Empty Categories: The empirical problem and the explanatory challenge

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

1. Syntax and silence

This course will deal with *Empty Categories* (EC), linguistic elements with grammatical function and semantic value but without any phonological realization.

Thus, this is a course about the linguistic properties of certain kind of silences.

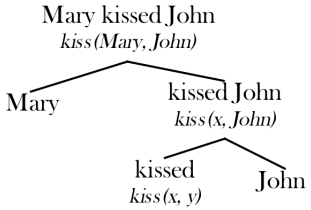
Three important questions we will try to answer in this course:

(i) How do we know there is something in silence?
(ii) Are all silences the same?
(iii) How can we explain these silences?

But before starting answering these questions, maybe some of you would be interested in an answer for a much more commonsensical question: why do syntactitians even bother in proposing and studying phonologically null entities?

(1) *Compositionality*
The meaning of a sentence/phrase is determined by its minimal parts and how they are combined.

(2) a. [$_{\text{AGENT}}$ Mary] kissed [$_{\text{THEME}}$ John].
b.



The structure in (2b)[1] allows explaining (or, at least, saying something about) (i) the word order of the sentence and (ii) its meaning. This follows from a very usual definition of Syntax:

(3) Syntax is a system connecting two cognitive domains, one involved in the processing of meanings and one involved in the processing of sounds.

This means that syntactic theory should provide some kind of *mapping rules* between the semantic representation (or Logical Form, LF) and the phonological representation (or Phonological Form, PF).

Consider, for example, the toy rule in (4).

(4) If V is a verb denoting an event involving a *Theme*[2] θ-role, then the complement of V should be interpreted as a Theme.

This is a good generalization: since complements in a SVO language are supposed to be at the right of the verb, (4) predicts the right thematic interpretation for a lot of sentences.

(5) a. Mary kissed [John].
b. Yoko Ono destroyed [The Beatles].
c. Bill ate [the sandwich I bought for you].
d. *etc*.

However, consider now the following cases, where *John* (or a pronoun replacing it) seems to have been moved from its canonical position and the complement position of the verb is occupied by "silence":

(6) a. John was kissed.
b. Who did Mary kiss?
c. John, Mary kissed.

There are basically two ways to deal with these sentences. The first one would involve assuming specific rules for each kind of sentence:

(7) If V is in an active sentence and denotes an event involving a Theme θ-role, then the complement of V should be interpreted as a Theme.

(8) If V is in a passive sentence and denotes an event involving a Theme θ-role, then the constituent agreeing with the auxiliary verb should be interpreted as a Theme.

(9) If V is in a wh-question kind of sentence and denotes an event involving a Theme θ-role, and the complement position of V is empty, then the wh-phrase should be interpreted as a Theme.

[1] In this handout I will use very simple trees. We will deal with more detailed representations in the future.

[2] A *Theme* is the participant of an event that is affected by the event or undergoes a change of state.

(10) If V is in a topicalized sentence and denotes an event involving a Theme θ-role, and the complement position of V is empty, then the topicalized constituent should be interpreted as a Theme.

Of course, the rules in (7) to (10) "capture" the patterns in (5) and (6). But are they a real explanation of the properties of human syntax or just a list of constructions you can say in English?

Is an inventory of possible pairings of sound and meaning an actual theory of the mapping between sounds and meanings?

The other way is assuming that there is something in fact in the complement position of *kiss* in (7) to (10). We know that some phonologically empty elements do exist in language:

(11) a. The child is in the stable.
b. The children are in the stable.
c. *The child are in the stable.

(12) a. The sheep is in the stable.
b. The sheep are in the stable.

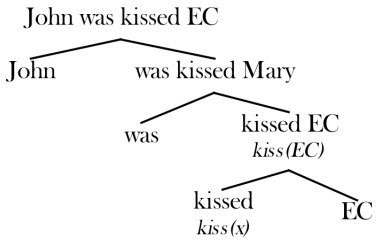
There must be a null plural morpheme attached to *sheep* in (12b).

Thus, it is not entirely crazy proposing the representations in (13), where an Empty Category EC occupies the complement position of the verb and it is connected somehow with the element in the left periphery of the sentence:

(13) a. $John_i$ was kissed EC_i.
b. Who_i did Mary kiss EC_i.
c. $John_i$, Mary kissed EC_i.

This implies that the EC is the element combining compositionally with the verb:

(14)



Now, it is necessary to explain how the EC and *John* are connected, because we know that *John* is interpreted as it would be occupying the complement position of *kissed.* Several languages show grammatical mechanisms that seem to connect a constituent at the "left" of the sentence and an element inside the VP. One of such cases is Clitic Left Dislocation (CLLD) in Romance languages:

(15) *Italian*
Gianni, lo conosciamo.
Gianni it know.we
'Gianni, we know him'.

In this case, *Gianni* is also interpreted as if it was occupying the complement position of the verb, but it is not: a clitic pronoun "connects" the thematic interpretation associated with this position with the noun in the left periphery of the sentence.

Thus, there is nothing intrinsically strange in assuming EC: (i) we know that there are some elements in language without phonological properties (e.g., zero morphemes) and (ii) we know that there are some mechanisms connecting elements inside the VP with elements in the left part of the sentence (e.g., CLLD).

So, we have a nice alternative story to the rules in (7) to (10). However, we still need to corroborate if our nice story is empirically adequate.

2. There must be something there…
Consider the sentences in (16):

(16) a. You read the book while I ate the pizza.
b. What did you read while I ate the pizza?
c. *What did you read the book while I ate?

Why is (16c) unacceptable? Let's try to explain this with a more detailed syntactic description:

(17) a. $What_i$ did you read EC_i [$_{ADJUNCT}$ while I ate the pizza].
b. *$What_i$ did you read the book [$_{ADJUNCT}$ while I ate EC_i]

At first sight, it seems we cannot connect an element in the left periphery of the sentence with an EC inside an adjunct. Something similar is observed with CLLD:

(18) *Spanish*
a. Al libro, lo leiste [$_{ADJUNCT}$ mientras comías la pizza].
The book it read.you while ate.you the pizza
'The book, you read it while you ate the pizza'.
b. *La pizza, leiste el libro [$_{ADJUNCT}$ mientras la comías].
The pizza read.you the book while it ate
'The pizza, you read the book while you ate it'.

Thus, by assuming the existence of ECs we can try to offer a unified analysis for these phenomena.

A traditional argument for ECs comes from *wanna-contraction*: sometimes the sequence *want+to* becomes *wanna*. However, the phonological process combining these elements seems to be interrupted by an EC in the subject position of the infinitive.

(19) a. Who_i do you want to see EC_i → Who do you wanna see?
b. Who_i do you want EC_i to do it → *Who do you wanna do it?
c. Who_i have you got to see EC_i → Who have you gotta see?
d. Who_i have you got EC_i to help you → *Who have you gotta help you?

Additional evidence for ECs comes from a phenomenon usually called Reconstruction: it is possible to interpret the constituents of a phrase in the left of the sentence as if they were actually in the position of the EC. For example, (20) shows that a pronoun can be bound by a QP only if it is in the c-command domain of the QP.

(20) a. $\underline{\text{No professor}_i}$ talked to a student of $\underline{\text{his}_i}$.
b. *A student of $\underline{\text{his}_i}$ talked to $\underline{\text{no professor}_i}$.

However, the pronoun *his* in (21) can be interpreted as if the whole phrase *which student of his* were occupying the position of the EC.

(21) Which student of $\underline{\text{his}_i}$ did $\underline{\text{no professor}_i}$ talk to EC?
*Which student of $\underline{\text{his}_i}$ EC talked to $\underline{\text{no professor}_i}$

As you may know, all these cases are supposed to involve a movement dependency between the EC and the phonologically realized antecedent: in (21a), for example, the wh-phrase used to occupy the position where the EC is, but it moved leaving behind a *Trace*, a particular kind of EC. However, there are other ECs that seem to be very different.

3. Other kinds of silences

3.1. NULL SUBJECTS (pro*)*

The ECs exemplified so far are in fact only one of the null constituents that have been proposed in the literature. Another well-known case of ECs involved the so-called *Null Subjects*.

(22) a. It rains
b. I run.
c. They ruined the painting.

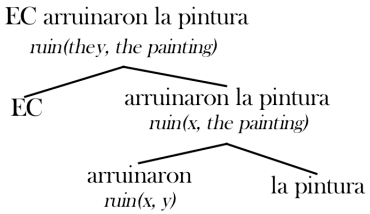
(23) *Spanish*
a. Llueve.
b. Corro.
c. Arruinaron la pintura.

The interpretation of (23b) is *referential* regardless of the fact of lacking an overt subject. Thus, it has been proposed that the syntactic representation for the sentences in (23) involve ECs in subject position:

(23) a. EC llueve [EXPLETIVE NULL SUBJECT]
b. EC corro [REFERENTIAL NULL SUBJECT]
c. EC arruinaron la pintura [REFERENTIAL NULL SUBJECT]

The kind of EC involved in null subject phenomena is usually assumed to be some kind of pronoun.

(24)



However, null subjects in pro-drop languages do not behave as actual pronouns in the same language. In fact, they seem to behave as English pronouns.

(25) $Mary_i$ read the letter carefully. She_i thinks she_i shouldn't answer back.

(26) *Spanish*
$Mary_i$ leyó la carta detenidamente. $Ella_i$ piensa que $ella_i$ no debe responder.

(27) *Spanish*
$Mary_i$ leyó la carta detenidamente. EC_i piensa que EC_i no debe responder

In (26), every new pronoun *ella* introduces a new referent. This does not happen in English, where every occurrence of *she* refers to *Mary*.

