DO-SUPPORT AND TIER SCANSION:
A MINIMALIST ACCOUNT OF HMC EFFECTS

HITOSHI AKAHANE
McGill University

In this article, we discuss HMC effects, notably the timing of verb movement and do-support in English, within the Minimalist framework. To derive HMC effects (and ultimately, RM) from a basic mechanism of Attract/Move, we introduce some autosegmental theoretic notions, such as feature organization, adjacency and tier scansion and combine them with Attract/Move. Analyzing do-support as an HMC effect, we argue that Neg(ative)\(^0\) has a feature which attracts V\(^0\)'s feature. We also discuss the historical change in the feature specification of English Neg\(^0\) and parametric differences in Neg\(^0\)'s feature. Finally, it is suggested that a more general condition replace Chomsky's (1995: Ch. 4) MLC.*

1. Introduction

Locality of head (X\(^0\)) movement has been vigorously discussed within contemporary generative grammar since Travis (1984) presented the generalization called the Head Movement Constraint (HMC). The HMC is defined as in (1) (for the original definition, see Travis (1984: 131)):

\[
\text{(1) An X}^0 \text{ may only move into the Y}^0 \text{ which takes XP as a complement in a standard X-bar structure.}
\]

A consequence of the HMC is to impose strict successive cyclicity on X\(^0\) movement. Several authors including Chomsky (1995: Ch. 2), however, have remarked that the HMC is not only particular to X\(^0\) movement but also empirically too strong. It is descriptively valid only for cases which fall under the Empty Category Principle (ECP) (cf. Baker * For helpful comments and suggestions I would like to thank Jonathan Bobaljik, Nigel Duffield, Masaharu Katoh, Mikinari Matsuoka, Ileana Paul, Glyne Piggott, Lisa Travis, and two anonymous EL reviewers. I am also grateful to Nancy Jones for suggesting stylistic improvements. All remaining errors and inadequacies are my own. This research was supported by FCAR grant 94-ER-0578.

© 1998 by the English Linguistic Society of Japan
In a similar direction, Rizzi (1990) suggests the possibility that the HMC is a subclass of Relativized Minimality (RM).

More recently, in the Minimalist Program (cf. Chomsky (1995: 307)), it is not entirely clear whether we can defend the HMC. Within this framework, Last Resort (or “Suicidal” Greed) given in (2) is proposed as a definitional condition on the operation Attract/Move:

\[
(2) \quad \text{Move } F \text{ raises } F \text{ to target } K \text{ only if } F \text{ enters into a checking relation with a sublabel of } K. \quad \text{(Chomsky (1995: 280))}
\]

Under Last Resort, movement into an H⁰ never takes place when there is no attracting feature in H⁰, and it can be skipped. As noticed, there would be no independent locality condition which requires strict successive cyclic movement on its own and, at the same time, is still compatible with (2). Note that Chomsky (1995: Ch. 4) proposes the Minimal Link Condition (MLC) as another definitional condition on Attract/Move, which can derive certain locality effects. I will discuss this in sections 3.2 and 6. Under the Last Resort with the MLC, long distance X⁰ movement can still ensue, however, unless a continuous sequence of targets exist which have features attracting the formal features of the X⁰.

In this article, we will consider phenomena which could be identified as HMC effects: such phenomena reflect some strict locality condition on X⁰ movement like the HMC. Among such phenomena, I will especially focus on the timing of verb movement and do-support in English, which apparently conceal HMC effects. Needless to say, the aim of this article is not to justify the old HMC or the old ECP within the Minimalist Program. Rather, in order to derive HMC effects (and ultimately, RM) from a basic mechanism of Attract/Move, I will introduce some notions such as feature organization, adjacency and tier scansion (a mechanism operative at some level in feature organization) which are well-established in autosegmental phonology, and combine them with recent Minimalist assumptions like feature movement. Analyzing do-support as an instance of HMC effects, I will argue that a functional category Neg(ative)⁰, intervening between T(ense)⁰ and V⁰, has a feature which attracts V⁰'s feature. I will also discuss the historical change in the feature specification of English Neg⁰ and parametric differences in Neg⁰'s feature in a comparative context. Finally, it will be suggested that the MLC be replaced in favor of another condition which invokes feature organization and tier scansion for Attract/Move.
2. Some Theoretical Assumptions

For the basis of the analysis, I will adopt some recent developments in the Minimalist Program, taking Chomsky (1995: Ch. 4) as a general background. First, I assume that economy considerations must be local (Chomsky (1995: Ch. 4; 1996), Collins (1997)): economy of derivation must be evaluated for each single step in a single derivation; thus, we could never compare two or more derivations nor look ahead in a single derivation. Pursuing this premise, we should eliminate the principles and theoretical constructs which reflect global economy. Among them are Procrastinate, the economy principle which prefers covert operations to overt ones, and Numeration, the repository of lexical items which fixes the ingredients of the derivation to evaluate global economy, both of which are assumed in Chomsky (1995: Ch. 4). Basically adopting Collins’ (1997) view, I do not maintain Procrastinate nor Numeration in the following discussion.

Elimination of global elements is strongly supported within the Single Level Model (S-Model) put forth by Groat and O’Neil (1996, henceforth G&O) (cf. Bobaljik (1995), Brody (1995)). In the standard Minimalist Program, syntactic computation is divided into overt and covert components by the operation Spell-Out. This division entails some serious asymmetries between the two components, as listed by G&O and others: e.g. locality of movement, (im)possibility of counter-cyclic movement, units of movement (features vs. categories). Under the S-Model, on the other hand, such asymmetries can be reduced since syntactic computation is all carried out in a single component and the same output serves as the input to both conceptual-intentional systems and sensorimotor systems. In the S-Model, there cannot be a global economy principle like Procrastinate simply because there is no covert component. For the same reason, it is not necessary to stipulate that lexical insertion takes place only in the overt component. This, in fact, weakens the motivation for a Numeration as the latter no longer has to guarantee that the LF and PF representations be compatible. Concerning lexical insertion, I assume that any lexical item comprising formal features is dealt with in the syntax: that is, lexical inser-

---

1 Chomsky (1995: 292) points out that an empty complementizer at the root in declarative sentences might be a possible exception to overt lexical insertion.
tion (including do-support) takes place only in the syntactic computation. For the above and other advantages, I will adopt the S-Model.

In arguing that syntactic computation proceeds in a unified way with no overt-covert asymmetry, G&O ascribe the apparent timing of movement (before or after Spell-Out) to the position to be pronounced in a chain (head or tail). Distinct from G&O, I will maintain the strong-weak asymmetry in Chomsky (1995: Ch. 4): category movement is driven by strong target features whereas feature movement is driven by weak target features. This will be needed in discussing $X^0$ movement, especially do-support, as will become clear as we proceed. As for feature movement, Chomsky (1995: 265) suggests that a formal feature F “carry along” other formal features (or free-riders) when it moves. He argues that such pied-piping of formal features is automatic, while broader pied-piping of categories with phonological features takes place only when it is required by PF convergence. Free-riders may, but need not, enter into a checking relation with features in the target. When we follow the spirit of Last Resort in (2), free-riders are not desirable. In addition, it is not clear why feature pied-piping must occur automatically. I argue rather that feature movement should only apply to the feature which is attracted by the target feature. If there is no dependency among formal features, we cannot expect any free-riders. But if there is some hierarchical dependency among features, we can expect the effect of feature pied-piping: that is, features dependent on a higher feature F are automatically pied-piped when F is attracted. Such hierarchical dependencies are traditionally captured by feature organization in autosegmental phonology. I argue that automatic feature pied-piping in syntax also follows from feature organization (see section 4.1).

2 As noted by Chomsky (1996), phonological features might be added after Spell-Out in morpho-phonological processes (cf. Halle and Marantz (1993)).

3 Chomsky (1995: 262) proposes the economy condition in (i):

\[(i) \quad F \text{ carries along just enough material for convergence.}\]

As noticed by Collins (1997), (i) is a global economy condition since it refers to (PF) convergence. In Chomsky (1996), it is suggested that category movement should be driven by bare output conditions (requirements from sensorimotor and perhaps other performance systems) rather than by the strength of formal features in targets. For now, I keep strong features but do not take broader pied-piping to be due to (i).
3. *Do*-Support as an HMC Effect

3.1. Emonds-Pollock Paradigm: Timing of Verb Movement

With the above assumptions, we address the problem of the Attract/Move approach to HMC effects. Before discussing the problem, we review a well-known paradigm which shows that English and French differ as to the relative position of a main verb in a sentence with respect to a negative/VP-initial adverb:

(3)

a. We don’t go to church.
b. We often go to church.
c. *On (ne) pas va à l’église.
d. *On souvent va à l’église.

