Auxiliary Adverb Word Order Revisited

Klaus Abels
Auxiliary Adverb Word Order Revisited
Auxiliary Adverb Word Order Revisited*

Klaus Abels

1 Introduction

This paper is about generalization (1) from Sag 1978. (1) says that in English an adverb or a floated quantifier cannot immediately precede the site of VP-ellipsis (VPE) or VP-fronting (VPF)—summarily E-site. For discussion see Baker (1971, 1981), Ernst (1983), and Sag (1978, 1980). The basic facts illustrating (1) for VPE are given in (2) and (3) and for VPF in (4) and (5).

(1) Sag's Generalization: *{Q/Adv} – E-site
(2) Fred has never been rude to Grandfather, but John {always has; *has always}.
(3) Some of them are working hard. In fact, they {all are; *are all}.
(4) He said he would study karate, and study karate he {surely has; *has surely}.
(5) They said they would study karate, and study karate they {all have; *have all}.

This paper suggests a new account of these facts, which deeply implicates Head Movement (HM). The present theory of HM has two crucial non-standard properties: (i) downward HM is allowed and (ii) a head's position depends largely on its (local) syntactic environment.

The paper is organized as follows. After discussing a problem with the standard theory of HM, section 2 introduces the theory assumed in the rest of this paper. Section 3 investigates Sag’s generalization in some detail, pointing out problems both with the generalization and existing accounts of it. Section 4 shows how the data can be accounted for under the present theory.

---

* Special thanks go to Željko Bošković, Howard Lasnik, Andrew Nevins, and, of course, Luisa Martí. I am also indebted to the audiences at a UConn Linglunch (spring 2001), the 24th GLOW Colloquium, the first TiLT workshop, the 26th UPenn Colloquium, and the XII Colloquium on Generative Grammar, where versions of this material were presented.

2 Head Movement

2.1 The The-higher-the-bigger-Theorem

Standard theories of HM assume (i) that HM proceeds upward and upward only, (ii) that HM obeys Travis's (1984) Head Movement Constraint, and (iii) that successive HM does not excorporate. Jointly (i) and (iii) entail (6).

(6) The-higher-the-bigger-Theorem:

For all heads $\alpha, \beta$ that are members of a single head-chain, if $\alpha$ c-commands $\beta$, then the features of $\beta$ are a proper subset of those of $\alpha$.

This prediction is wrong. English finite verbs violate (6): Although V does not raise to $T^0$ in English, V shows agreement and tense information. This violation of (6) is treated by invoking Affix Hopping (Chomsky 1957) or morphological merger under 'adjacency' (Bobaljik 1994, 1995a) (7). On the standard view, 'adjacency' tolerates intervening adverbs but not negation or arguments (for discussion cf. Abels, under review; Stepanov 2001).

(7) $\alpha_{\text{affix}} \beta \rightarrow \beta + \alpha$

Affix Hopping is suspect on several grounds. First, it is largely redundant with HM. Both processes target heads and both give rise to the same order: the higher head follows the lower one. Second, rules in natural languages are generally structure dependent. Affix Hopping, being a linear rule, then poses a learnability problem. Third, Bošković (2001) shows that Prosodic Inversion (which is homologous to Affix Hopping) fails to account for the facts it was designed for precisely because of its linear character.

---

1 (i) is often deduced from some version of Feengo’s (1974, 1977) Proper Binding Condition or the bottom-up nature of syntactic derivations (cf. e.g. Epstein 2001).
2 (ii) is often viewed as an instance of Rizzi’s (1990) Relativized Minimality.
3 In Chomsky’s (1995) system (iii) follows from the uniformity condition on chains. Roberts 1991, Bošković 1994, Matushansky 2002 among others assume that HM allows excorporation. I can’t address their arguments here and adopt the standard view.
But the problems do not end here. Some violations of (6) are not amenable to an Affix Hopping treatment. Consider the Vata examples (8)-(9) from Koopman (1984:74 ex. 76, 61 ex. 46). (8) shows that Vata has a head-final CP. The verb moves to T°. According to (6) information residing in C° cannot show up in T°. Yet, the fact that (9) is a relative clause is marked on the verb (REL in the gloss). The verb is not adjacent to C°; it is sandwiched between its arguments; Affix Hopping fails (cf. also Tuller 1986 for Hausa).

