Head movement, phrasal movement, and clitic doubling: 
towards a principled typology of locality conditions
Omer Preminger∗
University of Maryland

1. Empirical and theoretical backdrop

1.1. Head movement & phrasal movement: a tale of 
(near-)complementarity

Head Movement:

• The paradigm cases of head movement are maximally local
  ○ involving movement of a head to an immediately c-commanding 
    position—e.g.:

(1)  

(2) a. Have you \(t_1\) been watching?
    b. * Been\(t_1\) you have \(t_1\) watching?
  ○ this is Travis’ (1984) Head Movement Constraint

• Matushansky (2006): “head movement” = simple, plain-vanilla movement of a non-branching node + m-merger

(3)  

○ on Matushanksy’s account, head movement is maximally local 
  because m-merger is contingent on c-selection
  (more on this shortly)

Phrasal Movement:

• In contrast to head movement, phrasal movement looks like it cannot be maximally local:

(4) a. CP

(5)  

⇒ There is thus an apparent complementarity between the locality 
  conditions that apply to head movement, and those that apply to 
  phrasal movement.

cf.:

As we will see, the head-movement portion of Pesetsky & Torrego’s (5)— 
which is essentially Travis’ HMC—is something of an idealization;

∗A heartfelt thank you to Vera Gribanova, Boris Harizanov, Norbert Hornstein, Ruth Kramer, David Pesetsky, and Maria Polinsky, for comments and helpful discussion. All errors are my own.

There are cases of head movement that are not maximally local
- including (but not limited to):
  - many instances of clitic doubling / cliticization

One reaction to this is to toss out the baby with the bathwater, and simply abandon anything like Pesetsky & Torrego’s (5), or the insights it reflects;

Can we do better?

1.2. Clitic doubling

The term clitic doubling refers to the appearance of a reduced, pronoun-like morpheme whose features are determined by the features of a full lexical noun phrase elsewhere in the structure
- for example:

(6)  
\[ \text{Guraso-e-k parent(s)-ARTpl-ERG} \ni-ri \text{belarritako} \text{me-DAT} \text{earring(s)} \]  
\[ \text{ederr-ak beautiful-ARTpl(ABS)} \text{ bought} \]  
\[ \text{d-i-zki-da-te.} \text{(Basque)} \]  
(7)  
\[ \text{La} \text{i-oían listen.} \text{p.sc/a.sc/s.sc} \]  
\[ \text{3pl} \text{.3pl [a} \text{alguien que los ayudara].} \text{c.sc/l.sc} \]  
\[ \text{(*La}) \text{buscaban search.past.3pl A somebody comp cl.pl help.sbjunct} \text{‘They were looking for somebody who could help them.’} \text{P orteño Spanish; Suñer 1988]} \]

- it cannot be a matter of the clitic-doubling operation itself caring about these nominal properties (cf. (6))
- instead, these nominal properties regulate movement of the full noun phrase into a position from which clitic doubling is then obligatory (see, e.g., Diesing 1992, Sportiche 1998, Merchant 2006)
  - and note that Object-Shift is well known to be sensitive to exactly such factors (Diesing 1997, Diesing & Jelinek 1993, a.o.)

II. The doubled noun phrase behaves, for the purposes of locality, like an A-movement trace (Anagnostopoulou 2003)
- for example: as shown by Elordieta (2001), a.o., the dat argument in Basque ditransitives is structurally higher than its abs co-argument

(8)  
\[ \text{DP_{DAT}} \]  
\[ \ldots \]  
\[ \text{... ... ... ...} \]  
\[ \text{DP_{ABS}} \]  
consequently:
- a dat argument that has not been clitic-doubled intervenes in agreement targeting the abs (Preminger 2009);
- but in (6), the clitic-doubled dat ni-ri (“me-DAT”) does not intervene in (number-)agreement between the auxiliary and the abs belarritako ederr-ak (“beautiful-ARTpl(ABS)”) —
  - note the plural absolutive morpheme -zki on the auxiliary
  - because this dative argument has been clitic-doubled
  - note the 1sg.dat clitic -da- on the auxiliary

III. Working assumption:
There is no such thing as (syntactic) ‘cliticization’—only clitic doubling, what looks like cliticization is simply clitic doubling of pro.

---

2For extensive argumentation that these are indeed instances of clitic doubling, see Arregi & Nevins (2008, 2012), as well as Preminger 2009.
The locality conditions on clitic doubling are somewhere in between what we saw with head movement and phrasal movement.

