex sentences with an embedded clause), but that requires a distinct algorithm. As the term *recombination* makes clear, this is basically repeating the steps of a finished derivation, a kind of redundancy that Chomsky aimed to eliminate with the introduction of phases.

Additionally, it seems that phases fail to explain CED phenomena (both Subject and Adjunct Islands), while previous frameworks can. This is because originally DP (e.g. a subject) does not count as a phase. Therefore, nothing like PIC should prevent anything from moving out of the subject DP to the edge of the CP phase. One could stipulate that D is also a phase head, but this brings us to the problem mentioned above: how do we even know that something is a phase head?

The whole list of problems is impressive and much larger than the above. In fact, as Boeckx & Grohmann (2007) notice, there is very much in common between barriers and phases, so the latter are not a particularly new approach to locality in syntax. The only trade-off seems to be the conceptual superiority of phases, which is based on possibly favorable reduction

of resources used by the computational system. I say *possibly* because we currently do not know whether phases really achieve the goal of reduction of resources, and what are the resources that are supposed to be reduced, etc. Fortunately, there are promising alternatives.

4.2 Linearization and Multiple Spell-Out

The approaches above all have in common that the apparent locality constraints are explained either as direct constraints on movement (phases) or on the form the representations (and dependencies within them) must have (barriers). But syntax itself is not the only place to look for explanations. The modern theoretical work in linguistics is concerned with how the biological context of the faculty of language (e.g. a brain with limited resources, ways in which an expression can be externalized etc.) might explain what we previously thought to be specific to syntax. Chomsky (1999) formulated the Strong Minimalist Thesis:

... the human faculty of language FL [is] an optimal solution to minimal design specifications, conditions that must be satisfied for language to be usable at all... for each language L (a state of FL), the expressions generated by L must be “legible” to systems that access these objects at the interface between FL and external systems – external to FL, internal to the person. (Chomsky 1999: 1)

The thesis emphasized the rather obvious fact that whatever our faculty of language is, it must be so that we can think in expressions created by it and externalize them with whatever biological means we have. Especially, as concerns pronunciation it is easy to see what form our expressions must take before they are externalized – a string of ordered sounds. However, one of the most interesting observations in generative grammar is that linear relations between words do not seem to play much role in syntax. This implies that there must be a step which “converts” structured expressions created by syntax into strings – this process is referred to as **linearization**. Although probably unconcerned with the SMT, Kayne (1994) came up with Linear Correspondence Axiom (LCA) as a way to linearize structures. The pecu-

liarity of LCA consists in the fact that it only relies on structural relations. The slightly simplified version of LCA can be expressed in the following way:

(31) *Linear Correspondence Axiom*

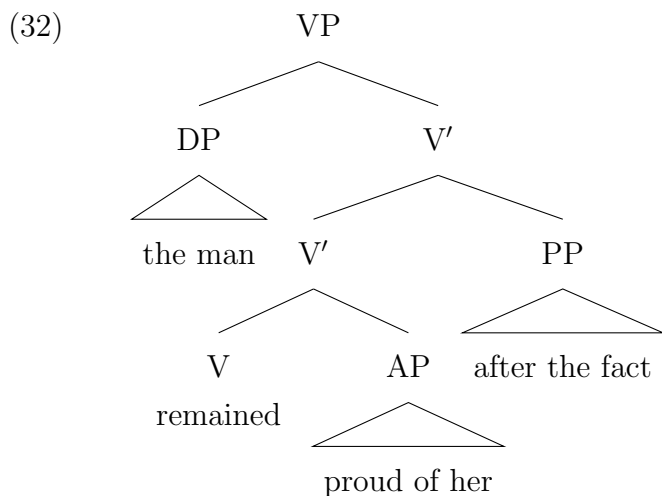
A lexical item α precedes a lexical item β iff α asymmetrically c-commands β .

Since, as Epstein et al. (1998) argued, c-command is the only fundamental syntactic relation in human syntax, LCA is a very simple SMT-compliant way to linearize structures. So, informally speaking, once a syntactic structure is built it is sent to whatever mechanism is responsible for its “conversion.” This process is usually called **Spell-Out**.

It is also worth asking whether “conversion” of structures to strings happens by step by step. Theoretically, it might be sometimes necessary to linearize a structure piece by piece than all at once. This is exactly the reasoning behind work by Nunes & Uriagereka (2000). First, Nunes & Uriagereka argue that if syntax is really economical (as we can see from Minimal Link Condition⁵) then restricting the number of Spell-Outs to one is an unmotivated stipulation. But are there cases where this is required? Their tentative answer is yes, because otherwise a structure could not be linearized in accordance with LCA.

Given the LCA, Nunes & Uriagereka observe that any derivation will necessarily involve stages where LCA cannot be applied because no c-command relation is established between some items. For example, in the VP below, all terminals of *remained proud of her* can be assembled within a single derivational stage by Merge. But PP must be assembled independently before it is merged with the rest of the structure. Because of that, when PP merges with V' there will be no c-command relation between the terminals inside PP and V'. LCA cannot apply in this case.

⁵The Minimal Link Condition states that derivations with shorter links are preferred over derivations with longer links.



The solution that Nunes & Uriagereka give is to allow Spell-Out (and Linearize) to apply to a piece of structure before it is merged. Then it essentially becomes a “giant compound,” a head-like word that can merge and enter new c-command relations. To actually make such a compound still accessible to syntax after it is linearized they assume that its syntactic representation is of the form $K = [\gamma \alpha, \beta]$, where γ is the *label* of the structure. Under this approach it becomes extremely easy to explain why extraction out of subjects or adjuncts should be impossible, without stipulating a constraint like CED (Subject Island / Adjunct Island): e.g. a wh-word is linearized before a C head with a Q-feature merges with the structure. Consider a derivation of the following sentence:

(33) * What did [that John drank ___] surprise Mary?

First, VP *surprise Mary* is built. In parallel, the subject which is the CP *that John drank what* is also built. At this point we cannot Merge the two structures, because the result will not be linearizable by LCA. So we linearize the CP subject and merge it with VP. Now all of the subject’s contents are in c-command relationship with all terminals in VP. Then, T is merged with the result. The subject possibly internally merges with T to satisfy some featural condition. After that, interrogative C head is merged. It bears a Q-feature that must be checked and *what* inside the subject is the only item that can check the feature. However, at this point the whole subject

is already linearized and whatever is **inside** the subject is not accessible for any operations, including movement, as the authors argue.

However, this approach has several problems. First of all, the authors rely on the notion of *label* to make a linearized object available to syntax. Basically, this means that a label is an entity distinct from the linearized contents of the object yet capable of representing it. How is the label chosen? If it is distinct from everything else what is it? While these questions are being pursued by some nowadays, I believe that it is nearly impossible to answer them using the standard methods of syntactic analysis that enabled progress up to this point. In this case, we cannot be sure that the argument holds if one of its premises is unclear.⁶

Another problem is the choice of the piece of structure to linearize. In the example above we assume that the adjunct PP is linearized to be able to merge with V' and enter in *c*-command relations with its terminals. But if that is the only reason then we could just as well linearize V' prior to merge with PP. If this happens, we would wrongly predict that objects of the verbs should be inaccessible for extraction (at least in cases like the one above). Finally, the authors have nothing to say about how to linearize items that mutually *c*-command each other: in the example above this is V' and PP. Nunes & Uriagereka simply mention that “The linear order between the lexical items of L and the lexical items of K will then be (indirectly) determined by whatever fixes the order of adjuncts in the grammar.”

Another alternative is developed by Fox & Pesetsky (2005). They also focus on linearization and how it is related to phasehood. They tackle a slightly different problem, but conceptually their goal is the same: to derive known constraints on movement in the most natural way possible. Linearization can be thought of as a process that establishes precedence relations between merged items, because these are needed to make a structure into an ordered

⁶This is not a significant problem, since we usually do not require *lexical items* to have a label to be able to Merge. In general, we can say that Merge applies to lexical items as *atoms* to form a larger constituent. Only this larger constituent needs a label, whose function is to essentially turn the constituent into another atom to enable further Merge. This is roughly the way the issue of labeling is explored in Hornstein (2009). Since linearized structures can also be treated as atoms, they do not need a label.

string.

