

Early Features*

Uli Sauerland

Massachusetts Institute of Technology

1 Introduction

The major point of the minimalist approach to syntax is that all syntactic movement processes receive a uniform description. Especially important is a set of conditions that uniformly restrict all movement operations. These conditions fall into three groups: first, economy-conditions, like *Shortest Move* and *Greed*; second, restrictions on the landing site, specifically those resulting from the definition of a *Checking-Domain*, and third, timing restrictions, that are associated with the notion of *Feature Strength*. In this paper, I will focus entirely on the third kind of restrictions: timing conditions, that decide at which point which operation has to take place.

The proposal will differ from the ones in (Chomsky 1993) and (Chomsky 1995) with respect to the treatment of strong features. The Chomskyan systems both make a distinction between three possible *Strength* values for a feature: strong, weak, and unmarked for strength. Within the computational system of (Chomsky 1993), these strength values are interpreted as

*The analysis presented in section 3 of this paper was conceived after a very fruitful appointment with David Pesetsky and presented first the next day in the MIT-seminar *The syntax of VSO-languages* run by Andrew Carnie and David Pesetsky. I would like to thank Jonathan Bobaljik, Noam Chomsky, Ken Hale, Heidi Harley, Alec Marantz, Colin Phillips, David Pesetsky, Hiroyuki Ura, and the audience at SCIL VII, especially Rajesh Bhatt, Michael Hegarty, and Kazuko Yatsushiro for stimulating discussion of an earlier version of this paper. Financially, this work was supported by the German Academic Exchange Service (DAAD) with a grant in the *Zweite Hochschulsonderprogramm HSP II/AUFE*.

following: strong features must be checked before Spell-Out for the derivation to converge; weak features must be checked before LF, where the separate principle *Procrastinate* gives preference to covert checking; and features unmarked for strength need not be checked at all. The interpretation of the strong/weak system is slightly changed in (Chomsky 1995), but the fundamental distinction between strong features corresponding to obligatorily overt checking, weak features corresponding to preferably covert checking, and unmarked features, which do not need to be checked at all, is maintained.

The system we will arrive at in this paper is different only with respect to interpretation of the strength value *strong*. I claim that, in the known examples of strong features, these actually only prefer to be checked overtly, but this is not obligatory. The equivalent of strong features in my proposal may be unchecked at Spell-Out, if this is enforced by other conditions on the derivation. Hence, I will call such a feature *early*, as there is a principle that forces such a feature to be checked at the earliest possible point of the derivation (cf. Pesetsky 1989b). The system argued for in this paper is hence referred to as the early/late system, though the properties of weak and late features are exactly the same in the Chomskyan strong/weak-system and the early/late-system proposed here.

The early/late-system I propose has a symmetrical nature to it that the strong/weak-system lacks. In the early/late-system, both strength values express preference: early features prefer overt movement; late features prefer covert movement. In contrast, strong features do not prefer overt movement, but enforce it, in the sense that they must obligatorily be checked overtly for a convergent derivation. The symmetry that the early/late system has, but the strong/weak system lacks seems to me to be a desirable property. It is expressed in the following definitions of the principles of Earliness and Procrastinate:

- (1) Earliness: At any point of the derivation, transformation A must not be made if there is a possible transformation B at this point, which would check a greater number of *early* features.

Procrastinate: At any point of the derivation, transformation A must not be made if there is a possible transformation B at this point, which would check a smaller number of *late* features.

In the following two sections, I will present empirical support for the proposed early/late system. In section 2, I will present an argument in favor of earliness based on data from object shift in Scandinavian languages. In section 3, I will briefly present an analysis of the typology of Casemarking systems, which makes crucial use of the early/late feature system (see Sauerland 1995). Further empirical support for the existence of Earliness was given in (Pesetsky 1989b), which also supports the proposal of this

paper despite some differences in the interpretation of Earliness. At the beginning of section 3 I will briefly repeat one of Pesetsky's (1989b) arguments from *wh*-movement and then point out a similarity between *wh*-movement and the distribution of Case morphology. In section 4, I spell out the economy condition proposed in this paper in precise terms. This will enable us to discuss potential counterexamples in subsection 4.2. Section 5 will be the conclusion.

2 Object Shift

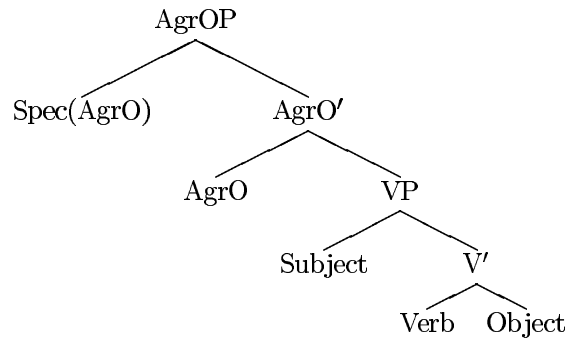
The first case which shows that the strong/weak system is inadequate was brought to my attention by Bobaljik (p.c.). The argument is based on the object shift paradigm in Scandinavian languages, which is illustrated in (2) using the case of Swedish object pronouns.¹ These object pronouns can appear outside of VP—to the left of negation—as in (2-a), but also inside of VP—to the right of negation—as in (2-c). However, the position is fully determined by the position of the verb. According to what is referred to as Holmberg's Generalization, the pronouns can undergo object shift, only if the verb moves, i.e. in matrix clauses. The contrast between (2-c) and (2-b) shows that in embedded clauses, where the verb does not move, the pronouns must remain inside VP, as shown in example (2-c). The reverse implication of Holmberg's Generalization also holds in Swedish: If the verb moves, the object pronoun must move out of VP as well. The contrast between (2-a) and (2-b) shows that in matrix clauses where the verb moves the object pronoun must move; sentence (2-b) where the object pronoun stays inside VP, is ungrammatical.

