



LUND UNIVERSITY

Michal Starke's Move Dissolves into Merge: A Theory of Locality

Beijer, Fabian

2003

[Link to publication](#)

Citation for published version (APA):

Beijer, F. (2003). *Michal Starke's Move Dissolves into Merge: A Theory of Locality*. (The Department of English in Lund: Working Papers in Linguistics). <http://www.englund.lu.se/research/workingpapers/index.phtml>

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Michal Starke's *Move Dissolves into Merge: A Theory of Locality*. A Summarizing Review.

FABIAN BEIJER

1 Introduction

After having presented the main ideas at various seminars, conferences and EGG summer schools¹ at least since 1999, Michal Starke completed his doctoral thesis in 2001. The thesis has not yet been published in book format². It is, however, available on the Internet as a pdf file at http://theoling.auf.net/papers/starke_michal. It is my personal opinion that Starke's thesis deserves more publicity and attention than it has hitherto received.

The main aim of this review article is thus to draw people's attention to a thesis I consider to be radical, inspiring, and of outmost theoretical importance. Another aim is to provide such a comprehensive account of the main ideas of the thesis that the interested reader not only acquires a decent understanding of the main ideas and arguments advocated by Starke, but also enough information on the contents of the thesis to be able to decide whether or not reading the entire thesis would in fact be worthwhile, given his or her specific field of interest.

The thesis under discussion contains the following ten chapters (preceded by an introduction):

1. Unifying Locality
2. SpecificQ-movement: Unifying Extractions out of Weak Islands with Standard Locality
3. Strong Islands: θ -across- θ
4. θ -movement as 'Long Scrambling'
5. The 'argument/adjunct' Asymmetry: R (θ , presupposition)
6. Clause-internal θ -movement
7. Interlude: Some Interpretive Illusions
8. Unifying Merge and Move
9. Purifying Merge of its 'specifier' Heritage
10. Overview

The present article is organised as follows: In section 2, I disclose the main aims of Starke's thesis, without going too much into the actual argumentation, and explain the motivation behind Starke's chapters 2 through 7 by accounting for the main ideas behind his theory of locality. In section 3, I give an account of the most important arguments in chapters 2 through 7. In section 4, I discuss Starke's chapter 8 ("Unifying Merge and Move"), the first chapter in which Starke proposes radical changes to modern, generative phrase and clause theory as we know it. In section 5, Starke's chapter 9 ("Purifying Merge of its 'specifier' Heritage"), his second radical proposal concerning simplification of modern syntactic theory, is outlined and discussed. In section 6, the review is concluded by a summary and some critical remarks.

¹ <http://egg.auf.net>

² Please note that Starke's chapter 9 is about to be published in Belletti (to appear).

2 Overview, Motivations and Locality

As given in the title of the thesis, Starke's main concern is locality, and the implications a unified view of locality will have on issues such as phrase structure and the nature of the phenomenon known as syntactic movement. The main idea is that given a proper theory of locality, and an appropriate understanding of the nature of the functional sequence, the syntactic operation Move will dissolve into the simpler, more economic, indispensable syntactic operation Merge, and X-bar theory will be simplified by dispensing of so-called specifiers, when these, just like heads, are allowed to project. This will be explained in more detail later on.

I take it that one of Starke's original ideas was that there should be no need for several different principles of locality in a minimal syntactic framework. In an ideal world, all principles of locality, such as Strong Islands, Relativized Minimality, Subjacency, Control, and Principles A and B of Binding Theory, could be reduced to only one principle. Starke argues that, in this respect, the present world is actually ideal, in that all restrictions on locality can be accounted for by a slightly modified version of Rizzi's (1990) Relativized Minimality (henceforth RM). Since locality is such an important part of movement, unifying locality is central to a proper understanding of movement.

Concentrating on the locality of *wh*-movement, Starke starts out by saying that there are three main generalizations to be accounted for, namely Weak Islands (henceforth WI), Extraction out of Weak Islands (eWI) and Strong Islands (SI)³. According to Starke, no unified account of these generalizations has been presented, despite numerous attempts (e.g. Chomsky 1986, Cinque 1990, Manzini 1992). Instead, the usual situation is that RM, or one of its close relatives (Minimal Link Condition, Shortest Move, Attract Closest), is used to explain WI, that Binding Theory is employed in order to account for eWI, and that SI is accounted for by some version of Chomsky's (1986) Barriers. As hinted at in the previous paragraph, Starke's aim is to show that an elaborated version of RM operating on a refined data-structure involving syntactic features organised in a feature tree, is all that is needed to account for all three generalizations.