Fox & Pesetsky (FP) aim to derive the constraints from the fact that any movement is a *revision* of established precedence relations. They base their investigation on the idea that successive-cyclic movement through certain positions is not only possible but required due to some factors. If Chomsky is on the right track with phases and cyclic Spell-Out is real then maybe there is something that **requires** movement to the so-called phase “edge,” instead of simply **allowing** it along other options (e.g. moving to a non-edge position). FP’s main hypothesis is that an item can move as long as movement does not lead to revision of precedence relations in an already (spelled out) linearized structure. If such movement occurs, a contradiction arises where a lexical item both precedes and follows another lexical item. This contradiction renders a sentence unacceptable. However, it is important to note that what they take to be actually spelled out is not a complement of a phase head as Chomsky assumes, but the whole phase, or, in the authors’ terminology, *Spell-Out domain*.

In particular, FP target the constraint that is known as Holmberg’s Generalization (HG).

(34) *Holmberg’s Generalization*

Object shift of an element α from the complement domain of a verb β occurs only if β has moved out of VP. (Zwart (1994))

Clearly the way that HG is stated is not particularly illuminating, because it does not shed light on deeper properties of human FL that we are trying to discover. FP suggest that HG follows from the fact that movement out of VP is basically a revision of word order inside VP and not all possible movements avoid ordering contradiction. Below are the abstract predictions of their hypothesis.

(35) $[D' \dots X \alpha [D t_X Y Z]]$

Ordering: $X > \alpha \alpha > Y \rightarrow X > Y$ (no contradiction)

(leftward movement from a left-edge position)

Let D and D' be Spell-Out domains in FP’s terminology. After Spell-Out domain D is linearized, X moves to a higher position (e.g. to have a feature

checked by α). When D' is spelled out the order of X and everything within D remains the same as if it never moved, so no contradiction arises and movement is licit. However, if Y moves instead of X then at the stage of Spell-Out of D' a contradiction arises since now Y *precedes* X while within D Y *followed* X .

- (36) * $[D' \dots Y \alpha [D X t_Y Z]]$
 Ordering: $Y > \alpha \alpha > X \rightarrow Y > X$ (contradiction with $X > Y$)
 (leftward movement from a non-edge position)

Finally, the logic of the hypothesis is such that if both X and Y move while preserving their relative order then again no contradiction arises and the movement is licit.

- (37) $[D' \dots X \dots Y \alpha [D t_X t_Y Z]]$
 Ordering: $X > Y X > \alpha Y > \alpha \rightarrow X > Y$ (no contradiction)
 (leftward movement from both edge and non-edge positions)

In (38) below we can see how restrictions on Object Shift are perfectly predicted by this hypothesis. Swedish is a verb-second language and allows a verb to move to C position (as in (38a)). As a result, Object Shift is allowed because eventually the relative order of V and the object is **not** altered.

- (38) Object Shift blocked by (unmoved) verb intervener
- a. Jag kysste henne inte $[VP t_V t_O]$
 I kissed her not
 - b. * ... att jag henne inte $[VP kysste t_O]$
 ... that I her not kissed
 - c. * Jag har henne inte $[VP kysst t_O]$
 I have her not kissed

FP conclude that it is natural to think of the phase edge as the only position such that movement through it avoids precedence contradiction. That is, there is nothing special (like mysterious EPP features) that attracts items to it. It is worth noticing that this approach overcomes one problem with phases that original accounts couldn't: it allows for the spelled out contents to still be syntactically accessible for further operations.

In general, FP’s approach is successful in deriving HG from more fundamental requirements and naturalizes the notion of phase but it also has drawbacks, mainly in the form of hidden assumptions. First of all, it relies on phases as defined by Chomsky, although as noted by Boeckx & Grohmann (2007) the rigid choice of phase heads is not justified. Second, they do not consider impossibility of linearizing two merged complex phrases according to LCA which motivated Nunes & Uriagereka’s work in the first place. Instead FP provide their own vision of the operation *Linearize* that uses a much more complex definition of LCA, undesirable in our case.⁷ Third, as a consequence of the first and the second problems, subjects are not Spell-Out domains for the authors. That is, a DP subject is neither a phase that must be spelled out nor a constituent that must be linearized prior to merge in order to avoid a situation where terminals of DP are not in a c-command relationship with the rest of the structure. This means that whatever moves out of a subject does not cause precedence contradiction and is therefore acceptable. Thus, FP fail to explain Subject and Adjunct Island data. Fourth, this approach implies (and even requires it to be secured) that precedence relations of each already spelled out domain be checked again and again every time new Spell-Out occurs. This is at least not very optimal and at worst can cause identity problems in sentences like *John wants to hit John*, where both *John < wants* and *wants < John* is true.

4.3 Summary

We have looked at how islands (in a broad sense, domains that constraint movement) in human syntax have been approached after Subjacency Condition. The promising richness of the barriers framework, from a different, minimalist angle turned into undesirable redundancy and complexity. The phase framework provides explanations in a much more conceptually elegant

⁷Basically, FP define what they call *Laws of precedence* and comment that they are “...agnostic as to the balance between language-specific and UG-determined aspects of these Laws; e.g. whether UG encompasses a “Head Parameter” or not.” For FP, specifiers in English precede complements by law, so the problem that Nunes & Uriagereka faced never arises.

manner, but fails to cover as much phenomena as barriers do, unless modified. Moreover, as noted by Boeckx & Grohmann (2007), there is a very strong parallelism between barriers and phases, meaning that the latter are much like barriers but in minimalist terminology. Since there are numerous problems that phases face, it is desirable to explore alternative strategies. This is what Nunes & Uriagereka (2000) and Fox & Pesetsky (2005) do in their attempts to derive CED (and Subject/Adjunct Island) effects and constraints on Object Shift, respectively. Both alternatives assume that the very obvious requirements on linearization of a two-dimensional syntactic structure are behind the observed constraints. However, despite both empirical and conceptual advantages each approach has nonoverlapping drawbacks. This means that there is still a room for improvement. If we agree with the authors at least that linearization is behind most observed constraints, then it is necessary to provide an explanation that will combine advantages of both approaches while avoiding the drawbacks.

Chapter 5

Island Effects as a Side Effect of Working Memory Limitations

If we agree that considering how our theories fit into the larger biological picture is a priority and that Minimalist Program is currently the best way to reinterpret all existing proposals and bridge linguistics with other natural sciences then existing explanations of island effects must also be properly reevaluated. While there is little doubt that structural configuration is an important factor to consider, we have seen that attempts to actually formulate constraints lead to stipulating entities and relations that hardly make sense in any science other than linguistics. Perhaps this is the way things are supposed to work and the so-called granularity mismatch problem⁸ can never be solved. However, with recent growth of insights from interdisciplinary research, we know that at least the very fundamentals of the linguistic theory like the Merge operation are good targets for theorizing about in fields like neuroscience or evolutionary biology. Still, it is very hard to imagine that more primitives of the theory of syntax can make their way into interdisciplinary research. In this case, today, to explain a phenomenon as purely syntactical would amount to explaining it with Merge – an extremely hard task. Fortunately, many phenomena do seem to be explainable if we are allowed to make reasonable guesses about the *interfaces* that are available to the faculty of language. By interfaces we mean those parts of encapsulated mechanisms of the brain that are open for “interaction” with other encapsulated mechanisms but define a strict format of such interaction. As we have seen in the previous chapter, there is hope that the need to linearize

⁸This is how Poeppel & Embick (2017) define the granularity mismatch problem: “Linguistic and neuroscientific studies of language operate with objects of different granularity. In particular, linguistic computation involves a number of fine-grained distinctions and explicit computational operations. Neuroscientific approaches to language operate in terms of broader conceptual distinctions.” As Hornstein (2009) says, “The most the present proposals can lay claim to is that the suggested basic operations are of the right kind. Nothing more specific can be claimed.” All present proposals are “*speculative* neuroscience.”

structures built by Merge can explain many constraints on movement. The works by Nunes & Uriagereka (2000) and Fox & Pesetsky (2005) are good examples of such reasoning. In this chapter I provide my own alternative that explains some island effects by the way linearization exploits working memory resources.

