Two Related Talks on Linearization Matters (*Uriagereka*)

1. Why is linguistic linearization backwards?

### 1.1. Conceptualizing the Squeezing Problem

- Higginbotham (1983:151): ‘one can, in point of fact, just make one sound at a time’, ‘a consequence of the applications of the laws of nature to the human mouth’.
- Levelt (1981:306) independently raises concerns along these lines, separating ‘knowledge-based’ from ‘process-related’ conditions that somehow meet, or even clash, in the phenomenon of language as we experience it.
- Tesnière 1953/1959 points out an ‘antinomy’ (sic) between what he calls structural order –which he claims is ‘multi-dimensional’– and speech, which is one-dimensional.

All of these authors, from various viewpoints, have hit on the intuition that interesting tension exists between hierarchical and linear aspects of the language faculty.

Graffi (2001:197) notes that Tesnière took this as ‘the squaring of the circle’ in language, whose resolution is the condition *sine qua non* for the faculty itself.

The Higginbotham/Levelt/Tesnière (HLT) insight is interesting if we remind ourselves of a difficulty emphasized by Townsend and Bever (2001), refining ideas from Halle and Stevens (1962). While syntactic conditions are generally thought to be bottom-up, linguistic behavior in parsing proceeds serially and incrementally from before to after.

We should thank Marantz (2005) for revamping the Derivational Theory of Complexity (DTC), as the Minimalist Program has the right stuff to try this route again. But from this perspective it is legitimate to worry about how competence and performance meet.

‘Commensurability’ is possible once we break down the structures for which the problem arises to a size that makes the problem not just solvable, but in fact *effectively* so.

The HLT insight is of great significance, and it in fact constitutes one of those points of convergence between grammatical theories. Since at least Gazdar et al. 1985, non-transformational models are concerned with linearization matters (see Ojeda 2006). If, broadly speaking, dominance and precedence relations have a separate nature, it is natural to ask how they are connected, so that dominance translates into precedence.

### 1.2 Boundary Conditions on the Squeezing Problem

Parsing assumptions:

- First lexical analysis, so let’s assume some version of Halle and Stevens’s (1962) Analysis by Synthesis (AbS) ‘hypothesize-and-test’ method.
- Next, parsing proper, understood as a syntactic analysis checking that the processed lexical tokens, form some grammatical expression –yet another instance of AbS, except at a higher level of abstraction, with a different time resolution.
As speech is 1D, it constitutes a natural channel for a simple list of words. Now while this may be relevant for purely ‘finite-state’ components of grammar, it comes out short.

Phrasal units demand, from the algorithm recognizing such computational entities, something in the league of stacked elements. How can the parser deal with such objects?

The fact that phrases are not encoded in the speech signal is surprising. The human language faculty could have been such that [[[grammars or parsers] [[add brackets] [every time] [they’re needed]]], for instance having speakers go ‘quote and quote’.

- Attention, memory or pattern recognition alone cannot achieve the real parsing task.

- In the opposite side of things, if speakers where telepathic, they would only need their grammar to share their thoughts. Then these linearization issues would not arise.

The AbS approach suggests that parsers ‘go hypothetical’ in these situations, by way of recipients using their own grammar as their best bet, which raises time-resolution issues.

While speech is appropriately linear, hierarchical constructs are not based directly on speech, but on words, morphemes, features, which combine in familiar ways to yield the appropriate hierarchies. There isn’t a simple mapping even between speech and those.

Some projections stem from morphemes or silent categories, grammatical elements are distributed over distant speech units, even showing up as semantically vacuous copies, gaps, with and without antecedents… All of that shows that the parser needs a grammar.

Indeed, the parser makes a cyclic use of the grammar. If the computations of language tracked one another, it might have been easy to keep everything in check. But since nothing is that simple, pockets of regularity are needed for things to be halted for evaluation (i.e. domains for decisions about gaps relating to their antecedent, or be uniquely determined as specific gaps, where copies are identified as copies of one other, where different token silent categories are identified for the purposes of projection, etc.)

So, one: Phases are timing units (Drury 2005), thus the timing-resolution matter for an AbS method of parsing at this level is resolved via computational cycles of this size.

Two: the parser needs access to Linearization mechanism L, which the grammar uses to squeeze phrasal (or more complex) internal constructs into a speech signal.

These are two different requirements, which separate the Derivation by Phase (DbP) program from the Multiple Spell-out (MSO) program –and make them compatible too.

