Somber Prospects for Late Merger[0]

Dominique Sportiche

Abstract: I discuss the phrase theoretic operation of *Late Merge* proposed in Lebeaux (1991) as it used in late adjunct merger or wholesale late merger. I show that under current theoretical assumptions about the nature of the *Merge* and *Move* operations, Late Merge overgenerates in ways that are difficult to overcome. Given that Late Merge, is otherwise demonstrably computationally unparsimonious (Kobele and Michaelis, 2012), I conclude that it should not be an available operation and suggest *Neglect* (Sportiche, 2016) as an alternative to generate the needed logical form representations.

Keywords: late merge, late adjunct merger, wholesale late merger, neglect.

1 Introduction

Late merger is a phrase structure theoretic composition operation proposed in Lebeaux (1991) and used by many authors ever since in a number of very influential pieces of work, for example Bhatt and Pancheva (2004), Chomsky (1993), Demirdache (2015), Fox and Nissenbaum (1999), Fox (2017), Hulsey and Sauerland (2006), Lebeaux (2009), Stanton (2016), Takahashi and Hulsey (2009). I will argue that this operation should not be allowed. This is not to say that the logical forms produced by late merger are undesirable. These LFs have the right desirable properties (e.g. to circumvent Condition C), but they should not be produced by merging late.

2 Why late merger

Lebeaux' s 1991 classic late adjunct merger proposal is motivated by the following asymmetry:[1]

(1) a. Which villages $\boxed{\text{near Picasso}_1\text{'s estate}}$ did he$_1$ visit

 b. *Which pictures $\boxed{\text{ Picasso}_1}$ did he$_1$ sell

 c. *Whose hypothesis $\boxed{\text{that Picasso}_1\text{ was a fraud}}$ did he$_1$ resent t

Here is the standard account in terms of late merger,[2] framed in a system that takes traces to be copies and phrase structures to be built derivationally. The boxed relative clause in (1-a) is taken to be an adjunct to the noun *villages*, the boxed phrase in (1-b), (1-c) a complement to the noun (*pictures* or *hypothesis*). The late adjunct merger analysis allows (1-a) to be derived in either of the following two ways:

(2) a. Which villages near Picasso$_1$'s estate did he$_1$ visit ~~which villages near Picasso$_1$'s estate~~

 b. Which villages $\boxed{\text{near Picasso}_1\text{'s estate}}$ did he$_1$ visit which villages

If the trace is a full copy of the moved phrase as in (2-a), the (unpronounced) bottom copy (crossed out) must be interpreted; this yields a condition C effect at LF. But late adjunct merger also allows the derivation in (2-b) with the boxed part late merged, that is adjoined to *villages* after wh-movement, thus circumventing Condition C. (1-a) with this derivation is fine.

Because the *of* phrase in (1-b) or the *that* clause in (1-c) is assumed to be a complement, it cannot be inserted (so[3]) late. Consequently, the bottom copy must contain *picture*, a complement (of the D complement of) *sell*, and *of Picasso* or the

that clause, which are complements of the nouns. The full representation of (1-b) for example must contain the substructure below, triggering a Condition C effect:

(3)   ...he$_1$ sell   pictures of Picasso$_1$

The availability of late merger non trivially increases the generative capacity of grammars. It has been established that the formal systems underlying the linguists' current grammars (e.g. Minimalist Grammars with copy, see Kobele, 2006) are computationally well behaved without late merger (Michaelis, 1998). Kobele and Michaelis (2012) shows that adding Late Merge increases the generative capacity of such grammars to supersets of unknown computational properties, possibly intractable. Although this not damning in itself, as some (unknown) constraint on late merging may in fact restrict generative capacity to demonstrably tractable sets, there is a proven sense in which late merger itself is a priori unparsimonious. Whether late merger should be allowed is thus in principle a substantive question.

But there is also evidence that late merger as used by the aforementioned authors should not be allowed: as we show, because late merger must be able to operate within islands or be unboundedly countercyclic, this allows the equivalent of island violating movements, or unconstrained parasitic gaps among many other pathological cases.