My approach is not meant as a critique of the above approaches. Rather, I seek to achieve roughly the same results by making less assumptions about the syntax. The reason is that any critique of a particular approach like phases is based on the latter's incapability of explaining certain phenomena. This is problematic in two ways: a) it is very likely that a particular approach is not even meant to cover the cases in question and b) the critique implies commitment to the idea that the set of phenomena in question is a natural class in the first place, while the very existence of the approach that is being criticized shows that it might not be so. As long as a particular approach conforms to the boundaries set by the current agreement on the nature of language, it cannot be invalidated simply by the argument from data coverage. I believe that we should not take it for granted that all the islands can be categorized together as islands. Rather, we should let the available theoretical distinctions decide whether a universal account of original island data is possible.

In chapter 3 I mentioned a study by Caplan & Waters (1999) that argues for existence of a certain "type" of working memory, distinct from verbal working memory which is traditionally measured by such tasks as serial recall or n-back task. This distinct type is basically encapsulated working memory that is only used by syntax. To use a computer metaphor, we can think of memory as a limited resource that can be used by different processes, but in such a way that memory used by one process is not accessible to some other process. This is what encapsulation means here. Further, it is reasonable to assume that only a limited amount of memory is available to each process, since we cannot allow a situation where some core brain processes cannot function because no memory is available.

Let us then assume that whenever syntax requires working memory resources, it can only use a limited amount of it. Using more than available

leads to observable consequences. Observable means either being immediately obvious to a speaker through judgement, or by analysing experiment data that reveals some tendency like superadditivity. I claim that island effects (and the related phenomenon of superadditivity) are precisely *the* consequences of overusing memory resources. However, it is inconceivable that syntactic operations, in the simplest case Merge alone, might require working memory to operate.⁹ Linearization process seems to be a process that might require some memory. Below I will show how it does so, and how it affects acceptability of certain expressions.

5.1 Methodology

First, it is crucial that we are clear about methods of investigation. If by working memory we understand encapsulated memory then it is very possible that it is not measurable by traditional methods, as Caplan & Waters (1999) argue. But if we cannot measure it at all, meaningful study becomes impossible. Thus, I would like to use an indirect approach that can be characterized as follows:

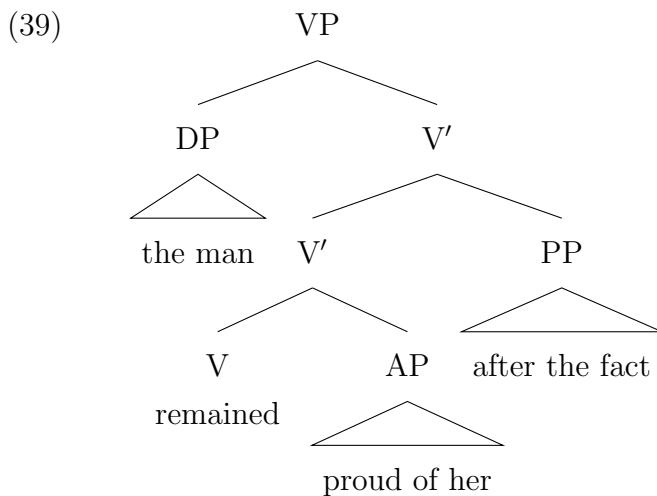
1. Define linearization and explain how/why it might access memory.
2. Identify a sentence S that is known to be unacceptable but violates no known grammatical constraint.
3. Assume the cause of unacceptability of S to be the usage of more memory resources than available.
4. Investigate other sentences S' (e.g. containing islands) and make sure that their linearization requires the same or more memory resources as S.

⁹One could conceptualize Merge as an operation defined over *workspaces*, which are collections of lexical items preselected for Merge. Arguably, a workspace requires some memory. Then, one could take advantage of phases to prevent workspaces from becoming too large. In fact, the idea of workspaces is very tightly tied to the idea of phases. If there is no phase-by-phase access to the lexicon then workspace size may approach infinity which is impossible because our brains are limited. If there are no workspaces then it becomes unclear what phases are for. We can assume that there are workspaces, but memory required by a workspace is released after a structure is “transferred” to other modules.

5. Conclude that unacceptability of S' is caused by the same factor as that of S .

5.2 Linearization flow

Let us start with a simple question: how would linearization occur if pieces of structure were not spelled out separately? Obviously, the linearization process would have to deal with the full representation (full derivational history). Since syntactic structures naturally have a “depth” dimension it is reasonable to assume that linearization proceeds from the most prominent point, or, technically speaking, the item which is not (asymmetrically) c-commanded by anything. Usually, this is a matrix C . Let us take a look at the tree structure from the previous chapter. Assume that everything above the VP is already linearized.



Here *remained proud of her* can be linearized since c-command is defined for every item. But if either V' or PP is not linearized before they Merge, linearization process would face a dilemma after the subject DP is linearized: should V' or PP be linearized first? For the sake of argument we assume that precedence relations between sister heads are determined according to the language (English, Japanese etc.).¹⁰ However, neither V' or PP are heads,

¹⁰Even if there is no Head Parameter there must be *something* that determines the order between heads. Since the number of categories is universally rather small (compared

both are complex and are assembled by Merge separately. Thus, terminals within them do not c-command each other. Therefore, to resolve the linearization dilemma one phrase has to be linearized first and the other one – immediately after. Now, it is reasonable to assume that while one phrase is being linearized, another one has to be *kept* somewhere. This another phrase is not ignored – its linearization is just suspended. Suppose that encapsulated syntactic working memory is exactly where a phrase can be put for waiting.

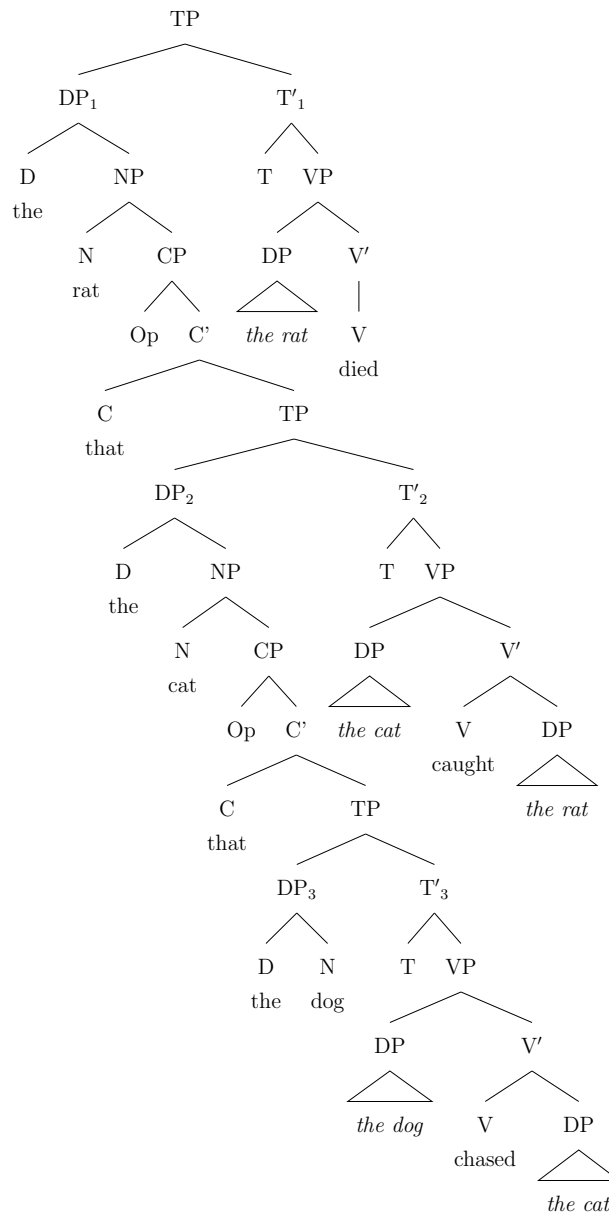
But how do we *measure* the memory usage if no traditional tests are reliable (as noted above)? The answer is simple: we can count how many phrases must be stored in memory *simultaneously* to linearize a certain sentence.¹¹ This is a number that can be calculated for any sentence whose structure we know. Let us look at some sentence that is known to be unacceptable but violates no known grammatical constraint. The most well known example is the center-embedding sentence:

(40) # The rat the cat the dog chased caught died.

to the number of lexical items), it is not unreasonable that speakers learn and use the idiosyncratic rules, where each rule dictates the order between a head of a certain category and its complement. For example, articles never merge with anything but nouns, so it takes only one rule of the form D>N to determine the order for all articles in all sentences.

