Head movement is an artefact of optimal solutions to linearization paradoxes[*]

Mark de Vos

Department of English Language and Linguistics, Rhodes University, South Africa
Email: m.devos@ru.ac.za

Abstract
Head movement, while endemic in natural languages, has long been a thorn in the sides of syntacticians as it does not seem to be logically necessary nor does it follow from first principles. I will argue that head movement is not only necessary – it is indispensable. It is an intrinsic part of the language-computational system. Converting two-dimensional "trees" into uni-dimensional linearizations is mathematically difficult and in doing so, losing information is a distinct possibility. If too much information is lost then it would prove difficult for a hearer or a child acquiring the language to infer the original syntactic information from the signal and the system would become unlearnable. Linearization is the strategy of choosing an optimal ordering and head movement is a logical response to an optimization puzzle.

Keywords: Linearization, head movement, optimality, LCA, PF interface

1. Introduction

In mathematical inquiry, mathematicians start with a set of axioms and then set about exploring whether the resulting worlds have interesting properties. For instance, one might explore a world where parallel lines never meet or one where parallel lines do. The results are Euclidian and non-Euclidian geometries respectively. Similarly, the field of topology is dedicated to the at-first-glance obviously nonsensical proposition that all surfaces are infinitely malleable and that discrete objects are defined by the number of holes they incorporate. Thus, a lemniscate shape could be either a figure of eight or, if you massage it enough, a teapot. In the same spirit, this paper will outline a few very basic axioms defining a syntactic world which, if you care to explore it, will turn out to require head movement as an inalienable property. The *effect* of head movement will follow from a general linearization algorithm which, in turn, is motivated by the need for the syntactic module to be able to communicate with the modules at the PF interface.

In section 2, I will briefly describe verbal head movement before defining an abstract system of linearization in section 3. This will enable me to demonstrate how head movement (i) creates more optimal linearization possibilities, and (ii) reduces the total number of linearization outcomes, thus making computation more tractable. Part II of this paper is logically separate and demonstrates how this system can be applied to a paradigm example of verbal head movement, namely French V-T movement. It will also be demonstrated that English T-lowering and do-support, and how these interact with negation, follow from the same analysis.

Part I

2.   What is head movement?

Head movement occurs when a head appears to be displaced from its base position. This is usually most visible in relation to some other category, such as an adverb or a subject. Thus in example (1a), a modal verb *must* appears right-adjacent to the subject *you*, and is assumed to be located in the head of T^0. In (1b), once the question transformation has been applied, the same modal appears left-adjacent to the subject and is now located in C^0. This is an example of T-C head movement.

(1)   a. You must have eaten
        b. Must you have eaten?
        c. *Have you must eaten?

Head movement is constrained by the Head Movement Constraint (HMC; Travis 1984). The HMC requires that a moved head may only move to the head position that immediately dominates it, implying that a moved head may not skip any intervening, dominating head. The only way a head may move past a dominating head is by first adjoining to it and then pied-piping the complex head thus created to the next head position[1]. As originally stated, the HMC captured an important insight, namely that a head can only move to a head which selects it[2]. Thus, it is not possible to move the auxiliary *have* across the intervening modal (1c).

Head movement is endemic to natural language and, in a real sense, is an essential part of the descriptive apparatus of grammars. Example (2) illustrates "short" verb-raising from V^0 to v^0, where the lexical verb moves across the indirect object. In Romance languages, lexical verbs raise from V^0 to v^0 to T^0 across adverbials and markers of negation, as in (3).

(2)   a. I will give Mary a book
        b. *I will Mary give a book

(3)   a. *Je mange quelque fois des escargots*
            I eat.1SG sometimes of.DET.PL snails
            'I sometimes eat snails' (French)
        b. **Je quelque fois mange des escargots*

Nevertheless, there are a number of well-documented problems which are raised by head movement. For instance, in Baker (1988), head movement by adjoining to a higher head does

---

[1] I do not assume any kind of excorporation of head-moved constituents.
[2] An X^0 may only move to a Y^0 which properly governs it.

not necessarily extend the tree upwards (but contra Matushansky (2006)); it thus violates the Extension Condition. Similarly, in Baker's system, since it adjoins to an existing head, it does not C-command its trace. This suggests very strongly that head movement is a different type of movement to A and A-bar movement.

In response, grammarians have tried several strategies to make head movement more amenable to understanding. One type of response is to argue that head movement is not exceptional at all but that it is a predictable member of the typology of movement, and to argue that it adjoins directly to the root of the tree, thus eliminating the extension and C-command problems (Matushansky 2005, 2006). Another group of responses argues that head movement does not exist and that putative head movement effects are derived by remnant movement of XPs (Hinterhölzl 2006, Koopman 2000, Mahajan 2000, Müller 2004). Others have argued that it is essentially a PF effect (Boeckx and Stjepanovic 2001; Chomsky 1995a, 2001; Harley 2004), although others have pointed to the fact that it can have semantic effects as evidence against a PF approach (Lechner 2005) and that it is LF driven (Ackema and Čamdžić 2003).

As each of these arguments waxes and wanes, the related question of the trigger for head movement has its own ebb and flow. Given its clearly parametric nature, it is difficult to turn to universal principles. A long tradition ascribes the trigger for head movement to "strong" morphology (Rohrbacher 1999) although the exact mechanism has remained elusive and the generalization itself is obviously weakened by the fact that languages with no verbal inflectional morphology, such as Afrikaans, exhibit head movement phenomena such as verb-second. Others have rejected the link between morphology and head movement (Alexiadiou and Fanselow 2000, Bobaljik 2002). One is thus left with the impression that the status and trigger of head movement remain elusive[3].

In this paper, I will show that, under certain assumptions, linearization of syntactic structures leads to paradoxes which can be resolved by creating complex feature bundles. In a sense, this sidesteps the question of whether head movement applies in the syntax or post-syntactically. It will just turn out that when structures are passed to the PF interface, the linearization requirements of that interface will be better served by a representation including complex feature bundles, no matter where they are derived. Thus, head movement follows as an inevitable consequence of linearizing a syntactic structure constrained by morphological resources within a given language.

3.   Assumptions about syntactic relations and linearization

In this section, I will outline some assumptions about syntactic relations and how they could be linearized.

3.1 Syntactic relations

I embed my analysis within Minimalist syntax (Chomsky 1991, 1993 *et seq.*) and within the Normalization/Relational vision of Minimalism outlined in De Vos (2008, 2013, 2014). I take it for granted that a syntactic structure consists of various relationships between features, words,

---

[3] See Roberts (m.s.) and Harley (2004) for excellent overviews.

phrases, etc. These include selection, Case marking, checking/deletion of ϕ features, etc. These are instantiated by only three mechanisms, namely MERGE, AGREE and MOVE[4].
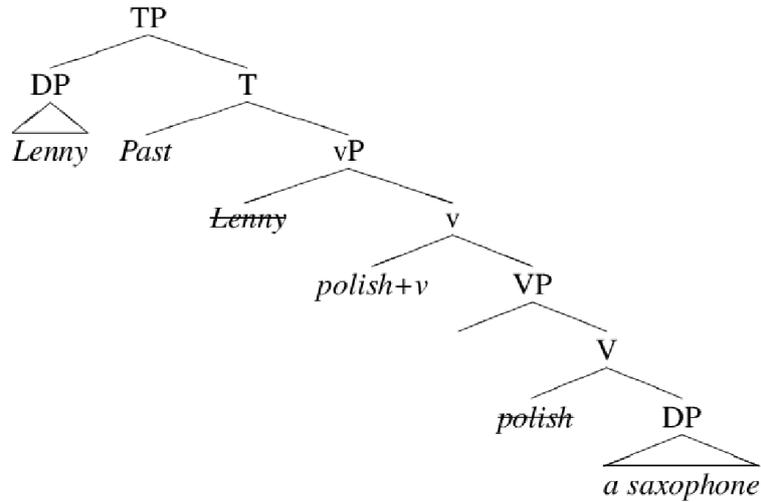
By the term "syntactic relation", I mean an unambiguous, pairwise relation, instantiated in narrow syntax, between a pair of features (or feature bundles), where one feature is the antecedent and one feature is a dependent. Furthermore, along with Chomsky (1995a) and many others, I take this to be a partial order of the form $\{p,\{p,q\}\}$ (see also Cornell 1996, Fortuny 2008, Halmos 1960, Kayne 1994, Kracht 2003, Langendoen 2003, Uriagereka 1999, Zwart 2009). This includes selection (instantiated by MERGE) and feature checking/agreement (instantiated by AGREE). I also assume that "syntactic relations" exclude semantic co-reference, variable binding, quantifier raising, polarity licensing, etc. until evidence emerges that derives them from MERGE, MOVE or AGREE[5]. I also assume that C-command is a derivational relationship, i.e. it is a function of hierarchy (Epstein and Seely 2006, Seely 2006) and is encoded, by definition, into AGREE and MOVE. Thus, while many constituents in a phrase structure are in some type of C-command relationship, it is not the case that all of them are in checking relationships; however, all constituents which are in checking relationships instantiated by AGREE and MOVE are also in C-command relationships[6].

