1 Introduction

A central claim of the Minimalist Program (Chomsky 1993, 1994, 1995) is that movement is triggered solely by the need to check features. In this squib I will argue that this analysis provides a natural

Parts of this squib were presented at Harvard University (1996), at the "Specifiers" Conference at York (1995) (together with Günther Grewendorf), and at the 25th meeting of the North East Linguistic Society (NELS) (1994). I would like to thank the audience for helpful discussions. Special thanks to Elisabeth Ravaoarimalala, Günther Grewendorf, Roger Bruno Rabenilaina, Charles Randriamasimanana, and an anonymous *LI* reviewer for valuable comments.
account for the formation of complex (multisegmental) categories—that is, for ‘‘cluster formation.’’ I assume the Cluster Hypothesis (CH).

(1) Cluster Hypothesis
A feature F that is attracted by K attracts a feature of the same type F.

(1) is only a descriptive generalization. In fact, (1) interacts with Attract F (see Chomsky 1995).

(2) Attract F
K attracts F if F is the closest feature that can enter into a checking relation with a sublabel of K.

(3) Closeness
β is closer to K than α if β c-commands α.

Given the definitions in (2)–(3), and given a structure like [. . . α . . . β . . . γ . . . ] where α asymmetrically c-commands β and β asymmetrically c-commands γ, α cannot attract γ because β is closer to α. However, β can attract γ and α can attract the complex [β β γ] (or [β γ β]) if, as stated in (1), the feature F of β that is attracted by α may itself attract the analogous feature of γ. As shown in section 2, independent evidence for cluster formation as a result of (1) can be gained from complex X^0 categories consisting of multiple verbs and multiple clitics. Furthermore, in section 3 I will argue that in multiple wh- and focus constructions, cluster formation applies overtly in languages such as Bulgarian and Malagasy and covertly in Japanese.

2 Head Movement and the Cluster Hypothesis
Assuming with Chomsky (1995, to appear) that (syntactic) head movement is triggered by a [+affix] feature, we can account for the distribution of several incorporation phenomena on the basis of (1). Consider first verb incorporation as discussed in Baker 1988:371. Baker notes that in Quechua, multiple verb incorporation, transformationally derived from a base like (5a), takes the form in (4b) (= (5b)) and not (4a). As pointed out by Baker, (4a) shows that it is impossible to incorporate the verbs ‘make’ in CP2 and ‘eat’ in CP3 directly and independently of each other into the matrix verb. Assuming that incorporation in (4) is triggered by a feature [+affix] (i.e., that ‘want’, ‘make’, and ‘eat’ realize a [+affix] feature), ‘want’ may not attract ‘eat’ independently of ‘make’ as in (4a). According to the CH, the only possible derivation is (4b) (= (5b)), where ‘eat’ (V3) is attracted by ‘make’ (V2) and the complex verb ‘eat + make’ is attracted by the matrix verb ‘want’ (V1). This derivation yields the well-formed (4b).

eat-want-make-1SG.OBJECT-3SUBJECT
‘It makes me feel like eating.’
b. Mikhu-chi-naya-wa-n.
eat-make-want-1SG.OBJECT-3SUBJECT
Clitic-climbing phenomena provide further evidence for the CH. Let us consider the following examples from Spanish. I follow Jaeggli (1986) and Sportiche (1996), who assume that clitics in Spanish are associated with an overt lexical NP/DP or pro. The former case represents the “clitic-doubling” construction.

(6) Juan y Maria la llamaron a ella/pro.
Juan and Maria CL.3SG.FEM called her
‘Juan and Maria called her.’

Furthermore, I assume that (as proposed by, e.g., Suñer (1988), Fernández (1989), Franco (1991), Runner (1991), Zubizarreta (1992), Sabel (1995), and Sportiche (1996)) clitics must be analyzed as heads of a functional projection (e.g., as Agr elements) or (as proposed by Torrego (1996)) they must be analyzed as heads of a DP complement selected by the light verb (as in . . . [vP [v] [DP [D] [VP DP [V] . . .]]).