(8) ãlÔ Ó wà sãkâ ë
who he-R want rice WH-Comp
Who wants some Rice?

(9) kO' (mômO') Ó ë ɗ -dâ -bØ zué sãkâ,...
manHIM-HIM he-R eat -PT -REL yesterday rice
the man who was eating rice yesterday, ...

To conclude, Affix Hopping is neither a desirable mechanism nor can it account for all the violations of the The-higher-the-bigger-Theorem. 4

2.2 Head Movement in Mirror Theory (Brody 2000)

To solve the problems from section 2.1, we assume that HM is post-syntactic and modular (Abels under review; Brody 2000). 5 The two modules producing HM effects are (i) Morphological Word Formation (MWF)—words are formed under the necessary but not sufficient condition that one head be the head of the complement of the other, and (ii) Positioning—a unique linear position for elements with multiple hierarchical positions is determined. We examine the modules in turn (for details Brody 2000; Abels, in press).

The heads joined by an arc form a word. In (10a) and (10b), ß° is the head of the complement of α°, and in (10b) γ° is the head of the complement of β° and MWF can take place. The condition that one head be the head of the complement of the other can be understood transitively. However, MWF of α° with γ° is blocked in (10c), because the intervening head β° is skipped. This captures the HMC and the ban against excorporation.

---

4 Chomsky (1993) voids theorem (6) by adopting strict lexicalism. This move avoids the problems with Affix Hopping, but it also denies the possibility of giving a syntactic account of Baker’s (1985) Mirror Principle (cf. Brody 2000 for discussion).

5 HM has always been late: Chomsky 1957; Fillmore 1965; Baker 1971. For recent discussion see also Boeckx and Stjepanović 2001; Chomsky 2000, 2001.
In (11a) and (11b) $\beta^o$ is again the head of the complement of $\alpha^o$. $\gamma^o$ is the specifier of $\beta P$. Therefore, MWF is allowed in (11a) but not in (11b).

Above we characterized the structural configuration head-of-complement-of as a necessary but not sufficient condition. To see why, consider some examples. Structure (10a) could be instantiated by taking $\alpha = v^o$ and $\beta = say$. These two can form a word: $\sqrt{say} + v$. (10b) could be instantiated by taking $\alpha = T^o$, $\beta = v^o$, and $\gamma = say$. These three heads can form a word: $\sqrt{say} + v + T$. The same configuration would also be instantiated by taking $\alpha = v^o$, $\beta = say$, and $\gamma = that$. These cannot form a word by morphological fiat: $*that + say + v$. Implicit in these examples is the assumption that the top to bottom order of syntactic heads determines the right to left order of morphemes. This captures the Mirror Principle (Baker 1985, 1988; Brody 2000).

Usually, Affix Hopping is assumed to be blocked by intervening negation but not by adverbs. This fact can be re-described in the present theory as a statement about MWF. Assuming that (12)—with $n't$ as the specifier of NegP—is the correct syntactic structure, we simply say that neg$^o$ can form a word with auxiliaries and T$^o$, but not with main verbs V: $*V + v + neg$.

(13) depicts the situation for adverbs. If adverbs are specifiers of dedicated functional heads (Alexiadou 1997; Cinque 1999), then these heads must be able to form a word with T$^o$, v$^o$ and V$^o$ (13a). Alternatively, if they are adjuncts, then no complications arise since no heads intervene (13b).
Let’s turn to the second module of the present theory of HM: Positioning. Labels aside, the structure assumed for a complex word like (11a) is identical to that assumed for an XP-chain (14). Complex words and movement chains give rise to a problem: What is the (unique) linear position of an item occupying several hierarchical positions simultaneously? Assuming that positions in a word/chain can be strong (s) or weak (w), the linearization problem has the general solution (16) in a single output syntax (Bobaljik 1995b; Brody 1995; Gärtner 2002; Groat and O’Neil 1996; Kayne 1998)

(16) Positioning Generalization
Pronounce an element E (a word or a chain) in the lowest position such that all higher positions P’ of E are weak.