3.2. CONTROL (PRO)

Consider the following sentence:

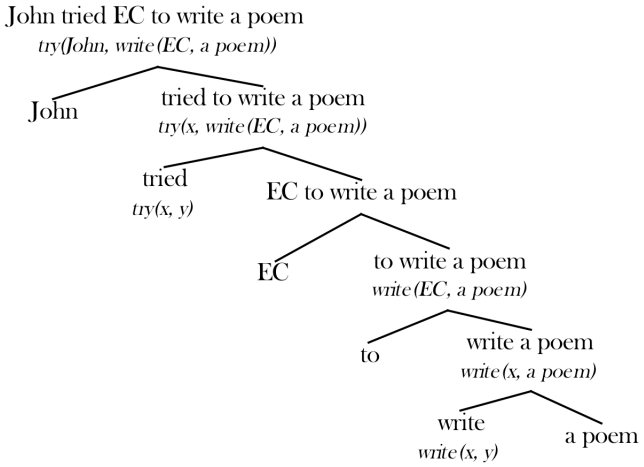
(28) John tried to write a poem.

There are two verbs here, *try* and *write*, assigning two θ-roles each. Who is the "tryer" and who is the "writer"?

Yes, both θ-roles seem to have the same denotation: *John.* Interestingly, *John* occupies a syntactic position where it can receive the "tryer" interpretation, but the "writer"-related position is (apparently) empty.

Once again, an EC to the rescue:

(29)



Thus, the EC receives the "writer" θ-role, and since it denotes somehow the subject of the matrix clause*, John* is interpreted both as the "tryer" and the "writer".

An important difference between Null Subjects and Control is that the EC involved in the later cannot refer by itself. If there is no proper antecedent for it, the interpretation of the EC is arbitrary.

(30) There is not likely EC_{arb} to be a riot.

3.3. ELLIPSIS

If we talk about silences, we cannot avoid talking about *ellipsis*, a grammatical mechanism that allows omitting the phonological realization of certain constituents. Ellipsis is an anaphoric phenomenon: you can only omit something that has been already mentioned or is really salient in the communicative context. Basically, it is possible to distinguish between two types of elliptical anaphora: *Surface Anaphora* and *Deep Anaphora*.

(31) *Surface Anaphora*
The elliptical site has internal complex structure. Thus, constituents are present in the syntactic derivation, but they do not receive phonological representation.

Two classic examples of Surface Anaphora are Gapping and NP-Ellipsis.

(32) *Gapping*
John can play the guitar, and Mary ~~can play the guitar~~ too.

(33) *NP ellipsis*
John was mentioned in fifteen repetitive examples that I had to make up but Mary was mentioned only in tree ~~repetitive examples that I had to make up~~.

(34) *Deep Anaphora*
The elliptical site is an EC (without internal structure).

Null Complement Anaphora is typically assumed to be a Deep Anaphora case.

(35) *Null Complement Anaphora*
I asked Bill to leave, but he refused.

This sentence looks a lot like a Surface Anaphora, but there are important differences. For example, Surface Anaphora requires *Syntactic Parallelism:* the antecedent must be syntactically identical to the constituents in the elliptical site.

(36) Nobody else would take the oats down to the bin
a. So Bill did. [VP ELLIPSIS]
b. So Bill volunteered. [NULL COMPLEMENT ANAPHORA]

(37) The oats had to be taken down to the bin.
a. *So Bill did. [VP ELLIPSIS]
b. So Bill volunteered. [NULL COMPLEMENT ANAPHORA]

The example in (37) shows that a passive form is not a proper antecedent for an active form. However, Null Complement Anaphora (and Deep Anaphora in general) is insensitive to this parallelism requirement.

Determining which silences are Surface or Deep Anaphora is an empirical enterprise. In fact, we have been assuming that all the "silences" we analyzed so far are some kind of Deep Anaphora, but maybe this assumption is totally wrong. This is something that we will try to answer in this course.

4. What are we going to discuss in this course exactly?

As seen, there are several empirical domains involving some kind of EC. In this course we will focus on two: the type of ECs typically related with *syntactic movement* (e.g., (6)) and Null Subjects. We will discuss several theoretical approaches to both phenomena and evaluate their empirical adequacy.

Leaving behind a trace of silence

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

1. The general idea once more

Last class we discussed pairs as these ones:

(1) a. Kramer punched the thief.
b. The thief was punched (by Kramer).

(2) a. George bought the stamps.
b. What did George buy?

And we notice that there is some kind of mismatch between the *form* of these sentences and their meaning. Schematically.

(3) a. AGENT punched THEME.
b. THEME was punched (by AGENT).

(4) a. AGENT bought THEME.
b. THEME did AGENT buy?

How is it the case that a constituent is pronounced in a position A (e.g., at the beginning of the sentence) but it is interpreted as occupying position B (e.g., the complement position of the verb)?

And we proposed a sketchy hypothesis explaining this: there is an *Empty Category* (EC) in the thematic position and it is connected somehow with an overt (i.e., pronounced) constituent.

(4) [The thief] punched [$_{\text{THEME}}$ EC]

Now, we will discuss an actual technical implementation of this general idea: *Trace Theory*.

2. A sketch of Trace Theory

This theory was first developed in the seventies (cf. Chomsky 1973, Fiengo 1974), and it was the "standard" approach to syntactic movement until, at least, Chomsky (1993). (So, now you know: this is actually really old stuff).

Basic idea:

(i) Syntax can *move* constituents from one place to another.
(ii) Once you moved a constituent, you need to put something in its original position.
(iii) So you put there a particular kind of EC, a *Trace*.
(iv) The Trace and the moved constituent are bound by an index.

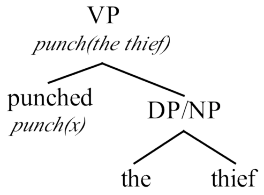
Let's go part by part. According to (i), Syntax has an operation that allows moving constituents. This operation was called *Move-α*.

(5) *Move-α (Chomsky 1982)*
Move any category anywhere.

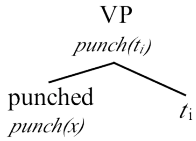
In the next section we will talk a little about the *unrestrictiveness* of this definition. For the moment, just remember you have an operation that allows moving constituents.

Regarding (ii) and (iii), we discussed last class how θ-roles are assigned to noun phrases in some syntactic positions. Past participles (the ones in passive sentences) usually assign a Theme θ-role to their complements, so they require an element in that position to, let's say, express their whole meaning:

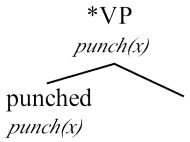
(6)



(7)



(8)



Thus, after moving an element, an EC, a trace, *must* occupy its place to avoid the unacceptable scenario in (8).

Now, regarding (iv), I have two options: lying to you and saying that indexes are required to avoid ambiguous scenarios as the one depicted in (9)…

(9) a. He was defeated *t* by the Bubble Boy.
b. By whom was he defeated *t t* ?

…or giving you a little taste of how things were in the eighties.

> *The 80s:*
> - The framework was called *Government and Binding* (GB, cf. Chomsky 1981), basically because those two notions were used to explain almost everything.
> - You had some very general and unrestricted rules (e.g., (5)) and a lot of conditions and principles constraining the possible outputs.

So, this was the real idea regarding ECs: (i) grammar has only one EC, (ii) you introduce that EC in empty positions and (iii) depending on Binding[1], that EC is going to be a trace, a PRO, etc.

(10) *Functional Determination of Empty Categories (adapted from Chomsky 1982:35)*
a. An EC is a variable (i.e., a trace of A-bar movement) if it is in a thematic position and is bound by an operator.
b. An EC that is not a variable and occupies a thematic position is an anaphor (i.e., a trace of A-movement).
c. An EC in a thematic position that is not a variable is a pronominal (i.e, a PRO) if it is free or bound by an antecedent with an independent θ-role.

- How do you know when two elements are bound?

When they share the same index.

- And how do you assign those indexes?

Randomly.

Thus, let's consider how (10) works by assuming the *Theta-Criterion*.

(11) *Theta-Criterion*
Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.

(12) Who$_i$ was punched EC$_{i/j}$? [Approx. meaning: *for what person x, x was punched*]

[1] Binding is the part of the grammar explaining the unacceptability of some correference patterns by positing some conditions on the locality of possible antecedents.

(i) *John$_i$ asked Mary to wash himself$_i$. [Anaphors should be locally bound]
(ii) *John$_i$ washed him$_i$. [Pronouns should not be locally bound]
(iii) *He$_i$ asked Mary to wash John$_i$. [R-expressions cannot have an antecedent]

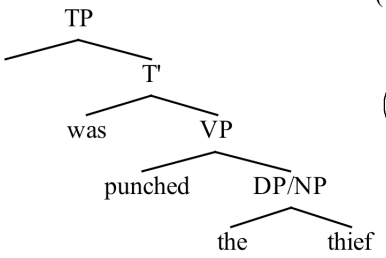
(13) [The thief]$_i$ was punched EC$_{i/j}$

(14) I$_i$ want to EC$_{i/j}$ punch someone.

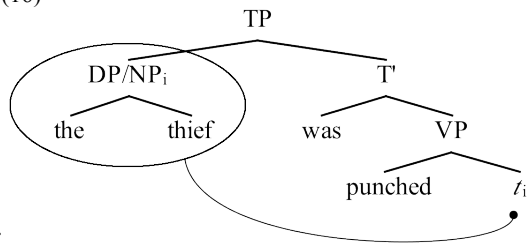
Since (14) cannot be accounted by Binding and (11) alone, it was necessary to postulate a theory governing the distribution of PRO: *Control Theory*.

Anyway, the result is a theory where a structure like (15) is transformed by Move-A to the structure in (16).

(15)



(16)



2. Movement goes up! Introducing the ECP

If ECs are assumed, it is necessary postulating some kind of theory explaining their distribution. Consider the following sentences:

(17) t_i was punched [the thief]$_i$

(18) t_i was punched who$_i$

What is wrong exactly with these sentences?

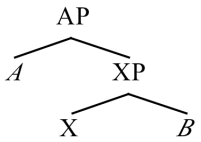
> *Syntactic movement* involves connecting a phonologically empty constituent downstairs and an overt constituent upstairs, not the other way round.

The principle stating this was called the *Empty Category Principle* (ECP).

(19) *ECP (simplest version ever)*
Traces must be governed.

As I already said, GB was mainly about two very important notions. One was Binding, and the other one was *Government*. It was a structural relation based on c-command. Basically, A governs B if A c-commands B and A is an adequate governor for B (e.g., Case-assignment was supposed to be based in a government relation, so V was an adequate governor for the accusative argument).