3   Simple late adjunct merger

Simple late adjunct merger must be able to be unboundedly countercyclic: a treatment of the asymmetry above[4] in terms of late adjunct merger requires not only countercyclic syntactic operations, but an unboundedly countercyclic syntax. The countercyclicity comes from the fact that the relative clause in (1-a) must be assumed to be late inserted in the position in which it is interpreted, namely inside

the complement structure of the determiner of the relative clause. This insertion is countercyclic since merger is not at the edge (of the relative clause phase). But the logic of this account requires that such late merger be unboundedly countercyclic. This is illustrated by the following kind of examples:

(4)   [Whose criticism of [Mary's rendition of (…) the claim [that you [formulated (…) the hypothesis [[that Henri [visited the villages near $Picasso_1$'s estate]]]]]]$_k$ did he_1 endorse t_k

Such sentences with the indicated coreference are well formed. This means that the boxed adjunct must have been late inserted. But it is late adjunction to an element (*villages*) which is a recursive complement of a complement of the main head noun *criticism.* This means that none of this intervening material can be (so) late inserted: as a result, the late merger of the adjunct must take place inside the four bracketed constituents all of which are phases. It should be clear that the example can be modified so that late merger be required in an arbitrarily deeply embedded constituent, yielding unbounded countercyclicity.[5]

Why is this undesirable: is the fact that unboundedly countercyclic syntactic operations are allowed undesirable? Not necessarily in itself. As we will see below, late merger yields structures that are readily legible, that is, readily interpretable at the LF or PF interface by standard interpretive rules. But once late merger's generative power is allowed, overgeneration ensues. Here is how.

First, since Chomsky (2004), or Starke (2001), it has been recognized that formally, the operations of Merge and Move were one and the same phrase structure theoretic operation, differing only on their domains. Standard Merge is the subcase operating on doubletons $\{\alpha, \beta\}$ where neither α nor β has been previously merged. Standard Merge is thus sometimes called First Merge. Standard Move is the partial

complement case where exactly one element, say α, of the doubleton has been previously merged or moved. It is thus called Remerge (of α) to β: α and β now form a new object γ. Late merge now can be seen as relaxing the definition of Standard Merge and allow first merge of say α, to and crucially into[6] a previously merged β. In (2-b) for example, the boxed adjunct has been late merged into *which villages* that had ben previously merged and to *villages*, which is a subpart of this previously merged phrase.

Such late merger of an adjunct is usually invoked in instances of late first merger of this adjunct. But there is in fact no reason why it couldn't be late **re**merge since there is no theoretical difference between the operations of first merge and remerge (only the sources of the merged material differ). This second option, which in effect fakes a movement dependency, predicts that adjunct movement can violate (probably) any island.

To see this, consider the following representation.

(5)   Near Paris John thinks … that you live t

This structure would be out if there was no trace, here t = (copy of) *near Paris*, as the adjunct is not interpretively related to anything. But there are now two ways of deriving this structure with a trace.

One is the standard way: first merge *near Paris* with *live*, then remerge it higher.

The other base generates the adjunct where it appears. There is surely nothing wrong with this: since there is no theoretical difference between first merge and remerge, clefted (or topicalized, or dislocated, .) adjuncts can in principle be first merged in their landing position[7] This would normally lead to uninterpretability, because as noted above, the adjunct is not interpretively related to anything.

But suppose that next *near Paris* is remerged late where t is, yielding (5). Since late (re)merging must be unboundedly countercyclic, this means that the dots ...in (5) can stand for anything e.g. the null string, yielding a well formed output.

These dots can also stand for weak islands, strong islands or what have you. The prediction is that adjuncts can freely escape any island, unboundedly.[8] another example, late remerge makes it possible to create ill-formed structures mimicking any type of parasitic gap structure whatever (even with PP s which do not allow parasitic gap structures):

(6)　　Near Paris John said [...that you live t] [...when a plane crashed t]

Here again, both t' s are generated by late merger and the dots could stand for whatever we choose.

One may wonder if these clearly undesirable derivations are not ruled out by other considerations making them e.g. LF or PF uninterpretable, and thus crashing at the interface. Late remerger in effect creates a representation identical to a representation created by movement but in reverse order so this is unlikely but to check this, let us be more explicit about the interpretive procedures.