¹¹Additionally, we could consider the number of lexical items in each stored phrase—the *size* of a phrase. For example, we could assume that larger phrases require more memory. Here and below, in order to isolate the number factor from the size factor, I treat all phrases as if they are of equal size. The motivation is that former explanations of islands too did not rely on the size of an island.

(41)



Let us construct a table that conveniently represents memory input/output operations. "Current" is the item/constituent that is being linearized and "Buffer" represents the status of buffer that temporarily holds a phrase in order for another to be linearized or holds none if there is no such need.

Step	Current	Buffer	Step	Current	Buffer
1	DP ₁ = the rat	T' ₁	14	TP	T' ₂ , T' ₁
2	D	T' ₁	15	DP ₃ = the dog	T' ₃ , T' ₂ , T' ₁
3	N	T' ₁	16	D = the	T' ₃ , T' ₂ , T' ₁
4	CP	T' ₁	17	N = dog	T' ₃ , T' ₂ , T' ₁
5	C'	T' ₁	18	T' ₃	T' ₂ , T' ₁
6	C	T' ₁	19	T	T' ₂ , T' ₁
7	TP	T' ₁	20	VP	T' ₂ , T' ₁
8	DP ₂ = the cat	T' ₂ , T' ₁	21	(DP ₃)	T' ₂ , T' ₁
9	D	T' ₂ , T' ₁	22	V'	T' ₂ , T' ₁
10	N	T' ₂ , T' ₁	23	V	T' ₂ , T' ₁
11	CP	T' ₂ , T' ₁	24	(DP ₂)	T' ₂ , T' ₁
12	C'	T' ₂ , T' ₁	25	T' ₂	T' ₁
13	C	T' ₂ , T' ₁	26

Table 1: Linearization of (41)

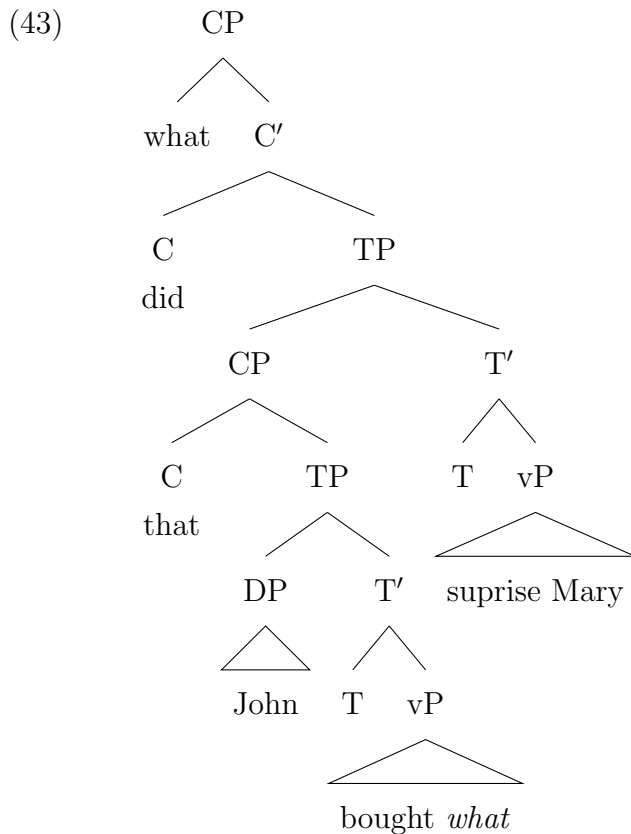
It is easy to see that the steps 15-17 are the steps that require the most memory resources. At this point we can hypothesize that storing **three** items in memory simultaneously causes the sentence to become unacceptable.¹² Can this logic be applied to phenomena that were traditionally explained by formalized syntactic constraints, like islands? Consider the following example of Subject Island:

- (42) a. That John ate the whole pizza surprised Mary.
b. *What did [that John ate ___] surprise Mary?

What makes this sentence different from the previous one is that apparent movement (internal Merge) is involved so we cannot be completely sure that movement is not the sole cause of unacceptability, as is usually (and reasonably) believed. Indeed, the identical sentence but without wh-movement is perfectly acceptable. Although it is difficult to imagine any other **cause**

¹²Again, either observable immediately or by carefully analyzing experimental data as in the case of Scandinavian languages.

of unacceptability, it is possible to reinterpret the **effect** of wh-movement in our framework: when the linearization process encounters a wh-word it temporarily stores it in memory buffer because it needs to make sure that this wh-word and the one in the canonical position are identical to avoid pronouncing it twice. Once it knows that two items are identical, they are removed from the storage.¹³ Let us look at the structure of (42b):



¹³I assume here that identity is checked in the simplest possible way—phonological identity. In fact, some languages that allow or require multiple wh-fronting, do not allow fronting of phonologically identical wh-words (like *what* and *what*). Bošković (1997) gives an example from Serbo-Croatian:

- (i) a. Šta uslovljava šta?
 b. *Šta šta uslovljava?

The same holds for its Russian counterpart: “**Čto čto obuslovalo?*” During linearization, both wh-words would be processed before their canonical positions are reached. If their phonological identity is checked, the lower one should never be pronounced. This is why presence of both wh-words as in the examples above leads to the feeling of unacceptability.

Now, let us construct the table of memory operations as we did for the center-embedding case. This time I omit obvious operations that do not require memory buffer: those that do not involve linearization of two complex phrases. Every step represents a situation where either an item has to be stored in memory buffer (like a wh-word) or where two complex phrases are merged (XP, YP) and a phrase has to be stored.

Step	Current	Buffer
1	C'	What
2	CP	T', What
3	DP	T', T', What
4

Table 2: Linearization of (43)

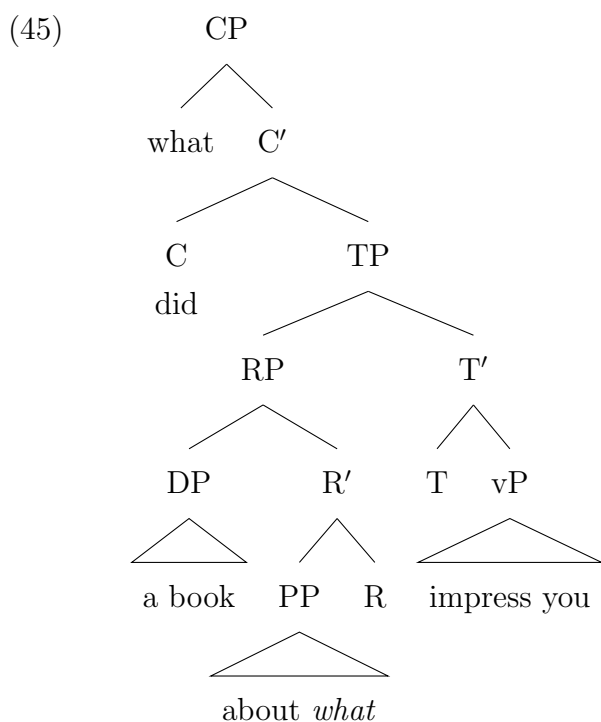
We can see that simultaneous storage of up to 3 items is required at step 3, when DP *John* is linearized.¹⁴ Note that all this time the wh-word *what* is kept in memory, because its canonical position is not reached yet. It is easy to see that were there no movement the sentence would not reach the (hypothesized) limit of 3 and be acceptable as predicted. Given the fact that this sentence is unacceptable and that it requires three items to be kept in memory we can argue that it is because of exceeding available memory resources like in the center-embedding case. Therefore, while retaining the intuition that movement is the cause of unacceptability, we reinterpret how exactly it leads to the island effect, without postulating any grammatical constraint.