---

[4] I take S-selection, C-selection and the selection of an appropriate theta argument to be included under the sobriquet of selection since they are all treated equally under the analysis proposed.

[5] I remain agnostic about whether these relations can be reduced to partial orders. First, it seems to me that these types of relations are very semantic in nature and thus it is not clear whether they are instantiated by MERGE, MOVE or AGREE. Secondly, they can also be inferred from C-command (or Armstrong's axioms (1974)) and do not necessarily entail a derivational relationship between the two categories. I leave it to further research to determine whether these are partial orders.

[6] In addition to these relations, there are conceivably many other possible relations that can be defined over a phrase-structure marker, with the most notable being hierarchical C-command relations. An anonymous reviewer queries whether C-command relationships should not be included as candidates for linearization. By and large, fundamental syntactic relationships all refer to features of some sort, e.g. selectional features, interpretable or uninterpretable features. These are all instantiated by the fundamental operations of MERGE, MOVE and AGREE. C-command emerges as a property of phrase structure as a result of these operations being applied (Epstein and Seely 2006, Seely 2006). In other words, C-command is already part of the definitions of MERGE, MOVE and AGREE. Consequently, if X moves to the specifier of a head Y^0, then by definition X must C-command Y^0 and everything dominated by Y^0. Similarly, if AGREE establishes a relation between constituents containing interpretable and uninterpretable features, then by definition there must also be a C-command relation between them. Consequently, to suggest that the syntactic relations in question do not include hierarchical information is incorrect: C-command is the syntactic means by which each relation is effected. There is an additional problem with including C-command information, namely that C-command relations can be defined over constituents that have no derivational, semantic or functional relation. For instance, a subject in SpecTP will C-command any object regardless of whether or not they are in a syntactic relationship to each other. As such, it gives rise to spurious possible relations which, while useful for syntax, may not be useful for linearization (cf. Kayne's (1994) discussion of non-terminal nodes). For these reasons, I will restrict my attention to relations based on the fundamental syntactic operations.

(4)   a. Lenny polished a saxophone
        b.

For a simple transitive sentence like (4a), a number of relationships (indicated by →) are instantiated during the derivation in (5): V^0 MERGEs with an object, *a saxophone,* then v^0 selects V^0 and MERGEs with it; v^0 selects an agentive subject in its external specifier and checks accusative case on the object by AGREE. T^0 selects and MERGEs with v^0. In turn, uϕ features on T^0 probe corresponding interpretable features on the subject and uT (i.e. Case) features on the subject probe iT on T^0 by AGREE. These relations are listed in (5) for ease of reference.

(5)   a. $T^0 \rightarrow v^0$ T^0 selects v^0
        b. $v^0 \rightarrow V^0$ v^0 selects V^0
        c. $V^0 \rightarrow O$ V^0 selects the object
        d. v^0(iCase) $\rightarrow$ O(uCase) v^0 assigns Case to the object
        e. $v^0 \rightarrow S$ v^0 selects the subject
        f. T^0(iT) $\rightarrow$ S(uT)[7] T^0 assigns Case to the subject
        g. S(iϕ) $\rightarrow T^0$(uϕ)   The subject checks ϕ features on T^0

These relations are all, mathematically speaking, partial orders. Selectional features trigger MERGE which builds an asymmetric structure of the form {p,{p,q}} where *p* selects *q*. Similarly, an interpretable/uninterpretable feature pair will trigger AGREE which instantiates an asymmetrical agreement relation between the pair where the interpretable feature determines the value/status of the uninterpretable feature[8]. Similarly, movement is equivalent to internal MERGE and is parasitic on a prior AGREE relation; it follows that it also instantiates a partial order. Given that these relations are all underlyingly defined as partial orders, I further assume the strongest hypothesis that they should be treated identically. This is expressed by the Relational Equivalence Axiom in (6).

(6)   **Relational Equivalence Axiom (REA):** All asymmetric, syntactic relations which are instantiated by MERGE, MOVE or AGREE will be treated as being formally

---

[7] Strictly, iT on T^0 checks uT on D^0 of the subject.
[8] Agreement is asymmetrical with respect to any single interpretable/uninterpretable feature pair.

equivalent insofar as they all instantiate partial orders of the form {p,{p,q}} (i.e. there should be no separate treatment for different types of relation).

The REA is a principle of methodological conservativity that acts like Occam's Razor by stripping away unwanted ancillary assumptions about the nature of the relations: it simply does not matter whether they are semantic theta roles or whether they are agreement relationships or whether they are selectional or specifier-head relations, etc. The null hypothesis is that the PF interface does not distinguish between them for the purposes of linearization.

3.2 Linearization of relations

Once a syntactic structure is derived, the question arises as to how to linearize such a structure. This is perhaps one of the most important questions in syntactic theory. The question of linearization ultimately reduces to the mathematical question of how to map a two-dimensional representation to a one-dimensional string. There are presumably many ways of arbitrarily making such a mapping, however the mapping is constrained by the fact that it must minimize information loss. What I mean by this is the following: a syntactic structure can be expressed by a string of words. This string, when pronounced, must be interpreted by a hearer who must be able to infer the original relations from the given string combined with a grammar. If the hearer cannot infer the relations from the string, then the string is effectively uninterpretable and becomes meaningless. In a similar vein, a child language acquirer must be able to infer the grammatical relations from text strings and Universal Grammar (UG) in order to be able to learn the grammar of his or her caregivers. Failure to infer this information will result in non-acquisition of the target grammar.

There are several such proposals for linearization, including mapping headedness parameters to linear orders, mapping asymmetric C-command to linear order via the Linear Correspondence Axiom (LCA; Kayne 1994), the tree-parsing proposal of Toyoshima (2013), Bury's (2007) approach where syntax underdetermines word order, and the Relational Linearization approach of De Vos (2008, 2013). In what follows, I will explore the implications of one of these approaches, namely the Normalization/Relational Linearization approach (De Vos 2008, 2013, 2014), for head movement. Since this approach is relatively new, I will give a brief outline of it here.

The approach outlined by De Vos (2013) positions itself as a mechanism for linearization (an alternative to the LCA (Kayne 1994)) which is compatible with, and informed by, the Minimalist Program. As such, the ontology of categories, the interfaces and syntactic operations of MERGE, AGREE and MOVE are consistent with standard theory. The Relational Linearization approach is dedicated to the naïve null hypothesis that, given the primacy of syntactic relations, these can be mapped to linear order in a one-to-one manner. Thus, (7) defines the Relational Precedence Axiom.
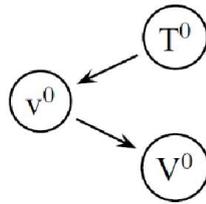
(7)   **Relational Precedence Axiom (RPA):** For any syntactic relation (indicated by $\rightarrow$), if $p \rightarrow q$ then p precedes q in linear order.

This is the naïve null hypothesis because at first sight it appears to be much too strong a claim to be taken seriously; it must surely be wrong. There are many ways of weakening it, for example, by discarding the REA (6): one could claim that different types of syntactic relations

could be linearized in different ways (e.g. selection relations could be linearized left-to-right while AGREE relations could be linearized right-to-left, etc.); one could also claim that syntactic relations must be mediated by MERGE and that linearization should apply to the resulting phrase structure (this would probably yield a system closely akin to the LCA). Yet, it is the very strength of the hypothesis that makes it a claim worth investigating and I will continue to assume it on this basis.
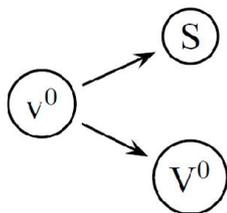
In practice, this means that for selection relations mediated by MERGE, the selector precedes the selectee; for agreement relations instantiated by AGREE, the site of the interpretable feature precedes the site of the uninterpretable feature in linear order. In order to see how this works, consider the simple schematic in (8) where the following transitive relations apply: $T^0 \rightarrow v^0 \rightarrow V^0$. This maps trivially to $T^0 > v^0 > V^0$. Note that I adopt the strong position that the RPA is an absolute condition which, if violated, causes the entire derivation to crash. Therefore, an order such as $v^0 > T^0 > V^0$ violates the RPA because, although T^0 selects v^0, it does not precede it. In what follows, therefore, I only consider linearization schemas which conform to the RLA, all others being ill-formed.