(20)



This is a simple version, of course. At the end of the eighties the notion was SO complex that it was necessary abandoning it and replacing it by bare c-command and feature-matching between A and B.

Anyway, a more explicit version of (19) is, then, (21).

(21) *ECP (second version)*
Traces must be antecedent-governed.

(22) *Antecedent Government*
A-antecedent governs B iff (i) A governs B, and (ii) A is coindexed with B.

Thus, the definitions in (21) and (22) allow us to explain the unacceptability of (17) and (18). However, these are not the only cases of traces in "wrong positions". Consider the following pattern.

(23) a. Who$_i$ do you think that John saw t_i?
b. Who$_i$ do you think John saw t_i?

(24) a. *Who do you think that t_i saw Bill?
b. Who do you think t_i saw Bill?

It seems to be the case that no trace may be immediately to the right of the (overt) complementizer *that.* This is the reason they call this phenomenon *That-trace Effect.*

Chomsky tried to derive the pattern in (23) by using the ECP. Obviously, this implied increasing the complexity of the definition in (21).

(25) *ECP (last definition for today)*
Traces must be antecedent-governed *or* lexically governed.

(26) *Lexical Government*
A head lexically governs its complements.

The idea is: a complement trace can be governed either lexically or by its antecedent, while a subject trace can only be governed according to (22). However, if a *Barrier*, a category that blocks government appears in between the subject-trace and its antecedent, that trace cannot be governed. Thus, a more complex definition of Government is necessary for this to work.

(27) *Government (adapted from Chomsky 1986:8)*
A governs B iff (i) A c-commands B, and (ii) there are no barriers between A and B.

Maybe you already noticed where I am going: (explanations based on) traces required a lot of assumptions and definitions. This is precisely the kind of "theoretical paraphernalia" that led to the *Minimalist Program*. But we will talk about that next class.

3. Revisiting the Surface and Deep Anaphora distinction

We have been assuming tacitly that traces are not Surface Anaphora. Basically:

> A trace does not have internal structure, so it is a case of *Deep Anaphora*.

Consider, then, the ambiguity of the next sentence:

(28) John wondered which picture of himself Bill took.

Who is *himself* referring to? What does Binding Theory tell us about these interpretations? (see footnote 1).

This kind of interpretative phenomena led to the postulation of *Copy Theory* (Chomsky 1993). According to Copy Theory, the underlying structure for (28) should be (29):

(29) John wondered [which picture of himself] Bill took [which picture of himself].

ONGOING RESEARCH IN GRAMMAR AND/OR COGNITION
HANDOUT 3 (28.04.2015)

The Copy Theory of Movement

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

1. On movement and chains

Before starting with Copy Theory, we need to be familiar with another GB notion: chains.

> Suppose we associate with each NP in S-structure a sequence (p_1, …, p_n) which, in an obvious sense, represents the derivational history of this NP by successive applications of Move-α; thus p_1 is the position of the NP itself; p_2 is the position (filled by a trace) from which it was moved to its final position; etc., p_n being the position (filled by a trace) occupied by the NP in D-structure. (Chomsky 1981: 45).

This concept of chain survived for several years to be defined in Chomsky (1986: 95) as "a reflection of a 'history of movement', consisting of the positions through which an element has moved from the A-position it occupied at D-Structure".

(1) a. $John_i$ seems t_i to have been killed t_i
b. CH = ($John_i$, t_i, t_i)

This notion was used to posit several conditions on syntactic movement: (i) Case was supposed to be assigned to chains, (ii) θ-roles were assumed to be assigned to chains.

(2) *Chain Condition (Chomsky 1986:171)*
In a maximal Chain C = {α_1, …, α_n}, α_n occupies its unique θ-position and α_1 its unique Case-marked position.

Notice that (2) restates the Theta-Criterion and the Case Filter in terms of chains.

(3) *Theta-Criterion*
Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.

(4) *Case Filter*
*NP if NP does not carry Case.

However, as Chomsky (1981: 45) observes, the whole concept of (movement) chain introduces a redundancy in the theory since they "can in fact be recovered from S-structure itself, given other properties of syntactic representation". Thus, for example, it is possible to postulate a mechanism of chain formation, *Form Chain*, based on part of the ECP (one of the conditions we saw last class):

(5) *Form Chain*
The collection of syntactic objects {α_1, α_2, … , α_n} form a chain iff every α_i antecedent-governs α_{i+1}.

This evident overlapping between movement operations and chains seems to suggest that one of these notions is redundant or, at least, derived from the other. In Brody's (1995) terms:

> Given that chains and Move α cover the same class of phenomena, we have an argument from conceptual economy against a theory that makes use of both concepts. If Move α is independently motivated, then the theory that uses chains to capture antecedent-trace type relationships is wrong; if a principle like Form Chain is independently necessary, then a theory incorporating Move α must be mistaken. (Brody 1995: 8).

Chomsky (1981, 1986) assumed that chains were derived representations from the more primitive operation Move-α. But in fact, he did not propose the concept to derive some property of movement dependencies but to establish some theoretical generalizations.

> We may think of a chain as an abstract representation of the phrase that is its head and assume that θ-roles and Case are assigned to chains. […] The chain is headed by a Case-marked position and terminates in a θ-position; Case is "transferred" from the head to the terminal position of the chain making the latter visible to receive the T-role that it, in turn, "transfers" to the argument that heads the chain. (Chomsky 1986: 96).

These properties of chains were extended and generalized to other kinds of syntactic dependencies where Move-α was not involved. To do so, Chomsky (1986) proposed the existence of a type of chain-like units that did not involve movement, the so-called *CHAINs*.

(6) $There_i$ is $[a\ man]_i$ in the window

2. The role of chains under Copy Theory

The main idea under *Copy Theory* (Chomsky 1993) is that movement does not exist as an independent operation. Syntactic movement is nothing but an epiphenomenon based on how interfaces (particularly, PF) interpret chains of non-distinct elements.

Yes, movement does not exist anymore. But we will keep using the term "movement" to refer (almost metaphorically) to this kind of phenomena.

(7) a. John seems ~~John~~ to have been killed ~~John~~
b. CH = (John, John, John)

When Chomsky proposed Copy Theory, he observed that this kind of theory would require two basic mechanisms: (i) a *Copy* operation and (ii) a phonological deletion procedure. He observes that such mechanisms are already required by grammar for independent reasons. Regarding (i), the Copy operation (or something similar) is necessary to explain morphological reduplication.

(8) *Pingelapese (Micronesian Language)*

a.		b.	
	kɔul ('to sing')		mejr ('to sleep')
	kɔukɔul ('singing')		mejmejr ('sleeping')
	kɔukɔukɔul ('still singing')		mejmejmejr ('still sleeping')

And regarding (ii), Chomsky observes that Ellipsis phenomena involve phonological deletion of phrases (notice that Chomsky is making a theoretical choice here: he is assuming that ellipsis involves surface anaphora).

(9) a. John said that he was looking for a cat, and so did Bill [say that he was looking for a cat].
b. John said that he was looking for a cat, and so did Bill.

Chomsky's idea is that both derivations in (7) and (9) require a deletion operation applying on chains formed by *identical elements*: "the antecedent and its copy are strictly identical and constitute a chain, if a chain is understood as (constructed from) a pair of terms (α_1; α_2) that are identical in constitution. It will follow, then, that the copy deletes" (Chomsky 1995: 253).

Thus, a sentence as (10) would involve the derivation in (11).

(10) John was killed.

(11) a. *Merge(killed, John)*
K = [$_{VP}$ killed John]
b. *Merge(was, VP]*
K = [$_{TP}$ was [$_{VP}$ killed John]]
c. *Copy (John)*
K = [$_{TP}$ was [$_{VP}$ killed John]]
L = John
d. *Merge (John, TP)*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed John]]]
e. *Chain Formation*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed John]]] CH = (John, John)
d. *Phonological Deletion of Copies*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed ~~John~~]]] CH = (John, ~~John~~)

So, basically, under Copy Theory:

Movement = Copy + Merge + Phonological Deletion

Notice that if there is no primitive movement operation, chains are no longer redundant. In fact, the way chains are manipulated in the phonological component "derives" movement.

3. Trace Theory vs. Copy Theory

But, is Copy Theory better than Trace Theory? The standard answer is "yes". There are several reasons:

3.1. RECONSTRUCTION

Last class we discussed briefly this example:

(12) John wondered which picture of himself Bill took.

And we proposed the following analysis for its ambiguity:

(13) John wondered [which picture of himself] Bill took [~~which picture of himself~~].

Compelling as it is, there are even more crucial examples. Consider the example in (14), which involves a violation of Condition C (a R-expression must be free).

(14) *He$_i$ believed an argument that John$_i$ is a genius.

Compare (14) with 15).

(15) *[Which argument that John$_i$ is a genius] did he$_i$ believe?

Now, tell me: which analysis allows explaining (15) straightforwardly?

(16) a. [Which argument that John$_i$ is a genius]$_j$ did he$_i$ believe t_j?
b. [Which argument that John$_i$ is a genius] did he$_i$ believe [which argument that John$_i$ is a genius]?

3.2. INTERPRETATION OF DISCONTNUOUS IDIOMS

We know that the VP *take pictures* means 'to photograph'. We cannot have that meaning for the verb *take* alone. So, how can we explain the meaning of (17)?

(17) How many pictures of John did you take?

Again, which analysis would you choose to explain (17)?

(18) a. [How many pictures of John]$_i$ did you take t_i?
b. [How many pictures of John]$_i$ did you take [~~how many pictures of John~~]

3.3. TRACES WITH PHONOLOGICAL CONTENT

Sometimes, in some languages, you can pronounce traces.

(19) *German*

Wen	glaubt	Hans	*wen*	Jakob	gesehen	hat?
Whom	thinks	Hans	whom	Jakob	seen	has

'Who does Hans think Jakob saw?'

(20) *Romani*

Kas	misline	*kas*	o Demìri	dikhlâ
Whom	think.you	whom	Demir	saw

'Whom do you think Demir saw?'