Specifically, the clitic and the full noun phrase must be *clause-mates*

- where the extension of ‘clause’ is sensitive to restructuring (a.k.a. ‘clause-union’)

- The locality conditions for clitic doubling are thus slightly more stringent than for “pure” $\phi$-agreement (see, e.g., Preminger 2009)

- My goal for today: try to develop a unified theory for the locality conditions that these three types of dependencies exhibit (head movement, phrasal movement, clitic doubling)

2. Cracks in the façade

There are some interesting problems lurking in the picture sketched in §1:

I. Given current, feature-relativized conceptions of minimality, the Head Movement Constraint should not exist (!)

- there should be nothing wrong with with head movement skipping a featurally-irrelevant head en route to a featurally-relevant one:

\[
\begin{align*}
&X^0 \quad \cdots \\
&f \in \text{feat}[X^0] \\
&Y^0 \\
&f \not\in \text{feat}[Y^0] \\
&Z^0 \quad \cdots \\
&f \in \text{feat}[Z^0]
\end{align*}
\]

- see, e.g., the discussion in Roberts 2010:179ff.

II. If Matushansky is right—and what we’ve called “head movement” is decomposable into 2 steps, the first of which is simple movement—then:

- in cases where the moving head doesn’t project in its base position, the movement in question looks like it violates Anti-Locality

\[
\begin{align*}
&vP \\
&\downarrow^0 \\
&\sqrt/ V \\
&\Rightarrow \\
&m-merger \\
&t \sqrt/ V
\end{align*}
\]

III. Long-distance head movement

- assuming that clitic doubling is an instance of head movement:

\[
\begin{align*}
&\text{clitic doubling quite straightforwardly violates the HMC} \\
&\text{that’s because clitic doubling probably never involves movement only as far as the immediately c-commanding head} \\
&\text{to see this, consider what it would look like if clitic doubling did comply with the HMC} \\
&\text{we would have expected the constituent structure in (11):}
\end{align*}
\]

---


(11) … \{auxiliary/TAM, \{(transitivity/voice), \{clitic, \overline{\{V\}^{\chi}}\}\}\} …

○ whereas what we actually find looks like (12):

(12) [L’as]-tu fait?
    [cl.-have]-you done
    ‘Have you done it?’

⇒ clitic doubling qua head movement is movement of D at least as far as \(v\), and often further still (\(T^0/Asp^0\))

• irrespective of clitic doubling, long-distance head movement can be shown to exist in, e.g., Breton:

(13) a. Lenn a ra Anna al levr.
    read-INF-PRT does Anna the book
    ‘Anna reads the book.’

b. Lennet en deus Anna al levr.
    read-PPRT has Anna the book
    ‘Anna has read the book.’

(14)

\[
\begin{array}{c}
\text{C} \\
\text{lennet} \\
\text{read-PPRT}
\end{array}
\begin{array}{c}
\text{TP} \\
\text{en deus} \\
\text{has} \\
\text{DP} \\
\text{Anna} \\
\text{vP} \\
\text{v} \\
\text{V} \\
\text{VP} \\
\text{DP} \\
\text{al levr}
\end{array}
\]

\[\text{Roberts 2010:194ff.}\]

\[\text{see Borsley & Kathol (2000), Borsley et al. (1996), Jouittea (2005), Roberts (2004, 2010), Schafer (1994) and Stephens (1982)}\]

\[\text{for discussion and, crucially, for arguments that this is indeed long head movement (rather than, say, remnant VP-fronting)}\]

Our current desideratum, then, looks something like this…

WANTED:
A theory that derives the complementarity between the locality conditions on head movement and on phrasal movement—and which includes some sort of “loophole” allowing head movement, in some instance, to be less-than-maximally local.

3. Structural conditions ≠ surface generalizations: head movement as the preferred mode of movement

• In this section, I will suggest the following view of syntactic movement:
  ○ the grammar always “wants” to do head movement;
  ○ but this is seldom possible
  ⇒ resulting in the apparent ubiquity of phrasal movement.

• Accepting this picture requires that we, as theoreticians, make a clear distinction between:
  (i) which surface configurations are more common than others
  (ii) which structural conditions are implicated in the proper characterization of a given linguistic property/process/phenomenon

• This is not really news; the importance of this distinction has been recognized in linguistic theory at least as far back as Kiparsky 1973

• Here’s an example from syntax:
  ○ in very few languages—if any—does the external argument stay literally as low as its theta position (say, [Spec,vP])
  ○ but, as we know, this could only be taken as an argument against the Predicate-Internal Subject Hypothesis\(^5\) if we conflated (i) with (ii)

• Similarly, the claim here is that head movement is the grammatically preferred mode of syntactic movement (in a sense to be made explicit)
  – even if it turns out that, on the surface, phrasal movement seems more common than head movement

In its abstract form, this is a type of setup that I’ve been exploring for a while now —

- e.g. in Preminger 2014 (216ff.), I proposed that the grammar “tries” to Object-Shift every [+specific] DP out of the verb phrase
  - but that the success of these attempts is subject to familiar constraints
    - e.g. superiority; further constraints on DP type; and, where applicable, Holmberg’s Generalization
  ⇒ meaning that very few of these attempts actually succeed, and result in movement

- and similarly, the grammar “tries” to move every [+definite] DP (and, in some languages, every ext arg) to subject position
  - but that the success of these attempts is subject to language-specific constraints on what can actually move to subject position
    - in, e.g., English, Hebrew:
      whichever noun phrase has been targeted by φ-agreement
    - in, e.g., Icelandic, Basque:
      the closest DP (irrespective of φ-agreement)
  ⇒ meaning, again, that very few of these attempts actually succeed and give rise to movement

4. Proposal

(15) **Minimal Remerge**

If some projection of X moves, it must be X⁰ / X_min — unless such a derivation is unavailable.