(16) says that an element is pronounced in the highest strong position. If there is no strong position, it is pronounced in the lowest position, i.e. the base position (cf. Brody 2000; Gärtner 1999; Abels, under review). The formulation allows for multiple strong positions within a chain or word.

The difference between wh-in situ and wh-movement can be captured easily. (15a) is the case of a wh-in situ language: the wh-element moves to SpecCP, but the high position is weak. According to the Positioning Generalization, the wh-chain is realized in the base position in (15a). In
(15b) on the other hand, the high position is strong and the wh-chain is phonologically realized in the high position.

The difference in terms of V-to-T movement between English and French (cf. Emonds 1978; Pollock 1989) is captured in the same way. In both language the verb and T form a word: #V+v+T#. In French T° is strong (#V+v+Tg#), but in English T° is weak (#V+vTw# — following Koizumi 1995; Lasnik 1995, I assume that v° is strong in English). This is the only relevant difference between the languages.

Note how this view solves the problems for theorem (6) discussed above. MWF obeys the strict locality usually associated with HM, MWF obeys the non-excorporation condition, but Positioning, crucially, does not give rise to the The-higher-the-bigger-theorem (cf. also Zwart 2001 for a similar proposal). 6

We can summarize the results so far as follows. The Mirror Theoretical view of HM does not entail the problematic theorem (6), i.e., downward HM is allowed. HM arises as the combined effect of post syntactic MWF (a possibility inherent in the standard theory, cf. Chomsky and Lasnik 1993) and post syntactic positioning.

3 *{Q/Adv} — E-site

This section discusses Sag’s generalization (1) in some detail. After noting a problem for (1), we discuss Oku’s (1998) and Sag and Fodor’s (1995) explanation of (1) and show that they run into additional problems. Before that though, we will broaden the array of data under consideration somewhat.

The auxiliary adverb order is not the only factor interacting with VPE. Baker (1971, 1981) notes that the stress level of auxiliaries also interacts with ellipsis and word order (cf. also Wilder 1997). The fact is that unstressed (tensed) auxiliaries never follow adverbs ((17) and (18)). 7

(17) John always {∀h[æ]s; * h[ə]s; * ‘s} been rude to grandpa.

---

6 This raises the question whether syntactic HM exists at all. In Abels’ (in press) formalization of Brody’s (1997, 2000) Mirror Theory, it follows as a theorem of phrase structure that syntactic HM cannot exist.

7 There is some discussion whether the converse also holds, i.e., whether stressed auxiliaries can precede adverbs. The answer seems to be ‘yes’ (Baker 1981 with reservations; Ernst 1983, Sag, 1980; Wilder 1997 - contra Baker 1971): (i)-(ii).

(i) John {∀h[æ]s; h[ə]s; √ ‘s} always been rude to grandpa.

(ii) √She said that they’d all read Moby Dick, and they HAD all read it.
(18) He said he would study karate, and study karate he surely $\{\sqrt[h]{\text{he Ya}}; s; *\text{he Ya}^*\}.$

A theory of auxiliary adverb word order has to account for this fact along with Sag's generalization.

### 3.1 Sag's Generalization

Recall Sag's generalization (1), which was exemplified above in (2)-(5). Is Sag's generalization empirically correct? Superficial counterexamples are easy to construct (19). Examples like (19) are only apparent counterexamples, however. Sag's generalization says that that adverbs and floated quantifiers cannot precede the E-site. In (19) the adverb actually follows the (silent) E-site (20). The fact that (20) is the correct analysis of (19) can be shown by replacing *slowly* by an adverb that cannot appear in VP-final positions such as *hardly* (21). Example (22), which is parallel to (19) is ungrammatical. (19) is therefore not a true counterexample to Sag's generalization.

\[
\begin{align*}
(19) & \checkmark \quad \text{John writes fast, but Peter does slowly.} \\
(20) & \quad \text{John writes fast, but Peter -s write slowly} \\
(21) & \quad \text{Ian's theory has {*surprised me hardly; }\sqrt{\text{hardly surprised me}}}. \\
(22) & \checkmark \quad \text{Ed's theory has utterly surprised me, but Ian's theory has hardly.} \\
(23) & \checkmark \quad \text{Ed's theory has utterly surprised me, but Ian's theory hardly has.}
\end{align*}
\]