The semantic interpretation of a standard movement involves the following abstract steps:

1. Merge to β a phrase α

2. Merge in a position that c-commands β, a copy of α

3. Prepend to the phrase that is the sister to the copy of α: λn,

4. Trace Convert (as in Fox, 2002) α, giving it the index n.

The Late Remerge derivations described above would go as follows:

1. Merge to β a phrase α

2. Merge in a position that β c-commands, a copy of α

3. Prepend to the phrase that is the sister to the copy of α: λn,

4. Trace Convert (as in Fox, 2002) α, giving it the index n.

These derivations are the same with respect to their semantics. The higher phrase is made a binder and the lower phrase is made the variable that is bound. The only differences in the derivations is where the copy is put: higher or lower. Countenancing Late Merge allows the copy to be put into the lower position, creating problems. Preventing Late Merge also forces the copy to be put in the higher position, thereby avoiding the problems.

Turning now to the PF interface, one could try to capitalize on the reasonable position advocated that (at least some) cyclicity effects on movement should be reduced to linearization/PF mapping constraints (cf. e.g. Fox and Pesetsky, 2005), with linearization itself proceeding phase by phase. Would the linearizability requirement distinguish a derivation of (5) with downward late remerge (in effect downward movement) from its derivation via standard remerge (upward movement)?

In cyclic linearization structures, linearization constraints are added at each phase. A movement dependency, two copies of the same item, would create contradictory requirements but it can be linearized as long as one if them is deleted (phonetically ignored, neglected). Now, an unboundedly deep late merged structure must also be linearizable. This means that all previous phasal information has been lost up to the highest constituent in the derivation at the point at which the late merged constituent is merged: so the structure visible to the linearization algorithm must be the entire structure (the very top, whose merger diagnoses the point at which Late Merge takes place, and the very bottom, where late merge actually occurs). But then, the same principle that allows moved elements to be linearized

7

(silent trace) applies in unboundedly deep Late Remerge too: both copies are simultaneously visible to the ordering constraint algorithm, but, as in standard movement cases, the bottom one can be phonologically ignored, allowing linearization.

Finally, one could also deny that remerging a phrase α higher or lower than α yields the same object, a movement "chain" in both cases. For example, Fox (2017) tentatively suggests a near equivalent of the following idea: for a relation between two copies to count as movement, Agree must have been involved in the remerging creating one of the copies.[9] This would indeed distinguish the two options since Agree is directional. Note first that there is no theoretical reason why Agree itself should be directional (rather than following from an appropriate ban on countercyclicity).[10] But this would not be sufficient to avoid the pathological cases allowed by Later Merger, as the case in (6) shows. Indeed, consider again(5), generated by standard movement of the adjunct *near Paris* leaving a trace t. The results is (7-a), where the adjunct agrees with say the Topic head in some feature, say *top*:

(7)   a.   [[Near Paris]$_{top-}$ [Topic$_{top+}$ [[John [...live t]]]]

        b.   [[Near Paris]$_{top-}$ [Topic$_{top+}$ [[John [...live t ...when a plane crashed ζ]]]]

Suppose now that there is additional structure as in (7-b), which could be part of an adjunct clause, of one or more wh-islands, a complex noun phrase or what have you. Late remerge of *near Paris* as a VP adjunct in the position ζ is allowed (and, if it matters, we can restrict this to cases where ζ is c-commanded by *near Paris* and not by t).

I conclude Late Merger of adjuncts should not be allowed.