- Informally, we can think of (15) as “move as little material as possible”
  - see §6, as well as Donati (2006:29–30)

⇒ Under what circumstances is head movement unavailable?
  - recall, from §2, that the answer cannot be: “when it is ruled out by the HMC”

Instead, I propose:
- **Bare Phrase Structure**⁷ + iterative downward search⁸ conspire to yield an A-over-A-like effect
  ⇒ ruling out most—but not all!—instances of head movement
  (for related proposals, see Hornstein 2009:72–74, Roberts 2010:33–40)

- Here’s how it works:
  - Bare Phrase Structure tells us that the label of the entire phrase (or “XP”) is nothing but the head itself
    - in other words, the distinction between X⁰ / X_min and XP / X_max can only be defined relationally—and crucially, not featurally
  - if movement is viewed from the perspective of the attractor (or in earlier versions of the theory, the perspective of the landing site) —
    - an iterative downward search, looking for a node to move (or remerge), will encounter the phrasal node first
  - and because the two are literally one and the same object:
    - there is no possible (featural) search criterion that would result in skipping of the phrasal label in favor of the head

⇒ Okay; but then, how do we ever get head movement…?

(16) **Principle of Minimal Compliance**: original version

[Richards 1998:601; see also Richards 2001]

For any dependency D that obeys constraint C, any elements that are relevant for determining whether D obeys C can be ignored for the rest of the derivation for purposes of determining whether any other dependency D’ obeys C.

(17) **Principle of Minimal Compliance**: my (bastardized?) version

Once a probe P has entered into a syntactic relation with a goal G, any locality constraints satisfied by that relation are thereupon irrelevant to subsequent relations between P and anything dominated or c-commanded by G (= anything dominated by the mother of G).

---

⁶There is also a certain affinity (in spirit, though, crucially, not in technical detail) between (15) and Chomsky’s (1995:262ff.) suggestion that—all else being equal—feature-movement is the preferred mode of syntactic movement.


• The PMC is necessary to account for locality effects involving, e.g., multiple-wh questions:

(18) a. * [Which book]k did the journalist spread the rumor that the senator wanted to ban tk?

b. ? [Which journalist]j [which book]k spread the rumor that the senator wanted to ban tk?

◦ looking at (18a–b), one might be tempted to proffer an explanation along the following lines —

— “(18b) is good because, in that example, the wh-phrase in the island doesn’t move out”

⇒ what Richards (1998, 2001) shows is that such an explanation is, at best, insufficient:

(19) a. * [Koja kniga]k razprostranjašes [mâlvata če which book spread the-journalist the-rumor that senatorât iska da zabrani tk]? (Bulgarian)

‘Which book did the journalist spread the rumor that the senator wanted to ban?’

b. ? [Koj žurnalist]j [koja kniga]k [tj razprostranjaše [mâlvata which journalist which book spread the-rumor če senatorât iska da zabrani tj]?

that the-senator wanted to ban

‘Which journalist spread the rumor that the senator wanted to ban which book?’

◦ once one wh-phrase has moved to [Spec,CP] of the interrogative clause, other potential movers to the same position are exempt from Subjacency-like conditions

[And yes, this pretty much means that attempts to derive syntactic islands from, e.g., (multiple-)spellout, are off track as well.]

⇒ And given Minimal Remerge + the PMC:

○ the first relation between a head H0 and some projection of X can only target the maximal projection of X (i.e., XP);

⇒ but subsequent relations a head H and some projection of X should be free from this constraint.

5. Deriving Anti-Locality from Minimal Remerge

Remember Anti-Locality? We can now derive it from the theory of (head-)movement just proposed.

• You might (justifiably) ask:

– Haven’t others already done this? Why do we need a new way to derive something that we can already derive by other means?

○ one reason is that you may find existing accounts of Anti-Locality unsatisfying (depending on who ‘you’ is)

⇒ but the more important reason is the relation between Anti-Locality and the locality of head movement

– accounts of Anti-Locality that have nothing to say about the locality of head movement force us to say some second thing to derive the latter

(And in linguistics, of course, 1 > 2)

○ plus, you’ll recall (from §2) that Abels’ Anti-Locality, taken verbatim, stands in conflict with Matushansky’s theory of head movement

– and as we’ll see, we want to hold on to the latter, because it holds the key to understanding clitic doubling

So here goes.

• Let H0 be a movement attractor, and let X be the head of H0’s complement

⇒ it follows that H0 is in a c-selection relation with X(P)

(indicated here as a “wavy line”)

(20)
We therefore have in place a structural relation between H⁰ and X(P), which conforms to the aforementioned A-over-A-like condition.

**NB:** This point, concerning c-selection being implementable under pure sisterhood, was in fact one of the selling points of Bare Phrase Structure to begin with.