Let us now turn to examples with multiple clitic climbing. Examples (7)–(8) contain two restructuring verbs, querer ‘want’ and permitir ‘allow’ (Aissen and Perlmutter 1983:366, 367). Four derivations are possible. The clitics may remain in the sentences in which they are base-generated (7a); the clitic from CP₃ may move to CP₂ (7b); the clitic associated with permitir ‘allow’ may move to CP₁ (7c); or both clitics may appear in the matrix clause (7d). I assume that no clitic movement is involved in (7a). The order verb-clitic in this example simply results from verb movement (i.e., left-adjunction of the infinitival to the clitic).

(7) a. [CP₁ Quiero [CP₂ permitirte [CP₃ hacerlo]]].
I.want to.allow-you.DAT to.do-it.ACC
b. [CP₁ Quiero [CP₂ permititelo [CP₃ hacer t]]].
c. [CP₁ Te quiero [CP₂ permitir t [CP₃ hacerlo]]].
d. [CP₁ Telo quiero [CP₂ permitir t [CP₃ hacer t]]].
As was the case with verb incorporation, the order of clitics is fixed. For example, *lo may not precede *te (8a), and the clitic cluster *te + *lo cannot be separated (8b).

(8) a. *[CP, Lo quiero [CP, permitirte [CP, hacer t]]].
    b. *[CP, Te quiero [CP, permitir t lo [CP, hacer t]]].

Let us assume that clitic movement is triggered by the same feature that triggers other incorporation processes such as verb incorporation in (4).¹ Then, in (7c) the matrix verb querer ‘want’ realizes a [+affix] feature. This feature is also realized on the clitic te, which is attracted and moves into the matrix clause. Now recall the CH: ‘‘A feature F that is attracted by K attracts a feature of the same type F.’’ The CH predicts that a clitic with a [+affix] feature like te may attract another clitic. This prediction is borne out. Te and lo may appear together as a complex head in CP₂ (7b) or CP₁ (7d). In (7b) lo right-adjoins to te before verb movement applies. In this case te realizes the [+affix] feature that attracts lo. In (7d) the matrix verb and te in CP₂ attract the closest element with a [+affix] feature. Te attracts lo, yielding (7b) at one step of the derivation; then the clitic cluster is attracted by the matrix verb. (7d) is thus derived in the same way as (4b). (8a) shows that a clitic may not skip an intervening clitic.² It is ungrammatical because the [+affix] feature of the matrix verb can only attract the closest clitic. This is either te alone as in (7c) (if only the verb and te realize the [+affix] feature) or the whole clitic cluster as in (7d) (if the matrix verb and both clitics realize the [+affix] feature). In (8b) both clitics and the matrix verb bear the [+affix] feature. The attracting verb therefore only ‘‘sees’’ the whole cluster, and only (7d) can be derived.³

¹ Analyzing verb incorporation and clitic climbing as syntactic processes that are driven by feature checking is conceptually motivated by the idea that Last Resort is a defining property of Move (Chomsky 1995:253, to appear). According to this idea, incorporation should be possible only if it is obligatory movement motivated by feature checking. I assume that the [+affix] feature is optionally realized and that, if it is realized, it must be checked. Note that the optional realization of a certain feature is the technical implementation of allowing optional movement within the minimalist framework. The same effect (i.e., allowing (optional) incorporation processes) is achieved in Baker 1988 by assuming special subcategorization frames (i.e., optional subcategorization properties) for X₀ categories or in Baker 1996 by assuming different base structures for incorporated and nonincorporated variants of a certain structure (see, e.g., Baker 1996:282). However, in contrast to these approaches, the feature-checking analysis has the advantage that it allows a unified treatment of X₀ and XP cluster formation in terms of the CH, as shown in section 3.