4 Wholesale late merger: same overgeneration problem

Wholesale late merger, which allows late merging of a complement under certain conditions (Bhatt and Pancheva, 2004, Takahashi and Hulsey, 2009) is illustrated below:

(8)   The conclusion that John$_1$ had cheated seemed to him$_1$ to be unfounded

Coreference *him/John* is possible, it is argued by proponents of this type of late merger (Lebeaux, 2009, Takahashi and Hulsey, 2009) that the DP headed by *the* is merged as argument of *unfounded*, without its NP *conclusion that John had cheated* and that this NP is late merged after the DP has raised to subject of the main clause. Agreeing with Lebeaux or Takahashi and Hulsey' s description, let us assume a timing constraint on such late merging: descriptively, this type of late NP merger is limited to applying to a D at the latest when this D is in a Case position (so that, for Takahashi and Hulsey, NP can check its own Case in a timely fashion). Now note that coordinate structures are among the islands within which late merging can operate:

(9)   a.   [[The pictures that John$_1$ likes] and [the books]], he$_1$ had to sell.

     b.   [[The conclusion that John$_1$ had cheated] and [the punishment]] seemed to him$_1$ to be unfounded

In both well formed examples, the boxed constituent must have been inserted late inside one conjunct of a coordinate structure to avoid a Condition C violation. The first example illustrates late adjunct merger, the second wholesale late merger.

Just like late adjunct merger, wholesale late remerge can fake (illegal) movement of e.g. NPs as in e.g. Split Topicalization available in a variety of

languages but not in English (van Hoof, 2006). As van Hoof (2006) illustrates, Split Topicalization in German can move an NP, stranding a determiner as in:

(10) $[_{CP}$ $[_{NP}$ Definitiver Beweis$]_k$ $[_{C'}$ ist bis jetzt $[_{DP}$ keiner t_k] gefunden
      definitive proof   is until now   no   found
 worden]].
 been
 'No definitive proof has until now been found'

In addition, van Riemsdijk (1989) shows that Split Topicalization exhibits the diagnostic properties of movement: it leaves behind a gap, is island sensitive and shows the connectivity effects with respect to binding and morphological case typical of movement dependencies. Given that NPs are precisely what Wholesale Late Merger merges late, it should be clear that we can duplicate with Wholesale Late Merger of NP the faking of movement across islands that we did with adjuncts, incorrectly predicting that Split NP Topicalization is unbounded.

 In English, we can also fake NP movement involved in the promotion analysis of relative clauses, yielding pathological structures. To illustrate, some NP can be remerged to some wh-D allowed in a relativized DP regardless of where this D is. Thus it is possible to relativize from inside an island (here a pied piped DP) without movement by remerging the NP book as in:

(11) a. I bought the book [[the topic of [which and this review]]$_1$ [you knew about t_1]]

   b. I bought the $[_{NP}$ book] [C_{rel} [the topic of $[_{DP}$ $[_{DP}$ $[_D$ which] $[_{NP}$ *book*] and $[_{DP}$ this review]]$_1$ [you knew about t_1]]

   c. I bought the $[_{NP}$ book] [C_{rel} [the topic of $[_{DP}$ $[_D$ which]]$_1$ [you knew about t_1]]

The relative clause in (11-a) would be merged to an external NP book (as in matching derivations – see Hulsey and Sauerland, 2006), and this NP is late remerged to which inside a coordinate structure as in (11-b). Pied Piping applies legally in (11-a): following Heck's 2009 analysis, Pied piping can occur in case the wh element triggering the pied piping (here *which*) is accessible via Agree to the C probe of the relative clause C_{rel}. In (11-a), it is accessible as shown by:

- the well formedness of (11-c) showing that the DP containing the conjunction is pied pipable and
- the fact that Agree is able to reach into coordinations (as shown by e.g. first conjunct agreement configurations). So there is nothing wrong with pied piping per se.[11]

Late remerge of the italicized NP *book* in (11-a) is not subject to the timing constraints on complement insertion (since it is already Case checked) and it can violate the coordinate structure constraint: it can thus be performed, overgenerating.

I conclude that no variety of late merger is available.[12] As mentioned at the outset, the LFs produced by late merge have the right desirable properties (to circumvent Condition C), but they should not be produced by late merger. How then are they produced?