(4)

a. *We go not to church.
b. *We go often to church.
c. On (ne) va pas à l’église.
d. On va souvent à l’église.

Following Emonds-Pollockian tradition (Emonds (1978), Pollock (1989)), in negative sentences, the position of *not (-n’t) and *pas is crucial for determining the position of a finite verb. Belletti (1990), Pollock (1989), Roberts (1993) and others suggest that the English negative suffix -n’t and the French negative clitic ne are generated in Neg⁰, and negative adverbs like English not and French pas in [Spec, Neg]. I essentially assume this, but I treat -n’t as an inflectional suffix attached to tensed auxiliary verbs (including do) and be in the lexicon (cf. Zwicky and Pullum (1983)). In (3) and (4), it is shown that a main verb follows a negative/VP-initial adverb in English whereas a main verb precedes an adverb in French.⁴ According to Emonds’ (1978) analysis, French has finite verb raising, but not affix lowering which has often been proposed for English. Assuming this contrast, Pollock (1989) argues that French finite main verbs overtly move to T⁰ in a structure like (5) while English counterparts with affix lowering do not:⁵

⁴ Contrastive contexts (e.g. go not to A but to B) are irrelevant in (4a, b). And, this article will not deal with French infinitives (see Pollock (1989)).

⁵ Although Agr plays an important role in Pollock (1989) and Chomsky (1995: Chs. 2, 3), I assume, following Chomsky (1995: Ch. 4), that there are no Agr projections. Furthermore, since the internal structure of VP is not crucial in the following discussion, I use VP for verb phrase (rather than vP, the layered VP suggested by Chomsky (1995: Ch. 4)) and V⁰ for its head, regardless of the type of verb.
(5) \([\text{TP} \text{DP} T^0 [\text{NegP} (\text{not/pas}) [\text{Neg'} \text{Neg}^0 [\text{VP} (\text{Adv}) V^0 \ldots \text{]]}]])\]

On the basis of Pollock's analysis, Chomsky (1995: Ch. 2) in turn argues that English main verbs covertly move to T^0 along with the inflectional affixes which are overtly lowered from T^0 to V^0. In the Minimalist Program, however, one of Pollock's (1989) and Chomsky's (1995: Ch. 2) core assumptions is rejected: under no circumstances is lowering permitted. Instead, the Minimalist Program adopts a strong version of the lexicalist hypothesis in which lexical items are already inflected when they are introduced into syntactic computation. Inflectional, or formal, features must enter into a checking relation with corresponding features of a functional category. Suppose that T^0, rather than Agr^0 (see fn. 5), is assigned a value for the strong-weak parameter which is originally proposed by Chomsky (1995: Ch. 2) (cf. Pollock (1989)). In the standard Minimalist Program, a strong feature must be checked before Spell-Out for convergence (for specific discussion, see Chomsky (1995: Chs. 3, 4) and others) while the checking of a weak feature is postponed until the covert component, as required by Procrastinate. Yet, as assumed in section 2, I reject Procrastinate, and the difference in the timing of verb movement is converted into the difference in movement unit, namely, category versus feature. As traditionally argued in the Minimalist literature, French chooses a strong inflectional feature (now in T^0) and English a weak one; the former triggers category movement and the latter feature movement. Thereby follows the difference in the relative order of a finite main verb in the two languages, as shown in (6):

(6) a. \([\text{TP we T}^0 [\text{NegP Neg}^0 [\text{VP often go to church}]]) \quad (=\text{(3b)})\)

\[\hline\text{..............................F}\]

b. \([\text{TP on va-T}^0 [\text{NegP Neg}^0 [\text{VP souvent t} \text{ à l'église}]]) \quad (=\text{(4d)})\]

3.2. Do-Support and Feature Movement

In the structures in (5) and (6), Neg^0 intervenes between T^0 and V^0. For polarity interpretation at or after LF, it seems plausible to assume the existence of NegP in affirmative sentences as well as in negative sentences.\(^6\) In (6), however, we tacitly ignore Neg^0, which allows us to

---

\(^6\) As observed by several authors (Pollock (1989), etc.), the affirmative adverb so as in (i), just like the negative adverb not, can occupy the Spec of a functional
consider that a finite main verb or its feature moves directly to T⁰. Skipping the intervening Neg⁰, the V⁰-to-T⁰ movement in (6b) violates the HMC. If the HMC (or some similar locality condition) constrains feature movement, it is also violated in (6a). In English negative sentences, on the other hand, do-support must be invoked when there is no finite auxiliary or be as in (3a). This fact might suggest that the HMC violation resulting from the feature movement from V⁰ to T⁰ is, in a sense, evaded by do-support. Here, let us consider how we could deal with do-support in terms of pure feature movement.

I am assuming that a weak target feature triggers pure feature movement (without pied-piping in Chomsky's (1995: Ch. 4) sense). Suppose that the feature moving from V⁰ to T⁰ in (6a) is the tense feature, [tns]. The T⁰'s feature attracting [tns] from V⁰ is [TNS] (hereafter, attractor features are given in upper case and attracted ones in lower case; when strength is at issue, strong features are presented like [Fs] and weak ones like [Fw]). In (6a), [tns] moves alone to check [TNSw] so that do-support does not obtain. In (3a), however, this seems to be blocked as illustrated in (7):

\[ \text{(7) } [\text{TP } T⁰ [\text{NegP Neg}⁰ [\text{VP ...V⁰...}] ] ] \]
\[ [\text{TNSw}] \leftarrow \cdots \text{*}\cdots [\text{tns}] \]

If [tns] movement is not blocked in (3a), examples like (8) without do-support should be acceptable, contrary to the fact:

\[ * \text{We not go to church.} \]

In Chomsky (1995: Ch. 4), the MLC, proposed as a definitional condition on Attract/Move, is considered to impose locality on movement, and is defined as in (9):

\[ (9) \]
\[ a. \ K \text{ attracts } α \text{ only if there is no } β, \ β \text{ closer to } K \text{ than } α, \text{ such that } K \text{ attracts } β. \quad \text{(ibid.: 311)} \]
\[ b. \ β \text{ is closer to } K \text{ than } α \text{ if } β \text{ c-commands } α. \quad \text{(ibid.: 358)} \]

Informally, where there is more than one candidate which the target K can potentially attract, K only attracts the highest one in the c-command configuration. With the MLC, the movement of V⁰'s [tns] to T⁰ will be blocked in case Neg⁰ also has [tns], as Neg⁰ is closer to T⁰ than V⁰. However, Neg⁰ does not seem to be specified for [tns]. If so, the category identical or akin to NegP, which might support the projections of Neg in affirmative sentences:

\[ (i) \text{ He did so faint.} \]
MLC does not account for the question of why \( \text{tns} \) movement is blocked in negative sentences, i.e., why (8) is ungrammatical.

Another question is why do-support is obligatorily applied in (3a). As a possibility, a feature in Neg\(^0\) attracts a feature of (main/auxiliary) verbs. This is the direction that we will pursue in this article. In French, finite verbs are likely to carry a feature to be checked against Neg\(^0\)'s feature. This is actually motivated by the fact that finite verbs in French can host the negative clitic *ne* as in (4c), although *ne* is optional or, perhaps, alternative with a zero form in colloquial French. For the *ne*-hosting ability, French verbs can have a feature, [\(+/-\text{ne}\)], which is checked against the feature in Neg\(^0\), \([+/−\text{NE}]\) ([\(+\text{NE}/\text{ne}\)] is specified in negative sentences and \([-\text{NE}/\text{ne}\)] in affirmatives). In English, negative suffix \(-n't\) can be attached to finite *be* and auxiliaries including *do* but not to finite main verbs. I then assume that *be* and auxiliaries can have \([+/(−)\text{ne}]\) to check against Neg\(^0\)'s \([+(−)\text{NE}]\) while main verbs cannot (cf. Roberts (1998)). With these assumptions, we now have the following structure for (3a):

\[
\begin{array}{c}
\text{(10) } \left[ \text{TP we T}^0 \left[ \text{NegP Neg}^0 \left[ \text{VP go to church} \right] \right] \right] \\
\quad \left[ +\text{NE} \right] \\
\quad \left[ \text{TNS}_w \right]
\end{array}
\]