And it is far easier explaining this by saying that sometimes you do not delete some copies than by proposing an operation replacing ECs with wh-pronouns.

3.4. SATISFACTION OF INCLUSIVENESS

Chomsky (1995) introduced the *Inclusiveness Condition*:

(21) *Inclusiveness Condition (Chomsky 1995:228)*
Any structure formed by the computation is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation, apart from rearrangements of lexical properties (in particular, no indices, bar levels in the sense of X-bar theory, etc.).

The Inclusiveness Condition must be understood (I think) as methodological requirement on descriptive technology: it states that grammar operates with lexical features (i.e., things you have in your mental lexicon). Thus, accordingly, any non-lexical material used to describe or explain linguistic phenomena must be abandoned.

As discussed last class, Trace Theory makes use of two types of non-lexical elements generated by the grammar itself, traces and indexes, thereby violating Inclusiveness. On the other hand, Copy Theory complies with this condition since it is supposed to rely exclusively on (copied) lexical material.

ONGOING RESEARCH IN GRAMMAR AND/OR COGNITION
HANDOUT 4 (5.05.2015)

Copy Theory: linearization and economy

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

1. Two questions for Copy Theory

Last class, we discussed the following derivation for the sentence *John was killed.*

(1) a. *Merge(killed, John)*
K = [$_{VP}$ killed John]
b. *Merge(was, VP]*
K = [$_{TP}$ was [$_{VP}$ killed John]]
c. *Copy (John)*
K = [$_{TP}$ was [$_{VP}$ killed John]]
L = John
d. *Merge (John, TP)*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed John]]]
e. *Chain Formation*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed John]]] CH = (John, John)
d. *Phonological Deletion of Copies*
[$_{TP}$ John [$_{T'}$ was [$_{VP}$ killed ~~John~~]]] CH = (John, ~~John~~)

Let's consider the step in (1d): there is a deletion operation applying on the lower copy of *John*. Why? Or in other words, how can we explain the pattern in (2).

(2) a. John was killed ~~John~~. CH = (John, ~~John~~)
b. *~~John~~ was killed John. CH = (~~John~~, John)
c. *John was killed John. CH = (John, John)
d. *~~John~~ was killed ~~John~~. CH = (~~John~~, ~~John~~)

We will discuss now Nunes' (1995, 2004) version of *Copy Theory*. Basically, he attempts to answer two questions:

(i) Why is it the case that (in general) a nontrivial chain cannot have all of its links phonetically realized?
(ii) Why is it the case that (in general) traces and not heads of chains are the links that are deleted?

Thus, question (i) asks for an explanation for the contrast between (2a) and (2c), while question (ii) asks for an explanation for the pair (2a) and (2b).

2. Question (i): Uniqueness

According to Nunes, the fact that only one copy is pronounced follows from the *Linear Correspondence Axiom* (LCA) of Kayne (1994). In a nutshell, the LCA is a mechanism of *Linearization*:

(3) [$_{XP}$ Z [$_{X'}$ X Y]] → Z … X … Y

(4) *Linear Correspondence Axiom (simplified)*
If α asymmetrically c-commands β in the syntactic representation, then α precedes β in the PF representation.

Asymmetrical c-command and linear orderings share some properties. One of those properties is *Antisymmetry*.

(5) *Antisymmetry*
If X precedes Y, then Y cannot precede X.

However, consider (5) for a PF representation like (6):

(6) John was killed John.

Here, *John* precedes *was* and, at the same time, *was* precedes *John.* Thus, (6) violates Antisymmetry.

The problem is solved, however, if we delete all the copies in the representation but one.

Therefore, to linearize syntactic representations it is necessary to reduce all movement chains to only one link. The mechanism in charge to do so is called *Chain Reduction.*

(7) *Chain Reduction (Nunes 2004: 27)*
Delete the minimal number of constituents of a nontrivial chain CH that suffices for CH to be mapped into a linear order in accordance with the LCA.

This definition allows explaining three of the four cases in (2).

(8) a. John was killed ~~John~~. CH = (John, ~~John~~) by (7)
b. *~~John~~ was killed John. CH = (~~John~~, John) STILL REQUIRES EXPLANATION
c. *John was killed John. CH = (John, John) by (7)
d. *~~John~~ was killed ~~John~~. CH = (~~John~~, ~~John~~) by (7)

For Nunes (1995, 2004), the phonological realization of one link in a movement chain follows from Linearization requirements imposed by the LCA.

3. Question (ii): Rank

It only remains to derive the unacceptability of (8b), the case where the lower copy receives phonological realization.

As it may be noticed, despite the fact that there is a "sameness" relation between the copies in a non-trivial chain, their formal features may be different.

(9) John$_{K}$ was killed John$_{uK}$.

The lower copy of *John* bears an uninterpretable Case feature K (*uK*), whereas the higher copy has already checked that feature with the T head. It is necessary some operation deleting the uFF in these kind of syntactic representation.

(10) *FF-Elimination (Nunes 2004: 31)*
Given the sequence of pairs σ = <(F, P)$_1$, (F, P)$_2$, ... , (F, P)$_n$> such that σ is the output of Linearize, F is a set of formal features, and P is a set of phonological features, delete the minimal number of features of each set of formal features in order for σ to satisfy Full Interpretation at PF.

Now, assume we try spell-out (8b). In that case we require:

(11) *List of PF operations to spell-out (8b)*
a. Chain Reduction on the higher copy
b. FF-Elimination on the uK feature in the lower copy.

Compare that to the operations required to spell-out (8a):

(12) *List of PF operations to spell-out (8a)*
a. Chain Reduction on the lower copy.

Since the higher copy already checked its uFF syntactically, it is not necessary to apply FF-Elimination to spell it out. Thus, pronouncing the higher copy in the chain is more *economical* than pronouncing lower copies. And since shorter derivations block less economical alternatives (e.g., Chomsky 1995), the sentence in (8b) is ruled out.

According to Nunes (1995, 2004), economy considerations predict that deletion of lower copies should be preferred over deletion of higher copies in a movement chain.

4. Explaining the exceptions

In sum, Nunes' (1995, 2004) system allows deriving the phonetic realization of the head of a movement chain and the deletion of the remaining copies. But more importantly, it also allows explaining the exceptions to *Uniqueness* and *Rank*.

Consider the following pattern:

(13) a. Ko šta kupuje?
Who what buys
b. *Ko kupuje šta
Who buys what

(14) a. *Šta šta uslovljava?
What what conditions
b. Šta uslovljava šta?
what conditions what

(15) a. Šta neprestano šta uslovljava?
What constantly what conditions
b. *Šta neprestano uslovljava šta?
What constantly conditions what

According to Bošković (2002), it seems that a language specific condition of Serbo-Croatian prohibits two adjacent homophones at PF. Thus, in (14) the lower copy in the chain must be pronounced to avoid this scenario.

(16) [šta[~~šta~~$^{i}{}_{wh}$ … uslovljava šta$^{i}{}_{uwh}$]]

In this case, there is no shorter successful derivation blocking the pronunciation of the lower copy. Therefore, it is a principled exception to Rank.

Now, consider some cases as the ones introduced last class:

(17) *German*
Wen glaubt Hans *wen* Jakob gesehen hat?
Whom thinks Hans whom Jakob seen has
'Who does Hans think Jakob saw?'

(18) *Romani*
Kas misline *kas* o Demìri dikhlâ
Whom think.you whom Demir saw
'Whom do you think Demir saw?'

As said, Nunes' system motivates the deletion of lower copies to comply with a linearization requirement: the presence of more than one copy violates the antisymmetry condition on linear orderings imposed by the LCA. So, at first it seems that this system cannot capture patterns involving multiple copies.

However, Nunes follows Chomsky's (1995) suggestion that *the LCA does not apply word-internally* (i.e., the LCA is blind to the internal structure of syntactic terminals). If the LCA ignores the elements inside words, it is predicted that if a link of a nontrivial chain were *reanalyzed* as part of a word it would not be necessary to apply Chain Reduction on it.

According to Nunes, an intermediate copy of *wen* in (17) is reanalyzed as part of an embedded complementizer C^0 through an application of the morphological operation *Fusion*.

(19) *Fusion (Halle & Marantz 1993: 116)*
Fusion takes two terminal nodes that are sisters under a single category node and fuses them into a single terminal node.

(20) [$_{CP}$ wen^{i} … [$_{CP}$ #wen^{i} + C^0# [$_{TP}$ … wen^{i} …]]]

Therefore, Fusion takes the interrogative pronoun *wen* and the embedded complementizer and combines them in a single terminal node #wen + C#.

Given that the LCA cannot access the segment delimited by the # signs, and since Chain Reduction does not apply vacuously, the reanalyzed copy of *wen* is pronounced.

Furthermore, this analysis makes two very strong predictions regarding the nature of the multiple copy phenomena: first, only cyclic movement-based copies in "intermediate positions" may produce the kind of configuration that is necessary to apply Fusion, so it is predicted that there is no multiple copy pronunciation involving the tail a chain.

(21) *German (Fanselow and Mahajan 1995, apud Nunes 2004: 39)*
Wen denkst Du *wen* sie meint *wen* Harald liebt?
who think you who she believes who Harald loves
'Who do you think that she believes that Harald loves?'

(22) *German (Nunes 2004: 39)*
**Wen* glaubt Hans *wen* Jakob *wen* gesehen hat?
whom thinks Hans whom Jakob whom seen has
'Who does Hans think Jakob saw?'

And second, since Fusion is a morphological operation, it is predicted that there is no multiple copy pronunciation of full phrases.

(24) *German (Nunes 2004: 39)*
**Wessen Buch* glaubst du *wessen Buch* Hans liest?
whose book think you whose book Hans reads
'Whose book do you think Hans is reading?'

In a nutshell, the patterns involving phonetic realization of traces are elegantly explained in this system.