⇒ Given the PMC (17), subsequent relations between H and X are no longer subject to this locality condition

- meaning it is now possible for H to attract the head of X alone (21a)

⇒ And, crucially, in situations where both phrasal movement and head movement are in principle possible, Minimal Remerge (15) will rule out phrasal movement:

(21) a. 

```
    v
   /\v
  /  \v
 /    v
H ~~~ "XP"
```

⇒ And, crucially, in situations where both phrasal movement and head movement are in principle possible, Minimal Remerge (15) will rule out phrasal movement:

b. 

```
    v
   /\v
  /  \v
 /    v
  "X"  "X⁰"
```

⇒ And, crucially, in situations where both phrasal movement and head movement are in principle possible, Minimal Remerge (15) will rule out phrasal movement:

• Importantly, on this view, there is nothing intrinsically wrong with moving the complement of H per se to [Spec,HP]

• And, indeed, in the case where the complement of H is non-branching, Minimal Remerge (15) is trivially satisfied

⇒ meaning we expect no Anti-Locality effects in this case

(22) THE MATUSHANSKYAN EXCEPTION TO ANTI-LOCALITY

⇒ Prediction: movement triggered by a head H⁰, if it is not the first relation between H⁰ and the moving projection, should always be head movement.

**Q:** Doesn’t this predict that movement of an agreed-with subject to [Spec,TP] will be head movement rather than phrasal movement?

**A:** Yes… unless movement to subject position is not a relation established directly between T⁰ and the moving element!

- **Preminger 2014 (129ff.):** movement to subject position relies on ϕ-agreement to establish the identity of the mover⁹ (rather than probing for a mover directly)

  - crucially, this was argued on completely independent grounds, having nothing to do with head movement vs. phrasal movement

6. An interlude: OT vs. economy

• Now that we’ve seen a little bit of how Minimal Remerge, repeated here, would work —

(23) **MINIMAL REMERGE**

```
M/i.sc/n.sc/i.sc/m.sc/a.sc/l.sc R/e.sc/m.sc/e.sc/r.sc/g.sc/e.sc
```

If some projection of X moves, it must be X⁰/Xmin — unless such a derivation is unavailable.

[=(15)]

⇒ we might be tempted to ask:

- Is (23) not the epitome of a violable constraint? And if so, does that mean we have entered the domain of Optimality Theory?¹⁰

---

⁹One potential problem with this involves quirky-subject languages (which do not instantiate the dependency of movement on ϕ-agreement), in those cases where the mover happens to be the same DP that was targeted for agreement. On the current proposal, one would be forced to assume that movement to subject position, in such languages, lands in a position that is not the specifier of the very same head responsible for ϕ-agreement. Ideally, one would want to derive this as something more than an accident; see §7 for further discussion.

Answer:
- Something like (23) certainly can be modeled in OT;
- **However**, OT is fundamentally ill-suited to handle cases of ineffability
  - where, for a given input, there is simply no well-formed output
  [because, in OT, no output is intrinsically ill-formed; it is only rendered so by the existence of an even-more-optimal alternative based on the same input]
- And syntax is rife with such cases (e.g. PCC effects, many of the strong island effects).

.jet
- I take this to indicate that OT is, at its most fundamental level, the wrong formalism for representing syntactic competence
  ⇒ if we want to incorporate Minimal Remerge in to our theory of grammar, it is preferable to conceive of it as an *economy* condition:
    (see Donati 2006:29–30 for a closely related proposal)

(24) **MINIMAL REMERGE**  
*[economy version]*
If $X_{\text{min}}/X^0$ is movable, move only $X_{\text{min}}/X^0$.

- In cases where the PMC has been satisfied with respect to the *A-over-A*-like locality condition —
  ○ the head can be moved
    — and therefore, following (24), the head must be what moves;
- But in all other cases, what moves will be the maximal projection.

[What about (phrasal) pied-piping? Following Cable 2007, 2010, I adopt the view that there is no such thing.]

7. Interim summary, and methodological assessment
- Phrasal movement and head movement look [kind of] complementarity in their locality constraints (following P&T 2001, minus the “kind of” part)
- Existing accounts of phrasal Anti-Locality on the one hand, and of the locality of head movement on the other, each don’t have much to say about the other’s domain of application
  - There is a way to derive Anti-Locality, and (a certain version of) the locality of head-movement, from the same thing — namely:
    ○ a preference for **Minimal Remerge** +
      *iterative downward search* (yielding an *A-over-A*-like effect) +
      the **Principle of Minimal Compliance** (Richards 1998, 2001)

As things stand, however, we are in danger of “methodological leakage”:
- I don’t know of a working, restrictive theory of *possible formal feature*  
  ⇒ consequently, there is also no working, restrictive theory of *possible agreement relation*  
  ⇒ In practice, if we observed an instance of head movement that was not maximally local…
    (e.g. from $Z^0$ to $X^0$, where $ZP$ is c-commanded by, but is not the immediate complement of, $X^0$, as in (25))