Since English main verbs cannot have \([+\text{ne}]\), *do*(\(-n't\)) with \([+\text{ne}]\) must be inserted into Neg\(^0\)[\(+\text{NE}]\): otherwise, unchecked \([+\text{NE}]\) will cause the derivation to crash. Here, as assumed in section 2, *do*(\(-n't\)) is regarded not as an expletive element in the classical sense which is inserted post-syntactically, but as a pure lexical item which comprises formal features to be checked syntactically. Suppose then that \(\text{don't}\), which is inserted into Neg\(^0\), carries \([\text{tns}]\) as well as \([+\text{ne}]\) while the main verb *go* does not carry \([\text{tns}]\):

\[
\begin{array}{c}
\text{(11) } \left[ \text{TP we T}^0 \left[ \text{NegP don't-Neg}^0 \left[ \text{VP go to church} \right] \right] \right] \\
\quad \left[ +\text{ne} \right] \left[ +\text{NE} \right] \\
\quad \left[ \text{TNS}_w \right] \left[ \text{tns} \right]
\end{array}
\]

In (11), \([+\text{NE}]\) checking is carried out. Subsequently, \([\text{TNS}]\) attracts \([\text{tns}]\), raising don't(-Neg\(^0\)) to T\(^0\). In claiming this, it is necessary to clarify whether \([+\text{NE}]\) (and \([\text{TNS}]\)) in English are strong or weak. I will address this question in section 4.3. We have another problem. Since \([\text{tns}]\) can be assigned to main verbs as well as *do*, we could consider structures like (12):

\[
\begin{array}{c}
\text{(12) } \left[ \text{TP we T}^0 \left[ \text{NegP don't-Neg}^0 \left[ \text{VP go to church} \right] \right] \right] \\
\quad \left[ +\text{ne} \right] \left[ +\text{NE} \right] \\
\quad \left[ \text{TNS}_w \right] \left[ \text{tns} \right]
\end{array}
\]
In (12), the main verb went has [tns] but do does not, irrespective of the (non)existence of nonfinite dummy do (instead of tenseless do, we can consider infinitival to which does not carry [tns] but [+ne]). If movement of the relevant feature by itself suffices, (12) should be possible, where [tns] alone moves to T0 from V0, just as in (6a). However, this is not the case. As the MLC only cares about the features which the target attracts, it does not provide account for the imposibility of (12) under the assumptions made in section 2. The problem here is that we have no specific locality condition which constrains feature movement.\(^7\) We could not find such a locality condition as long as we only look at the relevant features as in (12). To define locality for feature movement, we need to know which level is pertinent to the operation: for example, the category level (XP/X0), the lowest feature level, or some intermediate level. In the next section, we will see that what is crucial is not the lowest feature (nor the highest category) level, but the intermediate level right above, or one-node-up, the lowest, to define locality for feature movement. I will introduce some notions which are well-motivated in autosegmental theory of generative phonology.

4. Tier Scansion and HMC Effects

In this section, I will argue that do-support and other verb movement phenomena actually involve a certain locality condition, which derives HMC effects in a very interesting way. As I mentioned in section 1, I do not consider the HMC (or the ECP) an independent principle. I argue that a locality condition deriving HMC effects is included in the definition of Attract/Move, just like Chomsky's (1995: Ch. 4) MLC. However, the condition I propose differs from the MLC, since I will propose that it is sensitive to a certain level of feature organization.

\(^7\) One EL reviewer points out that (12) could be captured by Bobaljik's (1995) adjacency condition on affixation in Morphology. However, his adjacency seems rather questionable in ignoring certain intervening constituents like VP adverbs:

\[(i)\] Stanley T0[PST] completely ate his Wheaties.
The condition to be proposed concerns not only feature movement but also category movement. Moreover, it will generalize locality in syntax and phonology under the same (metatheoretical) notion *adjacency*, and ultimately substitute for the MLC (see section 6).

4.1. Scan as a Subcomponent of Attract/Move

Let us make the proposal more specific. The locality condition I propose below is built into the process of tier scansion (*Scan*), which is taken to be a subcomponent of Attract/Move. As practiced in autosegmental theory, suppose that features are hierarchically organized and, as argued by Archangeli and Pulleyblank (1987, henceforth A&P) and others, an operation (in the present case, Attract/Move) scans the level, or tier, of the relevant feature or class node grouping features. I explicitly interpret Scan in the following way:

(13) An operation whose trigger is F scans the tier immediately dominating F.

For Attract/Move, the feature which triggers the operation is in the target: the trigger and the target are roughly the same. In contrast, for feature spreading (*Spread*) in phonology, the trigger feature is not in the target: the trigger and the target are separated from each other (as we will see in (16)). (14) schematizes (13):

(14) \[ \text{Scan} \Rightarrow A \, B \, C \]

\[ \text{F} \quad \text{F} \quad \text{F} \]

The nodes A, B and C, which are on the same tier, *provide immediate access to* the tier of the feature F: put differently, they immediately dominate F. In accordance with (13), the tier to which A, B, C belong is scanned. Importantly, what the operation (e.g. Attract/Move) really affects is F rather than the node immediately dominating it: the latter is only scanned. The locality condition is defined as follows:

(15) An operation can apply only if \(\alpha\) is adjacent to target K on the scanned tier \(\tau\).

I will call (15) the *Generalized Locality Condition (GLC)* since it constrains both Attract/Move and phonological operations like Spread. In the autosegmental theory, the adjacency at issue is defined on a linear

---

8 (13) is adapted from A&P's minimal scansion (p. 25). A&P parametrize scansion as to whether an operation pertains to syllable structure (maximal scansion) or feature configuration (minimal scansion).
string: $\alpha$ is adjacent to target $K$ on the scanned tier $\tau$ iff there is no $\beta$ which intervenes between $\alpha$ and $K$ where $\alpha$, $\beta$ and $K$ are on $\tau$. Thus, in (14), if $A$ is the trigger and $B$ and $C$ are potential targets, $F$ under $A$ cannot spread to $C$ skipping $B$ because $A$ is adjacent to $B$, not to $C$. We can find such an example in back harmony in Mongolian, analyzed by A&P as in (16), where irrelevant tiers are omitted:

\[(16) \quad /z\ a\ x\ i\ r\ +\ e\ x/ \Rightarrow [zaxirax] \text{‘to direct’}\]

\[
\begin{array}{ccccccc}
\text{P} & \text{P} & \text{P} & \text{P} & \text{P} & \text{(P: Place)} \\
\text{Scan} \Rightarrow \text{S \quad S} & \text{(S: Secondary Place)} \\
\text{[−hi]} & \text{[−hi(gh)]} \\
\text{[+back]}
\end{array}
\]

In (16), $[+\text{back}]$ spreads from the first vowel $/a/$ to the third vowel $/e/$ skipping the second vowel $/i/$; consequently, $/e/$ ($[−\text{back}]$) becomes $[a]$ ($[+\text{back}]$) but $/i/$ remains unaffected. This is possible since Spread scans the $S$ tier immediately dominating $[+\text{back}]$ and on this tier, $/a/$ is adjacent to $/e/$ with $/i/$ unspecified for the $S$ node. This skipping of an intervener reminds us of the fact that $[\text{tns}]$ movement skips Neg$^0$ in English affirmatives, to which we will come back in section 4.4. In syntax, hierarchical relations defined by c-command is generally more prominent than linear relations (see Kayne (1994), among recent work). Thus, for syntax, I define adjacency in terms of c-command as in (17) where $\alpha$, $\beta$ and $K$ are taken as syntactic categories:

\[(17) \quad \alpha\text{ is adjacent to } K\text{ on the scanned tier } \tau \text{ iff } K \text{ c-commands } \alpha\text{ and there is no } \beta\text{ such that } K \text{ c-commands } \beta\text{ and } \beta\text{ c-commands } \alpha, \text{ where } K, \alpha\text{ and } \beta\text{ provide access to } \tau.\]

For phonological operations, $\text{precede or succeed}$ would be used instead of $\text{c-command}$, and $\alpha$, $\beta$ and $K$ are phonological units. Yet the notion adjacency involved here might be better understood at a meta-theoretical level rather than at a more empirical level.