- (2) a. Maria₁ gillar₂ honon₃ inte [t₁ t₂ t₃]_{VP}
 Maria likes him not
- b. ??Maria₁ gillar₂ inte [t₁ t₂ honon]_{VP}
 Maria likes not him
- c. at Maria₁ inte [t₁ gillar honon]_{VP}
 that Maria not likes him
- d. *at Maria₁ honon inte [t₁ gillar t₂]_{VP}
 that Maria him not likes

In the theory of (Chomsky 1993) the ungrammaticality of (2-d) is explained as a violation of the *Shortest Move* condition on movement. Object shift is explained as movement of the object from its original position as complement of the verb to the specifier of the functional projection AgrO, which is the Case-checking position for the object. The condition *Shortest Move* implies that this movement is only possible if the target position Spec of

¹Thanks to Ingvar Løfstedt for discussing these examples with me.

AgrO and the specifier of VP are in the same minimal domain. The way *minimal domain* is defined, the minimal domain of the V-head only includes Spec(VP) and Spec(AgrO) after the verb moves from its base position to adjoin to the Agr-head. So, the impossibility of object shift without overt verb movement finds a satisfactory explanation in the minimalist system. In fact, it is the paradigmatic example for the principle of equidistance in this system.



However,, the obligatoriness of object shift in Case the verb moves is a problem that Chomsky's (1993) analysis cannot account for. If we only have the strength values *weak* and *strong*, the paradigm in (2) doesn't find a satisfactory explanation. The three possibilities the *strong-weak* dichotomy offers for the object shift driving feature of the pronoun are: a) the feature is always *strong*; b) the feature is always *weak*; and c) the feature is either *weak* or *strong*. All three choices make wrong predictions: choice a) predicts that object-pronouns are not allowed in Swedish embedded sentences because the checking of a *strong* feature is impossible; choice b) predicts that in matrix clauses the pronoun must not be shifted, because overt shift would violate Procrastinate; finally, choice c) predicts that object shift in matrix clauses should be optional, not obligatory, because nothing prohibits the use of the pronoun with the *weak* feature in the matrix clause.

One might at this point begin to look for an solution to the Swedish puzzle in (2) outside of syntax. An alternative account of the Swedish paradigm that seems possible up to now, would be that the pronouns are clitic-like elements that have to surface in a position next to the verb. As Diesing & Jelinek (1995:131) point out, however, the obligatory shift of pronouns is also found in German embedded clauses, which are head-final. These are shown in (3). Here an account based on adjacency between pronoun and verb clearly fails, since the sentence (3-a) where adjacency holds is ungrammatical, whereas the sentence (3-b) is grammatical, though the pronoun and verb are not adjacent.

- (3) a. *... weil ich nicht sie streichle.
 because I not her pet
 b. ... weil ich sie nicht streichle.
 because I her not pet
 ‘...because I don’t pet her’

With the introduction of the strength value *early*, the Swedish puzzle in (2) is solved. For the exposition here, we can assume Chomsky’s (1993) equidistance account of the ungrammaticality of (2-d). But we have to assume that the feature driving object shift of pronouns is *early*. Intuitively speaking, this has the consequence that objects try to shift overtly, if this is possible, but do not cause ungrammaticality if they are spelled out in their base position. Hence, in matrix clauses, where verb movement enables object shift, object shift is forced by economy. In embedded clauses where the verb does not move to AgrO overtly, overt object shift is impossible. To rely on the equidistance condition is not necessary however; Bobaljik (1995) analyses (2) making use of a morphosyntactic account of Holmberg’s generalization.

3 Morphological Case

In this section, we will discuss an application of the *early/weak*-feature distinction introduced in the previous section to the typology of Case systems. This may not seem to be an argument because the distribution of morphological Casemarking is often not considered a part of syntax. Especially, Marantz (1991) and following him Harley (1995) propose that the different Case endings are all identical within the syntactic part of the derivation and only morphologically distinct. Before presenting my proposal to explain the case distribution, I will point out a formal similarity between the distribution of morphological Case and multiple *wh*-question data from Pesetsky (1989a, 1989b). This similarity argues for the claim that there is indeed a common element in the mechanisms that govern these two phenomena, which, as I will claim then, is the feature system.

3.1 *wh*-Movement and Case Morphology

Pesetsky (1989a) points out that there are two different kinds of *wh*-phrases: D-linked ones and not D-linked ones. The two kinds differ in their syntax and interpretation, though rarely in their morphology. The distinguishing semantic property of D-linked *wh*-phrases is that they can only be used to question a constituent, if the previous discourse already established a finite set of possible values for this constituent. This is shown in example (4).

- (4) Of Apple Performa, PowerMac, and Powerbook, which computer should I buy?

In the syntax, D-linking correlates with the possibility to resist superiority effects. The contrast between (5-a) and (5-b) shows the superiority effect with non D-linked *wh*-phrases. Only example (5-a) where the structurally higher *wh*-phrase moves overtly is grammatical. In recent work (Epstein 1992, Kitahara 1993) the superiority effect has been explained as a direct consequence of the economy condition on movement first proposed in (Chomsky 1991): The derivation for (3-a) where the higher *wh*-phrase moves to Spec(CP), is more economical than the one for (3-b) where the lower *wh*-phrase moves, because the movement in (3-a) is shorter. This is the explanation of the superiority effect I will assume in the following.