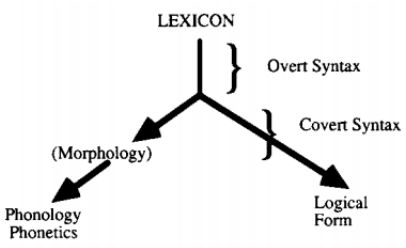
Single-Output Syntax

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

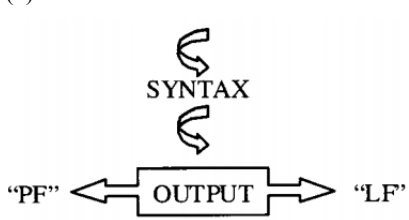
1. The basic idea

We will discuss the predictions of two different syntactic architectures.

(1)



(2)



The one on the left is the classically called *Inverted Y Model* (cf. Chomsky & Lasnik 1977). The one on the right has been referred as *Single Output Syntax* (Bobaljik 1995, 2002).

Consider the following sentences:

(3) Mary read every book.

A semanticist would say that the logical form of this sentence should look a lot like (4):

(4) For every book x, Mary read x.

So, how do we go from the sentence in (3) to the LF representation in (4)? Consider now the sentence in (5a) and its logical form in (5b).

(5) a. What book did Mary read?
b. For what book x, Mary read x.

Notice that there is a "semantic similarity" between (4) and (5b). In (5a) we have syntactic movement, so maybe we also have some kind of movement dependency in (3)…

May (1977) proposed a rule moving every quantifier phrase to the left of the sentence (more specifically to the projection we know today as TP) to form an operator-variable dependency. This operation was termed Quantifier Raising (QR).

QR is usually assumed to explain the following ambiguity:

(6) a. Every man loves some woman.
b. For every man x, there is some woman y, such as x loves y.
c. For some woman y, there is every man x, such as x loves y.

QR is assumed to involve syntactic movement since it seems to be constrained by island restrictions.

(7) a. They met someone [who knows Noam Chomsky].
b. *[Which linguistic celebrity]$_i$ did they meet someone [who knows t_i]?

(8) I know some woman who loves every man.

How do we explain the silent nature of QR? In the architecture in (1) this follows from a "timing" distinction between overt and covert movement: covert movement was supposed to apply after the syntactic representation was shipped to PF.

According to the architecture in (2): (i) syntax generates a single representation that is interpreted differently at the interfaces; (ii) movement operations are restricted to the syntactic component; (iii) the distinction between covert and overt movement follows from *Copy Theory* (Chomsky 1993).

(9) *Wh-Movement*

Syntax:	Which book	did Elaine read ~~which book~~.
LF:	For what book *x*	Elaine read the book *x*

(10) *QR*

Syntax:	~~Every book~~	Elaine read every book.
LF:	For every book *x*	Elaine read the book *x*

Notice that under Copy Theory is not entirely obvious how to derive the variable-like interpretation of the lower copy. Sauerland (1998), Fox (2002), Elbourne (2005), among others, propose an LF operation to deal with this problem:

(11) *Trace Conversion (Fox 2002: 67)*
a. Variable Insertion: (Det) Pred→ (Det [Pred λy(y=x)]
b. Determiner Replacement: (Det) [Pred λy(y=x)] → the [Pred λy(y=x)]

The mechanism in (11) transforms a quantifier phrase into a definite description with a variable inside (see the LFs in (9) and (10)).

In what remains of the class we will discuss a couple of analyses exploring the predictions of the architecture in (2).

2. Holmberg's Generalization

All Germanic languages but English display verb-second phenomena, where the verb raises to some high position. In these cases, a pronominal element undergoes "object shift" to the left periphery of the VP. This shift is obligatory except some dialects of Swedish[1].

(12) *Swedish*

Igår	läste	de	den$_i$ [$_{VP}$	inte t_i]
yesterday	read	they	it	not

'They didn't read it yesterday'.

(13) *Icelandic*

Hann	las	{ær}	ekki	{∗ær}
he	read	them	not	them

'*He didn't read them'.*

Object shift, however, is impossible when the verb has not moved from the VP. This observation is what has come to be known as *Holmberg's Generalization* (HG).

The data in (14) and (15) show that the verb must remain inside the VP in embedded clauses in Swedish.

(14) Jag tvivlar på [$_{CP}$ att [$_{IP}$ han [$_{VP}$ verkligen läste boken]]]

I	doubt	on	that	he	really	read book-the

'I doubt that he really read the book'.

[1] All the examples in this section are taken from Bobaljik (2002).

(15) *Swedish*
*Jag tvivlar på [$_{CP}$ att [$_{IP}$ han läste [$_{VP}$ verkligen boken]]]
I doubt on that he read really book-the
'I doubt that he really read the book'.

Thus, according to HG, a pronominal cannot undergo object shift in embedded clauses.

(16) *Swedish*
a. Det är troligt [att de [$_{VP}$ läste den]]
b. *Det är troligt [att de den$_i$ [$_{VP}$ läste t_i]]
it is probable that they it read it

According to Bobaljik (2002), the explanation for HG is similar to the account for *Do-Support* offered by Chomsky (1957).

(17) a. Sam eats ham.
b. *Sam not eats ham.
c. Sam does not eat ham.

The sentence in (17) is supposed to be unacceptable because of a violation of an adjacency requirement between the inflection and the verb in order to apply Morphological-Merger (for the sake of simplicity we will not discuss Bobaljik's treatment for adjuncts to the VP).

(18) a. [$_{TP}$ Sam [$_{T'}$ –s [$_{NegP}$ not [$_{VP}$ eat- ham]]]]
b. O----------*----------O ←*Adjacency Disrupted*
c. *do* ←*do-insertion*

According to Bobaljik, something similar happens in the Scandinavian languages: Morphological-Merger should "connect" the verb stem and the inflection.

(19) *Swedish*
a. Det är troligt [att de –te [$_{VP}$ läs- den]]
it is probable that they PST read it
O----------O ← *Successful Merger*

What would happen, however, if the pronoun *den* undergoes object shift?

(20) a. Det är troligt [att de –te den$_i$ [$_{VP}$ läs- t_i]]
it is probable that they PST it read
O-----*----------O ← *Adjacency Disrupted*

These languages do not introduce a dummy element like *do* to support the inflection. They do something different: since movement involves copies, the higher copy of the pronoun undergoing object shift is phonologically deleted and the lower is pronounced instead.

(21) a. Det är troligt [att de –te ~~den~~ [$_{VP}$ läs- den]]
it is probable that they PST it read it
O----------O ← *Successful Merger*

Thus, pronouns *always* undergo object shift in these languages, no matter if the verb moves or not. However, if the verb does not move, then the higher copy of the object pronoun must be deleted to allow applying Morphological-Merger.

3. Adjunct extraposition and Antecedent Contained Deletion

If overt and covert movements take place in the same cycle, they should interact somehow: e.g., overt operations are expected to follow covert operations sometimes.

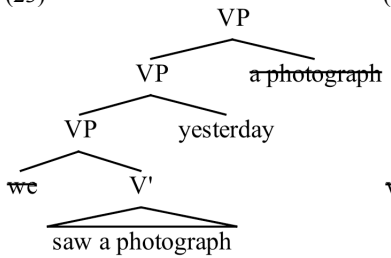
This prediction seems to be true for at least two empirical domains: adjunct extraposition and Antecedent Contained Deletion.

Fox & Nissenbaum (1999) provide an analysis for extraposition in these terms. According to them, the sentence in (22b) is derived by QR followed by *Late Merger* (cf. Lebeaux 1988).

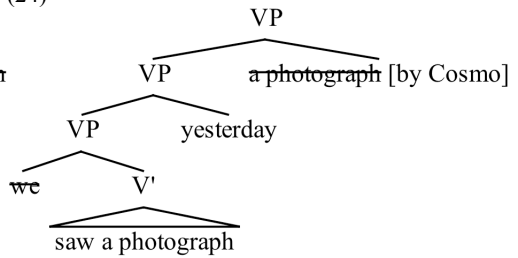
(22) a. We saw a photograph by Cosmo yesterday.
b. We saw a photograph yesterday by Cosmo.

The DP undergoes QR to VP (23), and the adjunct *by Cosmo* is merged to the NP (24).

(23)



(24)



This analysis allows deriving the lack of Condition C effects in adjunct extraposition.

(25) a. ??/*I gave him$_i$ a photograph from Cosmo's$_i$ collection yesterday.
b. I gave him$_i$ a photograph yesterday from Cosmo's$_i$ collection.

Also, it derives the fact that the NP seems to have wider scope than its surface position.

(26) *Adjunct-extraposition marks scope (Fox & Nissenbaum 1999:5)*
When an extraposed constituent (EC) is an adjunct, the scope of the source NP will be at least as high as the attachment site of EC.

Fox & Nissenbaum exemplify this property by using "free choice" *any*. They observe that *any* must appear in the scope of a modal operator like *look for* or *would* (27a). Therefore, the unacceptability of (27b) follows from (26).

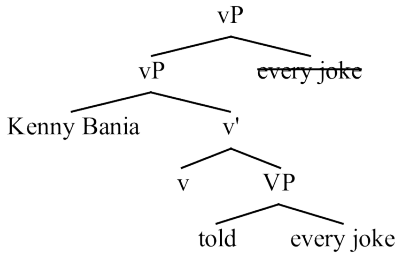
(27) a. Newman looked very intensely for *anything* that would help him against Jerry.
b. *Newman looked for *anything* very intensely that would help him against Jerry.

A similar analysis is proposed by Fox (2002) for *Antecedent Contained Deletion* (ACD).

(28) Kenny Bania told every joke Jerry did.

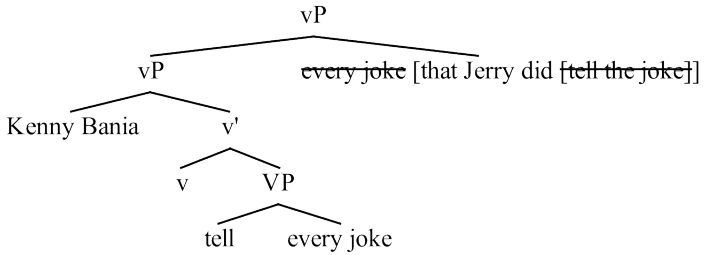
According to Fox, ACD involves two steps: first, QR of the quantified DP to VP.