In comparing the GLC with the MLC, $\text{adjacency}$ in the former can be paralleled with $\text{closeness}$ in the latter, both of which are defined in terms of c-command (see (17) and (9b)). Seemingly, they are similar. However, the GLC and the MLC are crucially different in that only the former refers to the scanned tier $\tau$. That is, whereas the MLC cares about the features themselves which the target attracts, the GLC cares about the tier that immediately dominates the target’s feature. In section 3, we saw $[\text{TNS}_w]$ in $T^0$ may attract $[\text{tns}]$ from the main verb,
ignoring Neg$^0$. We now hypothesize that not only T$^0$ and V$^0$ but also Neg$^0$ provide access to the same tier, which immediately dominates [TNS]. I label this tier Vr(elated) after V-relatedness, the formal property of verbs, in the sense of Chomsky (1995: Chs. 1, 3). For the MLC, such a hypothesis does not have any effect because the MLC only sees the relevant feature (e.g. [tns]). As for the GLC, this intermediate class node (Vr) tier will be critical as we will see below. (18) displays the feature organizations which I propose:

\[
\begin{array}{cccc}
T^0: & Vr & V^0: & Vr \\
& [PST] & [pst] & [\alpha NE]
\end{array}
\]

In (18), the attractor and the attractee features are assumed to be on the same tier: e.g. the [TNS/tns] tier. Furthermore, they are assumed to be immediately dominated by the same tier: e.g., [TNS] and [tns] are immediately dominated by Vr. [TNS/tns] dominate [PST/pst] (only specified in past tense), and [tns] also dominates $\phi$-features. These dependencies oblige [tns] to pied-pipe $\phi$-features as well as [pst] (but not [\alpha ne]) when it is attracted by [TNSw], and explain subject-verb agreement in finite clauses even without V$^0$ movement as in English. Those features dependent on [TNS/tns] will be omitted in the following, as they are not directly relevant.

4.2. Strong [NE]: V$^0$-to-Neg$^0$-to-T$^0$ in French

To see how our system works, let us first consider V$^0$-(to-Neg$^0$-)to-T$^0$ movement in French. I repeat (4c) as (19) below:

\[(19) \text{ On (ne) va pas à l'église.}\]

In section 3.2, I have suggested that Neg$^0$ has [+/-NE] which attracts [+/-ne] of a main verb. Supposing so, the derivation of (19) is as follows. As in (20), Neg$^0$ merges with VP creating NegP:

\[
\begin{array}{cccc}
\text{NegP} & \text{Neg}^0 & \text{V}^0 & \ldots
\end{array}
\]

\[
\begin{array}{cccc}
Vr & Vr & [\alpha NE] & [\alpha ne] \\
& [tns]
\end{array}
\]
In (20), both \([\text{tns}]\) and \([\alpha\text{NE}/\text{ne}]\) \((\alpha = + \text{ or } -)\) depend on \(V_r\), as assumed in (18). For the moment, let us simply say that category movement is triggered by strong features, to which we will return in section 4.3. French \(\text{Neg}^0\) will thus have \([\alpha\text{NE}_v]\). By definition, Attract/Move scans the tier immediately dominating \([\alpha\text{NE}_v]\), namely, the \(V_r\) tier. There is no intervening head which provides access to the \(V_r\) tier, so \(V_0\) is adjacent to \(\text{Neg}^0\). Hence, Attract/Move successfully applies. \(V_0^0\)-to-\(\text{Neg}^0\) movement demonstrated here is intended to occur not only in negative sentences but also in affirmatives, as in (21), since the value \(\alpha\) of \([\alpha\text{NE}/\text{ne}]\) takes either \(+\) or \(-\) (but see section 4.4):

\[
(21) \text{On va souvent à l'église. } (=4d)
\]

After \([\alpha\text{NE}_v]\) checking via \(V_0^0\)-to-\(\text{Neg}^0\) movement, \(T_0^0\) merges with \(\text{NegP}\):

\[
(22) \left[\text{TP} \; T_0^0 \left[\text{NegP} \; V_0^0\text{-Neg}^0_0 \left[\text{VP} \ldots \text{tV}\ldots\right]\right]\right]
\]

\[
\ldots \ldots \; V_r \ldots \ldots \; V_r \ldots \ldots
\]

\[
[\text{TNS}] \; \text{[tns]}\]

In (22), \(V_0^0\) is adjacent to \(T_0^0\) on the \(V_r\) tier so that \([\text{TNS}]\) can attract \([\text{tns}]\) in \(V_0^0\). If \([\text{TNS}]\) in (22) is strong as assumed in section 3.1, \([\text{tns}]\) pied-pipes the whole \(V_0^0\text{-Neg}^0\) (we will see another possibility in section 4.3). As a result, (23) obtains, which is the structure shared by (19) and (21):

\[
(23) \left[\text{TP} \; \text{DP} \left[\text{T} \; V_0^0\text{-Neg}^0_0\text{-T} \left[\text{NegP} \; t\text{Neg} \left[\text{VP} \ldots \text{tV}\ldots\right]\right]\right]\right]
\]

It should be noted that if \([\alpha\text{NE}]\) were weak, that is, if the whole \(V_0^0\) does not move to \(\text{Neg}^0\), \([\text{TNS}_v]\) could never attract \([\text{tns}]\) from \(V_0^0\) once \(\text{NegP}\) is embedded in \(\text{TP}\), because \(\text{Neg}^0\) intervenes between \(T_0^0\) and \(V_0^0\) on the \(V_r\) tier:

\[
(24) \left[\text{TP} \; T_0^0 \left[\text{NegP} \; \text{Neg}^0_0 \left[\text{VP} \ldots V_0^0\ldots\right]\right]\right]
\]

\[
\ldots \ldots \; V_r \ldots \ldots \; V_r \ldots \ldots \; V_r \ldots \ldots
\]

\[
[\alpha\text{NE}_w] \; \left[\alpha\text{ne}\right] \;
\]

\[
[\text{TNS}_v] \; \text{[tns]}\]

This is what we call an HMC effect, which is derived from the GLC but not from the MLC. At this point, one might wonder why there could not be any dependency between \([\alpha\text{ne}]\) and \([\text{tns}]\). Is it possible for \([\text{tns}]\) to depend on \([\alpha\text{ne}]\)? The answer is no: when \([\alpha\text{ne}]\) is
checked off, [tns] would be also removed from the structure before it is checked against [TNS]. Is it then possible for [a.ne] to depend on [tns]? Again, the answer seems negative. In (18), we have assumed that the attractor and the attractee features are immediately dominated by the same tier. If so, [αNE] cannot attract [a.ne] which is dominated by [tns]. Without pied-piping by [a.ne], [tns] remaining in situ cannot be attracted by [TNS], either (an HMC effect). Furthermore, infinitival to in English is assumed to have [a.ne], as it occurs in the environment “... [NegP not [Neg' to(-Neg0) [VP...”. But, it does not have [tns], so there should not be any dependency between the two features (the same might be true of French infinitives, which we do not discuss here).

4.3. Weak [NE]: Do-Support in English

In section 3.1, we have considered that English T0 has a weak feature, [TNSw]. As a potential question, what happens if Neg0 and T0 have weak features, [αNEw] and [TNSw]? In that case, just as seen in (24), we cannot expect that [TNSw] attracts [tns] from V0. [αNE] is weak so that it only triggers movement of [a.ne] without pied-piping the whole V0. [tns] still remains in V0. T0's [TNSw] unsuccessfully attracts [tns] because there is a blocker, Neg0 standing between T0 and V0 on the Vr tier. We should keep in mind that the GLC is relevant not only to category movement but also to feature movement, as mentioned earlier. Hence, the derivation will crash. For the same reason, (12) is ruled out: the intervening Vr node under Neg0 blocks [tns] movement from V0 even if we attempted to insert tenseless do (or to) to check [+NE].

One way to avoid this difficulty is to invoke do-support. In generative grammar, do-support has been often regarded as a kind of “last resort,” though “last resort” used in this classical sense should not be confused with Suicidal Greed in (2). In Chomsky (1995: Ch. 2), as seen in section 3.2, an affixed main verb cannot skip Neg0[+NE] in undergoing covert V0-to-T0 movement, namely, the HMC effect. Therefore, do is inserted directly into a certain functional category (e.g. T0), which picks up inflectional affixes on behalf of the main verb. In the Minimalist Program, however, the old HMC/ECP (as well as affix lowering) no longer exists. The present system with Scan in Attract/Move, on the other hand, makes it possible to maintain the HMC effect. Let us then see how do-support takes place.