² Several authors have argued that intermediate adjunction in the course of movement should be excluded in general (see, e.g., Hoekstra and Bennis 1989, Cinque 1990, Rizzi 1990, Sabel 1998). Note that, given this assumption, (8a) cannot be derived from (7b) since such a derivation would violate this constraint on movement.

³ As pointed out by a reviewer, in (i) and (ii) the clitic cluster is formed
3 XP-Movement and the Cluster Hypothesis

On the basis of the CH it can be argued that wh-elements in multiple wh-questions in Bulgarian and Romanian attract wh-phrases overtly. Furthermore, an analysis similar to the one proposed for Bulgarian can be suggested for ‘invisible’ wh-cluster formation in multiple wh-constructions in Japanese. In addition, in sentences with multiple focused elements, Malagasy shows focus cluster movement.

The hypothesis that wh-clusters are formed in multiple questions can be illustrated with examples from Bulgarian (9) (the same holds for Romanian; see Comorovski 1986, 1989). As is well known, the characteristic property of these multiple fronting languages is that every wh-word must be fronted in multiple wh-questions. This is shown by the contrast between (9a) and (9b) (from Rudin 1988).

\[(9)\]
\[
\begin{align*}
\text{a. } & \text{Koj ku}˘de \text{ misli}ﬂ [CP } \text{če } t e \text{ otišul t}]? \\
\text{who where think.2SG that has gone}
\end{align*}
\]

‘Who do you think (that) went where?’

\[
\begin{align*}
\text{b. } & *\text{Koj misli } [CP } \text{če } t e \text{ otišul k}˚de]?
\end{align*}
\]

who think.2SG that has gone where

According to Rudin (1988), all of the wh-phrases in multiple wh-constructions move to [Spec, CP], as in (9a) (see also Comorovski 1986, 1989, Bošković 1997). The evidence for this analysis is that, for example, Bulgarian allows long extraction of all of the wh-phrases (9a), unlike other multiple wh-fronting languages. In addition, ele-

---

as in (7b) and (7d). However, these examples are ungrammatical.

\[(i)\]
\[
\begin{align*}
\text{a. } & *[CP, Quiero [CP, permitir lote [CP, hacer t]]]. \\
& \text{I.want to.allow-it. ACC-you.DAT to.do}
\end{align*}
\]

\[
\begin{align*}
\text{b. } & *[CP, Lote quiero [CP, permitir t [CP, hacer t]].} \\
& \text{it.ACC-you.DAT I.want to.allow to.do}
\end{align*}
\]

They are impossible because the order of the clitics within the clitic cluster is due to parameterization. Indirect object clitics must precede direct object clitics in Spanish, whereas other languages show the reverse order. For example, Spanish (ii) differs in this respect from French (iii). (See, among others, Uriagereka 1988:chap. 3, 1995, 1996, for an analysis of this parametric difference between Spanish and French, which is compatible with the prediction made by the CH. Baker (1996:32) discusses a similar difference with respect to incorporation phenomena in polysynthetic languages; see also Baker and Stewart 1999.)

\[(ii)\]
\[
\begin{align*}
\text{a. } & \text{Juan se la presentara.} \\
& \text{Juan him her will.introduce}
\end{align*}
\]

‘Juan will introduce her to him.’

\[
\begin{align*}
\text{b. } & *\text{Juan la se/le presentara.}
\end{align*}
\]

\[(iii)\]
\[
\begin{align*}
\text{a. } & \text{Jean la lui présentera.} \\
& \text{Jean her him will.introduce}
\end{align*}
\]

‘Jean will introduce her to him.’

\[
\begin{align*}
\text{b. } & *\text{Jean lui la présentera.}
\end{align*}
\]

Note that the difference in clitic order between the two languages does not predict that (i) is grammatical in French, since clitic climbing is impossible in French, in contrast to Spanish (see, e.g., Kayne 1989, Sabel 1996).
ments such as parentheticals, clitics, and adverbs may not separate the sequence of fronted wh-phrases. Finally, as with verb incorporation and clitic climbing, the order of the fronted wh-phrases is fixed; that is, wh-cluster formation is subject to Superiority. The derivation of (10a) and (11a) is depicted in (12) (XP₂ = kogo ‘whom’ in (10a) and kak ‘how’ in (11a)). As (12) shows, the wh-cluster is formed before wh-movement to [Spec, CP] takes place (see the discussion below).