- (5) a. Who₁ did you persuade *t*₁ to read what?
b. ??What₂ did you persuade who to read *t*₂?

The corresponding examples with D-linked *wh*-phrases in (6-a) and (6-b) do not obey the superiority condition exemplified by (5). However, as (6-c) shows, in another respect (6) and (5) are similar; namely, in both cases Spec(CP) must be filled. The same point is illustrated in the examples in (7).

- (6) a. Which man₁ did you persuade *t*₁ to read which book?
b. Which book₂ did you persuade which man to read *t*₂?
c. *Did you persuade which man to read which book?
- (7) a. Chris asked which man₁ Martin believes *t*₁ read which book.
b. Chris asked which book₂ Martin believes which man read *t*₂.
c. *Chris asked Martin believes which man read Thomas Glynn.

In summary, Pesetsky (1989a) describes the following conditions on English question formation: first, Spec(CP) must be filled by exactly one *wh*-phrase; second, if there is a not D-linked *wh*-phrase, then the highest of the not D-linked *wh*-phrases moves to Spec(CP); otherwise, the highest of the D-linked *wh*-phrases moves.² Within the early/late feature system the following assignment of feature strength yields these generalizations: the *wh*-feature of not D-linked *wh*-phrases is early; the *wh*-feature of the Complementizer in questions is early; but the *wh*-feature of D-linked *wh*-phrases is late.

One argument for Earliness of (Pesetsky 1989b) is based on Russian. Languages like Russian do not show effects of the Doubly-filled Comp Filter – the restriction that only one *wh*-phrase may occupy Spec(CP) (see Koizumi

²Here I assume that, in examples, like (6-b) and (7-b) the *wh*-phrase that moves overtly is not D-linked. This semantic intuition is shared by most people I consulted.

(1995:chapter 6) for an characterization of this difference in terms of multiple specifiers). However, they exhibit D-linking effects as well, as Pesetsky (1989a) shows. Namely, the descriptive generalizations in Russian are: first, all not D-linked *wh*-phrases move to Spec(CP); and second, of the D-linked *wh*-phrases only the highest one moves and this only happens if all *wh*-phrases in the question are D-linked. Hence the same strength values as in the case of English are also at work in Russian, with the only difference being that in English the condition yielding the Doubly-filled Comp Filter is active.³ There is however one exception to the first empirical generalization in Russian, given in (8). Some independent condition prohibits overt *wh*-movement of phrases of the form *skol'ko* + N, as shown in (8-a). The *in situ* variant in (8-b) is fully grammatical. The argument is based on the fact that (8-b) is ambiguous between the D-linked and the non-D-linked interpretation. This shows that where an independent condition blocks extraction of non-D-linked *wh*-phrases, they can remain *in situ*. This is precisely what we predict, if we assume that the *wh*-feature of non-D-linked *wh*-phrases is early.

- (8) a. *Kto₁ skol'ko dollarov₂ t₁ zaplatil t₂ za ètu knigu?
 Who how-many dollars paid for this book
- b. Kto₁ t₁ zaplatil skol'ko dollarov za ètu knigu?
 Who paid how-many dollars for this book
 'Who paid how many dollars for this book?'

A paradigm that bears a similarity to that of *wh*-movement in English is given in (9). In Icelandic, usually the subject is marked with nominative Case. However, the class of subject experiencer verbs exhibits a different distribution of Casemarking: The subject is marked with dative Case, and the object is marked with nominative case. Finally, there are also examples where both the subject and the object appear in oblique Cases, namely (9-c), where I assume that the accusative Case of the subject is oblique following Yip *et al.* (1987).⁴

³Evidently, the explanation offered here has the deficiency that it does not link the syntactic behavior of D-linked *wh*-phrases to their semantic properties; a correlation that we find cross-linguistically. What I present here is merely a formal account of the syntactic distribution, which is completely compatible with the current syntactic framework. There is an obvious way to understand the behavior of D-linked *wh*-phrases in terms of semantic properties: Assume that the semantic mechanism or *wh-in situ* are Reinhart's (1994) choice functions, but contrary to what Reinhart assumes choice functions can only be computed on a finite domain given by discourse. From this assumption differences in the LF-distribution between D-linked and not D-linked *wh*-phrases would follow.

⁴For a detailed discussion of the Icelandic facts, I refer the reader to the papers of Zaenen *et al.* (1985) for the subject status of the quirky datives and Harley (1994) for evidence from transitive expletives that the nominative object is indeed in the object position, Spec(AgrO) on her account, and furthermore that the nominative Case on the object has properties of *Structural Case* in the sense of (Chomsky 1981).

- (9) a. Dagmamman bakaði brauðið. (Yip *et al.* 1987:Fig. 6)
the day-mommy_{NOM} baked the bread_{ACC}
‘The day-mommy baked the bread.’
- b. Calvini liki verkið (Harley 1994:(1))
Calvin_{DAT} like job-th_e_{NOM}
‘Calvin likes the job.’
- c. Mig iðrar þess (Yip *et al.* 1987:(9-e))
me_{ACC} repents this_{GEN}
‘I repent this.’