I have suggested in section 3.2 that auxiliary or dummy do(n't) car-
ries formal features like [±ne] and [tns]. It is selected from the lexicon and inserted to check [NEw] of Neg^0 as shown in (25):

(25) ([TP T^0] [NegP do(n't)-Neg^0 [VP ... V^0 ...]])

Do(n't) must be inserted into Neg^0 rather than T^0 since [±ne] carried by do(n't) can never be checked once it is inserted into or raised to a position higher than NegP. This is due to the C-Command Condition included in the definition of adjacency in (17). From Neg^0, do(n't) moves to T^0, and further to C^0 in interrogative sentences:

(26) [CP don't; [TP you ti; [NegP ti; [VP go to church]]]]

In this respect, do works parallel to other auxiliaries: modals, aspectual have, and be in finite forms. I assume that in negative sentences, these auxiliaries, carrying [±ne], start from Neg^0 or a lower position, e.g. Asp^0, placed between Neg^0 and V^0.

This analysis of do-support still needs explanation. If weak features are satisfied only with relevant features, insertion of the category do(n't) seems anomalous. Since we adopt strict local economy, LF(PF) convergence never triggers do-support. How can we insert do(n't) to check the weak feature of Neg^0? Before facing this question, we direct our attention to another question: In English, why can finite auxiliaries (including do) be raised to T^0 with [TNSw]? To answer this question, I revise the strong-weak asymmetry assumed by Chomsky (1995: Ch. 4) in the following way:

(27) a. Strong features trigger category movement either from θ-related positions (where predicates and their arguments are base-generated) or from non-θ-related positions;

b. Weak features trigger category movement from non-θ-related positions and feature movement from θ-related positions.

(27) might be deemed as a reinterpretation of Pollock’s (1989) transparent-opaque asymmetry. Pollock accounts for the English-French difference discussed in section 3.1 by relating it to a property of functional categories concerning θ-marking. Functional categories which are transparent to θ-marking allow movement of a main verb
into them. On the other hand, functional categories which are opaque to \(\theta\)-marking do not allow verb movement into them. I conjecture that the strong-weak asymmetry in (27), combined with the movement unit (category-feature) asymmetry, ultimately reduces to a kind of external requirement or bare output condition that \(\theta\)-relations must be maximally recoverable from the output representation. In other words, try to preserve \(\theta\)-relations in base positions for semantic interpretation. Empirically, (27) might be supported, for example, by Aoun and Li’s (1993) analysis of \(wh\) scope in languages like Chinese that lack obligatory \(wh\) movement. They argue that in Chinese, an empty question operator is base-generated in some non-\(\theta\)-related position and raised to \([\text{Spec}, C]\). This could be seen as a consequence of attraction by a weak feature in \(C^0\). For details, see Aoun and Li (1993). Conceptually, a requirement like \textit{Maximize }\(\theta\)-recoverability (\textit{Max-}\(\theta\)) seems fairly promising and worth pursuing. Leaving further justification for another occasion, let us consider questions raised above with (27).

By virtue of (27b), \([\text{TNS}_w]\) can raise auxiliaries to \(T^0\) from lower functional heads (=non-\(\theta\)-related positions) if no blocker intervenes (it could trigger direct insertion of (modal) auxiliaries into \(T^0\) from the lexicon without contradicting (27b)). Category movement from \(V^0\) (=\(\theta\)-related position) to \(T^0[\text{TNS}_w]\), as seen in (28), is disallowed unless \(V^0\) is raised to \(\text{Neg}^0\); in that case, \(V^0\) must have \([-\text{ne}]\) to be checked against \(\text{Neg}^0\)'s \([-\text{NE}_a]\):

\[
(28) \quad *[_{\text{TP}} \text{we go-}T^0 \left[\text{Neg}_P \text{Neg}^0 \left[\text{VP often t}_v \text{ to church}\right]\right]]
\]

Thus, in English declarative sentences, finite auxiliaries ultimately occupy \(T^0\) while finite main verbs do not. Contrastively, in French, whose \(\text{Neg}^0\) is assumed to carry \([-\text{NE}_a]\), the counterpart of (28) is possible:

\[
(29) \quad [_{\text{TP}} \text{on va-} \text{Neg}^0-\text{ }T^0 \left[\text{Neg}_P \text{t}_\text{Neg} \left[\text{VP souvent t}_v \text{ à l’église}\right]\right]]
\]

According to (27), \([-\text{NE}_a]\) triggers verb movement. Raised to \(\text{Neg}^0\), \(V^0\) can further move to \(T^0\), as in (29), whether \(T^0\)'s \([\text{TNS}]\) is strong or weak.

Supposing that English \([\text{NE}]\) is weak, we can insert \(do(n’t)[+\text{ne}]\) directly into \(\text{Neg}^0\), as in (25), to check \([+\text{NE}_a]\). This does not contradict (27b) since \(do(n’t)\) has not yet appeared in the phrase structure. If \(do(n’t)\) also carries \([\text{tns}]\), it is further raised to \(T^0\) as in (30), satisfying the GLC:

\[
(30) \quad [_{\text{TP}} \text{DP } [T' \text{ do(n’t)-} \text{Neg}^0-\text{ }T^0 \left[\text{Neg}_P \text{t}_\text{Neg} \left[\text{VP }\ldots \text{V}_0^0\ldots\right]\right]]]
\]

\([\text{aNE}_a]\) can trigger feature movement from \(V^0\) if \(V^0\) has \([-\text{ne}]\). If \(V^0\)
has [a ne] as well as [tns], do-support is not needed for [a NEw] checking, but the derivation never converges, manifesting an HMC effect. Hence, we should maintain that English main verbs do not have [a ne]. We will discuss this in section 5.

4.4. Underspecifying [NE]

As is familiar to us already, in a structure like (31), T0's feature can never attract V0's because V0 is not adjacent to T0 on the Vr tier:

(31) \[TP T^0 [\neg \neg^0 [VP ... V^0 ...]]] \]

\[\text{Vr} \quad \text{Vr} \]

\[\text{[a NE]} \quad \text{[tns]} \]

More precisely, if Neg0's feature is weak and only triggers feature movement from V0 without pied-piping the whole category or other features like [tns], T0 fails to attract V0's [tns] and thereby the derivation crashes. At this point, we want to consider what is happening in English affirmative sentences like (3b), repeated as (32):

(32) We often go to church.

In (32), the finite main verb does not move across the VP-initial adverb. Keeping the assumption that such a main verb stays in the base position, we expect the [tns] feature carried by go not to move to T0, as in (31). If this is true, the derivation should not converge, but (32) is clearly grammatical. To solve this contradiction, let us recall what is involved in Mongolian back harmony, cited in (16). There, in a configuration like /...a CiCe.../, [+back] spreads from the first vowel to the third vowel skipping the intervening vowel. This can take place because the S node of the intermediate vowel, a potential target for [+back] spreading, is absent with all features under it lexically underspecified. We suggest adopting this interaction between tier scansion and underspecification argued by A&P. Suppose therefore, that Neg0 is unspecified for Vr, as in (33):

(33) \[TP T^0 [\neg \neg^0 [VP ... V^0 ...]]] \]

\[\text{Vr} \quad \text{Vr} \]

\[\text{[TNSw]} \quad \text{[tns]} \]
Following A&P, if a node is absent with all dependent features underspecified, the relevant operation skips that node in scanning its tier. In (33), when \([\text{TNS}_w]\) attracts \([\text{tns}]\), scanning the Vr tier, Attract/Move skips \(\text{Neg}^0\), which lacks Vr. The skipped category does not enter into the calculation of adjacency bearing on the GLC. Hence, in (33), \([\text{TNS}_w]\) successfully attracts \([\text{tns}]\) from \(V^0\). In this case, what moves to \(T^0\) is only \([\text{tns}]\) because (27b) only allows feature movement from \(V^0\), a \(\theta\)-related position. In a word, the HMC effect is concealed by underspecification. We could say then that in a structure with \(\text{Neg}^0[\text{NE}_w]\), there are two ways to prevent the derivation from crashing: do-support and underspecification.