It is easy to see that these matters are live not just for parsing, but also for language acquisition, which at the very least presupposes some parser. In my view, that’s the strongest argument for a universal L, whether it is Kayne’s version or a different one.
1.3. The Finite State Limit on Phrase Structure

The strongest way to address the Squeezing Problem wouldn’t just be one that tackles a learnability consideration, but one that does so out of virtual conceptual necessity.

(1) The FS Limit on Phrase Structure
An exhaustively binary phrase-marker, none of whose branches symmetrically bifurcates, can be expressed in a FS fashion.

This statement is a sub-case of a more general equivalence from the literature on formal languages: Unary recursion (‘tail’ or ‘head’ recursion) is formally equivalent to iteration.

In more familiar Bare Phrase-structure (BPS) terms, (1) corresponds to the intuition that ‘tail’ recursion can be constructed without abandoning a derivational workspace.

The ‘Base Issue’ for the MSO program is whether we want to make something of that simple fact, and if so what. (1) invites the inference, within the general approach sketched above, that it is because of their FS limit that phrasal objects of the relevant sort linearize in terms of the very FS limit that corresponds to them. Note: no logical necessity. So:

(2) When \(x\) asymmetrically c-commands \(y\), \(x\) is linearly ordered with regards to \(y\).

HOWEVER: Epstein (1999) shows, this still doesn’t constitute specifically a linearization procedure –all that (2) does is put words in a row. Now compare:

(3) a. If \(x\) is linearly ordered with regards to \(y\), then \(x\) precedes \(y\).
   b. If \(x\) is linearly ordered with regards to \(y\), then \(x\) follows \(y\).

This, in effect, shows that the Linear Correspondence Axiom (LCA) is really as natural, in structural terms alone, as the Mirror LCA (MLCA). So why do we have (3a)?

1.4. Linearizing Context-sensitive Dependencies

(4) A Typical Wh-chain
a. PF: Which man did Fido bite Ø (LCA) a’. Ø Bite Fido did man which? (MLCA)
b. LF: [Which [Fido bite man]] b’. SAME.
c. Syntax: [Which man did [Fido bite t]] c’. SAME.

Both allow operators to c-command the variable, for compositionality reasons. But the LCA places c-commanding operators first in speech, while the MLCA places them last.

In both instances, when finding such context-sensitive dependencies, relevant parsers have to be alerted to a complex operation involving serious memory resources.

An LCA-based parser puts the operator on a memory buffer, until it finds the variable; an MLCA parser must instead put the variable in memory, in search of the operator.
Logical Space                  Grammatical Space

a. Variables  ASSOCIATED √  FREE √  ASSOCIATED √  FREE √
b. Operators  ASSOCIATED √  FREE √  ASSOCIATED √  FREE *

Operators must be associated to a variable, or else they are not interpreted; variables can be grammatically associated (bound) or not. Variable interpretation in each of those instances is different. Whereas operator-variable pairs presuppose some valuation process, free variables are interpreted in terms of complex contextual determinations.

An LCA-based parser that encounters an operator $\Omega(x)$ is in need of variable $x$; its computational demand is what it is (due to the operator-variable dependency), but \textit{resolvable within the confines of its derivational memory specifications}.

An MCLA-based parser that encountered a variable $x$ would be in an equivocal territory: $x$ may be associated to $\Omega(x)$, or a free $x$. The system won’t be able to resolve this open issue until the relevant search space where \textit{either option} can be satisfied is completed.

Baader and Frazier (2004) show how these matters affect regular parsers, in situations where variables could be equivocally interpreted as free or bound, reducing parsing speed. In a MLCA-based parser that difficulty would be the normal state of affairs, drastically reducing parsing ease –as compared to a corresponding LCA-based parser.

The argument just presented can be reproduced with chains of a simpler sort, involving minimalist Probe/Goal relations, but I will put that aside now in the interest of time.

1. 5. Linearizing Complex Branchings

Kayne resolves this matter by complicating the LCA, allowing the linearization of non-terminals. The relevant situation is as in (6), the problem arising, for instance, for the linearization of 4 and 6, for which c-command relations cannot be established:

```
(6)    1
       / \  
      2   3
     /   \ /   \ 
   4    5   6   7
  /…\   /…\  /…\  
```

In Kayne’s system, we can linearize 4 before 6 \textit{indirectly}, if we think of 2 as terminal, and thus as linearizable with respect to 6 –and we add a proviso to the following effect:

\textbf{(7) Linearization Induction}
If a non-terminal X dominates a terminal $y$, and X is linearized with respect to terminal $z$, then $y$ is linearized with regards to $z$.