The examples (9-a) and (9-b) show an effect that is comparable to the contrast between (5) and (6). (9-a) exemplifies the unmarked case of Icelandic Casemarking, and is hence comparable to the unmarked case of *wh*-movement in English, namely the case in (5) where the higher *wh*-moves. In (9-b) the nominative Case assignment to the subject is blocked by quirky Dative case, hence the object gets a chance to receive nominative Case. This is parallel to the case of *wh*-movement where the higher *wh*-phrase is D-linked, and hence does not have to move, with the effect that the lower *wh*-phrase can raise to Spec(CP). Let me try to also express the same parallelism, but this time in the terminology introduced by Marantz (1991) for the description of Casemarking. Marantz would call the structural, accusative Case in Icelandic a *Dependent Case*, because it is only assigned if nominative Case is assigned to another argument in the clause.⁵ Nominative Case would be the unmarked structural Case, which is assigned whenever neither the dependent structural Case cannot be assigned, nor an inherent Case. Applying the terminology of (Marantz 1991), we can now describe the distribution of not D-linked *wh*-phrases in English in the following way: The unmarked position of D-linked *wh*-phrases is in Spec(CP); the dependent position of D-linked *wh*-phrases is *in situ*. This parallelism between *wh*-movement and the distribution of morphological Case is what the formal system in the following chapter is designed to express.

3.2 The Lemmings Theory of Case

In this section, I will present an analysis of the distribution of Casemarking in a variety of languages that quite directly expresses the formal similarity between the syntax of D-linked and not D-linked *wh*-phrases and the distribution of nominative Case pointed out above. The analysis in this section is presented in more detail in (Sauerland 1995). Here I will only

⁵The same insight is represented by the concept of Case competitor in the account of Bittner & Hale (1994), though the two proposals of Marantz (1991) and Bittner & Hale (1994) are quite different in other respects. I do not know analysis of the Case distribution in other frameworks, but would not be surprised if the notion of a dependent Case is expressed there as well.

try to make the basic idea clear, and will give the predicted typology of Casemarking systems in the summary.

The analysis assumes that Case morphemes are independent heads in the syntax – an assumption that has been argued for by Lamontagne & Travis (1993) and Bittner & Hale (1993, 1994). Hence all nominal arguments are the complement of a K(ase)-head. As the basic clause structure, I assume that all arguments of the verb originate VP-internally. Since word order is not a primary concern in this section, I will for clarity make use of an unsplit INFL-head, just like in (Chomsky 1981). But nothing hinges on this assumption, and all I will be saying is equally compatible with a more articulated structure of the INFL-head. Incorporation will play an important role, and for now I assume that it is at least restricted by c-command, but probably additional restrictions hold, as is standardly assumed. A central part of the proposal is Mahajan’s (1994) idea that Case heads can incorporate into the inflectional complex of the verb, leaving behind a nominal phrase without any Case features. These nominal arguments without any Case features are those that surface with nominative or absolutive Casemarking. In effect then, the distribution of nominative Case in a clause is parallel to that of a trace *wh*-movement: The morphological presence of nominative Case indicates the absence of the Case head – and actually, as Bittner & Hale (1994) point out, nominative Case has, in many languages, zero morphology.

Two questions are still open. What determines the underlying distribution of Case morphemes? What mechanism governs the distribution of nominative Case? I will try to answer them in turn.

The underlying distribution of Case morphemes, I assume, is determined by thematic properties of the nominal arguments. Hence, the underlying distribution of Case in a clause should be universally the same as long as the θ -role assigning verb is the same. The agent θ -role is always represented by ERG-Case,⁶ patients are always headed by an ACC-Case morpheme, all other θ -roles bear various oblique Cases. NOM-Case is never present at this level, because it doesn’t correlate at all with a specific θ -role. I assume that the subjects of experiencer predicates have a different thematic role which is neither agent nor patient, but something like Pesetsky’s (1994) CAUSER-role. For a justification of this assumption see (Pesetsky 1994). Objects of experiencer verbs, I assume, get normal ACC-Case.

⁶I use the following abbreviations for Case: ERG for ergative, NOM for nominative and absolutive, which I assume are the same, ACC for accusative, DAT for dative, and OBL for oblique Case in general.

θ -role	underlying Case
Agent	ERGative
Patient	ACCusative
Subject Experiencer	OBLique
Object Experiencer	
Partitives, ...	

Of the nominal arguments in a clause, at most one of the Case morphemes incorporates overtly into the inflectional complex of the verb. Otherwise we find constructions with multiple nominative Case. This restriction to only one overt movement is similar to the Doubly-filled Comp Filter in the case of *wh*-movement. The mechanism governing the incorporation of the K-heads into INFL has to be feature checking, since I adhere to the minimalist mantra that all movement is governed by a certain, uniform mechanism. Namely, checking of the K-feature of the head against a K-feature in INFL. This is different from what Mahajan (1994) assumes, but his proposal that incorporation of the ergative morpheme in Hindi is blocked by the lack of adjacency has some obvious problems, for example, with head-final languages that are not ergative like Japanese and German. There will be essentially three K-features whose strength influences the distribution of nominative Case: namely, the ones of the ergative K-morpheme; of the accusative K-morpheme; and the K-feature of INFL. In some cases that I talk about in (Sauerland 1995) the strength of the K-feature of the OBL-morpheme matters.

Let us first discuss the Case systems of Icelandic and Niuean before I describe the general typology that this system gives us.