The analysis in this section could actually be also applied to French affirmatives. In (29), French \(\text{Neg}^0\) is assumed to have \([-\text{NE}]\), which triggers category movement from \(V^0\). Once \(V^0\) is raised to \(\text{Neg}^0\), it can be raised to \(T^0\) even if \([\text{TNS}]\) is weak. As another possibility, just as in (33), we could underspecify \([\text{NE}]\) and remove the Vr node from \(\text{Neg}^0\). If we adopt this alternative, French \(T^0\) must have \([\text{TNS}_x]\). Notice that both in English and in French, only \([-\text{NE}]\) is underspecified. Put differently, the value of \([\text{NE}]\) is only specified for the marked case, \([+\text{NE}]\). This seems meaningful. Such an economy-flavored principle, which would be called Minimize Feature Specification (Min-F), has been broadly accepted since the earliest stages of generative phonology (see Chomsky and Halle (1968), etc.). I consider that it enters the Minimalist Program quite naturally.

5. Historical and Comparative Implications

In the preceding section, we have seen that our system elegantly captures HMC effects, notably, do-support and related phenomena. Postulating \([\text{NE}]\) in \(\text{Neg}^0\), we explained HMC effects. In this section, on the basis of the above analysis, we will discuss HMC effects from historical and comparative perspectives, which will provide an interesting (though somewhat speculative) account for how do-support came to be introduced into English negative declarative sentences.

5.1. Jespersen’s Cycle

As outlined by Otto Jespersen (1924: 335–336), the development of English negative declarative sentences is considered to take the course
in (34):\(^9\)

\[(34)\]

\[\begin{array}{ll}
a. & \text{ic ne sece (nawiht/naht).} \quad \text{OE} \\
b. & \text{I ne seye not.} \quad \text{ME} \\
c. & \text{I say not.} \quad \text{LME\text{"~}(EModE)} \\
c'. & \text{I not say.} \quad \text{LME\text{"~}EModE} \\
d. & \text{I do not say.} \quad \text{EModE(\text{"~})}\(^{10}\) \\
e. & \text{I don’t say.} \quad \text{EModE\text{"~}} \\
\end{array}\]

The Old English (OE) and Middle English (ME) examples in (34) share basic properties with the French (4c). In (34a–c), finite main verbs precede a negative adverb. Moreover, OE and ME used the negative clitic ne, as in (34a, b), which became obsolete by the 15th century. Ishikawa (1995) notes that OE ne(/na) is certainly cliticized on a main verb as shown by V2 facts. In (35), drawn from Ishikawa (1995: 203), ne is pied-piped to \(C^0\) by the verb which it attaches to:

\[(35) \quad \left[\left[\text{CP } \text{donne} \left[\text{C'} \text{ ne gelyfd } \left[\text{TP he nanes sodes}\right]\right]\right]\right]

\text{NE believes he no truth 'Then he does not believe any truth.'}

French and ME ne behave the same way in “residual” V2 contexts.\(^{11}\) For these reasons, we could assume that in OE\text{"~}ME, Neg\(^0\) has \([+(-)NE]_{C}\), which drives overt V\(^0\)-to-Neg\(^0\) movement. V\(^0\) further moves up to T\(^0\) to satisfy [TNS\(_s\)] along the lines suggested in section 4.4 and to C\(^0\) in OE with V2. This first stage continues until Early Modern English (EModE) and still remains in very limited expressions.

From Late ME (LME) to EModE, the situation changes drastically. In this second stage, exemplified by (34d), \([+(-)NE]\) becomes weak, as discussed in section 4.3. Thereby, V\(^0\)-to-Neg\(^0\) movement is lost. Instead of raising a main verb, auxiliary do is inserted into Neg\(^0\) to check its feature.\(^{12}\) It is noteworthy that in EModE, the occurrence of

\(^9\) Jespersen (1924) does not mention the pattern (34c'). For the most part, (34) is adopted from Ishikawa (1995: 198).

\(^{10}\) Roberts (1993: 336, fn. 16; 338, fn. 21), citing Kayne (1989), notes that non-emphatic negatives with uncontracted not are ungrammatical in (colloquial) ModE/ Present-day English.

\(^{11}\) As for deriving the ordering of ne-V (not V-ne) in OE, ME and French, there can be various analyses. One might suggest that morphological realization should be determined by the mechanisms well discussed in Optimality Theory (cf. Prince and Smolensky (to appear)). However, because of limited space, I do not go further into details here.

\(^{12}\) Several authors (Ishikawa (1995), Kroch (1989: 236), Roberts (1993: 305), etc.)
auxiliary *do* was not limited to Present-day English (PE) contexts. As in (36), it was also used in environments such as non-emphatic, affirmative declaratives (see Ellegård (1953), etc.):

(36) Thus conscience *does* make cowards of us all (ˇ=stressed)

(Shakespeare *Hamlet*, II, i, 83; Roberts (1993: 240))

If [NE\text{w}] is specified for both + (= negative) and − (= affirmative) in this period, (36) surfaces as well as (34d). In LModE, *do*-support becomes unavailable in affirmative declarative sentences, apart from emphatic *do*. This decreasing tendency of non-emphatic *do* might be due to Min-F: in LModE/PE, the value of [NE\text{w}] is only specified for the marked case, i.e. [+NE\text{w}].

In the period of LME ~ EModE, we have another type of negative declarative sentences like (34c'). If *do*-support is obviated by feature movement, we could extend the analysis in section 4.4 to this example. Namely, such sentences can be derived by underspecifying [NE] and losing the Vr of Neg⁰. It should be noted that [NE] is a formal feature and it is not necessary to suppose this feature is responsible for semantic interpretation; we could have a semantic feature in Neg⁰ for polarity interpretation besides [NE], though we do not discuss it here. Owing to underspecification of [NE], no V⁰-to-Neg⁰ movement is triggered, but feature movement can occur from V⁰ to T⁰ triggered by [TNSₜₜ].

According to Roberts (1993), for earlier examples of this kind, there was some influence of Mainland Scandinavian (MSc) languages. In passing, in Present-day MSc languages, which do not possess any mechanism like *do*-support, we regularly observe that finite (main and auxiliary) verbs do not move across a negative adverb in embedded contexts. (37a) is such an example in Danish, drawn from Vikner (1995: 135):

consider that *not* shifts from XP occupying [Spec, Neg] to Neg⁰ sometime in the 16th or 17th century as the reduced form -n’t emerges. I, however, maintain the first assumption that *not* is base-generated in [Spec, Neg], while -n’t is lexically affixed to heads which are inserted in, or raised to, Neg⁰.

13 Pollock (1989: 420, fn. 49) proposes “Avoid *Do*,” an economy-flavored principle, which is to be subsumed under Min-F. Pollock regards Avoid *Do* as a subclass of Chomsky’s (1981: 65) “Avoid Pronoun.” If so, the latter might also derive from a more general economy condition, Min-F.

14 Vikner (1995) notes that when a clause is embedded under a bridge verb like ‘believe,’ the finite verb optionally raises high enough to yield V2 (Present-day MSc languages manifest V2 in matrix contexts).
(37) a. Det var godt [at han ikke købte bogen]
   b. *Det var godt [at han købte_i ikke t_i bogen]
   \hspace{1cm} it was good that he (bought) not (bought) book-the
   \hspace{1cm} ‘It was good that he didn’t buy the book.’

Platzack (1995: 113) analyzes inte, the Swedish counterpart of ikke, as occupying [Spec, Neg], based on the fact that the negative object DP is also placed normally in this position (I thank Mikael Vinka for bringing my attention to Platzack (1995)). The same seems true in Danish. Observe (38), a non-embedded example cited in Allan, Holmes and Lundskær-Nielsen (1995: 516) (the structure is my own):

\begin{equation}
(38) \quad [CP \text{ jeg havde } [\text{NegP } \text{ ingenting}_i [VP \text{ sagt } t_i]]]
\end{equation}

\hspace{1cm} ‘I had said nothing.’

If this analysis is correct, MSc languages will have NegP.\footnote{Scandinavian negative adverbs like ikke are often analyzed as left-adjointing to VP (Holmberg (1986), etc.). Similarly, English not in (34c') could be taken as left-adjointing to VP. The question of which analysis is more plausible is beyond the scope of this article, and I put it aside here. Note that (38) is not an object shift construction as it involves a finite auxiliary and a non-finite main verb.}

Assuming for now that negative items are in [Spec, Neg] and main verbs in VP in (37a) and (38), we apply the underspecification analysis to MSc languages as well as (34c').