(25)
```
  ``XP``
  
  ``Y``
  ``X``
  
  ``Y``
  ``Z``
  
  ``Z``
```

… we could always posit *some* unseen prior agreement relation between $X^0$ and $ZP$
  - satisfying the PMC w.r.t. the *A-over-A*-like locality condition
  - and enabling Minimal Remerge to exert its pressure, forcing subsequent movement to involve $Z^0$ alone (rather than $ZP$)

---

11Technically speaking, such cases can be handled in OT by adding ungrammaticality itself (a.k.a. the “null parse”) as a competing output. Such a move seems to me to be little more than an acknowledgement that OT is not suited for the relevant empirical domain (see Preminger 2014:170–174 for discussion).
Thus, we can’t (yet) generate concrete predictions about how local or non-local a particular instance of head movement will be.

Note, however, that the way we derived Anti-Locality (from Minimal Remerge, c-selection, and the PMC) is not susceptible to this concern.

However, if everything we’ve said so far is correct:
- it follows that long head movement —
  (head movement that is not maximally local / HMC-compliant)
  — is an “agreement detector”
- That’s because head movement of $Z^0$ to $X^0$ requires, given Minimal Remerge, a prior relation to be established between $X^0$ and $ZP$
  - if $ZP$ were in [Compl,X], c-selection would trivially fill that role;
  - but when $ZP$ is further away from $X^0$, c-selection won’t do the trick;
  $\Rightarrow$ if we nevertheless see head movement of $Z^0$ to $X^0$, it means (on this hypothesis) that $X^0$ has entered into some other relation with $ZP$.

Why is this important? i.e., why would we need an “agreement detector”? (Isn’t agreement universal?)
- e.g. even though English doesn’t look like it has object agreement, we are led to believe that it does—since it has ACC case—right?

8. When agreement really isn’t there
- As just alluded to, it is a common (mis)conception that underlyingly, all languages have the same syntactic infrastructure of agreement relations
  - and different languages express different parts of this infrastructure overtly (or none at all)
- And so it is very common in the syntactic literature to come across a statement more or less like the following:
  “Language $L$ has no morpho-phonologically expressed agreement with direct objects; but since there is ACC case on the object in $L$, we assume this is a reflex of agreement in $\varphi$-features between $v^0$ and the object.”

In this section, I will present an argument that this view is incorrect.

The Person Case Constraint (PCC):

    3.abs-v′sg.abs-1sg.dat-2sg.erg
    ‘You have sold the book to me.’

b. * Zuk harakin-ari ni saldu you.erg butcher-artsg.dat me(abs) sold [n-(a)i-φ-o-zu]
    1.abs-v′sg.abs-3sg.dat-2sg.erg
    ‘You have sold me to the butcher.’ [Laka 1996]

PCC-strong.
- 1st/2nd person direct object in the presence of an indirect object
- The PCC is a syntactic effect, not a morphological one
  - pace Bonet 1991, 1994, for example
- Evidence: (Albizu 1997, Rezac 2008)
  - Basque has two classes of 2-place unaccusatives
    - one class where DAT $\gg$ ABS, and one class where ABS $\gg$ DAT

(28) DAT $\gg$ ABS:

a. Kepa-ri bere buru-a gusta-tzen zako.
   Kepa-dat his head-artsg(abs) like-hab aux
   ‘Kepa likes himself.’

b. * Kepa bere buru-a-ri gustatzen zako.
   Kepa(abs) his head-artsg-dat liking aux
Head movement, phrasal movement, and clitic doubling: towards a principled typology of locality conditions

October 2015

Preminger

(29) \(\text{ABS} \gg \text{DAT} :\)

a. * Kepa-

\text{DAT} his \text{head}-\text{ART}_{sg}(\text{ABS}) \text{come-prog aux} ispilu-a-n.

\text{mitrof}-\text{ART}_{sg}(\text{ABS})-\text{loc}

\text{Intended: ‘Kepa is approaching himself in the mirror.’}

b. Miren-

\text{DAT} his/her \text{head}-\text{ART}_{sg}-\text{DAT talk-prt aux} mintzatu zaio.

\text{Miren}(\text{ABS}) his/her \text{head}-\text{ART}_{sg}-\text{DAT talk-prt aux} ‘Miren talked to herself.’  \text{[Rezac 2008:75; see also Elordieta 2001]}

\(\Rightarrow\) Crucially, only the \(\text{DAT} \gg \text{ABS}\) ones show the PCC:

(30) \(\text{DAT} \gg \text{ABS} :\)

a. Miren-

\text{DAT} sweet-\text{ART}_{pl}(\text{ABS}) like-\text{IMPF} \phi-zai-zki-o.

3.\text{ABS}-\sqrt{}-\text{pl. ABS}-3\text{sg.DAT}

‘Miren likes candy.’

b. */?? Ni Miren-

\text{DAT} like-\text{IMPF} 1.\text{ABS}-\sqrt{}-\text{sg. ABS}-3\text{sg.DAT}

‘Miren likes me.’