The descriptive generalization about Icelandic was the following. In Icelandic, agents always surface with NOM-Case; patients surface with NOM-Case only in two cases: if the subject is an experiencer, in which case the subject surfaces with dative Case; or if there is no subject at all – the case of unaccusative verbs. We can say then, that patients surface with nominative Case, if and only if no agent argument is in the same clause. The following table presents what I assume to be the distribution of underlying and surface Case in Icelandic for the relevant classes of predicates.

<i>Icelandic</i>	unerg.	unacc.	trans.	exp. pred.	exp. pred
underlying	ERG	ACC	ERG ACC	DAT ACC	DAT GEN
surface	nom	nom	nom ACC	DAT nom	DAT GEN

These facts can be accounted for by saying that ERG has an *early* K-feature, and the one on ACC is *early* as well in Icelandic. Hence, ERG never surfaces in Icelandic, and ACC, only if an ERG-agent is present. The settings are summarized in the following table.

K-feature on	ERG	ACC	INFL
Strength	early	early	late

The reason why in normal transitive clauses the ergative morpheme incorporates, but not the accusative, is I suggest, the *Shortest Move* condition. The subject is structurally higher than the object, and hence, the incorporation of the Case morpheme of the subject is preferred over that of the object.

Niuean is the equivalent of Icelandic among the ergative languages. In usual transitives like (10-a), the Case assignment is ERG on the subject, and NOM on the object. But if the object receives an oblique Case or is not present at all, the subject surfaces with NOM-Case. The oblique Case here is the so-called middle Case (MID) or the locative Case (LOC). This Case is used with the objects of perception and psych verbs in (11), or to give a partitive reading as in (10-b). A similar pattern is found in other Polynesian languages like Tongan and Samoan (see Chung 1978).

- (10) a. Koe kai he tama e talo. (Massam 1994:(4))
 Pres eat ERG we-EXCL NOM taro
 ‘We are eating up the taro.’
- b. Koe kai e tama he talo.
 Pres eat NOM we-EXCL LOC taro
 ‘We are eating of the taro.’
- (11) a. Onoono e tama ke he tau gata (Massam 1994:(3))
 look NOM child MID PL snake
 ‘The child is looking at the snakes.’
- b. Ita e faiaoga i a au
 angry NOM teacher LOC PERS me
 ‘The teacher is angry at me.’
- c. Manako nakai a koe ke he tau manu?
 like Q NOM you MID PI animal
 ‘Do you like animals?’

The relations between the underlying and surface Cases that I assume for Niuean are summarized in the following table.

<i>Niuean</i> underlying	unergatives ERG	unaccusatives ACC	transitives ERG ACC	perception/psych ERG LOC
surface	nom	nom	ERG nom	nom LOC

This system can be described by the following settings of strength-values for the K-features.

K-feature on	ERG	ACC	INFL
Strength	late	early	early

In general, if ergative Case has a *late* K-feature and accusative Case has an *early* one, the result will be overt incorporation of the accusative. Hence, all languages with this combination will be ergative languages. Niuean has, in addition, an *early* K-feature on tense, which ensures that, in all clauses with nominal arguments, one of them will surface with nominative Case. If one of the nominal arguments has a early Case feature, than this will be the one that incorporates, as in (10-a). However, if there is no argument with an underlying accusative Case, all arguments will have late Case features. In this case, the Shortest Move condition will favor incorporation of the highest one of the Case features. Hence, in the examples in (10-b) and (11), the subject surfaces with nominative Case.

The system of morphological Casemarking presented above predicts a typology of possible Case systems. Going through all the possible combinations of strength settings for the K-features on the relevant heads, we should exhaust the attested Casemarking systems. This is done in the table below. All the possibilities that are predicted are actually attested. Unfortunately, two redundancies arise, which, however, is not such a bad result compared to typological systems from other domains, like Idsardi's (1992) typology of word-stress assignment. Moreover all the case marking systems that are discussed in (Bittner & Hale 1994) are predicted by the typology, with the exception of classical ergative languages like Inuit and Warlpiri. In (Sauerland 1995), I present the possibility of capturing these languages by assuming that the K-feature of oblique Cases may be strong as well.

Parametrization			Underlying Case and Surface Case					Class
Tense	ERG	ACC	ERG	ACC	ERG ACC	OBL ACC	ERG OBL	
early	early	early	nom	nom	nom ACC	OBL nom	nom OBL	Icelandic
early	early	late	nom	nom	nom ACC	nom ACC	nom OBL	German
early	late	early	nom	nom	ERG nom	OBL nom	nom OBL	Niuean
early	late	late	nom	nom	nom ACC	nom ACC	nom OBL	= German
late	early	early	nom	nom	nom ACC	OBL nom	nom OBL	= Icelandic
late	early	late	nom	ACC	nom ACC	OBL ACC	nom OBL	Eastern Pomo
late	late	early	ERG	nom	ERG nom	OBL nom	ERG OBL	Basque
late	late	late	ERG	ACC	ERG ACC	OBL ACC	ERG OBL	Antekerrepenhe

4 Formalizing Earliness and Economy

In this section, I will formulate the Earliness principle more precisely than I have done so far using the language of ranked constraints of (Prince & Smolensky 1994). I will then point out that the analyses in this paper require a specific type of economy condition on the derivation; namely it must

be a set of ranked constraints. These constraints are evaluated at every step of the derivation. I will then go on to describe what a counterexample to this proposal would look like. I discuss the case of expletive sentences in English, which appears to fit this description if we accept certain other assumptions. However, as I will show, these other assumptions are very problematic and hence, the counterevidence itself is not as damaging for my proposal as it looked at first.