In fact, it is possible to give the same explanation even for *prima facie* simpler cases, like the one below:

- (44) a. What did you write [a book about ___]?
 b. *What did [a book about ___] impress you?

¹⁴I assume here that even single-word subjects like *John* are phrases. This is not at all unreasonable because in some languages (e.g. Greek) proper nouns actually require a definite article, so we can assume that in English it is some kind of phonologically null D.

At first glance it seems that there should not be many memory operations involved in linearization of this sentence. However, if we adopt Adger's (2012) analysis of the structure of PPs then the PP in our sentence is not the complement to N as is traditionally assumed.¹⁵ Rather there is a special functional head R that takes both PP and DP/NP as its arguments, where N is traditionally called a *relational nominal*. Then the tree structure looks like this:



Now let us construct the brief memory operations table and see whether memory usage reaches (or exceeds) the limit.

¹⁵Adger's argument is based on a large-scale cross-linguistic observation that can be stated as follows:

- (i) *PP Peripherality*
When AP modifiers and PP complements both occur to one side of N inside a noun phrase, the PP is separated from the N by the AP.

Step	Current	Buffer
1	C'	What
2	RP	What, T'
3	DP	What, T', R'
4	R'	What, T'
5

Table 3: Linearization of (45)

Again, we observe that simultaneous storage of up to 3 items is required (step 3). This leads us to believe that this is the primary cause of unacceptability even in this seemingly simple case.

Let us now turn to Adjunct Islands. Consider the following sentence:

(46) * Which movie did John have dinner [before Bill saw ___]?

Before we proceed, it is important to figure out the exact position of the adjunct clause. One of the best ways to do it is to elide a VP and see whether adjunct is elided completely with it. If it is then we can conclude that the clause is adjoined to VP, but if it isn't then it must be somewhere higher in the structure.

(47) He didn't [VP have any legitimacy] before this theatrical election and he will not [VP have any legitimacy] after (this theatrical election).

Contrast this with the following sentences where the adverbial is elided together with VP:

(48) a. John said he can [VP solve this problem really fast] and he did (solve this problem really fast).
b. * John said he can [VP solve this problem really fast] but he did (solve this problem) slowly.

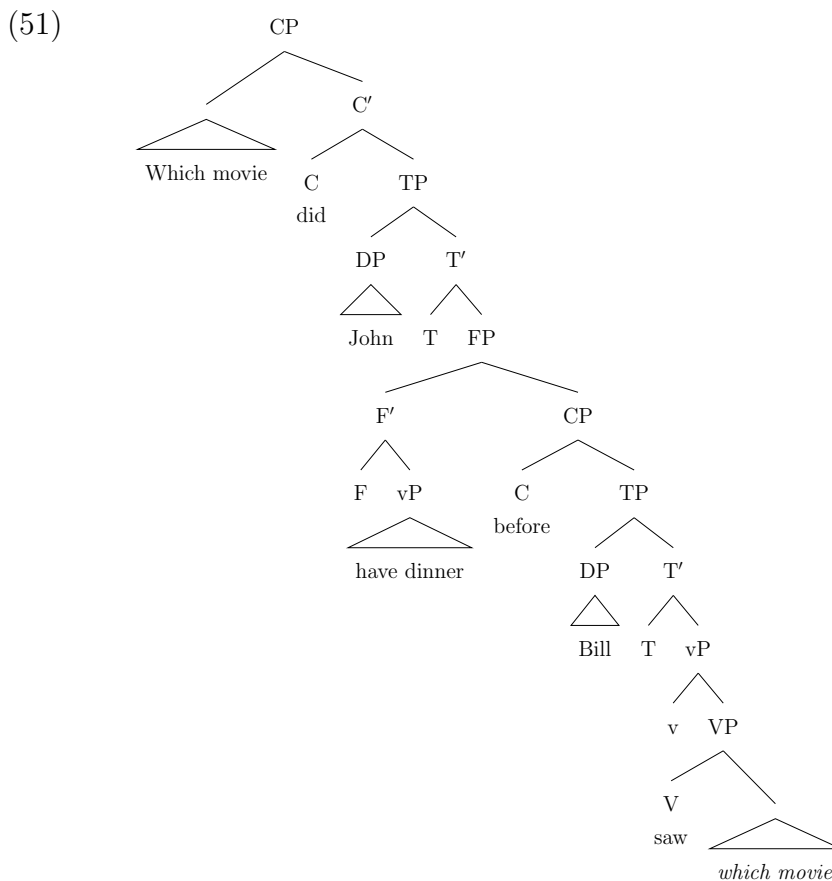
So, it is possible to elide VP without eliding the contents of the adjunct clause in (47). Then it must be positioned higher than VP in the structure. Given that it is a temporal adjunct, it is reasonable to assume that it is

adjoined to TP. But we must also guarantee that the matrix subject is able to bind the pronoun inside the adjunct clause.

(49) John_i came back soon after he_i left.

(50) *He_i came back soon after John_i left.

The problem is that usually we assume adjuncts to be merged higher than specifiers. In such a case *John* would be unable to bind the pronoun. But it is possible that adjunct is merged somewhere between TP and VP. For the issue at hand, it is irrelevant what the nature of the phrase that the adjunct adjoins to is. Let us dub it FP. This gives us the following structure for the sentence in (46):



Now we can construct the brief memory operations table as we did for previous cases:

Step	Current	Buffer
1	C'	Which movie
2	DP	T', Which movie
3	F'	CP, Which movie
4	vP	CP, Which movie
5

Table 4: Linearization of (51)

Interestingly, there is no step in the linearization which requires simultaneous storage of 3 or more items in memory buffer. Since it is unlikely that the structure is wrong, there are only two explanations: a) either our approach is wrong or b) unacceptability of extraction out of adjuncts is not due to memory limitations. I believe that we should opt for the latter.

To see why, it is important to mention a study by Truswell (2007). Truswell argues that in certain cases extraction out of adjuncts is possible and that all these cases share the same property. He proposes a Single Event Condition (SEC):

(52) **Single Event Condition**

A wh-chain is legitimate only if the minimal constituent containing the head and foot of the chain asserts the existence (in the actual world) of a single event.

In other words, movement from an adjunct is possible only if predicates in both matrix and adjunct clause belong to the same event. In other cases movement is not possible. Consider the following sentences:

(53) What did John arrive [whistling ___]?

(54) What did John drive Mary crazy [complaining about ___]?

Both sentences are acceptable because the predicates belong to the single event. In contrast, the predicates in the following sentence do not belong to the same event:

(55) * Which magic hat do you know Georgian [wearing ___]?

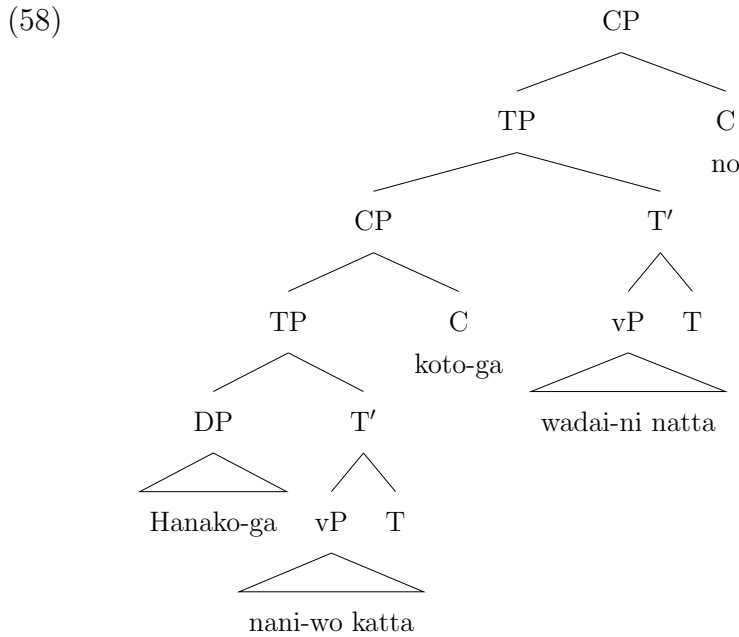
Rather than making an unjustified stipulation that single event predicates somehow alter the structure so that it becomes easier to linearize, it is reasonable to suppose that SEC is simply independent from what we looked at above. This is precisely the situation that I was talking about in the introduction. Quite obviously, the explanandum of SEC is neither the set of configurations that we call islands, nor even the subset. SEC does not presuppose existence of islands, but it accidentally does explain well one particular case that happens to be movement out of Adjunct Island. In principle, we could modify Subjacency Condition or our approach above to subsume SEC but that could lead to stipulating dubious syntactic operations, which we want to avoid. Instead, it is just as reasonable to accept SEC as an explanation of dependencies that involve adjunct clauses.