The next set of examples (from Baker 1981) are true counterexamples to Sag's generalization. (25) shows that *always* cannot appear clause finally. In (24) *always* thus immediately precedes the E-site in violation of (1). The same is shown for floated quantifiers in (26) and (27). The crucial factor seems to be the presence of negation in (24) and (26). Examples (28)-(30) show that matrix interrogatives can also violate Sag's generalization.

\[
\begin{align*}
(24) & \checkmark \quad \text{Fred has sometimes been rude to Grandpa, but he hasn't always.} \\
(25) & \quad \text{Fred has sometimes been rude to Grandfather, but he hasn't been rude to Grandfather always.} \\
(26) & \quad \text{Some of them are working on the assignment, but they aren't all.} \\
(27) & \quad \text{Some of the boys are working on the assignment, but they aren't working on the assignment all.} \\
(28) & \checkmark \quad \text{Some of the boys are working on the assignment. — Are they all?} \\
(29) & \quad \text{John hasn't gotten along with Grandpa lately. — }\sqrt{\text{Has he ever?}} \\
(30) & \quad \text{John hasn't gotten along with Grandpa lately. — ?* Has he gotten along with Grandpa ever?}
\end{align*}
\]
Sag’s generalization is too strong. Informally speaking, the auxiliary can be drawn away from its position immediately preceding the E-site by negation and by the interrogative complementizer.

3.2 The Radical Emptiness Account of Sag’s Generalization

Oku (1998) and Sag and Fodor (1995) attempt to derive Sag’s generalization. Although coming from very different theoretical viewpoints, the explanation is the same: Sag’s generalization holds because (at the relevant level of representation) the E-site is literally empty. If the E-site is empty, there is no structure the adverb and floated quantifiers could possibly adjoin to. Since adverbs and floated quantifiers need a host, they must adjoin higher, placing them to the left of the auxiliary. The account is appealingly simple.

The counterexamples to Sag’s generalization involving negation and matrix interrogatives, noted above, remain problematic. There are several further classes of examples that pose a challenge to the account of Sag’s generalization in terms of radical emptiness.

The first problem is posed by pseudogapping (cf. Levin, 1978, 1979/1986). Pseudogapping has been analyzed as a kind of VPE (Jayaseelan 1990; Johnson 1996; Lasnik 1995). In typical pseudogapping examples like (31a), there is clearly some structure present in the E-site: her essay. Yet the auxiliary must follow the adverb (31b).

(31) a. Joe quickly ripped up his paper, and Sue slowly did her essay.
   b. *Joe quickly ripped up his paper, and Sue did slowly her essay.

This paradigm cannot be accounted for by appealing to radical emptiness. To avoid this problem, proponents of radical emptiness would have to offer a convincing alternative analysis of pseudogapping.

The next two sets of examples (from Abels, under review) show that radical emptiness comes at the cost of enriching the theory of adjunction by ad hoc assumptions. Adverbs like completely cannot usually precede tensed auxiliaries (32). In fact Jackendoff (1972) claims that completely is within VP (cf. also Lasnik, in press; Oku 1998). However, under VPE the order completely > auxiliary becomes not only possible but obligatory as shown in (33). Under the radical emptiness assumption, (33b) is expected to be ungrammatical. But why is (33a) acceptable? If the range of categories an adverb can adjoin to is stable across constructions, the grammaticality of (33a) remains mysterious under the radical emptiness approach. A construction specific theory of adjunction sites must be invoked.
Some adverbs receive different readings depending on their position, as shown for *happily* in (34). Crucially the order *adv>**aux* forces the speaker oriented reading (34a). The VPE example (35) is a counterexample to the general pattern, since both readings are available. Radical emptiness demands that *happily* in (35) is adjoined in the same position as in (34a). This makes the false prediction that (35) only has the speaker oriented reading.

(34)a. Adv > Aux  
    John happily will return to his village.  
    √ speaker oriented reading, ?* manner reading

b. Aux > Adv  
    John will happily return to his village.  
    ? speaker oriented reading, √ manner reading

(35) John will return to his village and Bill happily will, too.  
    √ speaker oriented reading, √ manner reading

All of this suggests that the radical emptiness approach is wrong. Moreover, the data make sense if the base position of the auxiliary is above the relevant adverbs. The auxiliary would then move below the position of those adverbs in VPE contexts including pseudogapping, but could be attracted to a higher position if negation or the interrogative complementizer C₀ is present. The next section develops such an account.