(31) \(\text{ABS} \gg \text{DAT} :\)

Ni Peru-

\text{DAT} approach 1.\text{ABS}-\sqrt{}-\text{sg. ABS}-3\text{sg.DAT}

‘I approached Peru.’  \text{[Albizu 1997:21, Rezac 2008:73]}

\(\Rightarrow\) this shows that the PCC is fundamentally syntactic:

– the morphological “target forms” in (30b) and in (31) are identical

– and the distinction is in the hierarchical organization of arguments

\(\text{NB: As far as I can tell, this also renders impossible any meaningful account of the PCC in terms of ‘grammaticalization’ or ‘usage’ (see, e.g., Haspelmath 2004).}\)

(32)

\(\Rightarrow\) But the PCC is notoriously absent in environments that do not show overt \(\phi\)-feature agreement of some kind…

\(\circ\) this is so crosslinguistically (i.e., no PCC in languages without internal-argument agreement):

(33) ha-menahel-et ta-cig

\text{the-manager-F fut.3sg.F-introduce DAT them acc.me ‘The manager will introduce me to them.’}

\(\circ\) but also intra-linguistically (even in a language with PCC effects, they go away in, e.g., non-agreeing infinitives):

(34) a. Zuk

\text{you.erg me.dat book-ART}_{sg}(\text{ABS}) sold d-i-\phi-da-zu.

3.\text{ABS}-\sqrt{}-\text{sg. ABS}-1\text{sg.DAT}-2\text{sg.ERG}

‘You have sold the book to me.’

b. * Zuk

\text{you.erg butcher-ART}_{sg}.\text{DAT me(ABS) sold n-(a)i-\phi-o-zu}

1.\text{ABS}-\sqrt{}-\text{sg. ABS}-3\text{sg.DAT}-2\text{sg.ERG}

‘You have sold me to the butcher.’  \text{[=(26a–b)]}

(35) Gaizki iruditzen

\text{wrong look-impf} 3.\text{ABS}-\sqrt{}-\text{sg. ABS}-1\text{sg.DAT you.erg me(ABS)}

\text{harakin_aryi saltzea ].}

\text{butcher-ART}_{sg}.\text{DAT sold-NMZ-ART}_{sg}(\text{ABS})

‘It seems wrong to me for you to sell me to the butcher.’  \text{[Laka 1996]}
If the PCC is syntactic (a result of agreement + dative intervention);
And it is absent wherever we don’t see overt agreement with the verb’s internal arguments:
⇒ There must be no \( \phi \)-feature agreement at all (not even agreement that just “happens” to be null) in those environments where the PCC is absent.
— in other words:
○ where you don’t see agreement, there is no agreement.

This moves the relevant variation from PF (“object agreement in Hebrew just happens to be phonologically null”) to the feature-composition of functional elements (no unvalued \( \phi \)-features on Hebrew \( v^0 \))

Important, this is a kind of variation that is very easy to acquire:
○ starting assumption: no unvalued \( \phi \)-features on any functional head;
○ revise only in light of overt morpho-phonological covariance between a noun phrase and something else.

9. But what about clitic doubling?
● There is, however, a problem with the view just espoused —
○ and it comes from the distinction between clitic doubling (remember?) and actual \( \phi \)-agreement

\( \phi \)-agreement is a valuation relation:
○ a head \( H^0 \) (the “probe”) enters the derivation with a need for \( (\phi-) \)feature values
○ the computational system is “impatient,” and so it scans the already existing structure for appropriate feature values
  – yielding the familiar c-command condition on valuation
○ upon finding such values on some XP (the “goal”), those values get transmitted\(^\text{15}\) to \( H^0 \)
● clitic doubling, on the other hand, is not a copying of \( \varphi \)-values per se;
● it is the creation or copying of a \( D^0 \)-like morpheme alongside an appropriate host
  ○ notwithstanding the fact that the created morpheme is \( \phi \)-feature-matched (in the usual case) to the doubled argument
● For more on similarities and differences between the two processes, see Anagnostopoulou (2006, to appear) and references therein.

Why is this a problem for our tidy little story about the PCC?
● Because empirically, when we say “languages / environments that show overt agreement with internal arguments” —
  ○ we need this to include languages / environments with clitic doubling of (rather than \( \varphi \)-feat. agr. with) internal arguments, too.

ATTEMPT #1: It’s because clitic doubling is \( \varphi \)-feature agreement
● Doesn’t work because…
  ○ well, clitic doubling is not \( \varphi \)-feature agreement\(^\text{16}\)
    – syntactically (Anagnostopoulou op. cit.; see also Preminger 2009)

ATTEMPT #2: It’s because clitic doubling is movement, and all movement is conditioned by a previous agreement relation
● Doesn’t work because…
  ○ contra Chomsky 2000, 2001 [“\( \text{mvmt} = \text{Int. Merge} = \text{Agree} + \text{EPP} \)”,
    movement and \( \varphi \)-agreement are in fact two-way dissociable
    – see Zaenen et al. (1985) and Bobaljik (2008), among many others
⇒ there is no a priori reason to believe that a given instance of movement is preceded by agreement; that needs to be argued for.