(29)



Second, Late Merger of the adjunct relative clause to the NP in the higher copy. The VP inside the relative clause is elided under identity with the VP in the matrix clause.

(30)



Remember: Trace Conversion applies on the lower copies of A'-movement. This is actually what explains the identity between both VPs.

(31) *Matrix clause*
  a. *Syntax (adjunct omitted)*
     [$_{vP}$ every joke [$_{vP}$ Kenny Bania [$_{v'}$ v [$_{VP}$ tell [$_{DP}$ every joke]]]]]
  b. *LF (after Trace Conversion, adjunct omitted)*
     [$_{vP}$ every joke [$_{vP}$ Kenny Bania [$_{v'}$ v **[$_{VP}$ tell [$_{DP}$ the joke *x*]]**]]]

(32) *Adjunct relative clause*
  a. *Syntax*
     [$_{CP}$ OP joke [$_{TP}$ Jerry [$_{T'}$ did … [$_{VP}$ tell [$_{DP}$ OP joke]]]]]
  b. *LF (after Trace Conversion)*
     [$_{CP}$ OP joke [$_{TP}$ Jerry [$_{T'}$ did … **[$_{VP}$ tell [$_{DP}$ the joke *x*]]**]]]]

This analysis involves deriving several troubling properties of ACD, as *Tiedeman's Puzzle*. Consider the following pairs. Larson & May (1990) explain the cases in (33) claiming that the quantifier cannot "escape" the complement of *expect* due *Clause Boundedness*.

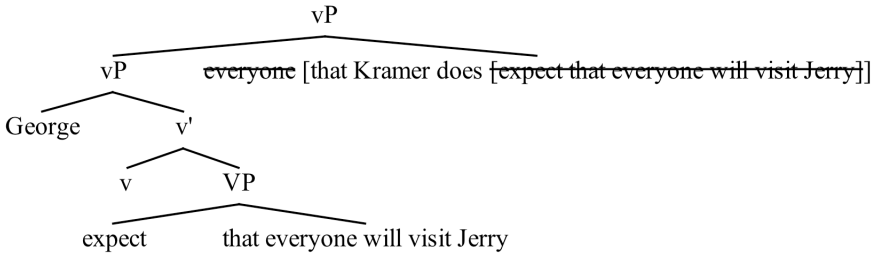
(33) a. *George expects [that everyone Kramer does will visit Jerry].
  b. *Newman said [that everyone Kramer did arrived].

However, Tiedeman (1995) observes that the sentences in (34) are acceptable.

(34) a. George expects [that everyone will visit Jerry that Kramer does].
  b. Newman said [that everyone arrived that Kramer did].

Fox's analysis allows deriving this contrast: it predicts that the chunk *(that) Kramer does* should follow the complement of the verb:

(35)



Hydras, forests and their linearization

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

1. Introduction

Until now we have discussed two possible analyses for syntactic movement: *Trace Theory* and *Copy Theory*. Both approaches rely on assuming that syntactic movement involves some kind of relation between two (or more) elements:

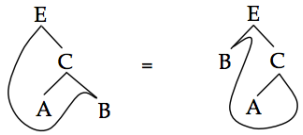
(1) John$_i$ was kissed t_i.

(2) John was kissed ~~John~~.

There is, however, a third alternative: assuming that movement involves *Remerging* of a constituent already present in the structure.

(3)

Merge (A, B) → C
Merge (B, C) → E



Think about it: if we assume that *Merge* is a free operation (as a lot of people think), then this possibility follows straightforwardly.

Notice that the element B in (3) is immediately dominated both by E and C. For this reason, this kind of approach to syntactic movement is usually termed *Multidominance.* Systems based on Multidominance avoid several inherent problems for Copy-based systems (as assuming the existence of a Copy operation, or dealing with "sameness"), but they present their own drawbacks.

We will briefly discuss the pros and cons of adopting a system allowing Multidominance. For that, we will follow De Vries´ (2009) presentation. You can find the paper in StudIP.

2. Types of Merge and their empirical motivation

De Vries introduces three logically possible "varieties" of Merge.

(4) *Merge(α,β)* → *γ constitutes*
- a. *first-time merge* iff α and β are independent roots before merger;
- b. *internal remerge* iff β is a root and α is included in β (or the other way around) before merger;
- c. *external remerge* iff β is included in some root δ, and α is an independent root (or the other way around) before merger.

The possibility of (4c), external remerge, leads to unconventional structures. However, this kind of syntactic dependency has been suggested several times in the literature. Remember the case in (5)?

(5) [What file]$_i$ did you save t_i without reading t_i?

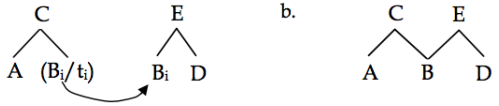
Nunes (2004) offers an analysis of *Parasitic Gaps* assuming Copy Theory. Basically, he proposes that there is a movement operation (i.e., Copy + Merge) between the adjunct and the main structure before the adjunct is adjoined to the later. Nunes terms this kind of dependency *Sideward Movement.*

(6) *Derivation for (5) involving Sideward Movement*
- a. *Two different trees K and L*
  K = [$_{ADJUNCT}$ without reading [$_{DP}$ which file]]
  L = [$_V$ save]
- b. *A copy of the DP is generated*
  K = [$_{ADJUNCT}$ without reading [$_{DP}$ which file]]
  L = [$_V$ save]
  M = [$_{DP}$ which file]
- c. *Merge the DP to the matrix structure*
  K = [$_{ADJUNCT}$ without reading [$_{DP}$ which file]]
  L = [$_V$ save [$_{DP}$ which file]]
- d. *Adjoin the Adjunct to the VP*
  K = [$_{VP}$ [$_{VP}$ save [$_{DP}$ which file]] [$_{ADJUNCT}$ without reading [$_{DP}$ which file]]
- e. *The derivation continues until the complement of save moves to Spec,C*
  K = [$_{CP}$ which file [$_{C'}$ did [$_{TP}$ you … [$_{VP}$ [$_{VP}$ save [$_{DP}$ which file]] [$_{ADJUNCT}$ without reading [$_{DP}$ which file]]]]]

Since the higher copy of *which file* c-commands its other two occurrences, both of them are phonologically deleted by *Chain Reduction*.

We may say that Sideward Movement is the "copy-friendly" version of external remerge. Its Multidominance version would be the strcture in (7b).

(7)

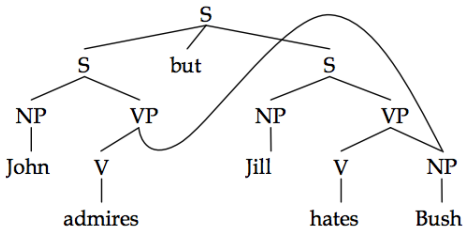


A classic example of external remerge in a Multidominance framework involves *Right Node Raising* (RNR) constructions according to McCawley (1982).

(8) John admires ____, but Jill hates Bush.

The analysis for RNR involves a structure like the one sketched in (9)

(9)



We will discuss in more detail the derivation for (9). For now, you should notice that external remerge is an operation that we, somehow, expect to exist in our system if Merge is actually *free.* Thus, we need to find a way to rule out some unwanted results of this assumption.

3. Constraining the outcome through cyclicity

Consider the standard definition of Merge offered in (10).

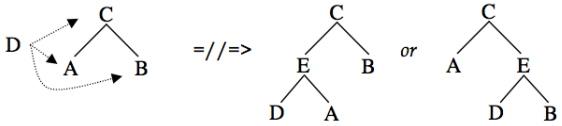
(10) *Merge*
Merge combines the SO α and β and creates a new SO [$_{\alpha P}$ α, β].

Also, we do not want new applications of Merge modifying previously generated structure: Merge is a structure-building and not a "structure-changing". This requirement is usually called *Cyclicity*: once you "created" something in a cycle you cannot change it in the next cycle. A condition on the cyclical nature of syntactic derivations is the *Extension Condition*.

(11) *Extension Condition (Chomsky 1993, 1995)*
Syntactic operations apply at the root, extending the tree.

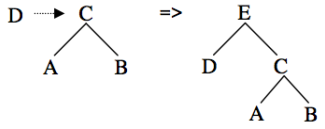
Thus, under (11), we rule out the combinatorial operations depicted in (12).

(12)

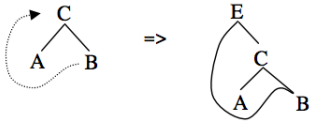


What we want to happen is exemplified in (13):

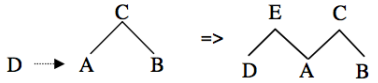
(13) a.



b.



c.



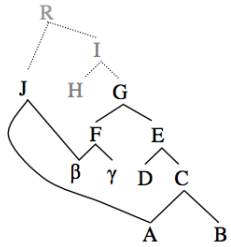
Now, consider the following scenario:

(14)

(Merge (β, γ) → F)

Merge (A, B) → C
Merge (D, C) → E
Merge (F, E) → G
Merge (A, β) → J

Merge (H, G) → I
Merge (J, I) → R



Since Merge creates, by definition, a new root by projecting a node, merging A and β creates a new root J. Potentially, the two roots can be combined in a new root R, licensing the structure. However, movement to embedded positions is typically considered ungrammatical.

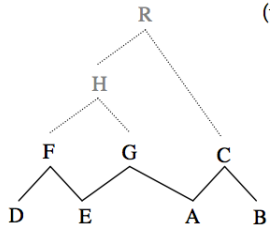
Consider now (15):

(15)

(Merge (D, E) → F)

Merge (A, B) → C
Merge (E, A) → G

Merge (F, G) → H
Merge (H, C) → R



Here, merging E and A creates a new root G, which can be combined with the matrix structure afterwards. However, (15) does not seem to correspond to any construction in any language. So, as (14), it must be excluded somehow.

(16) *Root Condition*
If α and β are selected as input for Merge, then α or β (or both) must be a root.

If syntax is a combinatorial mechanism (i.e., it combines multiple roots into a single root), (16) follows. It can also be understood as following from (11).