Contemporary theories of the syntactic computational system distinguish the derivation tree of a structure - the derivational history in terms of Merge and Move - from the maps from such a derivation tree to PF and LF representations (see Kobele, 2006 who introduces this for minimalist grammars, but the point is more general). The undesirable increase of generative capacity and the empirical problems we outlined are due to coding the origin of the desirable LFs into the derivational

process itself (by allowing late merger). Sportiche (2016) puts forth a proposal which does not assume late merger, but instead exploits what is already assumed concerning how the map from a derivation tree to an LF (or a PF) functions: informally speaking, such map can *neglect* to fully spell out the content of a trace. This way of reaching desirable LFs (or PFs) does not allow overcoming island violations (and does not *seem* to alter the generative capacity of the computational system, Gregory Kobele, p.c. and Thomas Graf, p.c., although a proof is needed, of course). It thus looks like a promising replacement to late merger.[13]

References

Adger, David, Alex Drummond, David Hall, and Coppe van Urk. 2016.
Deconstructing condition C reconstruction. Presented at the 47th Annual
Meeting of North East Linguistic Society (NELS 47), University of Massachusetts
Amherst.

Bhatt, Rajesh, and Roumyana Pancheva. 2004. Late merger of degree clauses.
Linguistic Inquiry 35:1–46.

Bruening, Benjamin, and Eman Al Khalaf. 2017. No argument-adjunct distinction
in reconstruction for binding condition C. Presented at 35th West Coast
Conference on Formal Linguistics, Calgary, Alberta.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view
from Building 20*, ed. Ken Hale and Jay Keyser, 1–52. Cambridge, Massachusetts:
MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond:
The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 104–131.
Oxford: Oxford University Press.

Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory essays
in honor of Jean Roger Vergnaud*, ed. R. Freidin, C. P. Otero, and M. L.
Zubizarreta. MIT Press.

Demirdache, Hamida. 2015. Resumption, epithets and trace conversion.
presentation at 25th Colloquium on Generative Grammar, IKER, Bayonne.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement.
Linguistic Inquiry 33:63–96.

Fox, Danny. 2017. Embedded late merge and the theory of displacement. Available at http://ling.auf.net/lingbuzz/003429.

Fox, Danny, and Jon Nissenbaum. 1999. Extraposition and scope: a case for overt QR. In *Proceedings of the West Coast Conference on Formal Linguistics 18*, ed. Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest, 132–144. Somerville, Massachusetts: Cascadilla Press.

Fox, Danny, and David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31:1–46.

Graf, Thomas. 2014. Late merge as lowering movement in Minimalist grammars. In *LACL 2014*, ed. Nicholas Asher and Sergei Soloviev, volume 8535 of *Lecture Notes in Computer Science*, 107–121. Heidelberg: Springer. URL https://doi.org/10.1007/978-3-662-43742-1_9.

Heck, Fabian. 2009. On certain properties of Pied-Piping. *Linguistic Inquiry* 40:75–111.

van Hoof, Hanneke. 2006. Split topicalization. In *The blackwell companion to syntax,*, ed. Martin Everaert and Henk Van Riemsdijk, volume IV, chapter 62, 410–465. Wiley and Sons.

Hulsey, Sarah, and Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics* 14:111–137.

Johnson, Kyle. 2016. Towards a multidominant theory of movement. Available at http://people.umass.edu/kbj/homepage/index_johnson.htm.

Kobele, Gregory, and Jens Michaelis. 2012. Cotags and acgs. In *Logical aspects of*

computational linguistics,, ed. Denis Béchet and Alexandre Dikovsky, volume volume 7351 of Lecture Notes in Computer Science, 119–134. Springer, Berlin.

Kobele, Gregory Michael. 2006. Generating copies: An investigation into structural identity in language and grammar. Doctoral Dissertation, University of California, Los Angeles, Los Angeles, CA.

Lebeaux, David. 1991. Relative clauses, licensing and the nature of derivations. In *Phrase structure, heads and licensing*, ed. Susan Rothstein and Margaret Speas, Syntax and Semantics 25. Academic Press.

Lebeaux, David. 2009. *Where does the binding theory apply?*. Linguistic Inquiry Monographs. MIT Press.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In *Proceedings, logical aspects of computational linguistics, LACL'98*, 179–198. Springer, New York.

van Riemsdijk, Henk. 1989. Movement and regeneration. In *Dialectal variation and the theory of grammar*, ed. Paola Benincà, 105–136. Dordrecht: Foris Publications.