(10) a. Koj kogo t vidja t?  
   who whom saw  
   ‘Who saw whom?’
   b. *Kogo koj t vidja t?  
   whom who saw

(11) a. Koj kak t udari Ivan t?  
   who how hit Ivan  
   ‘Who hit Ivan how?’
   b. *Kak koj t udari Ivan t?  
   how who hit Ivan

A wh-phrase may attract other wh-phrases in Bulgarian (for similar suggestions, see also Kraskow 1990, Saito 1992, Abe 1993a,b, Ackema and Neeleman 1998, Grewendorf and Sabel 1996, 1999, Grewendorf 2001). Examples such as (9a) may plausibly be derived by the CH, that is, by assuming that the [+wh] feature of kuj ‘who’ that is attracted by the [+wh] feature in C₀ attracts the [+wh] feature of kőde ‘where’ (the same holds for multiple wh-constructions in Romanian and Japanese; see below). A similar derivation with short extraction can take place when two wh-arguments are present as in (10a), or as in (11a) with a nonreferential adjunct and an argument. In (10) the [+wh] feature of kuj ‘who’ attracts the [+wh] feature of kőj ‘whom’ and then the wh-cluster is attracted by the [+wh] feature in C₀. Furthermore, in (11) the wh-adjunct is attracted by the wh-argument; then the resulting wh-cluster is attracted by the [+wh] feature in C₀. Assuming that the adjunct in this example is base-generated in a position lower than the argument, only (11a) can be derived according to (1).⁴

⁴ Examples (ia–b) show that in Bulgarian three wh-phrases may move together to [Spec, CP]. As pointed out by Billings and Rudin (1994), Richards (1997), and Bošković (1998), the order of the second and third wh-words is
Three questions arise at this point. First, given the condition according to which adjunction is possible only to nonarguments for \(\theta\)-theoretic reasons (see Chomsky 1986, 1995:330), why is it possible to adjoin \(\text{wh}\)-words to \(\text{wh}\)-arguments in the (a)-examples of (9)–(11)? Second, does \(\text{XP}_2\) c-command its trace in (12)? That is, how is it possible that \(\text{XP}_2\) is attracted by the [+\(\text{wh}\)] feature of \(\text{NP}_1\)? Third, consider the possibility that the adjunct is base-generated in a position above the \(\text{wh}\)-argument in (11b), a position from which it c-commands the \(\text{wh}\)-argument. If this is the case, why is an adjunct unable to attract other \(\text{wh}\)-phrases?

Let us begin with the first question. Given the VP-internal subject hypothesis and the assumption that object and subject NPs have to raise out of their \(\theta\)-positions into positions for structural Case assignment (i.e., [Spec, TP] for subjects and [Spec, vP] for objects), this problem in fact does not arise since \(\text{NP}_1\), the argument that represents the target of adjunction, is located in a non-\(\theta\)-position.