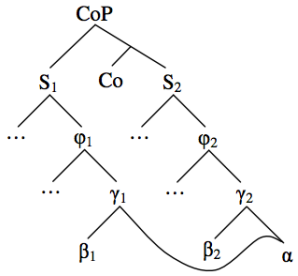
As previously said, RNR structures are frequently assumed to involve Multidominance. One property of RNR constructions is that they are insensitive to some restrictions on wh-movement.

(17) *What$_i$ does Mary like [men who sell t_i]?

(18) Mary likes [men who sell ___], but she hates [men who buy cars].

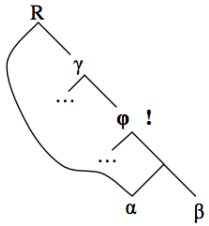
How can Multidominance explain this? The idea is that external remerge creates some kind of bypass between the two coordinated phrases before the locality boundaries are merged into the structures:

(19)



Of course, this does not imply that locality boundaries are not respected in regular movement constructions:

(20)



So far, so good, right? Not quite. Remember how Nunes´ (2004) version of Copy Theory relied a lot on Linearization mechanisms? Well, it seems that it is really difficult proposing a linearization mechanism in accordance with multidominant structures.

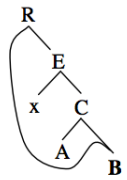
4. The linearization problem and a potential (and kinda inductive) solution

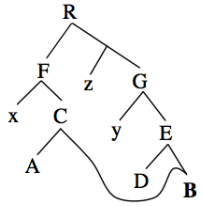
The main problem for the present discussion may be summarized in a pair:

(21) a. [Which violin] should this talented girl purchase ___?
b. The boy only admired ___, but the girl actually bought [this beautiful Stradivarius].

As you may notice, there is a difference regarding the position where the displaced constituent is actually pronounced. Thus, basically, we need a linearization mechanism explaining why the occurrences of B in (22) do not receive phonological representation in the same way.

(22)

a. 

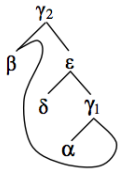
b. 

Some scholars have assumed that Merge determines linear order. Thus, according to them, the result of applying Merge(α,β) is an ordered pair <α,β>.

(23) *Uniformity of Mapping Hypothesis*
At the PF interface, generalized syntactic dependency is directly mapped onto phonological precedence, such that in a basic syntactic triad $<_\gamma$ α,β>, α will directly precede β.

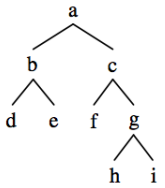
This approach to linearization, however, cannot be immediately extended to multidominant structures as (24). (Question for you: why?).

(24)



In order to deal with linearization of multidominant structures we need some other notions. The first one is *Tree Traversal.* It can be understood as a scanning procedure that creates a list of elements.

(25)



There are, at least, three types of traversal:

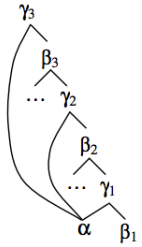
(26) a. *Preorder*
/abdecfghi/
b. *Inorder*
/dbeafchgi/
c. *Postorder*
/debfhigca/

We also need the following associated definition:

(27) *Current Mother*
The current mother of α is the most recently traversed mother during the linearization procedure.

Now, consider the following structure involving typical movement.

(28)

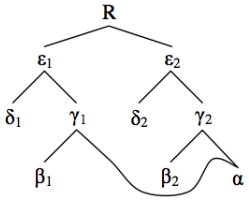


We know that we should pronounce the occurrence of α occupying the higher position. Notice that this occurrence is daughter of γ_3, the node dominating both γ_2 and γ_1. Thus, we can get the right result by postulating the following definition:

(29) *Spell-Out of Internally Remerged Nodes*
Linearize and α with more than one mother if the current mother dominates every other mother.

Now, we need a definition to Spell-Out externally remerged nodes. Consider (30):

(30)



Here, we know that we need to predict the pronunciation of the occurrence of α at the right. Crucially, both mothers of α are not in a dominance relation. However, there is a (linear) connection between them given by the ε nodes. Thus, the right result may be obtained with the following definition:

(31) *Spell-Out of Externally Remerged Nodes*
Linearize an α with more than one mother if
a. Every mother has been traversed, *and*
b. The current mother is not dominated by any other mother.
c.

Notice that definition (31) is the "exception" of the situation stated in (29) (i.e., (31a) will be achieved if the structure cannot comply with (29)). Therefore, it is not particularly difficult fusing both definitions in one:

(32) *Spell-Out of Remerged Nodes*
An α with more than one mother is linearized if and only if
a. The current mother is not dominated by any other mother, *and*
b. (i) every mother has been traversed, *or* (ii) the current mother dominates every other mother that has not been traversed.

5. Concluding remarks

Multidominance pros
- Prevents attributing ad hoc properties to traces and copies.
- Captures some very troubling constructions (e.g., RNR).
-

Multidominance cons
- "Graphical disadvantages"
- Linearization (really complex and, at the end, stipulative)

ONGOING RESEARCH IN GRAMMAR AND/OR COGNITION
HANDOUT 7 (02.06.2015)

A principled account of Non-Distinctiveness (I hope)

Carlos Muñoz Pérez
cmunozperez@filo.uba.ar

(This is a sketchy presentation of my own work. So comments and discussion are VERY welcome).

1. Introduction

According to Copy Theory there are no movement operations in language. The displacement property is explained in terms of collections of non-distinct elements in the syntactic structure being interpreted as "the same" element.

(1) John was kissed ~~John~~.

Thus, Non-Distinctiveness is maybe the most important property of movement dependencies under Copy Theory…

The problem is how can we define it.

One prominent view assumes that Non-Distinctiveness follows somehow from the Copy operation (cf. Chomsky 1995, Nunes 1995, 2004). Thus, if syntactic objects are assumed to carry some kind of distinctiveness markings, their copies will also carry the same index.

(2) a. *Derivational step α*
K = [$_{TP}$ was [$_{VP}$ kissed John$_1$]]
b. *Copy*
K = [$_{TP}$ was [$_{VP}$ kissed John$_1$]]
L = John$_1$
c. *Derivational step α+1*
[$_{TP}$ John$_1$ [$_{T'}$ was [$_{VP}$ kissed John$_1$]]]

Let's call this one the *Derivational Definition* of Sameness (DS). Basically, it involves marking as "the same" two elements since the syntactic derivation on.

This solution has been object of several criticisms during the years, basically because of its violation of the *Inclusiveness Condition* (cf. Chomsky 1995).

What I am going to propose here is redefining Non-Distinctiveness. My definition is different in two aspects: (i) it is a principled definition of Non-Distinctiveness (i.e., it is based on independent grammatical principles); and (ii) it makes different predictions, basically because it extends the empirical domain of Copy Theory beyond movement dependencies.

2. The system

Following Halle & Marantz (1993), it is assumed a *Late Insertion* model of grammar. Thus, syntactic terminals are taken to be just sets of features, each of them being a pair *attribute-value*, where the attribute denotes a *feature-class* (e.g., Category or Gender) and the value denotes a member of such a class (e.g., V, N, or MASC, FEM).

However, it seems that this assumption is not necessary for the interfaces: features at PF and LF are interpreted as instructions based on *privative* values that make no use of attributes (e.g., the noun *dogs* is interpreted as a plurality of entities without reference to the *Number* class). Therefore, a syntactic terminal consisting on the features {<Att$_1$,α>,<Att$_2$,β>} will be interpreted at PF and LF simply as the set {α,β}.

As an extension of this idea, a valueless feature <Att,__> will lack a representation at the interfaces, being *uninterpretable*. Thus, a syntactic terminal with the features {<Att$_1$,α>,<Att$_2$,β>,<Att$_3$,__>} will be interpreted as the set {α,β}.

Chomsky (2000, 2001): the operation Agree relates a Probe carrying an unvalued feature <Att$_1$,__> with an *active* Goal carrying a valued version of the same feature <Att$_1$,α>; ***the activity of the Goal*** is determined by an unvalued feature that will get satisfied after Agree

(3) *Activity Condition (Chomsky 2001)*
A goal G is accessible for Agree iff G has at least one uninterpretable feature.

I'm assuming that Agree participates in both A and A'-dependencies. The same kind of mechanism will be assumed for A-bar type of dependencies. Thus, DPs can carry an uninterpretable Case feature *κ* and also an uninterpretable left-peripheral feature *ω*.

(4) a. [$_{TP}$ T$_{\{<\varphi,__>,\ ...\}}$ [$_{VP}$ kissed John$_{\{<\varphi,3sg>,\ <\kappa,__>,\ ...\}}$]]
b. [$_{TP}$ John$_{\{<\varphi,3sg>,\ <\kappa,T>,\ ...\}}$ [$_{T'}$ was$_{\{<\varphi,3sg>,\ ...\}}$ [$_{VP}$ kissed John$_{\{<\varphi,3sg>,\ <\kappa,__>,\ ...\}}$]]]
c. [$_{TP}$ John$_{\{3sg,\ T,\ ...\}}$ [$_{T'}$ was$_{\{3s,\ ...\}}$ [$_{VP}$ kissed John$_{\{3sg,\ ...\}}$]]]

Notice that there is an *inclusion* relation between the occurrences of *John*: the set {3sg} is a subset of {3sg,T} (John$_{\{3sg\}}$ ⊆ John$_{\{T,3sg\}}$)). Such a relation will arise systematically for every new copy of a constituent (even if XP$_{\{...\}}$ = XP$_{\{...\}}$, the general case for cyclic movement), so it may be capitalized to define an interface mechanism to recognize non-distinct elements:

(5) *Non-Distinctiveness / Sameness*
Two constituents α and β are "the same" if:
a. α c-commands β,
b. the features of β are a subset of the features of α,
c. there is no δ between α and β being a proper subset of α or a proper superset of β.

This definition is based on (i) c-command (we cannot do anything without that), (ii) the Last Resort condition (encoded in (5b)) and (iii) a Locality consideration (5c). Thus, no "weird" mechanism is invoked.

This relation is ***transitive*** (as any other identity-type relation): if A and B are non-distinct, and B and C are non-distinct, then A and C are non-distinct.

Consider the analysis for these sentences.

(5) a. John was kissed.
b. John kissed John.
c. John said that John was kissed.