(8)

The following example is slightly more complex. In (9), two categories depend on v^0: little v^0 selects a V^0 complement and a subject in its specifier. Thus, according to the RPA (7), both S and V^0 follow v^0 but, crucially, there is, mathematically speaking, no ordering between S and V^0. This differs from standard, phrase-structure-driven approaches where the specifier is argued to be different to the complement. While this distinction may turn out to be justified, it represents an additional layer of assumptions and, for the moment, I beg the reader's indulgence: let us assume the stronger assumption – encoded by the REA in (6) – that there is no deep difference between selected specifiers and complements, and see how it turns out.

(9)

Since there is no mathematical ordering between S and V^0, it follows that this is consistent with two possible orders, namely $v^0 > S > V^0$ and $v^0 > V^0 > S$. Both of these orders obey the RPA, however, in both instances a dependent category is not immediately right-adjacent to its antecedent. In the

former, V^0 is not immediately right-adjacent to its selector v^0, but S intervenes. Similarly, in the latter, S is not immediately right-adjacent to its selector v^0, but V^0 intervenes.

At this point, it seems reasonable to propose a simple locality condition, the Relational Locality Condition outlined in (10) which ensures that dependents are placed as close as possible to their antecedents. By "locality", I assume the strongest possibility, namely that dependents should be strictly right-adjacent to their antecedents. If they cannot be strictly adjacent, as in (9), then they are annotated with an asterisk to indicate a single felicity violation of the condition. Although it is quite possible to envisage these violations as incremental, multiple or even statistical, I will adopt what I suspect is the more restrictive position that a violation is a once-off, polar occurrence: a dependent is either adjacent to its antecedent or it is not, thus triggering one felicity violation. Unlike the RPA, the RLC is a violable condition. Thus, the linearizations of (9) can be quantified as both equally optimal with one violation each (11)[9].

(10)   a. **Relational Locality Condition (RLC):** p should precede q as 'closely' as possible;
         b. p is 0-close to q if p is immediately left-adjacent to q; p is 1-close to q if there is one category, r, between p and q, etc.;
         c. if p is *not* 0-close to q, then q incurs a violation of the RLC, indicated as an asterisk on q[10].

(11)   a. $v^0 > V^0 > *S$ (9) 1 non-adjacency violation
         b. $v^0 > S > *V^0$ (9) 1 non-adjacency violation

4.   Linearization in action

Let us now turn to linearizing the relations comprising the sentence in (4).

(12)   Lenny polished a saxophone

         a. $T^0 \rightarrow v^0$
         b. $v^0 \rightarrow V^0$
         c. $V^0 \rightarrow O$
         d. $v^0(\text{iCase}) \rightarrow O(\text{uCase})$
         e. $v^0 \rightarrow S$
         f. $S(\text{i}\phi) \rightarrow T^0(\text{u}\phi)$
         g. $T^0(\text{iT}) \rightarrow S(\text{uT})$

Each of these relations must be linearized according to the RPA. As the linearization proceeds, it will be helpful to keep a tally of which relations have been linearized and which have not by striking them off the list. This will ensure that all and only these relations are linearized without introducing spurious relations or linearizing the same relation more than once.

---

[9] These optimal but imperfect linearizations raise the question of what would prevent the linearization system from simply deleting an offending category. Ultimately, a functional principle must come into play: deleting categories will result in their information being lost from the computation. The resulting structure will therefore not carry that information and a hearer will not be able to infer the existence of the deleted relation. Thus, in order to preserve informational integrity, such a system must incorporate a version of Full Interpretation (Chomsky 1995b) which prohibits spurious deletions and insertions.

[10] For the purposes of simplicity, I will use binary violations instead of categories incurring multiple violations. It may also be possible to mark the violations on p rather than q. I leave these to future research.

In this example, T^0 selects v^0 which selects V^0; by the RPA (7), these yield the linearization schema in (13a). V^0 selects its object and thus V^0 precedes the object, yielding (13b).

(13) a. $T^0 > v^0 > V^0$ $T^0 \rightarrow v^0$

                                                                              $v^0 \rightarrow V^0$

  b. $T^0 > v^0 > V^0 > O$ $V^0 \rightarrow O$
  c. $T^0 > v^0 > V^0 > {*}O$ $v^0(iCase) \rightarrow O(uCase)$
  d. $T^0 > v^0 > S > {*}V^0 > {*}O^{11}$ $v^0 \rightarrow S$
  e. $S > T^0 > v^0 > S > {*}V^0 > {*}O$ $S(i\phi) \rightarrow T^0(u\phi)$
  f. $S > T^0 > S > {*}v^0 > S > {*}V^0 > {*}O$ $T^0(iT) \rightarrow S(uT)$

Note, however, that v^0 assigns accusative case to the object. One option would be to locate a second copy of the object between v^0 and V^0. This would disrupt the strict adjacency of v^0 and V^0 established in (13a). Another option is to allow the RPA to be satisfied in a slightly non-local configuration where the object is placed to the right of V^0 but incurs a locality violation as in (13c).

Little v^0 selects a subject ($v^0 \rightarrow S$) which consequently follows v^0: if it is inserted directly right-adjacent to v^0 then it follows that it prevents V^0 from being adjacent to little v^0 (as established in (13a)). To represent the fact that these latter two categories cannot be strictly adjacent, I will use an asterisk as in (13d).

There are other possible spaces where the subject might be inserted, but each of them results in a locality violation. The following schemas and their locality calculations are included here:

(14) a. $T^0 > v^0 > V^0 > {*}O\ {*}S$
    b. $T^0 > v^0 > V^0 > {*}S\ {*}O$

Thus, although (13d) still conforms to the RPA (7), it does so at the expense of strict adjacency.

We now turn to the question of how to represent the various agreement relationships between the subject and T^0. Since the subject checks uϕ on T^0, it follows that the subject precedes T^0 (as in (13e)). Thus, the representation now includes two copies of the subject. Since the subject preceding T^0 does not interrupt any pre-existing adjacency relationship, no locality violations need to be included. Finally, T^0 checks uT (Case) features on the subject and thus T^0 must precede the subject. In fact, in (13e), it already does so. However, since each syntactic relation is taken to be mapped to a linearization pair, I will insert a copy of the subject adjacent to T^0 as in (13f).

Before proceeding, let us explore this issue more closely: when is it possible to insert a copy to obviate a RLC violation and when not? Presumably we would want to restrict situations where a category could be spuriously inserted multiple times, thereby removing all possible violations. This suggests we need to obey some form of Full Interpretation where each relation is spelled

---

[11] An anonymous reviewer suggests that, since the subject now intervenes between v^0 and V^0, the object might be accorded an additional violation as it is now two places removed from its case assigner. On the other hand, there has been no qualitative change in the representation: the object is still not 0-adjacent to its case assigner. For the sake of convenience, I assume a binary, once-off allocation of the violation rather than stacking up multiple violations. As far as I can see, nothing hinges on this decision.

out only once. In (13f), the subject has been inserted following T^0 in order to linearize the ϕ checking relation: $T \rightarrow S$. However, as pointed out, T^0 already precedes S in (13e). It is a legitimate question to ask why the additional copy ought to be inserted here; failure to do so would mean that one relation – namely, the iT $\rightarrow$ S(uT) relation – would not have been mapped to a linear order – a violation of Full Interpretation (cf. footnote 9). If this were to be the case, then the relation would effectively be lost since a parser/hearer would not be able to infer the existence of this relation from the linear order in (13e) alone. Put another way, the linearization schema in (13e) places the subject after v^0 for reasons independent of whether the iT checks uT on the subject; thus, using (13e) as a basis to infer the existence of the iT $\rightarrow$ uT relation is fallacious[12].