\(^{14}\)Recall that, as a working assumption, I take all instances of ’cliticization’ to be clitic doubling of pro (see §1.2).

\(^{15}\)I abstract away, here, from whether this ‘transmission’ amounts to copying (Chomsky 2000, 2001, a.o.), unification/feature-sharing (Frampton & Gutmann 2000, Gazdar et al. 1985, Pesetsky & Torrego 2007, Pollard & Sag 1994, a.o.), or something else entirely.

\(^{16}\)There are certainly diachronic connections between the two types of processes (cf. Gelderen 2011 and references therein); but we are interested here in the mental grammar, a notion that is necessarily synchronic.
ATTEMPT #3: It’s because clitic doubling is non-local movement of a head.

Since this movement is non-local, the probe ($v^0$) does not stand in a c-selection relation with the maximal projection of the goal (DP)

Absent some prior relation between the two, the proposal in §4 predicts that head movement would not be possible—only phrasal movement

⇒ Clitic doubling requires prior agreement between the probe ($v^0$) and the maximal projection of the goal (DP)

That’s why wherever one sees clitic doubling, the mechanisms of agreement, intervention, etc. are all active

○ and consequently, the PCC arises in those environments, as well.

This logic can obviously be utilized by us linguists; but it can also be utilized by the little language-acquirer…

If we’re assuming that clitic doubling and agreement are indeed different beasts, in the speakers’ synchronic grammar —

( and the work cited above leaves little choice but to assume this)

— then at some point in the course of language acquisition, the speaker must adjust their grammar to say:

“this instance of ‘agreement-like’ morphology is not $\varphi$-agreement; it’s a clitic”

⇒ Something must prevent the child from then assuming that it is just an instance of movement, sans any prior agreement relation

○ which, crucially, is a perfectly viable option in the general case

– cf. movement to subject position in Icelandic (Zaenen et al. 1985, Bobaljik 2008), also movement of a D projection.

If people ever ended up with such a grammar, we would have a language with $v^0$-level object clitics and no PCC —

○ but for the most part, that doesn’t seem possible

⇒ The framework espoused here (Minimal Remerge + locality + PMC) explains why.

10. Some remaining problems

10.1. Clitic doubling in Bulgarian

Harizanov (2014) shows that clitic doubling in Bulgarian behaves not like [head movement + m-merger], but like [phrasal movement + m-merger]

The “smoking gun” for that claim is this:

○ clitic doubling in Bulgarian can create new WCO violations w.r.t. material contained within the doubled noun phrase

(37) a. Ivan predstav[na vsjak[ækna]]k [njen[ajka]k badeš [sàpruj]]j minalatagodina
Ivan introduced to every woman her future husband last year
‘Ivan introduced to every woman$_k$ [her$_k$ future husband$_l$] last year.’

b. Ivan gô[3sgM] predstavi[na vsjak[ækna]]k [njen[ajka]k badeš [sàpruj]]j
Ivan po,3sgM introduced to every woman her future husband minalata godina
last year
‘Ivan introduced to every woman$_k$ [her$_k$ future husband$_l$] last year.’

(38) a. Ivan predstav[vsjak[ækna]]k [na nejnj[a]k badeš [sàpruj]]j minalatagodina
Ivan introduced every woman to her future husband last year
‘Ivan introduced every woman$_k$ [to her$_k$ future husband$_l$] last year.’

b. Ivan [mu] predstavi[vsjak[ækna]]k [na nejnj[a]k badeš [sàpruj]]j
Ivan io,3sgM introduced every woman to her future husband minalata godina
last year
‘Ivan introduced every woman$_k$ [to her$_k$ future husband$_l$] last year.’

To the best of my knowledge, this particular state of affairs does not arise in other clitic-doubling languages.¹⁷

¹⁷ Thanks to Karlos Arregi for discussion of this.
• These Bulgarian data are incompatible with, e.g., “Big DP” analyses of clitic doubling
   ○ in fact, they are incompatible with anything but a phrasal movement account of clitic doubling

**NB:** Sportiche (1996, 1998) proposed such an account, but that particular account
is incompatible with the facts we’ve already seen
   – in Basque, a clitic-doubled nominal is not an intervener, nor is the clitic itself (or its specifier) an intervener
   ⇒ Sportiche’s account, as formulated—essentially, covert movement of the nominal to [Spec,CliticP]—doesn’t work.

**Harizanov (2014):** clitic doubling in Bulgarian is run-of-the-mill phrasal movement + m-merger at the landing site
   ○ i.e., m-merger of a complete, branching, phrasal node into a structurally-adjacent head

   ![Diagram](39)

   ○ if you are familiar with Matushansky 2006 but not with Harizanov 2014, this should look weird
      – since, for Matushansky, m-merger was only even *defined* for two non-branching nodes in an immediate c-command relation
   ○ for Harizanov, the computational system basically “does the best it can” with this state of affairs —
      – it reduces the branching DP to its label (D^0), and then performs m-merger between the label and the host (X^0)

   ![Diagram](39)

   ⇒ why is this a problem for the current proposal?