4 The Account

The theory from section 2 accounts for the facts if the following additional assumptions are made: (i) Abstract heads need to be licensed. I will assume that one licensing mechanism for abstract heads is to form a word with overt material. Some principle regulating the distribution of empty heads is necessary if syntax is to have empirical content. Under the name of ECP, the principle regulating the occurrence of abstract elements was at the center of syntactic debate in the eighties. The interest has unfortunately subsided, but I will assume that word formation is part of an ultimate theory of licensing of abstract elements. (ii) Ellipsis and movement target only phrasal categories. (iii) Every phrase has a head (endocentricity).

The structure underlying simple clauses with adverbs is shown in (37).
The functional head hosting the adverb is designated as X° in the
diagram. X° can form a word downward with the verb giving rise to the two
words shown in (36a). In this case the resulting order is aux>adverb.

(36)a. #have+T°#  
     #study+AgrO°+v°+X°#

b. #X°+have+T°#  
     #study+AgrO°+v°#

However, X° can also form a word upward (36b). Since T°, have, and
X° are all weak, #X°+have+T°# is pronounced in the lowest of the three
positions, i.e., has follows the adverb. We can easily capture the fact that
auxiliaries are always stressed when following adverbials by assuming that
X° has the morpho-phonological reflex of blocking auxiliary contraction.8

Blaming X° for obligatory stress on the auxiliary derives certain excep­
tions from Baker’s generalization. Clearly, sentential adverbs like probably
can occur above negation (38a). Assume that negation is positioned right be­
low T° but above haveP (39b). (38a) then shows that probably is above ne­
gation and hence also above X°. With X° missing in (38b), aux can contract.

(37)

(38)a. √ John probably hasn’t forgotten about it.
   b. √ John probably ’s forgotten about it.

---

8 Given that distressing of the auxiliary is impossible before any E-site, X° might in
fact be implicated in licensing VPE.
We now turn to VPE. Suppose that VPE is PF-deletion. The underlying structure is always that shown in (37). In principle both word formation patterns in (36) are again available. As before, the order aux>adverb is derivable only if (36a) is chosen. However, if this pattern is chosen, VPE will elide part of a word: #study\text{+AgrO+v+X\#} will be partly elided: #study\text{+AgrO+v+X\#} or #study\text{+AgrO+v+X\#}. This can be ruled out in several different ways. First of all, X° might fail to be licensed in this configuration since it does not form part of a non-abstract word. Second, we might assume that there is no operation deleting parts of words. The adverb > aux order is obtained straightforwardly by choosing (36b) and eliding vP. No sub-word unit is elided now. X° is integrated into the auxiliary and thus licensed. We thus have an account of Sag’s generalization.

The counterexamples to Sag’s generalization (negation—(24) and (26) and interrogatives—(28) and (29)) can be explained if we assume that T° is strong when its complement is NegP and that C°Q° is strong. The structure for interrogatives is shown in (39a). Where other theories posit HM, the current theories posits word formation, i.e. T-to-C movement is modeled as word formation of T with C. Since vP is elided in (29), X° must form a word upward. The resulting word #X°+have+T°+C°\# correctly predicted to be pronounced in the highest strong position by (16), i.e. in C°.

(39)a. [CP C\text{Q_S [TP he T°_w [have_P have_w [XP always X°_w [vp ... ]]]]]]
   b. [TP he T°_S [NegP n’t Neg°_w [have_P have_w [XP always X°_w [vp ... ]]]]]

(39b) shows the structure of an example with negation, namely (24). Since vP is elided, X° must again form a word upwards with have, Neg°, and T°. By assumption Neg° renders T° strong, thus the resulting word #X°_w+have_w+Neg°_w+T°_s\# is pronounced in T°.