5.2. [ne] as an Inflectional Feature of Verbs

In section 4.4, I have remarked that do-support and underspecification of [NE] are (the) two ways to supplement the lack of V\(^0\)-(to-Neg\(^0\)-) to-T\(^0\) movement, which are nothing but HMC effects in disguise. In the development of English negative declarative sentences, why did complete underspecification as in (34c') precede do-support as in (34d)? As much argued in the literature (Kroch (1989), Roberts (1993), etc.), there might be two events involved in some way: the declension of verbal inflections (along with drop of the negative clitic ne) in ME and the development of auxiliary system in ModE.

In ME, the inflectional paradigm of verbs underwent some simplifications: in the present tense, the former distinction between strong and weak verbs disappeared; in the past tense, many verbs came to take the weak verb conjugation. As a consequence of further simplification in EModE, there remains almost no inflectional distinction with respect
to person and number (see (41)). The negative clitic ne also disappeared in ME. If ne is closely related to verbal inflection, or if it is actually a part of verbal inflection, its disappearance can be considered to prompt simplification of the verb conjugation. If so, it can be argued that the specification of [ne], which is connected with ne/-n't and its zero variant (see section 3.2), indicates the richness of verbal inflection. To be concrete, (39) is stipulated:

(39) If [ne] is specified, V₀ has rich inflectional features.

Following Rohrbacher (1994) and Vikner (1995), among others, I assume that richness is evaluated by the standard below:¹⁶

(40) [F]irst and second person are unambiguously marked in at least one tense in singular and/or plural. (Vikner (1995: 163))

(40) is true in ME(/OE) and French in contrast with PE and MSc, as shown in (41) (provided in the present indicative mood):

<table>
<thead>
<tr>
<th></th>
<th>French -er</th>
<th>PE</th>
<th>Danish (MSc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Sg</td>
<td>Sg [- ]</td>
<td>1Sg</td>
<td>Sg -(e)r</td>
</tr>
<tr>
<td>2Sg</td>
<td>1Pl [-5]</td>
<td>2Sg</td>
<td>Pl -(e)r</td>
</tr>
<tr>
<td>3Sg -eth</td>
<td>2Pl [-e]</td>
<td>3Sg -(e)s</td>
<td></td>
</tr>
<tr>
<td>Pl -e(n)</td>
<td>3Pl [- ]</td>
<td>Pl</td>
<td></td>
</tr>
</tbody>
</table>

Since ME(/OE) and French (main) verbs meet the necessary condition in (39), [ne] can be specified. Since PE and MSc main verbs do not, [ne] cannot be specified.

With regard to English main verbs, the first/second person distinction in singular forms was lost in the 17th century. As for past tense, this took place as early as the 16th century according to Pyles (1964: 205). Ne dropped in LME (around 1400). What does this time-lag mean? We can imagine that in the transition period, the state of language was mixed or halfway, so to speak, with respect to the inflectional system: that is, while it had been already becoming poor, the old inflectional morphology still (partially) remained. Perhaps there might have been some discrepancy between written and spoken languages as seen in Present-day French (cf. Vikner (1997)). One might interpret the drop of ne (or [ne]) as a sign of inflectional evolution or simplification which had been already in progress. Here, I argue that the latter was in-

¹⁶ Rohrbacher (1994) and Vikner (1995) link rich inflection to V₀-to-T₀ movement rather than V₀-to-Neg₀. They also contrast first and second person forms with infinitival and third person forms. See also Roberts (1993).
duced by the former. Suppose that in the mixed period in which (34c’) emerges, English [NE] becomes weak. As repeatedly seen in section 4, however, when V⁰ has [ne], [NEw] inevitably leads the derivation to crash. Under the GLC, Neg⁰[NEw] blocks [tns] movement from V⁰ to T⁰. To save this situation, we can resort to either underspecification of [NE] or do-support. But, before that, [ne] has to drop from V⁰; otherwise it causes the derivation to crash. Let us hypothesize that once Neg⁰ has [NEw], finite main verbs without [ne] come to be selected constantly, a sort of natural selection, which allows us the above two options (the theoretical status of diachronic natural selection remains an open question). Thus, the main verb in (34c’) is not specified for [ne]. If do-support is not yet available in this period, it becomes necessary to underspecify [NEw], dropping Vₐ. To recapitulate, the shift from [NEₐ] to [NEw] severs [ne] from main verbs, which in turn impoverishes the verbal inflection. Without do-support, [NEw] needs to be underspecified with its host Vₐ also unspecified. This lets [tns] bypass Neg⁰. Intriguingly, the shift from [NEₐ] to [NEw] seems to be linked with the shift from [TNSₐ] to [TNSw]. In the system proposed in this article, even with [TNSw], Neg⁰-to-T⁰ movement can follow V⁰-to-Neg⁰ movement which is triggered by [NEₐ] (section 4.3). This might indicate that the shift from [TNSₐ] to [TNSw] does not occur before the shift from [NEₐ] to [NEw] (cf. section 4.2). Such a linkage, which is derived from our system, can have important implications, though I do not undertake to investigate them here.

5.3. Balancing Two Universal Principles and the Rise of Do-Support

A little behind the declension of the verbal inflection, there was a remarkable development of the auxiliary system in ModE. Dummy do belongs to this auxiliary group. As still observed in PE, only auxiliaries (e.g. can, will) behave like ME/French (main) verbs: they occupy T⁰ in declaratives (C⁰ in interrogatives) and are able to carry the negative suffix -n’t. As a possibility, ModE auxiliaries are allowed to have [ne], unlike main verbs. I argue that this property of ModE auxiliaries is probably attributed to semantic change. As observed by many authors (see Roberts (1993), etc.), auxiliaries in earlier periods behaved like main verbs in taking logical arguments. Therefore, it would not be unreasonable to assume that in those periods, they were base-generated in θ-related positions as main verbs. After semantic change, auxiliaries are base-generated in non-θ-related positions, e.g.
Neg⁰, Asp⁰. We have assumed in the preceding discussion that Neg⁰ already has [NEw] in the mixed period (LME~EModE), whereby [ne] is not assigned to main verbs. Recall that owing to (27b), weak features trigger category movement from non-θ-related positions but not from θ-related positions. Connecting the hypothesis in section 5.2 with (27), we predict that in LME~PE, where Neg⁰ has [NEw], auxiliaries including dummy do can retain [ne] while main verbs cannot. On the other hand, in OE~EModE and French, where Neg⁰ has [NEs], main (as well as auxiliary) verbs can be specified for [ne]. If this is tenable, the similarity between PE auxiliaries and French (main) verbs is explained. Note, however, that (39) does not seem to apply to ModE auxiliaries like do. In this sense, English has created quite unique objects, distinct from ModE main verbs and French verbs.

In EModE, (34d) exists in parallel with (34c'). This means that in this period, [+NEw] triggering do-support and underspecification of [NEw] allowing feature movement from V⁰ to T⁰ were coextensive. The former option overwhelms the latter. Our analysis can provide an interesting account for this. As discussed in (27), movement from non-θ-related positions is favored over movement from θ-related positions, whether category or feature is moved. Agreeing with Max-θ, do-support is frequently observed in EModE both in non-emphatic affirmative declaratives (see (36)) and in negative declaratives. Yet an economy condition prefers underspecification of [NEw]. In the 17th century, negative sentences like (34c') declined in use as did do-support in non-emphatic affirmative declaratives. This could be interpreted as an outcome of balancing the two principles, Max-θ and Min-F.

The development shown in (34) is delineated by the diagram in (42):

```
   (42)   OE     ME     ModE     (PE)
   5~7C.  12C. 14C. 16C. 17C. 20C.
   (34a)  (34b)  (34c)(34d)(34e)
   [NEs]  [NEw]
   (34c')

underspecified [NEw]
```

Notice that MSc languages preserve the (34c') type in embedded contexts, as shown in (37a), not only because their main verb inflection is too poor (see (41)) to host [ne] but also because they did not develop an auxiliary system: their modals distribute not like ModE counterparts but like ModE main verbs. The situation in ModE tells us that a language L can have [NEw] in Neg⁰ to trigger do-support if and only if L
has a well-developed auxiliary system and does not assign [ne] to main verbs. I leave for the reader the question of what may happen next in French.

6. Is the MLC Good Enough?: A Theoretical Consequence

Thus far, we have seen that HMC effects are properly captured in Minimalist terms with some autosegmental concepts. When we pursue the restricted Minimalist assumptions in section 2, which seem quite reasonable, we recognize that a locality condition on Attract/Move is needed which can refer to a certain level of feature organization. As such a condition, I proposed the GLC. The GLC explains locality phenomena which have been analyzed as RM (including HMC effects and further relativizations of RM\(^{17}\)) and superiority effects. To see this, let us consider wh island and superiority effects like (43a, b):

(43) a. *Who\(_i\) do you wonder where John met t\(_i\)?
   b. *Who\(_i\) do you expect who\(_j\) to meet t\(_i\)?