Having derived the linearization schema (13f), it now needs to be provided with morpho-phonological content. Recall that a linearization schema is a linearization of syntactic relations as expressed through syntactic structure. At this juncture, PF rules of chain interpretation are applied. For example, the head of a chain of copies is usually spelled out in English with the remaining copies being given null phonological content (cf. Nunes 1999, 2004 and Bever 2003)[13]. I assume the same applies here, yielding (15b). In addition, each feature bundle will be matched with the most highly specified lexical item consistent with it in a model such as that of Distributed Morphology (Embick and Noyer 2001, Harley and Noyer 1999, Marantz 1997, Marantz and Halle 1993), ultimately yielding something like (15c).

(15)   a. $S > T^0 > S > *v^0 > S > *V^0 > *O$

Spelled out as:

b. Lenny > T^0 > Lenny > $*v^0$ > Lenny > $*V^0$ > $*O$
c. Lenny > will > t   > e > t   > polish > a saxophone

4.1   The effects of head movement on locality violations

At this point, let us explore, in abstract terms, the effects of head movement. Head movement serves the purpose of creating a complex set of features, i.e. of clustering together features on

---

[12] Both reviewers point out the importance of restricting the insertion of spurious copies. Such insertions would allow almost any linearization schema violations to be "rescued", and unrestricted insertions would violate Full Interpretation and would probably make the grammar too powerful.

[13] Since what is passed to the interface is a set of relations rather than phrase structure, the question may arise as to whether a chain defined over relations is identical to a chain defined over C-command. Given that C-command is already encoded in the relations of MOVE and AGREE, I assume that there is little substantive difference. That is, if X moves to a specifier (Spec) of Y and if iF on X determines uF on Y, then it follows from Armstrong's (1974) Axiom of Transitivity that X determines Y *and* all that Y determines, including the original copy of X. Thus, chains can be expressed in terms of relations too.

I concede that there is a possibility where a chain of linearization-induced movement might not be properly translatable into a relational notation. Suppose that P C-commands Q and that iF on P determines uF on Q. In this instance, let us suppose that Q does not move to Spec of P but remains *in situ* and features are checked by AGREE. In this instance, Q may be spelled out left-adjacent to P as a function of spell out rather than movement. Then, it might be the case that the Q chain will not be well-formed: note, however, that Q would always determine its own copy trivially. This is an issue that needs further exploration. Examples of this may be when iCase on v checks uCase on the object without the object moving to Spec$v$P. The framework outlined here makes an interesting prediction: should object shift occur under these circumstances, it would be predicted to not have the properties of a syntactic chain (e.g. it would not induce scope effects since such effects could not be deducible from Armstrong's axioms (1974)). Interestingly, it appears that object shift, as it occurs in mainland Scandinavian, has exactly these types of properties. I acknowledge that it is not possible to flesh out these types of issues in this paper.

a single head (or feature set) that would otherwise have been contained in two different feature sets. In the following linearization schemas, I have compared the schema with no feature clusters (16a) with a schema where v^0 and V^0 are clustered (16b), and with a schema where T^0, v^0 and V^0 are clustered (16c). This simulates the effects of no head movement, short v-V raising and V-T movement, respectively.

(16) a. S T^0 S $*v^0$ S $*V^0$ $*$O   3 violations

  b. S T^0 S $*$ $\boxed{v+V}^0$ S $*$O   2 violations

  c. S $\boxed{T+v+V}^0$ S $*$O   1 violation

Example (16a) incurs three locality violations as explained in the step-by-step outline in the previous section. Consider (16b): because v^0 and V^0 are in the same feature cluster, an XP which is adjacent to the complex feature cluster is deemed strictly adjacent to both features. In (16a), V^0 incurred an adjacency violation because it needed to be adjacent to v^0 but could not do so because the subject intervened. However, in (16b), since v^0 and V^0 satisfy the RLA (10) within the cluster itself, one violation is obviated. Thus (16b) is more optimal than (16a). Head movement has thus, in this instance, served to create a more optimal linearization schema.

Consider (16c): the locality of the selectional relationships between T^0, v^0 and V^0 are all satisfied within the complex feature cluster. Since the subject is adjacent to the cluster, it is also adjacent to T^0 (a Case-assignment relationship) and to v^0 (a selectional relationship). This obviates two locality violations. Thus in (16c), head movement has served to create a complex feature set which results in more optimal linearization schema.

This is a particularly powerful and important result. In narrow syntax, the effect of head movement is exactly to cluster features together into complex feature bundles, barring excorporation. The above results demonstrate that clustering of features into larger feature bundles serves to make linearizations more optimal and also reduces computational load by reducing the total number of numerations. Therefore, we have arrived at an explanation for why head movement happens – it is a means of resolving linearization paradoxes.

4.2 Constraints on bundling and deriving the Head Movement Constraint

Feature clustering is a powerful mechanism and we must now ask what constraints operate on these bundling operations. For instance, what prevents a derivation from simply bundling *everything* into a single feature bundle, thereby creating a structure with zero adjacency violations?[14] The first constraint on clustering is morphological: once features have been bundled and the linearization schema is passed to the morphological module, the resulting feature bundle is matched to various morphs as per Distributed Morphology (Embick and Noyer 2001, Harley and Noyer 1999, Marantz 1997, Marantz and Halle 1993). Thus, for any feature bundle, if a language has a morphological form that is specified for the features in question, then the feature bundle is spelled out with that form, otherwise the Elsewhere condition applies (Kiparsky 1973). The process of matching a feature bundle to a morphological specification is

---

[14] This arguably does occur in polysynthetic languages. I assume that these languages have the morphological resources to spell out complex feature bundles of this type.

handled by the Subset Principle in (17) which has been articulated many times (e.g. Harley and Noyer 1999:5, Van Koppen 2005:14).

(17)   **Subset Principle:** Spell Out of a syntactic feature bundle occurs by matching the features to the specification of a "phonological shape". Insertion occurs when the morph matches all or a subset of the features in the bundle. Where there is more than one candidate morph, the most highly specified wins.

Here is an example of how the Subset Principle operates:

(18)   a. $[+V^{Say} + T^{-Finite}] \approx$ (to) say
        b. $[+V^{Say} T^{3sg}] \approx$ (she) says
        c. $[T^{3sg}] \approx$ (she) does   [do-support]

Suppose that a feature bundle includes a verbal feature with associated lexical denotations of *saying* as well as a non-finite T^0 feature (18a). The morphological component would match this feature bundle to a lexical item, namely the infinitive verb *(to) say*. Similarly, if the feature bundle included a T^0 feature specified for 3SG, then it could be matched to the finite verb *says* (18b). Finally, if the feature bundle consisted solely of a T^0 feature specified for 3SG without any lexical verbal features, then it is possible, in English, to match it with a dummy verb *does* which is able to lexicalize verbal finiteness on T^0 without carrying any lexical verbal information. Naturally, this is a morphological resource that is available in English for historical reasons, but is crucially absent in Germanic V2 languages, for instance. Consequently, we can claim that, in addition to the Subset Principle, the available morphological or lexical resources of a language constitute an important constraint on feature bundling. In answer to the question posed above as to what prevents a derivation from bundling *everything*, the answer lies in the fact that a language invariably will lack a morph to spell out such a mega feature bundle[15].

Another likely constraint on bundling of features which has wide empirical support is the Head Movement Constraint (HMC; Travis 1984). As noted previously (cf. section 2), the HMC was originally formulated as a restriction on head movement, namely that a head may move to the head immediately dominating it. Usually, the HMC is read in conjunction with a ban on excorporation which together entail that a head may not "skip over" any intervening heads, but must necessarily adjoin to each head successively and then may pied-pipe to the next higher position. It was subsequently incorporated into the principle of Relativized Minimality (Rizzi 1991). Effectively, what this means is that head movement is dependent on the existence of a selection relationship between the feature bundles in question, i.e. a head (feature bundle) may only undergo head movement to a head (feature bundle) that immediately selects it.

From the perspective of the current analysis, the HMC also plays an important role in constraining the bundling of features: features (or bundles of features) may only be bundled in the process of a derivation if they are in a selection relationship. Thus in (19), where A selects

---

[15] An anonymous reviewer argues that Elsewhere effects will ensure that a bundle of features is always spelled out and that consequently the lexicon will not necessarily constrain bundling. However, s/he also points out that in Distributed Morphology, competition between different realization occurs at the morpheme level, not the word level. Thus, Elsewhere effects can still apply in competition for morphemes, even if they do not apply at the word level. This suggests that the notion of 'lexical blocking' that I utilise here falls strictly outside the Distributed Morphology framework. I will leave it to future research to flesh out these implications.