Let us go back to Subject Islands. In chapter 2 I mentioned wh-in-situ languages like Japanese or Chinese. These languages exhibit an unusual behavior called argument-adjunct asymmetry. Consider the following sentences:

- (56) Hanako-ga nani-wo katta koto-ga wadai-ni natta no?
 Hanako.NOM what.ACC bought that drew attention Q
 ‘What did that Hanako bought draw attention?’
- (57) * Hanako-ga naze ie-wo katta koto-ga wadai-ni natta
 Hanako.NOM why house.ACC bought that drew attention
 no?
 Q
 ‘Why did that Hanako bought a house draw attention?’

Even though no movement occurs in both cases, argument wh-words are allowed to occur inside the island structure, and adjunct wh-words are not. The Subjacency Condition makes it very clear that only movement can influence the acceptability outcome of a sentence, so, without additional assumptions, the argument/adjunct asymmetry phenomenon is unexplainable by the Subjacency Condition. We could imagine some kind of covert movement that occurs after the sentence is spelled out, but that would be a process orthogonal to generating data. Instead, we should see whether available theoretical

distinctions allow us to explain this phenomenon without Subjacency Condition. Fortunately, in our approach, no movement is just as significant as movement, since linearization process changes depending on where exactly the wh-word is in the structure. Let us represent structures of both sentences with a tree.



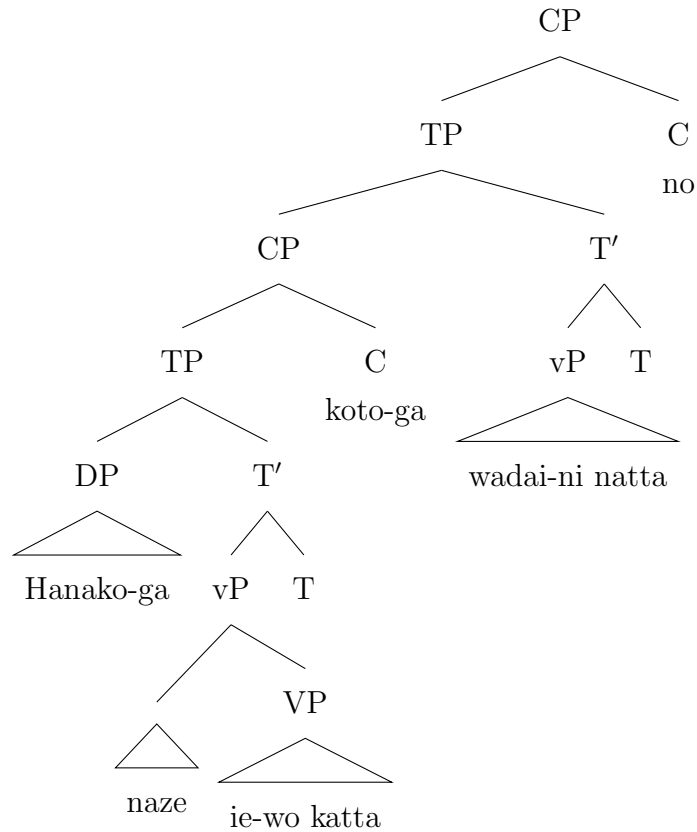
First, we need to construct a table of memory operations in (56) to see whether acceptability of this sentence is at least consistent with memory usage.

Step	Current	Buffer
1	CP	T'
2	DP	T', T'
3	T'	T'
4

Table 5: Linearization of (58)

No step in linearization of this sentence appears to require simultaneous storage of 3 items (maximum 2). Below is the tree structure of (57):

(59)



This sentence is unacceptable and now we have a chance to check whether this is caused by excess memory usage.

Step	Current	Buffer
1	CP	T'
2	DP	T', T'
3	naze	T', T', VP
4	VP	T', T'
5

Table 6: Linearization of (59)

In this case, step 3 indeed requires simultaneous storage of 3 items, which we hypothesised exceeds the limit. So, we observe essentially the same pattern as in English: a sentence becomes unacceptable when it uses too

much memory and when it does not—it is perfectly acceptable.

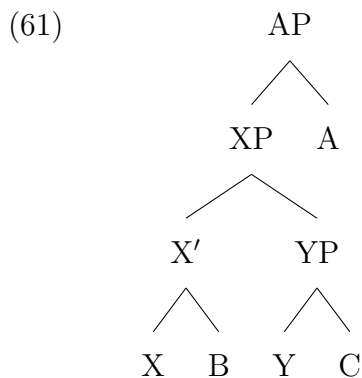
But there is a problem in our analysis of (57). Any sentence that includes an adjunct in the same position as *naze* is predicted to be judged unacceptable. For example, the following sentence:

- (60) Hanako-ga kodomo-no tame-ni ie-wo katta koto-ga
Hanako.NOM children.GEN sake house.ACC bought that
wadai-ni natta.
drew attention.
'That Hanako bought a house for her children drew attention.'

The sentence is perfectly acceptable. There is not much room for speculation here, so it is reasonable to consider the possibility that the so-called argument/adjunct asymmetry might not refer to a single phenomenon. SEC also does not seem to apply in this case. Search for the exact cause of unacceptability of (57) is out of the scope of this paper. However, we must still guarantee that other sentences with non-wh-adjuncts are acceptable under our approach. Up until this point I relied on an implicit assumption that the word order in a sentence, a product of linearization, reflects not only the language-specific rules which determine the order between sister heads, but also the order in which complex constituents are linearized (steps at which one of the constituents has to be put in buffer memory). For example, we know that in SVO languages like English subjects normally precede VPs. Suppose that this is because DP is linearized before T', which is temporarily stored in memory. Indeed, if the choice was completely optional, we would derive word orders that are unnatural for English (e.g. sentences where T' precedes DP).

In fact, my assumption goes even deeper. For example, subjects in Japanese also precede VPs. Also, we know that in Japanese C head (complementizers or question particles) follows the rest of the structure. But we do not want to say that this means that subjects are linearized *before* C in Japanese, which is contrary to our initial assumption that linearization starts with the *highest* item in the structure. This sort of confusion might arise if one equates linearization with pronunciation, but they are essentially two different processes. What I mean is that we cannot simply rely on the global word order

of a particular language to determine the order of linearization. But I do believe that we can rely on local word order, which is basically adjacency. Consider the following structure:



Suppose that we are dealing with a language where A always follows its sister, and X and Y precede their sisters. This information is *not enough* to determine the final outcome, because no rule specifies the relative order of X' and YP. However, we can hypothesize that this missing information is obtained *during* linearization. For example, it could be that X' precedes YP if YP had to be stored in memory in order to linearize X'. But it could also be vice versa—this is an empirical question. What we *do* know is that the relative order of A and the terminals of either X' or YP must be determined exactly at the step where one is being processed and the other one is being stored in memory—there is no other chance, because after the whole structure is linearized the information about phrasehood is gone by definition. I will assume that if the observed order is X>B>Y>C>A (YP is adjacent to A), it means that XP had to be stored in memory, and if the observed order is Y>C>X>B>A (X' is adjacent to A) then YP had to be stored in memory.

Let us reanalyze (57) in the light of the above discussion. We can see that the next highest constituent that is adjacent to C *no* is T' and not the CP subject. This means that T' was linearized first and CP was stored in memory. Likewise, T' of the CP subject clause is linearized before DP *Hanako-ga*. The memory operations table now looks like this:

Step	Current	Buffer
1	T'	CP
2	T'	DP
3	VP	naze, DP
4	naze	DP
5

Table 7: Linearization of (59)

There is no step which requires storage of three or more pieces of structure in memory anymore. At least now we guarantee that identical structures with non-wh-adjuncts are correctly judged as acceptable.