Now if 2 linearizes w.r.t. 6 in (6), 4, which 2 dominates, will also linearize w.r.t. to 6.
Once we’re willing to establish the claim in (7), there is no easy way of stopping it from overgeneralizing. Thus, in (6) non-terminal 3 asymmetrically c-commands 2, hence should linearize with regards to 2, and then as per (7), since 3 dominates 6, 6 should linearize with respect to 2; but we concluded above that 2 linearizes with respect to 6!

Both Kayne (1994) and Chomsky (1995) have ad hoc ways of avoiding this paradox.

1. 6. Towards a MSO Model
Uriagereka (1996, 1999) suggests an alternative. Suppose that, under certain circumstances, Spell-out can be *multiple*. Economy considerations alone would generally rule out such a move, if a single rule application is preferred over multiple ones (spell-out being a rule). However, those apply only *up to convergence*.

That situation obtains for structures as in (6). Suppose only the LCA as in (3a) exists in the grammar –not the extension in (7). Then (6) couldn’t converge. But we could break (6) into linearizable chunks, *each obeying the FS Limit for which the LCA does work?*

Contradictory linearizations do not arise, since for them we must have chunked (6) as in:

(8) a. 

```
1
/ \ 2 3
/ \ 4 5
/…\ 6 7
```

b. 3

```
/ \ 6 7
/…\
```

Now observe what happens when we replace the numbers by familiar symbols:

(9) a. 

```
TP
/ \ DP
/ \ / \ D NP
/ \ T VP
/…\ /…\
```

b. 

```
TP
/ \ DP T’
/ \ / \ T VP
/…\ /…\
```

(10) a. 

```
TP
/ \ DP T’
/ \ / \ D NP T VP
/…\ /…\
```

b. T’

```
/ \ T VP
/…\ /…\
```

The tree-division in (9) *doesn’t tamper with any lexical projection;* in contrast, the tree division in (10) does. Observe this by using BPS notation:
In (12) T in (a) and |T| in (b) do not match, so the overall tree in this instance cannot be recovered (T’ in (29b) comes out as a maximal projection because in that phrase-marker it is not immediately dominated by any other category of the same lexical type).

In (11) this difficulty doesn’t arise. |DP| in (a) and (b) is the same sort of category (a maximal projection in each instance) because we haven’t tampered with its projection in the tree division, and thus at no point does |DP| become a new sort of projection.

This will happen whenever we divide the phrase-marker in such a way that both divided chunks result in a maximal projection.

That is close in spirit to Chomsky’s (1995) stipulation (which prevents the computational system access to intermediate projections); however, we need not say anything specific about intermediate or any kind of projection. The issue is simply one of appropriately matching the separate chunks in the divided tree, a natural and independent demand for any system sending bits of structure to interpretation in different derivational cycles.

That identity-of-categories issue, however, must be highlighted:

The Address Issue
Whenever a phrase-marker K is divided into sub-components L and M for K to meet LCA conditions of multiple Spell-out, the daughter phrase-marker M must correspond to an identical term M in L.

2. Re-examining Sub-Extraction: A Response to Stepanov

2. 1. Derivational Vortices

Derivational Current (Definition)
A set of phrasal representations is a derivational current if and only if all its symbols can be expressed in FS fashion without information loss.
Any given sentence may contain more than one derivational current within. Currents can be tributaries to other (super-)currents, and contain within them tributary (sub-)currents that terminate on them. I will call main whichever current is not a tributary of any other in any given sentence, and which makes it all the way up to a matrix expression.

(15) Derivational Vortex (Definition)
A current is a derivational vortex if and only if it gets spelled out.

In the end every main current constitutes its own derivational vortex, but the important point is that, within it—in critical regions that cannot be directly integrated into the main current—there can be arbitrary many independent sub-vortices.

Currents attempt to be singly spelled out—the most economical alternative. The issue, then, is what happens beyond main currents, in those points that cannot be linearized within the standard derivational dynamics (by reduction to the FS limit).

The whole intuition of the MSO system is that, as we have no way of compressing these extensions ‘beyond the current’ into linearizations ‘within the current’, we must do the job prior to their very association to the current they end up being tributaries of: when they are still individual currents that, in themselves, can be independently spelled-out.