• Now, so far, you might say “okay, so Bulgarian has phrasal movement where other languages have head movement” —

**• Because Bulgarian *does* exhibit the PCC (Harizanov 2014, Migdalski 2006, a.o.); consequently:**
   ○ if the PCC arises via the mechanisms of agreement and dative intervention;
   ○ and doesn’t arise in scenarios where those mechanisms are not active;
   ⇒ then the relevant mechanisms must be active in Bulgarian.

• If that’s so; and if, in the absence of actual φ-agreement, it is non-local head movement that clues in the language acquirer that there must have been agreement at play —
   ○ *how can the child acquiring Bulgarian figure out that their language has internal-argument agreement and, therefore, that it has the PCC?*

Some speculation:

• It is not clear to me (and at least as far as the 2014 paper is concerned, to Harizanov either) how the acquirer of Bulgarian figures out that clitic doubling in their language is fed by phrasal (rather than head) movement
   ○ what I’m fairly confident in saying is that it’s not based on the amelioration of WCO effects under clitic doubling

• Recall that the acquisition story for unvalued φ-features was:
   ○ you start by not positing them;
   ○ and you only posit them in the face of compelling evidence, such as:
      – morpho-phonologically expressed covariance between a noun phrase and something else
      – non-local head movement of D

**• Bulgarian, at first glance, would certainly *look* like a language that had unvalued φ-features on H^0**

• Suppose that once you posit unvalued φ-features on H^0, there’s no “turning back”;

- 13 -
If whatever it is that clues in the acquirer that Bulgarian is different, i.e., that it uses phrasal movement to feed clitic doubling, is rather subtle (=late) —
  ◦ it might be too late to ‘revoke’ the unvalued $\varphi$-features on $v^0$
    ⇒ you still get the PCC

10.2. The locality of clitic doubling

- As noted in §1.2, the locality conditions on clitic doubling are more stringent than those on phrasal (A-)movement
  ◦ the clitic and the full noun phrase must be clause-mates

    Why would that be?
  ◦ after all, once $\varphi$-agreement has “unlocked” the DP vis-à-vis
    the Principle of Minimal Compliance —
      the $D^0$ head should be able to move just as far as any other
      projection of D can
  - More accurately:
    ◦ since the attractor has to first enter into an agreement relation with
      the DP, the locality conditions should be the intersection (=minimum)
      between \text{LOC-COND}[A-mvmt] and \text{LOC-COND}[\varphi-agr]
    ◦ which amounts to, roughly:
      anything up to a DP, PP, or finite CP boundary
    ⇒ so why does clitic doubling have to be more local than that?

Some speculation:

- Clitic doubling behaves, in various ways, like pronominalization
- The mechanism relating the clitic to the full noun phrase, in syntax, is movement
- Suppose that, alongside this, the clitic is subject to a semantic condition
  ◦ requiring that the clitic stand in a binding-theoretically sanctioned
    coindexation relation with the full noun phrase

And whereas syntax trades in things like DPs, PPs, or finite CPs —
  ◦ semantics trades in things like ‘complete predication domain’
  ◦ This may give rise to the clausemate condition (or something close to it).

11. Conclusion

- At first approximation, the locality restrictions on phrasal movement and on head movement stand in a complementary relation to one another
  ◦ \text{Anti-Locality} vs. the Head Movement Constraint
    (Abels 2003 and Travis 1984, respectively)
- However, there are exceptions, involving head movement that is not maximally local:
  ◦ most (if not all) instances of clitic doubling / cliticization
- I presented a proposal on how this picture can be derived, based on the following premises:
  ◦ Minimal Remerge (“if $X_{\min}/X^0$ is movable, move only $X_{\min}/X^0$”)
  ◦ locality (the $A$-over-$A$-like condition; Hornstein 2009, Roberts 2010)
  ◦ the Principle of Minimal Compliance (Richards 1998, 2001)
- I showed that this proposal derives Anti-Locality, as well as the Head Movement Constraint in the usual case
- While allowing head movement to be less-than-maximally-local under particular circumstances
  ◦ specifically, when a previous relationship has already been established
    between the attractor and the head’s maximal projection
- I then showed how this solves a nagging problem concerning the PCC
  ◦ and its relation to the presence/absence of unvalued $\varphi$-features on $v^0$
- Finally, I discussed a few challenges that the proposal faces, and offered some speculations on how they might be addressed:
  ◦ the underlyingly phrasal nature of clitic doubling in Bulgarian
    (Harizanov 2014)
  ◦ the more stringent locality conditions on clitic doubling, when compared with agreement and (A-)movement

$\text{Recall: the extension of ‘clause’ here is sensitive to restructuring/‘clause-union’; see fn. 3.}$
References


Head movement, phrasal movement, and clitic doubling: towards a principled typology of locality conditions

- 16 -