B selects C, well-formed bundles are those in (19b): (A,B), (B,C) and (A,B,C). In each instance, there is a selection relation between the bundled items. Also note that, at this stage, it does not matter whether A, B and C are atomic features or whether they are themselves feature bundles. However, a bundle (A,C) is ill-formed because neither A nor C selects the other. In this way, the HMC acts to constrain the creation of feature bundles in the derivational component of the grammar.

(19)   a. $A \rightarrow B \rightarrow C$

  b. $\boxed{A+B}^0$; $\boxed{B+C}^0$; $\boxed{A+B+C}^0$;

  c. $*\boxed{A+C}^0$

While the HMC is sufficiently supported empirically to allow it to be adopted as a stipulation, I think we can go one better. The HMC can be derived from the assumptions I have already outlined within the current framework, in other words, the HMC does not need to be stipulated but follows from first principles. To this end, consider (20), based on the relations in (16). (20a/b) represents HMC-licit combinations equivalent to short v-V raising and V-T raising, respectively.

(20)   a. $S\ T^0\ S\ *\boxed{v+V}^0\ S\ *O$ 2 violations

  b. $S\ \boxed{T+v+V}^0\ S\ *O$ 1 violation

  c. $S\ \boxed{T+V}^0\ S\ *v^0\ S\ *V^0\ *O$ 3 violations

  $\longleftarrow\!\!\times\!\!\longrightarrow$

Now, suppose that T^0 and big V^0 are bundled together, excluding little v^0 from the resultant bundle (20c). The result is, in effect, a HMC violation. I have illustrated this informally with an arrow to demonstrate how big V^0 has apparently skipped a head. The resulting linearization schema is listed in (20c). The subject precedes T^0 according to the same logic as the previous examples; since T^0 determines the Case of the subject, a copy of the subject also follows the T+V feature bundle. Importantly, however, although T^0 selects little v^0, they cannot be adjacent because a copy of the subject is already adjacent to T^0 (i.e. one violation of locality). Similarly, a copy of the subject follows little v^0 where it intervenes between little v^0 and V^0, incurring another violation of locality. (Note how a copy of big V^0 is required after little v^0, even though big V^0 is bundled with T: in a sense, it appears that bundling T^0 and V^0 in violation of the HMC has little impact on the number of locality violations.) Therefore, (i) head movement/ feature bundling serves to make linearization patterns more optimal, and (ii) HMC-violating movement/bundling does not lead to more optimal linearizations. Consequently, the effect of the HMC is derived from linearization considerations.

This result is important because it demonstrates that a representation which violates the HMC is less optimal than one which does not. Consequently, the effect of the HMC can be derived within the current framework and it does not need to be stipulated independently. Moreover, the system inherently provides a rationale for head movement – something that was lacking in the Principles & Parameters and Minimalist frameworks – namely the need to optimize linearization.

Part II

At the beginning of this paper, I asked you to accompany me on a journey into a theoretical space where linearization of syntactic relations occurs via a one-to-one, pairwise mapping from relations to linear order. I have demonstrated that partial orders introduce points of tension where orderings become less optimal. I have also demonstrated that under this system, bundling of features functions to make linearizations more optimal. In effect, what this means is that head movement and the HMC are deep properties of the language computation system, following from principles of linearization. In the following sections, I will outline how head movement applies in a number of instances and how it is sometimes blocked by the morphological properties of the languages in question.

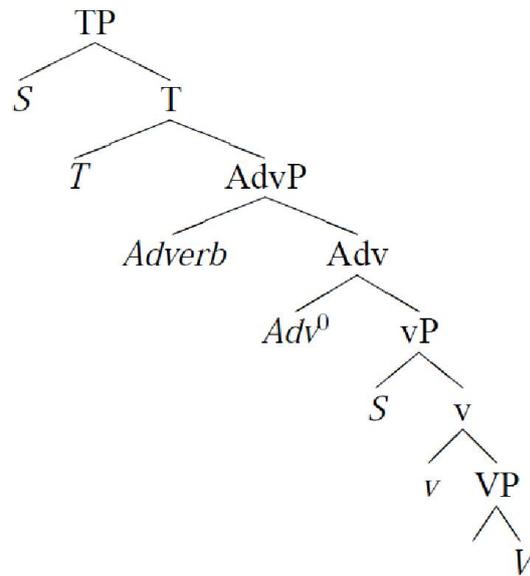**5.   V-T raising: The differences between English and French**

The parametric difference between verb movement in English and French (21) is well known: in French, a finite verb occurs to the left of an adverbial as a result of V-T raising; in English, the finite verb must occur to the right either because V-T raising is covert or because tense lowers from T^0 to V (Chomsky 1995b); or, in a lexicalist system, the pre-existing features of the verb are checked *in situ* with AGREE.

(21)   a. *Je   mange* <toujours>
              **Je* <toujours> mange*
         b. *I   eat <always>
              I <always>   eat

5.1 French V-T raising

Cinque (1999) proposes that adverbs are phrases selected for the functional hierarchy as illustrated by the tree in (22). In this particular case, T^0 selects Adv^0 which selects vP. The adverbial is located in the specifier of AdvP which has a null adverbial head (Cinque 1999). Thus, AdvP acts like any other functional head. The relevant relations are listed in 22(a-g).

(22)

a. $Adv^0 \rightarrow AdvP$
b. $T^0 \rightarrow Adv^0$
c. $Adv^0 \rightarrow v^0$
d. $v^0 \rightarrow V^0$
e. $v^0 \rightarrow S$
f. $T^0 \rightarrow S$
g. $S \rightarrow T^0$

Linearization of these syntactic relations proceeds as follows: one resulting linearization schema is listed in (23a). From the RPA (7), $T^0 > Adv^0 > v^0 > V^0$. However, since both v^0 and the adverbial in SpecAP are selected by Adv^0, an ordering paradox arises between them. Since it is not possible that both can be right-adjacent to Adv^0, one of them must be in a non-local position. This is indicated on the constituent by an asterisk. Similarly, the subject is part of a multivalued dependency: it is selected by v^0 and is also assigned Case by T^0, as well as checking ϕ features on T^0. The linearization schema in (23a) has three violations of the RLA (10).

However, an alternative linearization schema would be to cluster selectional relationships into a single feature bundle, as outlined in the previous section, yielding (23b). The advantage of this schema is that the selectional features of T^0, Adv^0, v^0 and V^0 meet the requirements of the RLA (10) by virtue of being local within the feature bundle. This feature bundle can be matched to a corresponding morphological form in the lexicon. Consequently, the linearization schema with the fewest locality violations is spelled out.

This schema results in the verb preceding the adverb because the ability to be clustered is dependent on the features in question being heads, as is standard in head movement accounts. Thus, it is possible for a null adverbial head Adv^0 to be incorporated into the verb-head cluster but the adverbial phrase itself must remain outside the complex head. However, since Adv^0 is

incorporated inside the complex head, it follows from the RPA (7) that the adverb must follow the complex verbal head. This results in a French-type V-T order[16].

(23)   a. S T^0 S $*Adv^0$ Adv $*v^0$ S $*V^0$ 3 violations

   b. S $\boxed{T+v+V+Adv}^0$ Adv $*$S   1 violation

   Spelled out as:

   S $\boxed{T+v+V+Adv}^0$ S *Adv*

   Je mange   *t   souvent*

5.2 English *v-in-situ*

I have outlined a system that explains head movement, however, the RPA (7) is a very strict assumption that ensures all movements are "overt". Although this assumption may be too strong, I will maintain it in order to push the proposal to its limits. English thus poses a challenge to a strict interpretation of the RPA (7) because English has only v^0-V raising and no V-T movement. To derive English word order, I will capitalize on a theoretical distinction with respect to the syntax and semantics of adverbs.