4.1 Earliness, Shortest Move, and Procrastinate

In section 2 and 3, I made use of three conditions on derivations: Earliness, Shortest Move, and Procrastinate. These three conditions interact and determine at each step which transformation is the next to apply. Of the transformations that are possible at any point of the derivation, one is selected as the next step by the computational system that satisfies the three conditions best in a sense to be made precise. As possible transformations, I see here any merger operations and movement operations that check at least one feature which needs to be checked, hence has to be either early or late. The restriction of possible movements to feature checking movements is equivalent to Chomsky's (1993) principle of Greed.

The definitions of the three economy conditions that govern every step of the derivation are:

- (12) Earliness: At any point of the derivation, transformation A must not be made if there is a possible transformation B at this point, which would check a greater number of *early* features.
- Procrastinate: At any point of the derivation, transformation A must not be made, if there is a possible transformation B at this point, which would check a smaller number of *late* features.
- Shortest Move: At any point of the derivation, transformation A must not be made if there is a possible transformation B at this point, which makes a shorter move.

Merger operations are optimal with respect to *Shortest Move*, because there no movement at all is taking place. However, most applications do not check any features. It follows that Checking of Early features by Merge or Move precedes applications of Merge which do not check any features. Late features on the other hand get checked after applications of Merge which do not check any features, because checking of late features is something the derivation tries to avoid. Checking of late features by movement is postponed until the point, when there are no further possible operations that do anything else but checking late features—in a *Zugzwang* situation.

Furthermore, I assume that a derivation terminates and interpretation applies when there are no more possible moves. Determining the point of

Spell-Out however is not quite as easy, since the first plausible guess turns out to be empirically wrong. Namely, it seems natural to assume that Spell-Out applies as soon as all lexical items participating in the derivation are combined into a single phrase marker. This, however, is falsified in cases where the last overt transformation is a movement e.g. *wh*-movement to Spec(CP) in English questions. What seems to be the case is that checking of early features is preferred if the derivation has a choice between checking an early feature and Spell-Out. We see now that the status of Merge operations that don't check any features and Spell-Out with respect to the conditions in (12) is the same. Both are operations that check no features, hence are neutral with respect to Earliness and Procrastinate, and do not involve movement, hence Shortest Move does not apply. This predicts that movement that checks early features always precedes applications of Merge and Spell-Out, whereas movement that checks only late features will follow it. So, on my account, there are three kinds of transformations: Merge, Move, and Spell-Out. Spell-Out has the additional requirement that it can only apply when the phrase marker has a single root node. From these assumptions, it follows that Spell-Out applies precisely at the right point; namely, after all lexical items in the derivation were combined into a single phrase marker and after all the early features of the root node that can possibly be checked have been checked.

The three conditions on the derivation conflict in a number of cases, e.g., if a movement checks an early and a late feature. I assume that there is a universal hierarchy among the three conditions. Using the language of (Prince & Smolensky 1994), this hierarchy is Earliness \gg Procrastinate, Shortest Move, as I will argue below.⁷ The relative ordering of Procrastinate and Shortest Move, as I will also show below, does not matter. Though I use the idea of ranked constraints from Prince & Smolensky's (1994) work on phonology, there is an important difference between the proposal here and all the applications of constraint ranking in phonology I have seen so far. Namely, in the phonological examples, constraint ranking relates two levels, input and output of phonology, with each other. I, on the other hand, assume that the constraints are evaluated at each step of the syntactic derivation and the transformation that is the best candidate is executed as the next step of the derivation. Actually, Prince & Smolensky (1994:4) note already that this kind of theory, which they call *harmonic serialism*, is a possible way of making use of the formal concept of ranked, recursively evaluated constraints they provide.

⁷The ranking $A \gg B \gg \dots$ can be calculated as follows: Firstly, of all candidates find those that satisfy A best. Secondly, from the result of the first step find all the candidates that satisfy B best. If there are further lower ranked constraints, apply the same procedure as before again. The candidates that remain are the candidates that satisfy the family of ranked candidates $A \gg B \gg \dots$ optimally. In the ranking that I state here the notation *Procrastinate*, *Shortest Move* describes a tie between the two constraints. As I will show below either ordering of the two constraints is also compatible with the data.

That the universal hierarchy Earliness \gg Procrastinate, Shortest Move that I introduced above is the only empirically correct one can be seen quite easily. Let us, as is customary, argue for each ranking in turn, beginning with Earliness \gg Shortest Move.

Looking at example , we can see immediately that Shortest Move \gg Earliness would be empirically incorrect. Specifically, if Shortest Move was ranked higher than Earliness, movement of the *wh*-phrase *who* would be less optimal than non-movement. Hence, further Merge operations and Spell-Out, both of which are neutral or even more optimal with respect to the conditions in (12), would have to take place before the *wh*-phrase moves. In fact, the ranking Shortest Move \gg Earliness would predict that there is no overt movement.

(13) Who₁ does every phonologist like t_1 ?

The argument for the ranking of Earliness \gg Procrastinate relies on the discussion of *wh*-movement and D-linking in section 3.1. There, I concluded that, in English, D-linked *wh*-phrases have a late *wh*-feature, whereas non-D-linked ones have an early *wh*-feature. Furthermore, the *wh*-feature of the Question-Complementizer is early in English. Based on this conclusion, we can see from example (4), repeated as (14-a), that the ranking Earliness \gg Procrastinate must hold. In , we see that a D-linked *wh*-phrase has to move overtly to Spec(CP). This movement, I argue, checks the early feature of Comp and the late feature of the *wh*-phrase. That this movement takes place overtly is predicted only by the ranking Earliness \gg Procrastinate. The reverse ranking Procrastinate \gg Earliness predicts that (14-b) should be the output, not (14-a), since Procrastinate prefers covert checking of the late *wh*-feature of *which computer*.