Concerning the second question, for reasons that will be discussed later, let us assume that the [+\(\text{wh}\)] feature is inherently associated with \(\text{D}^0\) in Bulgarian (and Japanese; see the discussion below) and that it projects up to DP (this must be assumed anyway, to account for pied-piping possibilities; see the discussion of (18)–(19)). Then, D attracts via its projection path; that is, DP is the attractor in examples with \(\text{wh}\)-clusters. Given that the higher DP c-commands the lower one, Attract F may apply (this also opens up the possibility that certain non-DPs may function as attractors). Hence, the [+\(\text{wh}\)] feature that is realized on the maximal projection of D attracts another [+\(\text{wh}\)]

---

\(i\) a. \([\text{CP} \text{Koj kogo kakvo } [C_{\text{e}} \text{ e pital}]]\)?
   who whom what AUX asked
   ‘Who asked whom what?’

b. \([\text{CP} \text{Koj kakvo kogo } [C_{\text{e}} \text{ e pital}]]\)?
   who what whom AUX asked

Given that Bulgarian allows objects to be freely ordered inside VP, Billings and Rudin (1994) attribute the variation in (i) to the scrambling possibilities of this language. Ackema and Neeleman (1998) point out that these facts might also be explained by assuming different underlying orders of the objects. Whatever analysis is adopted, both analyses are compatible with the idea of deriving (ia–b) in terms of the CH.

Furthermore, it is well known that referential \(\text{wh}\)-adjuncts behave like \(\text{wh}\)-arguments with respect to extraction properties. This phenomenon is also observed in Bulgarian with \(\text{wh}\)-cluster formation. In contrast to nonreferential \(\text{wh}\)-adjuncts such as \(\text{kak}\) ‘how’ and \(\text{za\'sto}\) ‘why’, referential adjuncts such as \(\text{kade}\) ‘where’ and \(\text{koga}\) ‘when’ may attract \(\text{wh}\)-elements (see Billings and Rudin 1994).

For an alternative analysis of multiple \(\text{wh}\)-constructions for Bulgarian (and Romanian) in terms of multiple specifiers, see Comorovski 1989:123ff. and Richards 1997.
feature, and this gives rise to the phenomenon of XP cluster formation. Note that the c-command situation is exactly as with head movement. No c-command problem arises if the immediately dominating segment of the cluster that is formed via adjunction either counts for establishing c-command or, alternatively, does not block c-command (for discussion, see Baker 1988:449, fn. 10, Chomsky and Lasnik 1993:522).

Let us now turn to the third question. Why is it impossible for an adjunct to attract other wh-phrases, as in (11b)? This inability is probably related to the fact that nonreferential adjuncts (in contrast to arguments) lack a position for a variable, as several authors have pointed out (see Tsai 1994, Reinhart 1995, Chomsky 1995:n. 65).

Note, furthermore, that adjunction to all wh-elements (adjuncts and arguments) in [Spec, CP] is excluded if we assume the copy theory of movement (Chomsky 1995) and the Uniformity Condition on chains (UC); that is, “Chains are uniform.” The UC allows something to be adjoined to the head of a trivial (one-membered) chain. On the other hand, given the copy theory of movement, the UC excludes adjunction to the head of a nontrivial chain since this would render the head of the chain distinct from its other parts, resulting in a nonuniform chain.5

On the other hand, A-movement is not analyzed as copy movement in Lasnik 1998 largely for binding-theoretic reasons. Thus, it follows that elements in A-positions are potential targets for cluster formation. In addition, no UC-related problem arises concerning head movement, as long as adjunction applies to the head of a trivial chain. To summarize, it follows that adjunction to an element in [Spec, CP] is not permissible: such adjunction would render the chain of this element nonuniform since its copy does not have an element adjoined to it. As far as the well-formed examples already discussed are concerned, the UC is respected.

Following Rudin (1988), I have argued that in overt multiple fronting languages such as Bulgarian, the fronted wh-elements constitute a complex wh-phrase in [Spec, CP]; contrary to Rudin, however, I assume that this complex must be formed prior to wh-movement to [Spec, CP]. This analysis correctly predicts the properties of multiple wh-questions in languages such as Bulgarian: (a) the order of the fronted wh-elements is fixed, (b) nothing may intervene between the fronted wh-phrases, and (c) long-distance movement of multiple wh-phrases is possible in these languages.