In contrast to a Cinque-style approach to adverbials, the standard view of adverbs is that they are adjuncts outside the backbone of the functional hierarchy. Exemplars of this approach are Ernst (2002) and Nilsen (2003). For example, within the approach of Ernst (2002), adverbial adjuncts "attach to" the nodes in a tree with which they are semantically compatible, i.e. adjuncts select their hosts[17]. In fact, this property does not distinguish Ernst's system from that of Cinque: in both, an AdvP might select a vP – or another relevant node – and this is important for their semantics. What differentiates the approaches is that in Cinque's framework, the AdvP is itself selected by a dominating node in the functional hierarchy whereas in Ernst's framework, the AdvP remains unselected. Given that these two adverb theories both seem to be supported by substantial evidence, and since this latter relationship does not seem to have any semantic import, I propose that it be parameterized: in French, AdvPs are selected by a dominating node; in English, they are not[18].
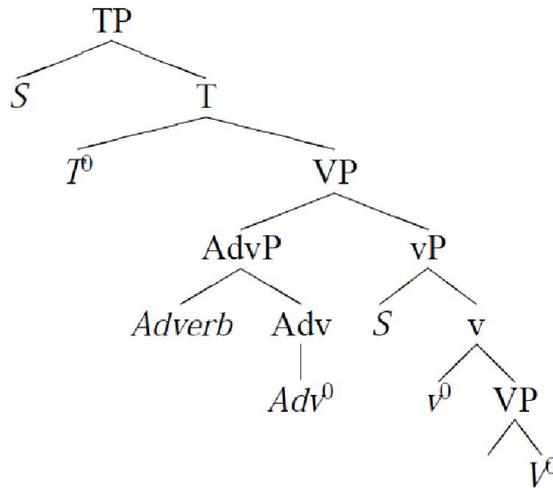
---

[16] It may be objected that the derivation requires a null adverbial head to be incorporated into a verb. In fact, this issue exists within standard models of V-T raising anyway: the HMC requires that V^0 incorporates into the head of AdvP on its way to T^0. Thus the model I am proposing here is not substantially different to the standard theory in this regard. In addition, there is no intrinsic reason why verbs cannot incorporate adverbial semantics as is illustrated by English verbs of walking which arguably incorporate manner adverbial heads, e.g. *amble*, *promenade*, *strut* vs *walk.*

[17] An anonymous reviewer points out that within a Bare Phrase Structure (Chomsky 1995a) approach, if any category selects its host it must also project. This entails that adjuncts do not select their hosts. Unfortunately, the means of formally representing an adjunction relationship is beyond the scope of this paper. It seems clear to me that one alternative, namely that the adjunct is selected by the host, is unlikely given that adjuncts do not behave like other selected categories such as, say, complements. This leaves open the possibility that, for instance, the adjunct is not related by a partial order at all (which might undermine the premise of Bare Phrase Structure) or that the host and adjunct mutually select each other, thus mutually projecting. Unfortunately, given the lack of consensus on this issue in the literature, I do not want to commit myself on any of these issues in this paper, although they do impact on the analysis. Rather, I choose to work with the accepted tree-structure for adjuncts, namely that they do not project.

[18] For space reasons, I will not offer any evidence that this is the case other than that the difference seems to be purely formal, making no semantic difference, and that it obtains the correct word order. Indeed, there may be

With this proviso in place, a syntactic tree of an English vP adverbial is represented in (24), and whose syntactic relationships are represented beneath it.

(24)

a. $T^0 \rightarrow v^0$
b. $Adv^0 \rightarrow Adv$
c. $Adv^0 \rightarrow v^0$
d. $v^0 \rightarrow V^0$
e. $v^0 \rightarrow S$
f. $T^0 \rightarrow S$
g. $S^0 \rightarrow T^0$

If one focuses on the relevant relationships between T, v^0 and the adverb, it can be seen that both T^0 and Adv^0 select v^0 and thus, by the RPA (7), both must precede v^0. However, since only one of them could be left-adjacent to v^0, the resulting linearization schema will yield a violation of the RLA (10): $T^0 > Adv^0 > *v^0$. Similarly, as for all the previous examples, the subject induces a series of locality violations. One possible linearization is illustrated in (25a).

(25)   a. S T^0 S Adv^0 Adv $*v^0$ S $*V^0$ 2 violations

b. S Adv^0 Adv $*$ $\boxed{T+v+V}^0$ S   1 violation

Spelled out as:

$S\ Adv^0\ Adv$ $\boxed{T+v+V}^0 S$

I   always eat   t

As outlined at the beginning of this article, when selecting heads are clustered into a single feature bundle, the number of violations is correspondingly reduced as in (25b). Adv^0 selects the adverbial in its specifier and is spelled out immediately adjacent to the adverbial; Adv^0 selects v^0 but cannot be adjacent to it because of the intervening adverbial; similarly the subject checks ϕ features on T^0 but cannot be adjacent to the subject because of the intervening Adv^0

---

alternative ways of deriving head movement in the system I am proposing; I would prefer to leave these as questions for future research.

and adverbial. When (25b) is spelled out, the adverbial precedes the inflected verb, thus mimicking lack of verb-raising in English[19].

The analysis thus derives the distinction between French and English from a difference in the syntax of adverbs[20]. This makes the analysis substantially different to traditional head movement analyses which emphasize the verb's dynamism. Aside from the fact that it derives head movement from general linearization principles, one advantage of this analysis is that it derives tense-lowering effects in English. This contrasts with the traditional head movement approach which, in the absence of head movement, also requires tense lowering or similar additional machinery.

6. Negation and head movement

The story of English and French verb-raising would not be complete without a discussion of how negation and head movement interact.

6.1 French negation

French negation is different to other French adverbials because it includes a negative prefix on the verb with a negative adverbial phrase following the verb. In colloquial usage the negative prefix is sometimes dropped.

(26)   *Je ne mange   pas/plus/rien/aucune*
       I NEG eat.1SG.PRES not/more/nothing/at all
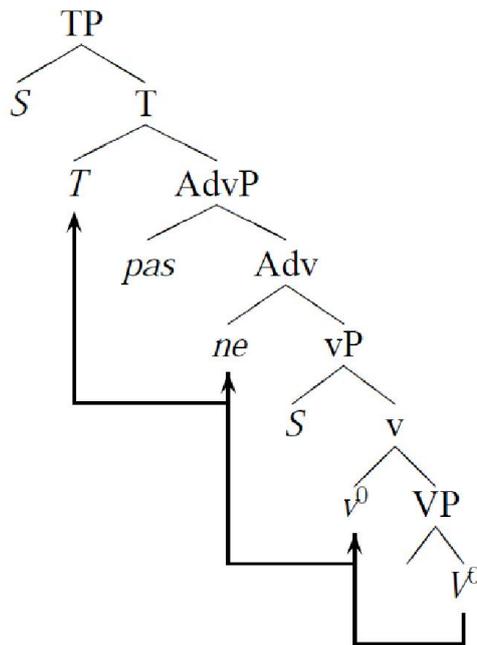   'I am not eating (anymore/anything/at all)' (French)

(27)   *Je sais   pas*
       I know.1SG.PRES not
       'I don't know' (French)

The version of the standard analysis is illustrated by (28) where V^0 adjoins to the negative prefix and then raises to T^0. The structure of French negation is thus largely identical to the structure for French adverbs except that Adv^0 is null while Neg^0 *ne* has morphological content. Thus, we can reasonably characterize the syntactic relationships involved as: T^0 selects Neg^0 which selects an overt negative adverb (*pas*, *rien*, *aucune*, etc.) in its specifier; Neg^0 also selects vP.

---

[19] One might ask why Adv^0 is not clustered into the complex head as it was with French. In French, Adv^0 was selected by T^0 and thus it *had* to be in the complex head; were it not, the complex head would not be able to be constituted in the first place. In English, Adv^0 is not selected by T^0 and therefore it is possible to create a complex head that excludes Adv^0. An additional question that arises is why Adv^0 is necessarily excluded in English. One answer might be that its inclusion is not motivated and therefore its inclusion could violate a "least effort" principle. Another type of answer might be that if Adv^0 is included it may yield additional violations. I leave this for future research.

[20] An anonymous reviewer queries whether this is a trivial difference insofar as in English, T^0 selects v^0 but in French, T^0 can select either Adv^0 or v^0. It seems to me that this is an inevitable consequence of adopting a functional hierarchy approach to adverbials – whenever a particular adverbial is absent, a selecting head must necessarily select the "next category down". On the other hand, if one assumes that all functional heads are present in the syntax, even if they are not filled, then the problem dissipates. Another possibility is that selection operates on a subset principle (perhaps derived via Armstrong's axioms (1974)) whereby if T^0 selects Adv^0 and Adv^0 selects v^0 then T^0 also transitively selects v^0.