(14) a. Of Apple Performa, PowerMac, and Powerbook, which computer₁ should I buy t_1 ?
b. *Of Apple Performa, Powermac, and Powerbook, should I buy which computer?

Finally, let us see why there is no way to decide the ordering of the conditions Procrastinate and Shortest Move. Essentially, this follows from the assumption that only early and late features can be checked and hence trigger movement, whereas features unmarked for strength need not be checked. In a hypothetical example deciding between the two orderings, two operations, A and B, need to be possible at some point of the derivation with the following properties: operation A checks more late features, but involves shorter or less movement than operation B; operation B checks fewer late features than A, but involves a longer movement. Such cases are possible, however it follows from the strong/weak dichotomy that the more

late features a transformation checks, the fewer early ones it checks, and vice versa. Hence Earliness favors operation B over operation A. And since Earliness is higher ranked than Shortest Move, we will not be able to decide the relative ranking of Shortest Move and Procrastinate.

4.2 Potential Counterevidence

In this section, I present an apparent counterexample to the Earliness proposal: movement to subject position in English. I will then go on to show, that the assumptions that make this example a counterexample themselves are problematic.

Let us first consider what a potential counterexample would formally look like. An example of a strong feature *F* which could not be early would have the following properties. First, we would have to see that indeed overt movement is obligatory, as shown by examples where it is possible. Second, and this would be the case where the predictions of the strong/weak and the early/late system differ, we would need an ill-formed example sentence where overt movement is impossible for some independent reason, but covert movement is possible. On the Earliness system, it is predicted that such an example should be fine, because the system predicts that in this case covert checking should be good enough to satisfy the demands of the early feature. However, if the early feature is really a strong feature, it would be predicted that only overt checking/movement could satisfy the feature in question.

Chomsky (p.c.) points out that the EPP/D-feature in English seems to be such a counterexample to the proposed Earliness. As is well known, it is obligatory in English to fill the subject position of all clauses by Spell-Out. The condition that applies in English is referred to as the Extended Projection Principle (EPP). However, Modern Irish does not seem to obey such a restriction, as McCloskey (1994) argues based on evidence from salient unaccusatives as in . McCloskey (1994) even finds good arguments that the subject position in clauses like in is not filled by an invisible *pro*, as it is commonly assumed in other languages, e.g. Romance null subject languages and Icelandic impersonal passives and weather verbs. Hence, McCloskey (1994) claims that Modern Irish does not obey the Extended Projection Principle.

- (15) a. *Laghdaigh ar a neart* (McCloskey 1994)
decreased on his strength
'His strength decreased.'
- b. *Chuaigh agam*
went at-me
'I succeeded.'

Chomsky (1995) and (p.c.) propose a different analysis of EPP effects. Specifically, he assumes that English has a strong D-feature in the functional projection T. Irish, on the other hand, is analysed to have a weak D-feature on T. This has the consequence, that in English a subject position—a Specifier of TP—always needs to be overtly filled by a DP, to satisfy the requirement imposed by the strong D-feature of T. In Irish, on the other hand, it is sufficient if a covert movement checks the strong feature of T.

This accounts for the grammaticality difference between the acceptable Irish examples in (15) and their ungrammatical English counterparts in , if we assume that in the Irish examples the D-feature of Tense can somehow be covertly checked by movement of the D-feature of *a neart* in (15-a) and *agam* in (15-b).

- (16) a. *Decreased on his strength.
b. *Went at me.

In addition, let us assume that there is a condition active in English that blocks overt extraction out of the PP to the subject position. We have to make this assumption, since otherwise sentences like the ones in are predicted to be grammatical.

- (17) a. *[His strength]₁ decreased on *t*₁.
b. *Me/*I*₂ went at *t*₂.

Now, we have a counterexample to Earliness. If the feature that drives movement to the subject position in English was early, we predict that the sentences in (16) should be fine. This is so, because overt movement is blocked of the DPs *his strength* and *me* is blocked, as we see in (17). Hence, the sentences (17) where the DPs remain *in situ* as they do in Irish should be fine, contrary to fact.

I believe however, that the theoretical assumptions underlying Chomsky's argumentation are too problematic to make his argument convincing. First of all, it is not clear what the reason for the ungrammaticality for the examples in (17) is. In fact, we could assume that the movement out of the PP to the subject position is possible, but the moving DP cannot satisfy the Case requirements of both positions: the complement of a preposition and the subject position. If this was the case, we could assume that Earliness drives overt movement in the derivation of (16), but this movement causes a feature mismatch among some other features. This would consequently cause the derivation to be crash.

Secondly, another assumption of (Chomsky 1995) and previous work is that the English expletive *there* checks only a D-feature. As Martin Hackl (p.c.) points out, if we adopt this about *there*, it would follow that the sentences in should be grammatical, because the expletive *there* satisfies the EPP-

feature of T.

- (18) a. *There decreased on his strength.
b. *There went at me.

5 Conclusion

In this paper, I proposed that there are two possible values of feature strength: early and weak. The economy condition that applies at every step of the derivation consists out of three parts, which are the three conditions paraphrased in the following:

- (19) Earliness: Check early features as soon as possible.
Procrastinate: Check late features as late as possible.
Shortest Move: Prefer shorter movements.