There is evidence that the situation that occurs overtly in Bulgarian multiple wh-fronting constructions is analogous to the one that occurs covertly in Japanese multiple wh-in-situ questions; Japanese seems to be a multiple wh-fronting language as well, except that the attracted wh-phrase is an invisible copy of the in-situ wh-phrase.

5 The constraint on movement mentioned in footnote 2 rules out the possibility that an element is first adjoined to the foot of a chain before it adjoins to the head of the chain in [Spec, CP], thereby creating a uniform chain.
Hence, the CH, if operative at all in a language, is supposed to be parameterized with respect to whether it holds for visible or invisible movement. As pointed out by Watanabe (1991), Groat and O’Neil (1996), Pesetsky (1998), and Chomsky (to appear), among others, there is independent evidence for invisible movement in the overt syntax in Japanese. The idea is that the difference between visible and invisible movement in the overt syntax is traced back to principles that determine pronunciation. Assuming that movement leaves copies of the moved element, ‘‘overt movement,’’ as it is traditionally called, is interpreted as movement that carries the phonological features of the moved element to the head of the movement chain, whereas ‘‘covert movement’’ leaves the phonological features behind in the position of the foot of the chain. The following discussion provides further evidence for this analysis and the CH. It relates to island phenomena and exploits an assumption made by Saito (1994), Abe (1993a,b), and Takahashi (1994), who base their analysis of multiple wh-constructions in Japanese on the idea that a wh-phrase can be covertly adjoined to another wh-phrase.

The first piece of evidence is provided by the contrast between (13) and (14) from Maki 1994. (13) shows that Japanese wh-in-situ displays wh-island effects. (14) differs from (13) in that in the former the matrix subject is changed to the wh-phrase dare-ga, with the result that the Subjacency violation disappears.

(Please tell me [Q John remembers [whether Mary bought what]].)

(14) [TP Dare-ga [CP Mary-ga nani-o katta kadooka] who-nom Mary-nom what-acc bought whether oboeteiru ka] osiete kudasai.
(Please tell me [who remembers [whether Mary bought what]].)

It follows from the CH that the more deeply embedded wh-phrase is attracted by the higher wh-phrase. On these grounds, the (invisible copy of the) embedded wh-argument adjoins to the matrix subject dare-ga in (14) without crossing two TP (or IP) nodes, followed by movement of the wh-cluster to [Spec, CP]. In (13), on the other hand, the first possible landing site that licenses the embedded wh-argument is [Spec, CP] of the matrix clause. Thus, the wh-argument must cross two TP nodes to reach its final position, giving rise to a Subjacency violation.

Since Huang 1982 it has been well known that extraction of arguments out of islands is far better than extraction of adjuncts. Different analyses of this fact have been proposed by Lasnik and Saito (1984,
1992), Chomsky (1986), Rizzi (1990), and Cinque (1990). Rizzi (1990), for example, tries to explain the fact that adjunct extraction must meet different locality requirements than argument extraction by assuming that only adjunct traces must be “antecedent governed.” Now consider the examples in (15). They show that Japanese wh-adjuncts are subject to island constraints, where (15a) illustrates the case for complex NP islands and (15b) for adjunct islands. Unlike wh-arguments, wh-adjuncts are allowed to occur within these islands. Interestingly, if the wh-adjunct is preceded in the same clause by a wh-argument, the example improves considerably. Compare (15a) with (16a) and (15b) with (16b), respectively (examples from Saito 1994).


(16) a. ??John-wa [NP[TP nani-o naze katta] hito]-o John-TOP what-ACC why bought person-ACC sagasiteru no? looking-for Q (Q John is looking for [the person [that bought what why]?)