Suppose that (43a) and (43b) have in common the abstract structure in (44), where the O(perator-related) nodes immediately dominate [WH\(_s\)] and [wh] ([wh] is +Interpretable, hence undeletable as argued in Chomsky (1995: Ch. 4)):

(44) \[
\begin{array}{cccc}
[CP C^0 [TP \ldots [wh \ldots[who_i\ldots]]]] \\
\ldots O \ldots O \ldots O \\
[WH_s] \ldots [wh] \ldots [wh]
\end{array}
\]

In (44), who\(_i\) is not adjacent to the matrix C\(^0\) on the O tier, because wh (where/who\(_j\)) intervenes between them. Thus, the GLC prevents [WH\(_s\)] from triggering movement of who\(_i\), ruling out (43a, b).\(^{18}\) Chomsky’s MLC provides a similar account for these phenomena. Thus, the GLC explains what the MLC explains.

\(^{17}\) As for further relativizations of RM, for example, X\(^0\) is relativized in Baker and Hale (1990) by its being lexical or functional, and in Roberts (1993) by its being A or A’-type.

\(^{18}\) (43b) seems acceptable when who\(_j\) is D(iscourse)-linked in the sense of Pesetsky (1987). In that case, we could exploit underspecification and remove the O node under who\(_j\).
However, the MLC does not seem to work like the GLC in other cases. Let us consider a super-raising example in (45):

(45) *\([TP \ he \ seems \ that \ [TP \ it \ is \ likely \ [TP \ he \ to \ win]]\)"

As proposed by Chomsky (1995: Ch. 4), T^0 has a strong feature, [EPP], which attracts the [D] of DP. In (45), he is moved to the matrix [Spec, T] to check [EPP]. Obviously, this is a violation of the MLC since it in the intermediate [Spec, T], asymmetrically c-commanding he (=t_{he}), is not attracted. Hence, (45) is barred. Compare, then, (45) with (46):

(46) *\([TP \ it \ seems \ that \ [TP \ it \ is \ likely \ [TP \ he \ to \ win]]\)"

(45) and (46) are intended to have the same interpretation and ingredients (not Numeration) as the grammatical sentence it seems that he is likely to win. In (46), it, inserted in the intermediate clause, moves to the matrix [Spec, T] and satisfies the MLC (it is possible for the same DP to be attracted more than once for the +Interpretable property of [D]). This time, while it moves to the matrix [Spec, T] for [EPP], the Case feature [nom(inative)] of he and the Case assigning feature [NOM] in the matrix T^0 are not checked. Note that [nom] of it is already checked against [NOM] of the intermediate T^0 and both the features are deleted because they are —Interpretable. Following Chomsky (1995: Ch. 4), [NOM] is weak. Suppose that [NOM_w] triggers movement of [nom]. In (46), then, [NOM_w] in the matrix T^0 should be able to attract [nom] alone from he without violating the MLC since there is no [nom] intervening between them. As a result, the derivation should converge, but it is not so. Needless to say, RM (as well as the HMC) cannot be assumed as an independent principle since it is nothing more than a set of epiphenomena derived from a more primitive principle.

Our account is as follows. There is a tier, call it Nr(elated), which immediately dominates [NOM_w] and [nom]. The structure in question will look like (47):

(47) \([TP \ T^0 \ seems \ that \ [TP \ it \ is-T^0 \ likely \ [TP \ he \ to \ win]]\)"

--- Nr --- Nr --- Nr --- Nr ---

[NOM_w] [nom]

—Interpretable features such as [NOM_w/nom] are deleted once checked. We should distinguish feature deletion (Delete) in syntax from feature underspecification in the lexicon: Delete only affects rele-
vant features, and structures above them are maintained, as commonly practiced in autosegmental theory. Thus, in (47), \( it \) and the intermediate \( T^0 \) retain \( N_r \) nodes. In this structure, \( he \) is not adjacent to the matrix \( T^0 \) on the \( N_r \) tier. So, the GLC does not allow the matrix \( T^0 \) to attract \([\text{nom}]\) from \( he \), whether or not \( it \) moves to the matrix \([\text{Spec}, T]\) to satisfy \([\text{EPP}_s]\); even if \( it \) moves and \( N_r \) in its trace is not visible to \( \text{Scan} \), the \( N_r \) under the intermediate \( T^0 \) still blocks the attraction of \([\text{nom}]\) from \( he \). Without the GLC, it is unlikely that the checked intermediate \([\text{NOM}_w]\) interrupts feature attraction.

One anonymous EL reviewer points out that \( N_r \) itself could be an independent formal feature. If so, can the MLC provide an account? For the MLC to work, we have to assume that \([\text{NR}]\) in \( T^0 \) attracts \([\text{nr}]\). In (47), \([\text{NR}]\) in the intermediate \( T^0 \) checks \([\text{nr}]\) of \( it \). Consequently, they will be deleted unless they are \(+\text{Interpretable}\). However, they cannot be \(+\text{Interpretable}\), as checking is needed (recall that \([\text{EPP}_s]\) attracts a subject DP, so we do not need to assume \([\text{NR}_s]\)). One might speculate that \([\text{NR}]\) or \([\text{nr}]\) (or both) remain visible to syntactic computation after deletion for the feature is not completely erased, as noted by Chomsky (1995: §4.5.2). The same holds even without \([\text{NR/nr}]\): \([\text{NOM}_w/\text{nom}]\) is visible after deletion. We have argued that \( N_r \) is a class node and is not affected by deletion of the feature which it immediately dominates (this also applies to \( V_r \) and \( O \)). Under the GLC, an intervener on the relevant tier can be a blocker whether or not it has the feature which is attracted by the target. Therefore, as seen in (47), we have blockers without assuming the Delete/Erase distinction, which is more or less dubious. This would suggest that the GLC replace the MLC.

7. Conclusion

We have discussed locality in feature attraction, with a focus on \( X^0 \) movement. The assumption which has played the most important role throughout the article is that pure feature movement is primarily applied to the feature which is attracted by the target. Under this assumption, we have argued that the Minimalist Program cannot properly capture HMC phenomena such as English \( do \)-support. Similarly, there could also be the possibility that other RM phenomena like super-raising escape from explanation under the Minimalist Program. To elucidate these phenomena, we have proposed the GLC, a de-
finitional condition on Attract/Move, which can naturally replace the MLC. The GLC operates not on features as such but on the tiers which immediately dominate the relevant features. Therefore, under the GLC, even without the precise feature which is attracted by the target, an intervener can block (feature) movement if it provides access to the tier which immediately dominates the target feature. Thus, with the intervening head Neg⁰ which provides access to the Vr tier, movement of [tns] from V⁰ to T⁰ is impossible unless V⁰ moves to Neg⁰. To prevent this HMC effect, do-support or underspecification is invoked. From a diachronic point of view, such an approach can offer an interesting explanation of the rise of do-support in English negative sentences. The GLC constrains both feature movement and category movement. This is advantageous when we try to reduce overt-covert asymmetries. In exploring a mechanism to capture the locality of (feature) movement, we have also suggested unifying syntactic and phonological principles of locality. The GLC appears to be such an overarching principle, which covers syntax and phonology, belonging to the broader computational system. As well as the GLC, automatic feature pied-piping requires feature organization to be involved in syntax. If we are on the right track, Minimalist syntax needs to scrutinize feature organization and related mechanisms.¹⁹

REFERENCES


¹⁹ One anonymous EL reviewer points out that positing tiers appears problematic if we assume Chomsky’s (1995: 228) inclusiveness condition: no new objects are added in the course of computation apart from rearrangements of lexical properties [of items selected for the computation]. Whether inclusiveness itself is sustained at all calls for further research. Just like X-bar structure, however, if we take tier not as a real object but as a kind of metaphorical notation to designate levels in feature organization (at least in syntax, and maybe in phonology), the problem of whether tiers are compatible with inclusiveness or not will disappear.


Department of Linguistics
McGill University
1001 Sherbrooke St. W.
Montreal, Quebec H3A 1G5
Canada
e-mail: bkm5@musicb.mcgill.ca