A compressed Merge structure is no longer phrasal: after Spell-Out, the phrase marker that has been thus compressed is, as it were, a giant lexical compound, whose syntactic terms are interpretable but not accessible to movement, ellipsis, and similar operations. In this form of the proposal, the spelled-out phrase marker behaves much like a word, so that it can associate with the rest of the structure, keeping its label after Spell-Out.

Technically, phrase marker \{\alpha, \{L, K\}\} compresses through Spell-Out into an object \{\alpha, <L, K>\}, formally equivalent to \{\alpha, \{\{L\}, \{L, K\}\}\}. Since this is not a syntactic object (in the sense defined in (18) below), it can behave only as a whole; i.e., a frozen expression. Syntactic objects, in the sense of Chomsky (1995), take two forms:

(16) a. Base: A word is a syntactic object.
b. Induction: \{\alpha, \{L, K\}\} is a syntactic object, for L and K syntactic objects and \alpha a label.

(16b) is obtained through Merge and involves a labeling function that Chomsky argues is projection. What is relevant for our purposes is how a label is structurally expressed:

(17) Within a syntactic object, a label \alpha is not a term.

(18) K is a term if and only if (a) or (b):
a. Base: K is a phrase marker.
b. Induction: K is a member of a member of a term.
(18a) is obvious, and (18b) is based on the sort of object obtained by merging K and L: one set containing K and L, another containing {L, K} and label $\alpha$ – i.e. {a, {L, K}. This whole object (a phrasal unit) is a term, by (18a). Members of members of this term (L and K) are also terms, by (18b). Label $\alpha$ is a member of the first term, hence not a term.

So now consider the collapse of {a, {L, K}} into the unit {a, <L, K>}, equivalent to {a, {{L}, {L, K}}}, as part of its squeezing into a linearized object. By (18b), {L} and {L, K} are terms. However, {L, K} is not a syntactic object, by either (16a) or (16b). Therefore, the set {a, {{L}, {L, K}}} cannot be a syntactic object by (16b); if it is to be merged higher up, it can be a syntactic object only by (16a); that is, a word.

Less technically, **linearizing a current destroys its phrasal base.**

This does not mean the objects whose ‘phrasal scaffolding’ has fallen apart are not recognizable – that would be terrible for interpretive purposes, either in PF or in LF.

But while the ‘cards’ may be still there, the ‘card-castle’ that supported them in a given phrasal array is gone. The issue is how this affects further computations. In principle it may be possible to treat the whole bunch of ‘fallen cards’ as a structure-less unit, or even determine that ‘the cards are there’, as individual entities; but it will not be possible to establish further operations that crucially require the scaffolding as such, including the precise identification (technically, a ‘context’) of any given card within the bunch.

We deduce the fact that the elements dominated by any $\gamma$ in any given current precede whatever $\gamma$ precedes. That $\gamma$ should precede or be preceded by the other elements in its own current was already shown, part of the motivation for the FS limit on phrase-structure and how, under specific assumptions pertaining to the DTC, this corresponds to a given PF order (if additional parsing considerations are introduced).

The fact, in turn, that the elements dominated by $\gamma$ act as $\gamma$ does within its own current is a consequence of the fact that $\gamma$ has been spelled out separately from the current it relates to, and is thus its own derivational vortex. The elements dominated by $\gamma$ cannot interact with those that $\gamma$ interacts with, in the main current. Their place in the structure is as frozen under $\gamma$’s dominance as would be the place of the members of a compound $\gamma$ or (in the radical instance) elements that have ‘gone to performance.’

**2. 1. Conditions on Sub-extraction**

That ‘dynamically bifurcated’ access to interpretive components makes predictions for what are viable derivations, from the point(s) of squeezing onwards. The most interesting one bears on the characterization of classical restrictions on extraction domains, which in essence must be complements (Huang (1982)). The central contrast:

```
(19)   a. [...]X [...]t...[
    e.g., *who did you see [a critic of t]?
  b. [...]t...] X...
    e.g., *who did [a critic of t] see you?
```
The problem is that whatever licenses (19a) (Minimal Link Condition, Last Resort, the A-over-A requirement etc.) should also license (19b).

From the MSO perspective, a complement is different from any other dependent of a head: the elements a complement dominates are *within the same current* as the head selecting this complement; this is not true for the elements a non-complement dominates.