If the wh-phrases in (16a) and (16b) are extracted from the islands one by one, we would expect these sentences to be as ungrammatical as (15a) and (15b) because of the island sensitivity of adjuncts. We can therefore conclude that the wh-phrases in (16) form a wh-cluster before exiting the island and that ultimately it is only one (complex) wh-phrase that moves to the matrix [Spec, CP]. To sum up, the CH enables us to provide a uniform account of multiple wh-constructions in Bulgarian and Japanese.6


It is obvious that language-specific properties are responsible for the for-
The CH could also be applied to focus movement constructions in the Austronesian VOS language Malagasy. Consider (17b), derived from (17a). Keenan (1976) argues that no is a focus particle in Malagasy, and Pensalfini (1995) assumes that no is located in C0 (see Chung 1998:chap. 7 for a similar analysis of the analogous construction in Chamorro, another verb-initial Austronesian language). Hence, focus movement as in (17b) can plausibly be identified with movement to [Spec, CP] (or [Spec, FocP]) in Malagasy.

(17) a. Manasa ny lamba Rabe.
    wash the clothes Rabe
    ‘Rabe washes the clothes.’

   b. $\text{[CP Rabe [C′ *(no) [TP manasa ny lamba t]]]}$.  
   Rabe (FOCUS) wash the clothes

As shown in (17b), if focus movement applies, no is obligatory. Furthermore, only one constituent may be preposed in front of no. However, as Keenan (1976) observes, multiple fronted focused elements seem to form one complex constituent, as in (18b) and (19b), derived from (18a) and (19a), respectively.

(18) a. $\text{[CP Manasa [DP ny lamba] [TP amin’ ity savony ity]}$
    wash the clothes with this soap this Rabe].
    Rabe
    ‘Rabe washes the clothes with soap.’

   b. $\text{[CP Amin’ ity savony ity ny lamba [C′ no}$
    with this soap this the clothes FOCUS
    $\text{[TP sasan t t dRabe]]}$.  
    PASS.wash Rabe

   c. $\text{*[CP Ny lamba amin’ ity savony ity [C′ no}$
    the clothes with this soap this FOCUS
    $\text{[TP sasan t t dRabe]]}$.  
    PASS.wash Rabe

formation of different types of XP and $X^0$ clusters. For example, what parametric property gives rise to the possibility of wh-cluster formation? Differences among languages with respect to multiple wh-constructions are often argued to be due to different morphological properties of wh-phrases (Cheng 1991, Watanabe 1991:52ff., Grewendorf and Sabel 1996, 1999, Grewendorf 2001). One possible explanation could rely on Cheng’s (1991) and Nishigauchi’s (1990) observation that unlike in English, where no wh-cluster formation can be assumed and wh-elements remain in situ at LF (see Chomsky 1995), in Bulgarian and Japanese wh-words are used to form indefinites if particular affixes are added. Cheng concludes from this difference that wh-words in the multiple wh-fronting languages Bulgarian and Japanese do not have inherent quantificational force. This fact is assumed to be responsible for obligatory fronting of all wh-phrases. Following this line of argument, the difference between languages such as Bulgarian and Japanese on the one hand and languages such as English on the other hand can be rephrased in terms of whether the head of a wh-phrase may contain a [+ wh] feature that functions as an attractor, as argued in Grewendorf 2001.
(19) a. [CP Manantena Rasoa [CP fa manasa ny lamba
    hope Rasoa that wash the clothes
    amin’ity savony ity ianao].]
    with this soap this you
    ‘Rasoa hopes that you wash the clothes with soap.’

b. [CP Amin’ity savony ity ny lamba [CP fa
    with this soap this the clothes FOCUS
    antenain-dRasoa [CP fa sasanao tt]].]
    PASS.hope-Rasoa that PASS.wash-you

c. *[CP Ny lamba amin’ity savony ity [CP fa
    the clothes with this soap this FOCUS
    antenain-dRasoa [CP fa sasanao tt]].]
    PASS.hope-Rasoa that PASS.wash-you