(Sorry, I'll sketch the analyses it in the blackboard).

Notice that this system builds "chains" based on the featural content of constituents; it does not say anything about the Copy operation (which I am still assuming). Notice also that there are no "syntactic" clues of what a "chain" can be, so the interfaces must scan *independently* the syntactic representation in search for non-distinct constituents.

Let's call this the *Representational Definition* of Sameness (RS). Let's explore some of its predictions.

Copy operation	LF-chain	PF-chain	Phenomenon
YES	YES	YES	Movement
YES	YES	NO	Multiple Copies
YES	NO	YES	Head Movement???
YES	NO	NO	What-Constructions (partial copying)
NO	YES	YES	Reconstruction asymmetries
NO	YES	NO	Resumption???
NO	NO	YES	Null-Subjects
NO	NO	NO	Trivial Chain (no movement)

2. Multiple Copies and What-constructions

In some languages it is possible to pronounce two (or even more) links of the same chain.

(6) *Romani (McDaniel 1986)*

Kas	misline	***kas***	o Demiri	dikhlâ?
Who	think	who	Demir	saw

'Who do you think Demir saw?'

(7) *German (*McDaniel 1986*)*

Wen	glaubt	Hans	***wen***	Jakob	gesehen	hat?
Who	thinks	Hans	who	Jakob	seen	has

'Who does Hans think Jakob saw?'

Some properties:

(8) *German (Fanselow and Mahajan 1995)*

Wen	denkst	Du	*wen*	sie	meint	*wen*	Harald	liebt?
who	think	you	who	she	believes	who	Harald	loves

'Who do you think that she believes that Harald loves?'

(9) *German (Nunes 2004)*

**Wen*	glaubt	Hans	*wen*	Jakob	*wen*	gesehen	hat?
whom	thinks	Hans	whom	Jakob	whom	seen	has

'Who does Hans think Jakob saw?'

(10) *German (Nunes 2004)*

**Wessen Buch*	glaubst	du	*wessen*	*Buch*	Hans	liest?
whose book	think	you	whose book		Hans	reads

'Whose book do you think Hans is reading?'

Following Nunes' (2004) analysis, I will assume that there is a morphological reanalysis between the wh-pronoun and an embedded C involving an application of the operation *Fusion*.

(11) *Fusion (Embick 2010: 78)*

$[_x\ \alpha] \frown [_y\ \beta] \rightarrow [_{x/y}\ \alpha,\beta]$

where α and β are features of X and Y.

When subordinate C and *wen* fuse, the result is a syntactic terminal with a set of features containing both the features of *wen* and C:

(12) $[_{CP}\ \mathbf{wen}_{\{Q,\ v,\ D,\ \varphi,\}}\ [_{C'}\ C_Q \ldots [_{CP}\ \mathbf{\#C + wen\#}_{\{C,\ v,\ D,\ \varphi,\}}\ [TP \ldots [_{VP}\ V\ \mathbf{wen}_{\{v,\ D,\ \varphi,\}}]]]]]$

According to (5), there are two chains in (12). And, as normal, the head of each chain is pronounced in both cases.

(13) a. $CH_1 = (\mathbf{wen}_{\{Q,\ v,\ D,\ \varphi,\}})$

b. $CH_2 = (\mathbf{\#C + wen\#}_{\{C,\ v,\ D,\ \varphi,\}}, \sout{\mathbf{wen}_{\{v,\ D,\ \varphi,\}}})$

Thus, in the system I am presenting here, multiple copy phenomena imply multiple chains at PF (but not at LF).

A more complex kind of pattern is the one attested by cases of non-identical doubling (or What-Constructions, cf. Fanselow 2006) in Dutch. The following data (and most of the analysis, in fact) is from Barbiers et al (2010).

(14) *Neuter and non-neuter wh-pronouns (Overijssel)*

Wat	denk	je	**wie**	ik	gezien	heb?
What	think	you	who	I	seen	have

'Who do you think I saw?'

(15) *Non-neuter and (non-neuter) relative pronouns (North-Holland)*

Wie	denk	je	**die**	ik	gezien	heb?
Who	think	you	rel.pron	I	seen	have

'Who do you think I saw?'

(16) *Neuter and (non-neuter) relative pronouns (Overijssel)*

Wat	denk	je	**die**	ik	gezien	heb?
What	think	you	rel.pron	I	seen	have

'Who do you think I saw?'

These are the only possible doublings in Dutch. Any other logical combination is unacceptable.

(17) ***Wie** denk je **wat** ik gezien heb?

Who think you what I seen have

'Who do you think I saw?'

(18) ***Die** denk je **wie** ik gezien heb?

rel.pron think you who I seen have

'Who do you think I saw?'

(19) ***Die** denk je **wat** ik gezien heb?

rel.pron think you what I seen have

'Who do you think I saw?'

Thus, the only possible orders are the ones sketched in (48).

(20)

a.	**wat** (neuter pronoun)	…	**wie** (non-neuter pronoun)
b.	**wie** (non-neuter pronoun)	…	**die** (non-neuter relative pronoun)
c.	**wat** (neuter pronoun)	…	**die** (non-neuter relative pronoun)

To provide an explanation, Barbiers et al. (2010) provide an analysis of these pronouns according to the following featural composition.

(21) a. ***wat*** = indefinite numeral (N)

b. ***wie*** = *wat* + φ-features (gender, G)

c. ***die*** = *wie* + definiteness (D)

I'll follow Cheng's (2001) analysis of this kind of constructions. According to her, there is a Partial Copying of a proper subset of the features of the original occurrence of the wh-pronoun/phrase.

(22) $[_{CP}\ \mathbf{wat}_{\{N,\ Q\}}\ [_{C'}\ C_Q\ [_{CP}\ \mathbf{wie}_{\{\varphi,N\}}\ [_{C'}\ C\ [_{TP} \ldots\ \mathbf{wie}_{\{\varphi,N\}}\ ]]]]]$

If you assume DS, the wh-elements in (22) will form a chain. This is what Barbiers et al. (2010) propose, so they have problems explaining (i) why *wie* is pronounced (they assume Nunes' 2004 explanation based on Fusion) and (ii) a semantic difference between movement and what-constructions we will see in (23). According to RS in (5), we have two chains here both at PF and LF. At PF we have two elements being pronounced, so pronunciation of *wie* is explained straightforwardly. Regarding LF, consider the following:

(23) Wie denk je niet dat zij uitgenodigd heeft?
who think you not that she invited has
'Who don't you think she has invited?'

(24) *Wat denk je niet wie zij uitgenodigd heeft?
what think you not who she invited has

Negation intervenes between *wat* and *wie* in what-constructions, but such a phenomenon does not occur in regular movement. This is a mysterious fact if we assume that chains and movement go together. But the present approach predicts this kind of asymmetry: if *wat* and *wie* do not form a chain, they must be "connected" by some other type of dependency. If it is a semantic dependency, it is expected that some scope-bearing element may disrupt the connection.

Proposal: since you do not have a chain at LF relating the wh-pronoun in the thematic position and the interrogative complementizer, these Dutch apply a wh-in-situ strategy. So the phenomenon is similar to cases as the following.

(25) *French*
a. *Jean ne mange pas quoi?
Jean Neg eat not what
b. Qui'est-ce que Jean ne mange pas t_i?
What that Jean NEG eat not
'What doesn't John eat?'

3. Reconstruction asymmetries

It has been frequently observed that copies of A-movement seem to bleed Condition C (cf. Chomsky 1995).

(25) [The claim that John$_i$ was asleep]$_j$ seems to him$_i$ to be correct t_j.

The acceptability of (25) is unexpected under Copy Theory, and it may be even considered surprising if sentences as (26), where a copy of A-movement is reconstructed ($\underline{t_k}$), are taken into consideration.

(26) [His$_i$ picture of the president$_j$] seemed to [every man]$_i$ $\underline{t_k}$ to be seen by him$_j$ t_k to be an intrusion.

This kind of asymmetry has been interpreted by Lebeaux (1988) as the absence of the relevant R-expression in some traces/unpronounced-copies (e.g., the NP [$_{NP}$ *claim that John was asleep*] would be missing in the original position of the subject in (25)). The problem is, then, explaining how different elements, some of them lacking a constituent, may form a movement chain.

Takahashi & Hulsey (2009) capture the pattern in (25) and (26) by proposing that some NPs are *countercyclically* merged with their Ds through *Wholesale Late Merger* (WLM): a D head may be merged in an A-position and undergo movement (7a); its NP restrictor may be merged after that (27b).

(27) a. [$_{DP}$ The] seems to him to be [$_{DP}$ the] correct.
b. [$_{DP}$ The [$_{NP}$ claim that John$_i$ was asleep]] seems to him$_i$ to be [$_{DP}$ the] correct.

We do not want counter-cyclic operations in our grammar. Under a derivational approach to syntax, syntactic structure is just the history of application of Merge (e.g., [V OBJ] is a constituent and not [SUBJ V] just because V and OBJ combine first).

There is, nevertheless, another possibility. According to (5), a D head without a NP complement can form a chain with a base-generated full-DP in a Case-marked position.

(28) [$_{TP}$ DP$_{\{T,D, Per, Num, Gen\}}$[$_{T'}$ T^0 … [$_{YP}$… D$_{\{D, Per\}}$ …]]] CH = (DP, D)

The derivation in (28) is strictly cyclic and generates the same syntactic representation as (27), explaining in exactly the same way the lack of Condition C effects in (25).

Regarding (26), a derivation like (29) is proposed. Here, the full-DP is merged in the c-command domain of the Case assigner, above the pronoun and below the quantifier, and from there it moves to the Case-marked position. Again, these three elements form a chain in virtue of (5).

(29) [$_{TP}$ DP$_{\{T,D, Per, Num, Gen\}}$ [$_{T'}$ T^0 … [$_{XP}$ DP$_{\{D, Per, Num, Gen\}}$ … [$_{YP}$… D$_{\{D, Per\}}$ …]]]]
CH = (DP, DP, D)

Since there is a R-expression in the second occurrence, reconstruction is predicted in this position, as it is attested in (26).