As a result, sub-extraction from a complement occurs within the same current, whereas *sub-extraction from a non-complement must happen across a vortex*, with paradox arising. If a non-complement is spelled out independently from its head, any extraction thereof will involve material from something that is not a syntactic object. If the non-complement is not spelled out, it will not be possible to squeeze its elements into speech.

2.2. Possible Subject Sub-Extractions?

Stepanov (2007) discusses putative counter-examples to Huang’s claim. His position is that, while sub-extraction from adjuncts is unavailable, the same is not true for subjects.

Let’s avoid irrelevant conditions: (i) in contexts that could be recovered as ‘aboutness’ inquiries (the Prolepsis analysis); (ii) in languages that make liberal use of empty pronominals in argument positions that may be related, as ‘gaps’, to various antecedents; (iii) particularly if these elements could be inside ‘extraposed’ sentential subjects.

(i) THE PROLEPSIS ISSUE

(20) a. Who is there a picture of on the wall?
   b. Of whom is there a picture on the wall?
   c. Of whom do you think there is a picture on the wall?
   d. I (often) think of politicians that there is a picture (of them) on every wall.

It is hard to decide whether (a) involves sub-extraction *from a subject*, as that depends on what the correct analysis of existential constructions is. Now most languages (those without preposition stranding) would actually ask question (a) as in (b). At this point we cannot easily avoid entering the realm of prolepsis, as is explicit in (c) if given the sort of underlying structure in (d). Although in (b) there is no pronounced psychological predicate to anchor the ‘aboutness’ phrase, such a predicate may be implicit in the mere assertion. Luckily this confusion can be avoided by invoking *sentential* subjects, and sub-extracting from them familiar dependents with regular case-markings.

(ii) THE EMPTY PRONOMINALS ISSUE

All of the examples Stepanov’s article presents, excepting German, correspond to pro-drop languages. Stepanov does acknowledge a potential difficulty, but dismisses it on the basis of an asymmetry: in the languages he studies, whereas a crucial gap inside a subject is possible, comparable examples with a gap inside an adjunct are ungrammatical:

(21) a. Mary, it is clear [that she was there].
   b. Mary, [I left because she was there].
In the irrelevant analysis, a pronoun is merely bound by a left-dislocated topic base-generated in the left-periphery of the clause. Already here some speakers find (21b) less eloquent than (21a). Consider, more relevantly, the matter in Spanish:

(22)  
\[
\begin{align*}
\text{a. } & \text{María, está claro [que __ estaba allí].} \\
& \text{María is clear that she was there} \\
& \text{‘María, it is clear that she was there.’}
\end{align*}
\]

\[
\begin{align*}
\text{b. ??? María, [sali porque __ estaba allí].} \\
& \text{María left-I because was there} \\
& \text{‘María, I left because she was there.’}
\end{align*}
\]

Now the ‘non-sequitur’ effect gets accentuated in (22b), and the sentence with a gap inside an adjunct sounds quite bad. But this could simply be because determining the reference of a null pronoun is a complicated matter, as Frascarelli (2007) argues, recalling observations that go back to Calabrese (1986).

Note that the contrast in (22) gets attenuated with an overt pronoun:

(23)  
\[
\begin{align*}
\text{a. } & \text{María, está claro [que su hijo estaba allí].} \\
& \text{María is clear that her son was there} \\
& \text{‘María, it is clear that her son was there.’}
\end{align*}
\]

\[
\begin{align*}
\text{b. ??María, [sali porque su hijo estaba allí].} \\
& \text{María left-I because her son was there} \\
& \text{‘María, I left because her son was there.’}
\end{align*}
\]

(23a) sounds less natural than (22a), but (23b) improves, vis-à-vis the comparable (22b).

Whatever the reason is for these contrasts, the methodological point is simple: this is a tricky spot to test whether the oddity of, say, (22b) is because of irrelevant reasons (the licensing conditions of pro inside adjuncts), or because it disallows a bona-fide extraction that (22a), in contrast, permits. Stepanov claims that an expression comparable to the Spanish (22a), the Palauan (24a), constitutes a valid instance of extraction –because an expression comparable to the Spanish (22b) is impossible:

(24)  
\[
\begin{align*}
\text{a. Mary [a kltukl [el kmo ng-oltoir er a John __ ]]} \\
& \text{Mary R-clear comp R-3sg-imp-love John} \\
& \text{Mary, [that ___ loves John] is clear.}
\end{align*}
\]

\[
\begin{align*}
\text{b. *[A di mla se’elik el se’al] a ku-rael] er a party le u’ul ng-mla er ngii.} \\
& \text{my old boyfriend IRR-1sg-left party because 3sg-was there} \\
& \text{My old boyfriend, I left the party [because ___ was there].}
\end{align*}
\]
But how do we know this contrast is any more serious than the one in (22), which has never been analyzed in terms of sub-extraction? And more centrally: as Palauan is a pro-drop language, just how can we even prevent an analysis in the terms suggested for (22)?