5.3 Other islands

So far we looked only at Subject and Adjunct Islands. It is, in principle, possible to extend the analysis to other configurations that are believed to be islands. But it becomes even harder to isolate other factors that have been mentioned in previous studies, both theoretical and experimental. For example, unacceptability of movement from wh-islands might really be a result of similarity-based intervention during processing, and this idea is supported by recent findings as we saw in chapter 3. In fact, if we actually look at memory operations for a sentence where a wh-word is extracted out of a wh-island, we will see that simultaneous storage of 3 items is never required. There is no correlation between memory usage as it is conceived in our approach and the length of a sentence (number of clauses), as long as the clauses are verb complements. There is correlation, however, between memory usage and how “deep” a *single* clause can be. In technical terms, what matters is not the total number of memory operations that occur during linearization but how much memory must be used during a particular step. This is the primary reason why Subject Islands pattern with center-embedding sentences. From this perspective, Adjunct Islands, which we found to be better explained by more semantic factors such as events, pattern with other cases.

Chapter 6 Discussion

In chapter 2 I reviewed the history of island effects. I showed how Subjacency Condition was conceived and its initial success in explanation of the class of unacceptability phenomena that involve movement out of island configurations. I also showed that it does not hold well cross-linguistically without modifications such as parametrization.

In chapter 3 I reviewed parsing theories. The phenomenon of superadditivity observed in experiment results indicates that factors other than the dependency length and the type of a clause might be present. One possible explanation is that this factor is working memory. However, no relationship between WM capacity of an individual and acceptability judgements was found. Some believe that traditional methods of measuring WM do not apply to syntactic processing which requires some kind of encapsulated working memory.

In chapter 4 I reviewed modern approaches in generative grammar to the problem of locality (and island effects, in particular), such as barriers and phases. The barriers framework is empirically strong but introduces severe complexity to the theory. The phases framework is conceptually superior to previous approaches, but has some empirical drawbacks. I also reviewed two alternatives to the phase framework that deal with CED phenomena and Holmberg's Generalization by focusing on linearization requirements at the interfaces.

In chapter 5 I developed another alternative. Following Nunes & Uriagereka (2000), I hypothesized that linearization of a sentence might involve complex cases where the simplest version of LCA cannot be applied. Again, following the authors I choose not to complicate the very definition of LCA as Fox & Pesetsky (2005) do. Where our accounts differ is that I also try to avoid multiple Spell-Out and instead investigate how encapsulated syntactic memory that I mentioned in chapter 3 can be exploited to resolve the complex cases. Buffer memory, seen as a temporal storage, where phrases can be put while the other is being linearized, seems to aid linearization in a rather natural way. However, it is also natural that nothing in our brain can use unlimited

memory. Linearization is not an exception and must operate within some limit. To find this limit, I took the well-known case of center-embedding, which is judged as unacceptable despite that no grammatical constraint is violated. It appears that exactly those instances of center-embedding are unacceptable which require simultaneous storage of 3 or more pieces of structure (2 and less are fine). I showed that Subject Island effects pattern with center-embedding effects, because 3 or more pieces of structure must be held in buffer memory simultaneously. What matters is not the total number of accesses to memory but a number of pieces that must be stored at a particular step. This is in line with intuition that given enough time humans can produce and parse infinitely long sentences.

Interestingly, my approach does not capture the similarity of Subject and Adjunct Islands that have been traditionally explained by a single constraint – CED. I believe that it reinforces the point that I made throughout the paper that what counts as island data largely depends on what kind of constraint we can formulate using available theoretical distinctions. We cannot allow the same theoretical distinctions that we had when Subjacency Condition was proposed. In my minimal setting, Subject and Adjunct “islands” simply do not count as the same kind of data. Similarly, distinctions that CED requires, namely X-bar-theoretic complements and non-complements, are also unavailable in this minimal setting. Additionally, given the fact that under certain semantic conditions movement out of adjuncts is possible, CED is both too strong and incorrect. My position is that it is perfectly acceptable to explain original island data with several, possibly unrelated explanations, if that allows us to make as few assumptions about syntax as possible. Thus, Adjunct Islands, for example, are better explained by Single Event Condition (SEC). But SEC does not work well with Subject Islands, which are better explained by limitations on memory in my approach.

The approach developed in chapter 5 has limitations as well. I took it for granted that identity of lexical items other than wh-words is somehow known to the external systems. But in fact, it is not at all obvious that an external system knows that two instances of *John* in the sentence “*John hit John*” are distinct. Chomsky (2007) suggests that something like phase-level

memory “...suffices to determine whether a given pair of identical terms Y , Y' was formed by Internal Merge. If it was, then Y and Y' are copies; if it was not (i.e., it was formed by External Merge), Y and Y' are independent repetitions.” This remains an open issue in generative grammar. If we could somehow reconcile our view of memory with what Chomsky calls phase-level memory then the issue can be successfully solved.

Further, following Nunes & Uriagereka (2000), I also assumed that it is known how exactly structural sisters must be linearized. It is possible that linearization additionally follows some limited word order instructions for each language. But it is unclear whether these instructions can influence the choice of the phrase to be put in buffer memory. If they do, they could alter the outcome of linearization and thus make a sentence acceptable, which would otherwise be unacceptable. For example, as Stepanov (2007) argues, not all languages prohibit movement out of subject islands. I showed how the choice of the phrase to be put in buffer memory leads to observable differences in Japanese. Another example of acceptable movement out of a subject is from Stepanov (2007):

- (62) [Op_i [Ahmet-in t_i git-me-si]-nin ben-i uzdugu] ev.
 Ahmet.GEN go-INF-AGR.GEN I.ACC sadden house.
 ‘The house [which [that Ahmet went to ___] saddened me]’

The structure of this sentence is basically a reverse of its English counterpart. If we assume that the order of linearization steps reflects the actual pronunciation order, we will see that T' in English is linearized after the subject of the relative clause, while in Turkish it is linearized before. The consequence of this is that storage of 3 pieces of structure in memory is never required. This is consistent with our prediction that such a sentence is thus acceptable. Further empirical investigation of languages with word order different from English is required.

During the discussion of memory limitations, I treated all phrases as if they are of equal size in order to isolate the number factor from the size factor. However, it is not unreasonable that linearization might be influenced by the size of a phrase as well. Consider the following sentences:

- (63) a. That John bought a new car surprised everyone.
b. *What did that John bought surprise everyone?
c. It surprised everyone that John bought a new car.
d. What did it surprise everyone that John bought?

We can treat (63c) as equivalent to (63a) but with the “large” CP subject linearized after T' and the expletive *it* inserted at the interface. Then, even if we need to move a *wh*-word from within the CP subject, memory limit is never exceeded just like in Japanese and Turkish examples. As predicted, (63d) is acceptable. The fact that *it* is not merged in syntax explains its meaninglessness. While rather controversial, I believe that it is worth searching for more of similar cases.

Finally, I said nothing about ellipsis. It is well-known that movement from elided “islands” sometimes avoids the constraints and is acceptable. This is called “island repair.” One example is given below.

- (64) That he'll hire someone is possible, but I won't divulge who (*that he'll hire is possible). (Fox & Lasnik (2003: 2))

This kind of ellipsis is called *sluicing*. Unlike with VP-ellipsis, whole TP is elided by sluicing. The fact that such sentences are acceptable fits well with my approach: if there is no need to pronounce something then there is no need to linearize it either. Then possible memory limitations are also avoided and as a result the sentence is acceptable. However, VP-ellipsis seems to be much less forgiving than sluicing. I believe that it is worth analyzing island repair from the perspective developed in chapter 5.

Chapter 7 Conclusion

The unacceptability phenomena collectively known as island effects are partially explainable by natural limitations on syntactic working memory. In particular, those sentences that require three or more pieces of structure to be held in buffer memory during a certain linearization step are rendered unacceptable. The sentences that require storage of less than three pieces of structure are acceptable. Island data is not uniform with respect to currently available theoretical distinctions and is better explained by independent theories. CED appears to be incorrect and must be replaced by the approach developed in this paper for movements out of subjects. For movement out of adjuncts, the more semantic Single Event Condition seems to be the better alternative. Analysis of island phenomena (or the absence of them) in non-English languages appears to lead to more natural conclusions if the word order of these languages is also seriously taken into account. Some well-known facts such as the possibility of movement from extraposed sentential subjects or island repair by sluicing seem to be compatible with the perspective developed in this paper. Further investigation of these matters might provide support for the idea that syntax is largely unrestricted, while its products are restricted by biological constraints that apply to the systems with which syntax interfaces.