The assumption that the strength or weakness of a head-position is not an inherent property of that head but is determined by its local context needs justification. Although this assumption is non-standard, it is not particularly strange. In fact for XP-chains it is the standard assumption. Whether a position in a wh-chain, for example, counts as weak or strong is determined by the host not the wh-element itself: the attractor determines strength.\footnote{For additional discussion of this point cf. Abels (in press).}

I assume generally that whether a head position counts as strong or weak is determined under MWF locally by an inherent property of the head of its complement. Thus if \(\alpha°\) and \(\beta°\) form a word and \(\beta°\) is the head of the complement of \(\alpha°\), then \(\beta°\) determines whether \(\alpha°\) counts as strong. This leaves the lowest head in a word without a value for strength. This is irrelevant...
though, as a glance at (16) reveals. Whether the lowest position in a word or a chain is strong or weak never matters. (16) asks whether positions higher than P are weak or strong. Whether P itself is weak or strong is not relevant.

We now turn to pseudogapping. The pseudogapped version of (37) is

Fred surely has karate. Karate can survive deletion only if vP and AgrOP survive deletion, i.e. if only studyP is deleted. Since v° and AgrO° are abstract, they need a licensor. The following word formation is forced: #AgrO° +v°S+X°W+haveW+T°W#.10 Together with (16) this gives the correct result: adverb > aux > object. A question arises at this point: Why is it impossible to form the two words #AgrO°+v°S+X°W+haveW+T°W# and #study# in the absence of ellipsis? If it were possible, the sentence John surely has karate studied would be acceptable. The answer is that #study# is morphologically ill-formed. Words are entered into the syntax as bare roots without category information. Morphology can only interpret words with categories. Category is assigned to the root by v° (Marantz 2001). Thus, in sentences where the root is not elided, it has to form a word at least with v°. This morphological constraint is irrelevant if the root is elided.

The remaining cases are straightforward. Low adverbs (e.g. completely and happily_manner) are inserted in SpecYP below v° (40). In non-elliptical sentences the root forms a word with Agr°, Y° and v°. Agr°, in this case, makes Y° strong not v°. The correct word order Sue completely solved the problem is derived. In elliptical sentences AgrOP is elided. The remaining heads form a word: #Y°+v°+T°#. According to (16), this word (i.e. do, the spellout of lone T°) is realized in Y°, correctly predicting the order adverb > auxiliary.11

(40) [TP Sue T°W [vP Sue v°W [YP completely Y°S [AgrOP ...]]]]

The account presented here makes a further prediction. If the root of the verb is not elided but moved, low adverbs like completely should always move along and never stay behind. Higher adverbs should have more freedom. Both expectations are borne out as (41) and (42) show.

(41) He said he would completely solve the problem, and
   (\^completely) solve the problem he (?*completely) did.

10 In light of the preceding discussion, AgrO° does not have a value for strength. AgrO° makes the next higher head (v°) strong, hence, verb > object order. Having strength determined extrinsically by the next lower head accounts nicely for Case adjacency in English (cf. also (40)), a fact that's otherwise hard to explain.

11 Alternatively, low adverbs might be adjoined to vP. This would not change the account. The text version is preferable, because of its uniform treatment of adverbs.
(42) He said he would always love her, and (always) love her he (always) did.

Sag’s generalization, the counterexamples to it, and even the problems for radical emptiness all fall under the present, simple theory.

5 Conclusion

Two main hypotheses make the account offered in this paper possible: (i) Theorem (6) of HM is rejected; and (ii) MWF and Positioning are post-syntactic processes. Affix hopping was shown to be inadequate to the task of dealing with violation of the The-higher-the-bigger-Theorem. The idiosyncrasies of Affix Hopping are reinterpreted as morphological selection.

The fast argument against the existence of traces of movement and ellipsis (Sag and Fodor 1995), i.e. radical emptiness, is dispelled. There is necessarily some abstract structure present at the ellipsis site. The present theory makes precise what it may mean for HM to be a PF-phenomenon (Chomsky 2000) without duplicating syntactic movement operations in PF.

References

Pensalfini and Hiroyuki Ura, 41-64. Cambridge, MA: MITWPL, Department of Linguistics and Philosophy, MIT.


AUXILIARY ADVERB WORD ORDER REVISITED


Department of Linguistics
University of Connecticut
337 Mansfield Road
Storrs, CT 06269-1145
klaus.abels@uconn.edu