(28)

    a. $T^0 \rightarrow ne^0$
    b. $ne^0 \rightarrow pas$
    c. $ne^0 \rightarrow v^0$
    d. $v^0 \rightarrow V^0$
    e. $v^0 \rightarrow S$
    f. $S \rightarrow T$
    g. $T \rightarrow S$

By the RPA (7), a possible linearization schema is represented in (29a). Since both the subject and Neg^0 are in syntactic relationships with T, one of them cannot be adjacent to T and incurs a violation by the RLA (10); the negative adverb is right-adjacent to Neg^0 and thus intervenes between Neg^0 and v^0, yielding another violation. As for the other examples, clustering of selected heads into a single complex feature bundle reduces locality violations and creates a more economical linearization schema (29b). This complex feature bundle is matched to morphological representations in the lexicon: it so happens that French negation has a bound morpheme that affixes to verbs and thus the negated complex head can be spelled out.

(29)   a. $S\ T^0\ S\ *ne^0\ pas\ *v^0\ S\ *V^0$ 3 violations

        b. $S\ \boxed{ne+T+v+V}^0\ S\ *pas$ 1 violation

        Spelled out as:

        $S\ \boxed{ne+T+v+V}^0\ S\ pas$

        Je ne sais   *t pas*

It is also possible to create an alternative linearization schema which also yields only a single locality violation (30). In this schema, two separate complex feature bundles are created: the leftmost one clusters Neg^0 and T^0 while the rightmost one clusters the remaining verbal features. $ne+T^0$ corresponds to a negated auxiliary and if it is in the numeration it will be chosen: the complex feature bundle is matched against the morphological resources in the lexicon. If the

numeration does not include a negated auxiliary, then this linearization schema will crash and yield a null result. This reliance on morphological resources in order to spell out optimal linearization schemas will play an important role in English negation.

(30)   a. S ne+T^0 S *pas v+V^0 1 violation

   Spelled out as:

   S $\boxed{ne+T}^0$ S *pas $\boxed{v+V}^0$

   Je n'ai   t   pas etudié
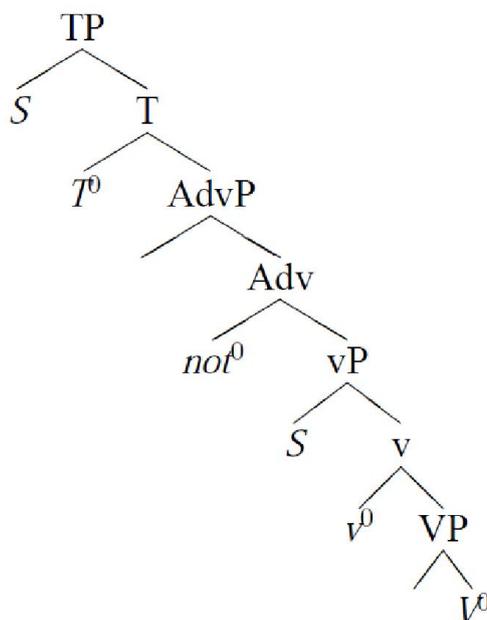
   'I didn't study'

6.2 English negation and T-lowering

English negation behaves very differently to other English adverbials: negation induces blocking effects and do-support which other adverbials do not.

(31)   a. James saw the peaches
        b. James <*saw> not <*saw> the peaches
        c. James <u>did</u> not see the peaches

The fundamental insight into this phenomenon is that in English, negation is a head and thus interrupts the relation between the tense and the verbal heads. In other words, English negation is very similar in its underlying structure to French negation which also sports a negation head. The difference is that (i) English negation is either a free morpheme (*not*) or a bound morpheme (*n't*) by lexical stipulation, and (ii) that while French Neg^0 selects a negative adverbial in its specifier, English lost negative concord in Standard English in the Middle English period (Frisch 1997, Ingham 2006). The structure of English negation is taken to be similar to the tree in (32).

(32)

a. $T^0 \rightarrow not^0$
b. $not^0 \rightarrow v^0$
c. $v^0 \rightarrow V^0$
d. $v^0 \rightarrow S$
e. $S \rightarrow T$
f. $T \rightarrow S$

With respect to linearization, the RPA (7) yields a linearization schema where $T^0 > Neg^0 > v^0$; the subject is, as usual, involved in a multivalued dependency which forces some locality violations by the RLA (10). One possible linearization schema is illustrated in (33a). As with the previous examples, clustering of the selected heads into a single complex head reduces the range of possible violations – in this instance, to zero. Unfortunately, this complex head would be equivalent to a negated lexical verb and English lacks the morphological resources to spell out such a verb (cf. *eatn't, *walkn't, *given't, etc.). I will return to this question below.

(33)   a. S T^0 S *not v^0 S *V^0 2 violations

         b. S $\boxed{T+neg+v+V}^0$ S   0 violations

Spelled out as:

S $\boxed{T+neg+v+V}^0$ S

*James eatn't   t

In the absence of being able to match the complex head to an appropriate morphological form, the linearization schema in (33a) fails to converge and alternative linearizations must be sought, even though they may be less optimal in terms of locality violations. Interestingly, there are *two* alternative linearization schemas, both of which have one violation.

The first of these is illustrated in (34) where the verbal heads are clustered excluding T^0 and Neg^0. Since T^0 is not immediately adjacent to Neg^0, one violation of locality results. English has the morphological resources to spell out this linearization schema: the v+V cluster is matched to an infinitive verbal form; Neg^0 is matched to a freestanding negation and T^0 is matched to the semantically underspecified verb *do*, yielding a do-support construction[21].

(34)   a. S T^0 S *not $\boxed{v+V}^0$ S   1 violation

Spelled out as:

S T^0 S not $\boxed{v+V}^0$ S

James did t not eat   t

The second of these is illustrated in (35) where, by analogy with (30), T^0 and Neg^0 are clustered. English has the capacity to spell out such a cluster because it includes in its morphological resources a dummy verb *do* which is sufficiently underspecified to be inserted. More importantly, English also allows negation to be spelled out as an affix on tense with the form *didn't*. This allows this linearization schema to converge as a do-support construction with

---

[21] Note that other languages which lack *do* are unable to spell out T^0 independently in this way.

contraction[22]. Interestingly, because both these linearization schemas incur a single violation, the analysis predicts that the contracted and non-contracted constructions should both be grammatical – which they are. The analysis thus not only derives the existence of do-support but also explains the co-occurrence of contracted and non-contracted negation.

(35)   a. S T+not ^{0} S * $\boxed{v+V}$ ^{0} S   1 violation

   Spelled out as:

   $S \boxed{T+not}\,^0\,S\,\boxed{v+V}\,^0\,S$

   James didn't t eat t

Let us also consider another possible linearization schema, this time representing "lowering". In (36), the verbal heads and tense are bundled together, excluding negation. Since negation is selected by T^0, it must follow the bundle; however, since the subject must also follow the bundle, one violation is incurred. Negation must also precede v^0 and this prevents the subject from being immediately adjacent to the bundle, thus incurring a second violation. Consequently, a "tense-lowering" linearization is less optimal than do-insertion.

(36)   a. S not * $\boxed{T+v+V}$ ^{0} S *not   2 violations

   Spelled out as:

   $S \qquad not *\boxed{T+v+V}\,^0\,S\,*not$

   *James not ate   not

7.   Conclusion

This paper explored some of the implications of a relational view of syntax (De Vos 2008, 2013) with respect to head movement in order to ascertain whether the approach has the potential to raise interesting questions and shed new light on old ones.

This paper had two parts: in Part I, I explored the status of head movement and, adopting a particular view of syntactic relationships, I argued that head movement, far from being a theoretical embarrassment, is an intrinsic part of grammar. Linearization of complex syntactic relationships is mathematically difficult insofar as there may not be trivial solutions to all linearization problems, but different linearizations may be more or less optimal. Against this backdrop, the creation of complex feature bundles serves to resolve linearization paradoxes and therefore to derive more optimal linearization solutions. Head movement is thus simply a formal mechanism to create these complex feature bundles and the HMC follows from general principles – an attractive theoretical result.

---

[22] The occurrence of forms such as *I'll*, *she'd* and *there's* may result from bundling of features from the subject and the modal or they may also be the result of phonological reduction. In the case of forms like *she'd* and *there's*, they are (i) restricted to certain lexical items (**the dog'd chewed the bone*), and (ii) do not co-vary with agreement respectively (*Where's my keys*; *there's lots of books on the table*). This suggests that these forms are morphologically conditioned. On the other hand, forms like *I'll* are more likely to be a result of phonological reduction since they are not restricted in their distribution: *The dog'll probably chew the bone*. I leave this to future research.