References

- Adger, David. 2012. *A syntax of substance*. Cambridge, MA: MIT Press.
- Bever, Thomas G. 1988. The psychological reality of grammar: a student's eye view of cognitive science. In William Hirst (ed.), *The making of cognitive science: Essays in honor of George A. Miller*, 112–142. Cambridge, MA: Cambridge University Press.
- Boeckx, Cedric. 2008. Islands. *Language and Linguistics Compass* 2(1). 151–167.
- Boeckx, Cedric & Kleantes K. Grohmann. 2007. Remark: putting phases in perspective. *Syntax* 10(2). 204–222.
- Bošković, Željko. 1997. Superiority effects with multiple wh-fronting in Serbo-Croatian. *Lingua* 102(1). 1–20.
- Bošković, Željko. 2002. A-movement and the EPP. *Syntax* 5(3). 167–218.
- Bošković, Željko. 2015. From the complex NP Constraint to everything: on deep extractions across categories. *The Linguistic Review* 32(4). 603–669.
- Caplan, David & Gloria S. Waters. 1999. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22(1). 77–94.
- Ceplova, Markéta. 2001. Minimalist islands: restricting P-features. Ms. MIT.
- Chametzky, Robert. 1996. *A theory of phrase markers and the extended base*. Albany, NY: SUNY Press.
- Chomsky, Noam. 1964. *Current issues in linguistic theory*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1968. *Remarks on nominalization*. Bloomington, IN: Linguistics Club at Indiana University.

- Chomsky, Noam. 1973. Conditions on transformations. In Stephen Robert Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 232–286. New York: Holt, Rinehart & Winston.
- Chomsky, Noam. 1977. On wh-movement. In Peter Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71–132. London, New York, San Francisco: Academic Press.
- Chomsky, Noam. 1986a. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1986b. *Knowledge of language: its nature, origin, and use*. Westport, CT: Greenwood Publishing Group.
- Chomsky, Noam. 1993. *Lectures on government and binding: The Pisa lectures*. Berlin: Walter de Gruyter.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1999. Derivation by phase. *MIT Occasional Papers in Linguistics* (18).
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT press.
- Chomsky, Noam. 2002. *On nature and language*. Cambridge, UK: Cambridge University Press.
- Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's minimalism and the view from semantics*, 1–29. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistics theory*, 133–166. Cambridge, MA: MIT press.

- Chomsky, Noam, Ángel Gallego & Dennis Ott. 2017. Generative grammar and the faculty of language: insights, questions, and challenges. Ms. MIT, Universitat Autònoma de Barcelona, and University of Ottawa. To appear in *Catalan Journal of Linguistics*. Downloadable at <http://ling.auf.net/lingbuzz/003507>.
- Collins, Chris. 2002. Eliminating labels. In Samuel Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 42–64. Oxford, UK: Blackwell Publishers.
- Engdahl, Elisabet. 1980. Wh-constructions in Swedish and the relevance of subadjacency. *Cahiers Linguistiques d'Ottawa* 9. 89–108.
- Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998. *A derivational approach to syntactic relations*. Oxford, UK: Oxford University Press.
- Epstein, Samuel David, Hisatsugu Kitahara & T. Daniel Seely. 2015. *Explorations in maximizing syntactic minimization*. Abingdon, UK and New York, NY: Routledge.
- Epstein, Samuel David & T. Daniel Seely. 2002. Rule applications as cycles in a level-free syntax. In Samuel Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*, 65–89. Oxford, UK: Blackwell Publishers.
- Epstein, Samuel David & T. Daniel Seely. 2006. *Derivations in minimalism*. Cambridge, UK: Cambridge University Press.
- Fox, Danny & Howard Lasnik. 2003. Successive-cyclic movement and island repair: The difference between sluicing and VP-ellipsis. *Linguistic Inquiry* 34(1). 143–154.
- Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31(1-2). 1–45.
- Gallego, Ángel J. & Dennis Ott. 2015. *50 years later: reflections on Chomsky's Aspects*. Cambridge, MA: MITWPL.

- Han, Chung-hye & Jong-Bok Kim. 2004. Are there “double relative clauses” in Korean? *Linguistic Inquiry* 35(2). 315–337.
- Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30(1). 69–96.
- Hornstein, Norbert. 2009. *A theory of syntax: minimal operations and universal grammar*. Cambridge, UK: Cambridge University Press.
- Hsu, Chun-chieh Natalie. 2008. Revisit relative clause islands in Chinese. *Language and Linguistics* 9(1). 23–48.
- Huang, James C.-T. 1982. *Logical structures in Chinese and the theory of grammar*: Ph. D. Dissertation, MIT dissertation.
- Ishizuka, Tomoko. 2009. CNPC violations and possessor raising in Japanese. Ms. UCLA.
- Jha, Amishi P. & Gregory McCarthy. 2000. The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. *Journal of Cognitive Neuroscience* 12(Suppl. 2). 90–105.
- Jurka, Johannes, Chizuru Nakao & Akira Omaki. 2011. It’s not the end of the CED as we know it: revisiting German and Japanese subject islands. *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 124–132. Cascadilla Proceedings Project Somerville, MA.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2003. Antisymmetry and Japanese. *English Linguistics* 20(1). 1–40.
- Kluender, Robert & Simone Gieselmann. 2013. What’s negative about negative islands? A re-evaluation of extraction from weak island contexts. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 186–207. Cambridge, UK: Cambridge University Press.

- Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633.
- Kuno, Susumu & Kenichi Takami. 1997. Remarks on negative islands. *Linguistic Inquiry* 28(4). 553–576.
- Kush, Dave & Filippa Lindahl. 2011. On the escapability of islands in Scandinavian. Ms. University of Maryland.
- Kush, Dave, Terje Lohndal & Jon Sprouse. 2018. Investigating variation in island effects. *Natural Language & Linguistic Theory* 36(3). 743–779.
- Miller, George A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63(2). 81–97.
- Nunes, Jairo. 1999. Linearization of chains and phonetic realization of chain links. In Samuel Epstein & Norbert Hornstein (eds.), *Working minimalism*, 217–249. Cambridge, MA: MIT Press.
- Nunes, Jairo & Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3(1). 20–43.
- Phillips, Colin. 2013. Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28(1–2). 156–187.
- Phillips, Colin, Nina Kazanina & Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22(3). 407–428.
- Poeppel, David & David Embick. 2017. Defining the relation between linguistics and neuroscience. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 103–118. Abingdon, UK and New York, NY: Routledge.
- Rizzi, Luigi. 1982. Violations of the wh-island constraint in Italian and the subjacency condition. *Journal of Italian Linguistics* 5(1). 157–191.

- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, MA: MIT Press.
- Ross, John Robert. 1967. *Constraints on variables in syntax*: Cambridge, MA: MIT dissertation.
- Sprouse, Jon & Norbert Hornstein. 2013. Experimental syntax and island effects: toward a comprehensive theory of islands. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 1–17. Cambridge, UK: Cambridge University Press.
- Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88(1). 82–123.
- Sprouse, Jon, Matthew Wagers & Colin Phillips. 2014. Deriving competing predictions from grammatical approaches and reductionist approaches to island effects. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 21–41. Cambridge, UK: Cambridge University Press.
- Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1). 80–126.
- Torrego, Esther. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1). 103–129.
- Truswell, Robert. 2007. *Locality of wh-movement and the individuation of events*: UCL (University College London) dissertation.
- Wagers, Matt. 2013. Memory mechanisms for wh-dependency formation and their implications for islandhood. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 161–185. Cambridge, UK: Cambridge University Press.
- Zaksas, Daniel & Tatiana Pasternak. 2006. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *Journal of Neuroscience* 26(45). 11726–11742.

Zwart, Jan-Wouter. 1994. On Holmberg's Generalization. *Language and Cognition* 4. 229–242.