Given that the order of the fronted elements is fixed ((18b) vs. (18c)
and (19b) vs. (19c)) and that Malagasy does not allow scrambling
(Keenan 1976), Keenan postulates the ‘‘Bodyguard Condition,’’ which
can be understood as saying that a constituent may move ‘‘on the
back’’ of another constituent.7 In terms of the analysis developed
above, (18b) and (19b) involve a further instance of the CH. In these
examples it is the [ + focus] feature in C0, the [ + focus] feature on the
DP ny lamba ‘the clothes’, and the [ + focus] feature on the PP amin’ity
savony ity ‘with this soap’ that interact in terms of the CH.

4 Conclusion

In this squib I have argued that the feature-checking analysis assumed
in the Minimalist Program provides a natural account for the formation
of complex categories, that is, for the process of ‘‘cluster formation.’’
Further, I have argued that the CH provides the basis for a uniform
account of multiple XP-movement constructions in languages such as
Bulgarian, Japanese, and Malagasy in that [ + wh] and [ + focus] fea-
tures in these languages trigger the formation of XP clusters. Further
evidence for the CH comes from the formation of complex X0 cat-
ergies such as clitic and verb clusters, which I have shown to be subject
to constraints similar to those constraining XP cluster formation.

7 The wrong order in (ib), (18c), and (19c) is allowed only with comma
intonation after the first element (ic). In the case of (ic) we are clearly not
dealing with one fronted constituent (compare (ib)).

(i) a. Aiza ianao no mividy ny vary?
    where you FOCUS buy the rice
    ‘Where do you buy the rice?’

b. * Ianao aiza no mividy ny vary?
    you where FOCUS buy the rice

c. Ianao, aiza no mividy ny vary?
    you where FOCUS buy the rice
References

Abe, Jun. 1993a. Binding conditions and scrambling without A/A′
distinction. Doctoral dissertation, University of Connecticut,
Storrs.
Abe, Jun. 1993b. The economy condition on dependency. In ESCOL
′93, 1–12. CLC Publications, Cornell University, Ithaca, N.Y.
Ackema, Peter, and Ad Neeleman. 1998. Optimal questions. Natural
Aissen, Judith L., and David M. Perlmutter. 1983. Clause reduction
in Spanish. In Studies in Relational Grammar 1, ed. David M.
versity Press.
Baker, Mark, and Osamuyimen T. Stewart. 1999. Verb movement,
objects and serialization. Ms., Rutgers University, New Brun-
swick, N.J.
Billings, Loren, and Catherine Rudin. 1994. Optimality and Superior-
ity: A new approach to overt multiple-wh ordering. In Formal
Approaches to Slavic Linguistics: The College Park Meeting,
cations.
Bošković, Željko. 1997. Multiple wh-fronting and economy of deriv-
University of Connecticut, Storrs.
Cheng, Lisa. 1991. On the typology of wh-questions. Doctoral disserta-
tion, MIT, Cambridge, Mass.
In The view from Building 20: Essays in linguistics in honor of
Sylvain Bromberger, ed. Kenneth Hale and Samuel Jay Keyser,
Chomsky, Noam. 1994. Bare phrase structure. MIT Occasional Papers
in Linguistics 5. MITWPL, Department of Linguistics and Phi-
losophy, MIT, Cambridge, Mass. [Also published in Govern-
ment and Binding Theory and the Minimalist Program, ed.
Chomsky, Noam. 1995. Categories and transformations. In The Mini-
in language, ed. Michael Kenstowicz. Cambridge, Mass.: MIT
Press.
Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles
and parameters. In Syntax: An international handbook of con-
temporary research, ed. Joachim Jacobs, Arnim von Stechow,
SQUIBS AND DISCUSSION


Kraskow, Tina E. 1990. Discourse-linking and the wh-island extraction
asymmetry. In ESCOL '90, 154–165. Department of Linguistics, Ohio State University, Columbus.