In Part II, I attempted to develop some of the implications of the proposed analysis of V-T raising. The resulting insights were applied to English and French with the caveat that these languages are parameterized with respect to the way adverbials are treated. In French, adverbials follow Cinquean assumptions while, in English, I argued that they be treated as adjuncts. Although this is a stipulated difference, it should be borne in mind that it is semantically trivial and that both theoretical approaches have merit. If one adopts these assumptions, then a number of head movement effects follow, including (i) the trigger of head movement being linearization, (ii) adverbial positioning in both languages, (iii) interactions with negation and do-support, (iv) morphological blocking effects, and (v) so-called "tense-lowering" in English.

References

Ackema, P. and C. Čamdžić. 2003. LF complex predicate formation: The case of participle fronting in Serbo-Croatian. *UCL Working Papers in Linguistics* 15(1): 131-175.

Alexiadiou, A. and G. Fanselow. 2000. On the correlation between morphology and syntax: The case of V-to-I. In J.-W. Zwart and W. Abraham (eds.) *Studies in comparative Germanic syntax: Proceedings from the 15th workshop on comparative Germanic syntax (Groningen, May 26-27, 2000)*. Amsterdam: John Benjamins. pp. 219-242.

Armstrong, W. 1974. Dependency structures of data base relationships. In J. Rosenfeld (ed.) *Proceedings of the 1974 IFIP Congress* 74. pp. 580-583.

Baker, M. 1988. *Incorporation. A theory of grammatical function changing*. Chicago and London: The University of Chicago Press.

Bever, T. 2003. Deconstructing functionalist explanations of linguistic universals. In A. Carnie, H. Harley and M. Willie (eds.) *Formal approaches to function in grammar*: *Papers in honor of Eloise Jelinek.* Amsterdam: John Benjamins. pp. 333-352.

Bobaljik, J. 2002. Realizing Germanic inflection: Why morphology does not drive syntax. *Journal of Comparative Germanic Linguistics* 6: 129-167.

Boeckx, C. and S. Stjepanovic. 2001. Head-ing toward PF. *Linguistic Inquiry* 32: 345-355.

Bury, D. 2007. Verb movement and VSO-VOS alternations. *UCL Working Papers in Linguistics* 19: 77-92.

Chomsky, N. 1991. Some notes on economy of derivation and representation. In R. Freidin (ed.) *Principles and parameters in Comparative Grammar*. Cambridge, MA: MIT Press. pp. 417-454.

Chomsky, N. 1993. A Minimalist program for linguistic theory. In K. Hale and S.J. Keyser (eds.) *The view from building 20: Essays in linguistics in honour of Sylvain Bromberger*. Cambridge, MA: MIT Press. pp. 1-52.

Chomsky, N. 1995a. Bare phrase structure. In G. Webelhuth (ed.) *Government and binding theory and the Minimalist program*. Cambridge, MA: Basil Blackwell. pp. 383-439.

Chomsky, N. 1995b. *The Minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. 2001. Derivation by phase. In M. Kenstowicz (ed.) *Ken Hale: A life in language*. Cambridge, MA: MIT Press. pp. 1-52.

Cinque, G. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

Cornell, T. 1996. Model-theoretic syntax. *GLOT International* 2: 3-7.

De Vos, M. 2008. Deriving narrow syntax from principles of lexical organization. *Lingua* 118: 1864-1899.

De Vos, M. 2013. Afrikaans mixed adposition orders as a PF-linearization effect. In T. Biberauer and M. Sheehan (eds.) *Theoretical approaches to disharmonic word orders*. Oxford: Oxford University Press. pp. 333-358.

De Vos, M. 2014. The evolutionary origins of syntax: Optimization of the mental lexicon yields syntax for free. *Lingua* 150: 25-44.

Embick, D. and R. Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32(4): 555-595.

Epstein, S. and D. Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.

Ernst, T. 2002. *The syntax of adjuncts*. Number 96 in the *Cambridge Studies in Linguistics* series. Cambridge: Cambridge University Press.

Fortuny, J. 2008. *The emergence of structure in syntax*. Amsterdam: John Benjamins.

Frisch, S. 1997. The change in negation in Middle English: A NegP licensing account. *Lingua* 101: 21-64.

Halmos, P.R. 1960. *Naive set theory*. Princeton: Van Norstrand.

Harley, H. 2004. Merge, conflation and head movement: The First Sister Principle revisited. In K. Moulton and M. Wolf (eds.) *The proceedings of NELS 34*. Amherst, MA: GLSA, University of Massachusetts.

Harley, H. and R. Noyer. 1999. Distributed morphology. *GLOT International* 4(4): 3-9.

Hinterhölzl, R. 2006. *Scrambling, remnant movement, and restructuring in West Germanic*. Oxford: Oxford University Press.

Ingham, R. 2006. Negative concord and the loss of the negative particle *ne* in late Middle English. *Studia Anglica Posnaniensia* 42: 1-21.

Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kiparsky, P. 1973. "Elsewhere" in phonology. In S. Anderson and P. Kiparsky (eds.) *A festschrift for Morris Halle*. New York: Holt, Rinehart and Winston. pp. 96-106.

Koopman, H. 2000. Prepositions, postpositions, circumpositions, and particles. In H. Koopman (ed.) *The syntax of specifiers and heads*. London: Routledge. pp. 204-260.

Kracht, M. 2003. *The mathematics of language*. Berlin: Mouton de Gruyter.

Langendoen, D. 2003. Merge. In A. Carnie, H. Harley and M. Willie (eds.) *Formal approaches to function in grammar: Papers in honor of Eloise Jelinek*. Amsterdam: John Benjamins. pp. 307-318.

Lechner, W. 2005. *Interpretive effects of head movement.* Available online: http://ling.auf.net/lingbuzz/000178 (Accessed 24 August 2014).

Mahajan, A. 2000. Eliminating head movement. *GLOW Newsletter* 44: 44-45.

Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2): 201-225.

Marantz, A. and M. Halle. 1993. Distributed morphology and the pieces of inflection. In K. Hale and S.J. Keyser (eds.) *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press. pp. 111-176.

Matushansky, O. 2005. Moving a-head. In K. Hiraiwa and J. Sabbagh (eds.) Minimalist Approaches to Clause Structure, *MIT Working Papers in Linguistics* (50). Cambridge, MA: MIT Press. pp. 1-45.

Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37(1): 69-109.

Müller, G. 2004. Verb-second as vP-first. *Journal of Comparative Germanic Linguistics* 7(3): 179-234.

Nilsen, Ø. 2003. Eliminating Positions: Syntax and Semantics of Sentential Modification. PhD thesis, Utrecht Institute of Linguistics.

Nunes, J. 1999. Linearization of chains and phonetic realization of chain links. In S. Epstein and N. Hornstein (eds.) *Working Minimalism*. Cambridge, MA: MIT Press. pp. 217-250.

Nunes, J. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Rizzi, L. 1991. *Relativized minimality*. Cambridge, MA: MIT Press.

Roberts, I. 2014. *Head movement and the Minimalist program.* Available online: www.ling.cam.ac.uk/people/roberts/headmovement.doc (Accessed 25 January 2014).

Rohrbacher, B. 1999. *Morphology-driven syntax. A theory of V-to-I raising and Pro-drop.* Amsterdam, Philadelphia: John Benjamins.

Seely, D. 2006. Merge, derivational C-command, and subcategorization in a label-free syntax. In C. Boeckx (ed.) *Minimalist essays.* Amsterdam: John Benjamins. pp. 182-217.

Toyoshima, T. 2013. A traversal parameter at the PF interface: Graph-theoretical linearization of bare phrase structure. In T. Biberauer and M. Sheehan (eds.) *Theoretical approaches to disharmonic word order.* Oxford: Oxford University Press. pp. 359-388.

Travis, L. 1984. Parameters and Effects of Word Order Variation. PhD thesis, Massachusetts Institute of Technology.

Uriagereka, J. 1999. Multiple spell-out. In S. Epstein and N. Hornstein (eds.) *Working Minimalism.* Cambridge, MA: MIT Press. pp. 251-282.

Van Koppen, M. 2005. One Probe-Two Goals: Aspects of Agreement in Dutch Dialects. PhD thesis, University of Leiden Centre for Linguistics.

Zwart, J.-W. 2009. Structure and order: Asymmetric merge. In C. Boeckx (ed.) *The Oxford handbook of linguistic minimalism.* Oxford: Oxford University Press. pp. 96-118.