Sauerland, Uli. 1998. On the making and meaning of chains. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Sportiche, Dominique. 2016. Neglect. Available at http://ling.auf.net/lingbuzz/002775.

Sportiche, Dominique. 2017. Relative clauses: Promotion only, in steps. Available at http://ling.auf.net/lingbuzz/003444.

Stanton, Juliet. 2016. Wholesale late merger in Ā-movement: Evidence from preposition stranding. *Linguistic Inquiry* 47.1:89–126.

Starke, Michal. 2001. Move dissolves into merge: A theory of locality. Doctoral Dissertation, University of Geneva. Available at http://ling.auf.net/lingbuzz/000002.

Tada, Hiraoki. 1993. A/A-bar partition in derivation. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Takahashi, Shoichi, and Sarah Hulsey. 2009. Wholesale late merger: Beyond the A/A′ distinction. *Linguistic Inquiry* 40:387–426.

Department of Linguistics, UCLA

dominique.sportiche@ucla.edu

Notes

[1]Even though they have been recently questioned, see Adger et al., 2016 or Bruening and Khalaf, 2017, such asymmetries are robust for many speakers (including me) in many languages, but not for all speakers (in any language?): this suggests the presence of at least one uncontrolled variable.

[2]The premises of this account have not gone unchallenged, see e.g. Sportiche (2016). If these premises are wrong, the particulars underlying this argument for Late Merger disappear but its logic remains (see e.g. Fox and Nissenbaum, 1999, Hulsey and Sauerland, 2006 or Fox, 2017.)

[3]Lebeaux also proposes that (the content of) complements can be late inserted under limited circumstances, a precursor to wholesale late merger discussed below.

[4] Or other such cases, for example the unbounded countercyclic cases of Late Merger motivated in Fox (2017) on the basis of extraposition.

[5]This conclusion (as well as the cases in Fox, 2017, cited in footnote 4) contradicts some claims in the literature, e.g. Tada, 1993, pp.63-70, or Sauerland, 1998, which argue that the depth of embedding at which Late Merger can apply is limited.

[6] By 'merged to' here and right above, I means that β projects in this case, which must be true if previously built structures can be added to but not changed: this is a weaker version of Chomsky's 2008 No Tampering Condition, which would independently bar late merger of any kind. By 'merged into' I mean that α is merged to a subconstituent of β.

[7]But if there is skepticism about this, the cases in (6) and (7-b) can be used.

[8]Technically, we have only shown that Late Merge can operate at any depth of a string of complements. To eliminate a potential objection, the point about island violations can be made with an (unbounded) sequence of wh-islands, that is, a nested

string of complements, which are strong islands for adjuncts. This said, allowing late merge into complements but not into adjuncts does not seem to be an option worth exploring.

[9]For Fox, 2017, this step is necessary to allow linearization, so this is in effect a subcase of the previous discussion.

[10]Essentially, and ironically, Graf, 2014, proposes to encode Late Merger in a computationally more parsimonious way by in effect making Agree bidirectional, thus allowing downward movement.

[11]Of course, (9-a) must be excluded. If Pied piping is possible why is it out? I argue elsewhere (Sportiche, 2017) that all relatives involve promotion of the head and consequently that such cases are excluded by standard constraints on movement - here the coordinate structure constraint (thus reducing (part of) pied piping theory to movement theory, unlike e.g. Heck, 2009, who reduces it to Agree theory).

[12]Although I will not discuss this in detail here, late merger as normally conceived which allows unboundedly countercyclic applications, is also incompatible with models using multidominance to represent movement. Indeed, in such models, movement is standardly represented by assigning two distinct mothers M1 and M2, one c-commanding the other, to a phrase P which is interpreted as having moved from M1 to M2 (but see Johnson, 2016, for an alternative). Late Merge would require P to strictly include a subconstituent at M1 without containing it at M. This is not possible: a given phrase cannot both include and not include a particular subconstituent.

[13]See also Graf, 2014, which is primarily concerned about generative capacity, and which explicitly recasts late merge as downward movement in a different formal system than standard Minimalist Grammars, but in a way that is computationally parsimonious. Such an approach could in principle be enriched to prevent the kind of overgeneration we discuss, and may yield another alternative.