Again, the issue of topics and their base-generation in the left-periphery can be controlled for: either by involving elements that disallow the binding of (null) pronominals or those that cannot survive in the left periphery (e.g. negative quantifiers).

However, pro-drop languages create a far more serious issue: they disallow overt pleonastics, one of the safest tests to determine whether a sentential subject has been either ‘extraposed’ or left inside the VP –as an associate of the pleonastic in the TP spec.

(iii) THE SUBJECT EXTRAPOSITION ISSUE

The problematic matter can be illustrated even with English examples as in (25) and (26):

(25)  
(a) [??] Syntax is what [that my kids could study __ ] would worry me.  
(b) ? Syntax is what it would worry me [that my kids could study __ ].

(26)  
(a) [??] She read more books than is obvious [that he read __ ].  
(b) ? She read more books than it is obvious [that he read __ ].

When the subject sentences containing the gap are ‘extraposed’, relevant extractions improve –although they still seem less than perfect. The problem: Imagine now a language where we didn’t have the crucial it in (25) and (26) (boldfaced in the examples) to tell us whether the clause we are attempting to sub-extract from is inside or outside VP. How do we know whether we are dealing with examples of the (a) or the (b) type? In English word-order helps: if the sentence is literally extra-posed (thus to the right of the verb), we know it cannot be in the specifier of TP. But many languages are not so helpful.

To illustrate this concern in detail, consider the following three examples, from Turkish, Navajo and Japanese, respectively –again with glosses as reported by Stepanov:

(27)  
(a) [Op] [Ahmet-in ti git-me-si]-nin ben-i üz-dü-ğ.-ü]  
   Ahmet-gen go-inf-agr-gen I-acc sadden-past-comp-agr house
   ‘The house [that Ahmet went to __ ] saddened me.’

(b) ? Łééchqį̂ iisxi-(n)i  
   giį̂ shi-į̂ bééhózin-igiį̂ naha’i.  
   dog perf.3.kill (something)-nom me-with is-known-rel imp.3.bark
   ‘The dog that I know to have killed (something) is barking.’

   Lit: ‘The dog that (the fact) that _ killed something is known to me, is barking.’

(c) [Op] [Mary-ga ti yonda no]-ga akirakana yorimo John-wa takusan-no hon-o yonda.  
   Mary-nom read that-nom is-obvious than John-top many-gen book-acc read
   ‘John read more books than [that Mary read __ ] is obvious.’

11
Compare those loaded glosses to the following free translations into English:

(28) a. ? (This is) the house that it has saddened me [that Ahmet went to __ ]
b. ? The dog [that it is known to me [ ___ has killed (a cat)]] is barking.
c. ? John read more books than it is obvious [that Mary read __ ]

Now these examples are much better even in English –the ‘extraposition’ effect. Precisely where is the sentential subject sitting in Stepanov examples?

In the Turkish instance the preverbal object indicates that the sentential subject is higher, possibly outside the VP. But even that is not definitive, because the sentential subject could have been remnant moved to the left periphery without affecting the general issue (sub-extraction could have preceded the left-ward clausal displacement):

(29) The house [which, [that Ahmet went to __ ], it saddened to me].

An analysis as in (29) for (27a) –with a null pleonastic in place of it in (25)– would make the improvement in grammaticality of no more interest than that of (26b) vis-à-vis (26a).

I have similar worries about Stepanov’s Hungarian and Russian examples (his (27) and (29), respectively), or for that matter the sorts of comparable Greek examples that Spyropoulos and Philippaki-Warburton (2001) report, Greek also being a pro-drop language. One would need to understand in detail why these cannot be analyzed with the sentential subject low within the VP and successive adjustments of the remnant and/or scrambling sort (perhaps interacting with focus in the case of Hungarian) being responsible for surface order. In fact, Stepanov’s positive proposal in his paper basically assumes just that for all instances of successful sub-extraction from subjects.