Of the three economy conditions Earliness is higher ranked than the other two. Ranking, I understood in the sense of Prince & Smolensky (1994), so the other two conditions can be violated in order to satisfy the Earliness condition.

The evidence for this proposal comes from cases, where a feature that in general triggers overt movement is prohibited from moving overtly by an independent condition. In these cases the strong/weak feature system of (Chomsky 1993) predicts ungrammaticality, because a strong feature remains unchecked at Spell-Out. The early/late system predicts grammaticality, because an early feature can remain unchecked at Spell-Out, if there was no possible transformation that would have checked the early feature. I presented empirical evidence from three domains: object shift of Swedish object pronouns in section 2; *wh*-movement and D-linking in Russian in section 3.1; and morphological Case marking in 3. In all three cases the evidence decided in favor of the early/late feature system.

References

- BITTNER, MARIA, & KENNETH LOCKE HALE. 1993. Ergativity: Towards a theory of a heterogenous class. Rutgers University and MIT ms.
—, & —. 1994. The structural determination of Case. *Linguistic Inquiry*. Forthcoming.
BOBALJIK, JONATHAN D. 1995. *Morphosyntax: The Syntax of Verbal Inflection*. Cambridge: Massachusetts Institute of Technology dissertation.
CHOMSKY, NOAM. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Berlin: Mouton de Gruyter.

- . 1991. Some notes on economy of derivation and representation. In *Principles and Parameters in Comparative Grammar*, ed. by Robert Freidin, 417–454. Cambridge: MIT Press.
- . 1993. A minimalist program for linguistic theory. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. by Kenneth Locke Hale & Samuel Jay Keyser, number 24 in Current Studies in Linguistics, chapter 1, 1–52. Cambridge: MIT Press.
- . 1995. Chapter four: Categories and transformations. MIT ms.
- CHUNG, SANDRA. 1978. *Case Marking and Grammatical Relations in Polynesian*. Austin: University of Texas Press.
- DIESING, MOLLY, & ELOISE JELINEK. 1995. Distributing arguments. *Natural Language Semantics* 3.123–176.
- EPSTEIN, SAMUEL D. 1992. Derivational constraints on \bar{A} -chain formation. *Linguistic Inquiry* 23.235–259.
- HARLEY, HEIDI. 1994. Abstracting away from abstract case. In *Proceedings of NELS 25*, ed. by Jill Beckman, Amherst. GLSA.
- . 1995. *Subjects, Events, and Licensing*. Cambridge: Massachusetts Institute of Technology dissertation.
- IDSARDI, WILLIAM JAMES. 1992. *The Computation of Prosody*. Cambridge: Massachusetts Institute of Technology dissertation.
- KITAHARA, HITSATSUGU. 1993. Deducing ‘superiority’ effects from the shortest chain requirement. In *Harvard Working Papers in Linguistics*, ed. by Höskuldur Thráinsson, Samuel D. Epstein, & Susumu Kuno, 109–119. Cambridge: Department of Linguistics, Harvard University.
- KOIZUMI, MASATOSHI. 1995. *Phrase Structure in Minimalist Syntax*. Cambridge: Massachusetts Institute of Technology dissertation.
- LAMONTAGNE, GREG, & LISA TRAVIS. 1993. The syntax of adjacency. In *Proceedings of WCCFL 6*, 173–186, Stanford. CSLI.
- MAHAJAN, ANOOP. 1994. The ergativity parameter: *have-be*-alternation, word-order and split-ergativity. In *Proceedings of NELS 24*, 317–331, Amherst. GLSA.
- MARANTZ, ALEC. 1991. Case and licensing. In *Proceedings of ESCOL*, ed. by Germán F. Westphal, Benjamin Ao, & He-Rahk Chae, 234–253, Ithaca. Cornell WPL.
- MASSAM, DIANE. 1994. Case without functional projections. University of Toronto, ms. To appear in the proceedings of the CLA.
- MCCLOSKEY, JAMES. 1994. Subjects and subject-positions in Irish. University of California at Santa Cruz ms.
- PESETSKY, DAVID. 1989a. *Wh*-in-situ: Movement and unselective binding. In *The Representation of (In)definiteness*, ed. by Erich Reuland & A. ter Meulen, 98–129. Cambridge: MIT Press.
- . 1989b. Language particular processes and the earliness principle. MIT ms.
- . 1994. *Zero Syntax: Experiencers and Cascades*. Cambridge: MIT Press.
- PRINCE, ALAN, & PAUL SMOLENSKY. 1994. Optimality theory: Constraint interaction in generative grammar. Rutgers University and University of Colorado at Boulder. Forthcoming at MIT Press.
- REINHART, TANYA. 1994. *Wh*-in-situ in the framework of the minimalist program. OTS Working Papers 94/03, Utrecht University, Utrecht.

- SAUERLAND, ULI. 1995. The *lemmings* theory of Case. In *Proceedings of SCIL VII*, ed. by Jason Boro, Laura Conway, Pai ling Hsiao, & Elizabeth Laurençot, Cambridge. MIT Working Papers in Linguistics. To Appear.
- YIP, MOIRA, JOAN MALING, & RAY JACKENDOFF. 1987. Case in tiers. *Language* 63.217–250.
- ZAENEN, ANNIE, JOAN MALING, & HÖSKULDUR THRÁINSSON. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3.441–483.

Author's address: Uli Sauerland
Room 20C-231B
MIT, Cambridge MA 02139
sauerlan@mit.edu