One way to avoid that loophole in the analysis would be by providing verbs that disfavor an ‘extraposition’/VP-internal interpretation for the clause. Compare:

(30) a. [That we lost our pet] saddened me.
b. It saddened me [that we lost our pet].

(31) a. [That we lost our pet] caused us to yell.
b. ? It caused us to yell [that we lost our pet]

Psychological predicates like \textit{sadden} sit more comfortably with sentential ‘extraposition’ than causative predicates, thus the relative unnaturalness of (31b). Concomitantly, corresponding sub-extractions from each sort of predicate fare as in (32):

(32) a. ? Which pet did it sadden you [that you lost ___ ]?
b. ?? Which pet did it cause you to yell [that you lost ___ ]?
This may be because the verbs which best allow for the sentential ‘extraposition’ process are those that Belletti and Rizzi (1988) analyzed as being unaccusative, in the sense that their surface subject (the theme of the psychological emotion) is an underlying object. That would make sub-extractions from such themes unremarkable.

The list of the verbs that Stepanov uses: be known, obvious, clear, sheer pleasure, any trouble (to someone); to please, surprise, sadden, impress (someone); to hear (in the sense of ‘reportedly’), and to turn up. One wonders why all test contexts should be of this sort, and why this couldn’t be controlled by using expressions for which an unaccusative analysis is not reasonable, starting with causatives (or other cases of event relatedness, that we lost began/started/lead to/triggered... finished/wrapped up/completed a riot).

2.3. Genuine Subject Sub-extractions.
We can now construct stronger tests with the new controls.

Uriagereka (1988:118) observes: (I) Sub-extraction from subjects of unaccusative predicates is better than from corresponding subjects of (in)transitive predicates; (II) Sub-extraction from post-verbal subjects is better than from pre-verbal ones. From the MSO perspective, (I) is expected, if subjects of unaccusative predicates are underlyingly part of the main current. We want to concentrate on (II). There has never been much doubt that sub-extraction from displaced subjects of any sort – particularly if sitting on the TP specifier – yields ungrammaticality (see Ormazabal et al. (1994), Takahashi (1994)). This has nothing to do with the MSO model (those works give a Uniformity treatment).

The issue of interest here arises when subjects sit lower within VP, specifically in their theta-position, and more concretely when this site is not part of the main current. The Spanish example in (33) illustrates a relevant experimental condition:

(33) Qué partido [te hizo gritar [(el) que hayas perdido t]]
What game you made scream the that have.you lost
‘What game has it made you scream that you lost?’

Here the verb cannot have a reasonable unaccusative analysis, and although the subject is surely not in the TP specifier (in terms of any reasonably simple analysis), it is also not part of the main current, which assembles te hizo gritar ‘made you scream’.

My feeling hasn’t changed in two decades: (33) is not perfect, but it is also not as bad as it would be if the subject were in its canonical subject position. That is the response I get from other linguists about similar sub-extractions in their native languages, when these apply as in (15) (and note the question mark that Stepanov himself reports at least for his Navajo example in (15b)). I take it that, in the absence of more calibrated experiments, these sub-extractions could be worse, they just sound more or less odd as they are.

We have the evidence we have, which at this time is the question mark(s) for subject sub-extractions. If these are not attributable to other factors, we can take the MSO approach
to provide a line of reasoning for why they are there: sub-extractions from elements under those conditions are not part of the main current, and the rest follows.

If something makes me feel confident about this overall pattern is an old, observation:

(34) a. * Cuando [te hizo gritar [(el) que hayas perdido el partido t]]
    when you made scream the that have you lost the game
    ‘When has it made you scream that you lost the game (for that reason)?’

    b. * Cuando [dijiste [ que perdiste el partido t]]
    when said you that lost you the game
    ‘When did you say you lost the game?’

Since Huang (1982) first studied this matter, no one has ever produced a result challenging another basic observation of his: that adjunct sub-extraction is never possible from inside any island, including a subject clause.

From the MSO perspective the matter is simple: the system predicts a ‘weak island’ for bona-fide subject sub-extraction. ‘Weak islands’ are, well, weak for argument extraction; but they yield impossible adjunct/trace dependencies (see Cinque (1990)). This explanation would be unavailable if there wasn’t an island in these contexts.

2.4. Sub-extraction from Adjuncts
Stepanov (2007) emphasizes that sub-extractions do not happen from adjuncts, giving examples like (35) to make his general case:

(35) ?* Who did Mary cry after Peter hit t ?

But to be fair, we should give these sub-extractions as fair a chance as we did sub-extraction from subjects. More carefully controlled examples are as follows:

(36) Ese es el partido que le hizo llorar [el perder __ ].
    this is the game that him made cry the to lose
    ‘That is the game that it made him cry losing.’

(37) a. ?Ese es el partido que lloró (amargamente) [por haber perdido __ ].
    that is the game that cried he bitterly for to have lost
    ‘That’s the game that he cried (bitterly) after having lost’.

    b. ??Ese es el partido que [por haber perdido __ ] lloró (amargamente).
    that is the game that for to have lost cried he bitterly
    ‘That’s the game that after having lost he cried (bitterly).’

(38) a. ??Ese es el partido que insultó a todos [por haber perdido __ ].
    that is the game that insulted he to all for to have lost
    ‘That’s the game that he insulted everyone after having lost.’
Speakers I have consulted find these progressively worse, quite generally.

More unexpectedly perhaps, consider English versions of these examples:

(41)  
  a. ? That’s the game that he cried (bitterly) [after having lost __ ].  
  b. ?? That’s the game that [after having lost __ ] he cried (bitterly).

(42)  
  a. ?? That’s the game that he insulted everyone [after having lost __ ].  
  b. ??? That’s the game that [after having lost __ ] he insulted everyone.

(43)  
  a. ?? That’s the game that he cried (bitterly) [because we had lost __ ].  
  b. ??? That’s the game that [because we had lost __ ] he cried (bitterly).

(44)  
  a. * That’s the game that he insulted everyone [because we had lost __ ].  
  b. * That’s the game that [because we had lost __ ] he insulted everyone.

The response I received from native-speaker linguists included comments like “I’m sorry to report that these sound okish to me –I may be losing my intuitions!”; and since I asked for grades, the best examples in the series (for example (41a)) was assigned a B (in the usual grading system). This is not the common wisdom.

So one can make adjunct sub-extractions quite bad, in English and Spanish alike, for instance if one extracts from a ‘very high’ adjunct, such as a porque ‘because’ domain. Even there, several factors ameliorate the judgment. For example, things improve with unaccusative verbs (vis-à-vis full transitives), and in fact also if the gap is in final
position (all the (a) examples above, vis-à-vis the (b) ones). Moreover, sub-extraction of an adjunct, in all of these instances, is terrible. I give just the example with a minimal chance to be grammatical (the equivalent of the Spanish (36) with adjunct displacement), but the point can be raised about all the other examples above, which readers can try:

(45) *Ese es el momento en que lloró [por haber perdido el partido].

that is the moment in that cried he for to have lost the game

‘That’s the moment when he cried after having lost the game’.

As the English gloss of this example indicates, this sentence is sharply ungrammatical if the moment under discussion is taken to be the one in which he lost the game.

In sum: we find a familiar pattern in sub-extraction from adjuncts: (more or less) deviant sub-extractions of argument elements and terrible sub-extractions of comparable adjuncts. With one caveat: the observations hold both for English and for Spanish (unlike what we saw for subjects). That may be simply because adjuncts need not (in fact generally cannot) occupy the TP specifier—so this emphasizes again the idea that very bad sub-extractions from the TP specifier are at right angles to this whole discussion.

So we must again conclude that neither subjects nor adjuncts are part of what I have called the main current, which immediately entails a difficulty for sub-extraction. This is very clear when what sub-extracts is an adjunct (out of an adjunct or a subject), but a relative marginality can also be perceived when what sub-extracts is an argument.

Unfortunately, genuine subject sub-extraction is simply untestable in English. Situations with pleonastics—which would count as valid tests, the pleonastic taking the TP spec—are almost certainly irrelevant in this language, since it does not have transitive expletive constructions. However, the fact that some sub-extractions from adjuncts do pattern as expected even in English suggest that the overall conclusion is correct.

It remains to be seen why even sub-extractions from adjuncts improve with unaccusative verbs and when adjuncts are placed last. Both of these points suggest some sort of reanalysis, which somehow results in the adjunct possibly joining the main current. There is an analysis by Belletti (2005), involving focalization followed by remnant movement that actually has the right results, but I leave showing you that for another talk.