What the generalization WI boils down to is the observation that it is impossible to move a quantificational element⁴, such as a *wh*-word, across another quantificational element, including negation (1 and 2), focalised elements (3), other *wh*-elements (4), and quantified adverbials (5), while the *wh*-movement is unproblematic when no such element intervenes (6)⁵:

- (1) *How don't you think that I should cook this stuff <how>?
- (2) *How do you think that I shouldn't cook this stuff <how>?
- (3) *How do you think that, THIS STUFF, I should cook <how>, not those eggplants over there?

³ The generalizations in question will of course be further discussed below, but to remind the reader of what they refer to, some linguistic examples might be useful:

WI:	* How do you wonder [what to repair ^t _{what} ^t _{how}]?	(Manzini 1997: 135)
eWI:	What do you wonder [how to repair ^t _{what} ^t _{how}]?	(Manzini 1997: 136)
SI:	* How have you found someone who would fix it ^t _{how} ?	(Cinque 1990: 1)

⁴ Note that Starke is not pleased with this term, since elements normally referred to as 'quantifiers' are not included, but he uses it anyway, for lack of a more appropriate one.

⁵ All examples in this paper are identical to Starke's, unless otherwise specified. The angle brackets indicate the position of Q1 before it is moved across Q2. Since the same relation can be illustrated with other means, Starke is not consistent in his use of angle brackets.

- (4) *How do you wonder why I should cook this stuff <how>?
- (5) ?*How should I often cook this stuff <how>?
- (6) How do you think that I should cook this stuff <how>?

On an abstract level, the ungrammatical configurations in (1) through (5), can be represented as (7), i.e. a member of the feature class (henceforth FC) Q (i.e. Q_1) cannot move across a member of the same feature class (i.e. Q_2):

- (7) * $Q_1 \dots Q_2 \dots <Q_1>$

Throughout the rest of this paper, bear in mind that Q refers to a FC, i.e. a class of features, not to the feature itself, and, consequently, that WI concerns FCs, not the features themselves.

Another lesson that (7) is intended to teach us is that grammar is restricted by an abstract anti-identity constraint, telling us that two items of the same class cannot “overlap”, or that (8) is ungrammatical if class α and class γ are identical, and γ intervenes between the two instances of α :

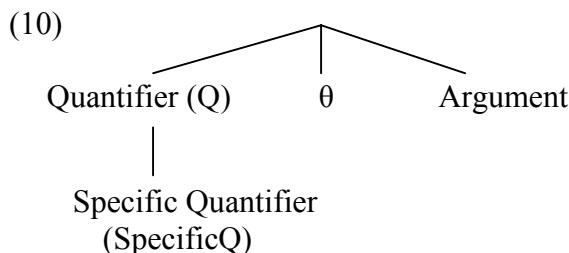
- (8) * $\alpha \dots \gamma \dots \alpha$

An important thing to note here is that, since an element α can be a member of more than one class, and since (8) can be grammatical with regard to one class, but ungrammatical with regard to another, the anti-identity constraint must be sensitive to what type of relation is being built. Consequently, Starke's first formulation of the definition of RM must be as given in (9):

- (9) X-relating two occurrences of α is legal only if α is a member of X and there is no γ such that γ is a member of X and γ intervenes between the two occurrences of α .

Note that, for the time being, to relate two occurrences of α can be interpreted either as a movement operation, or as the creation of a chain relation.

As already mentioned, Starke considers FCs to be ordered in syntactic feature trees (henceforth FT). Each node in the tree defines a FC containing itself and every node it dominates. In the FT (10), for instance, the feature Quantifier defines the FC Quantifiers containing Specific Quantifiers:



The nature of the FT gives us new situations to be handled by RM (the anti-identity principle), such as cases when the intervener belongs either to a subclass, or to a superclass, of the elements to be related. Let us imagine a class C , to which α , but not x belongs. As already stated above, the configuration in (11) is then grammatical, while the one in (12) is ungrammatical:

- (11) $\alpha \dots x \dots \alpha$
 (12) $* \alpha_i \dots \alpha_j \dots \alpha_i$

If we then suppose that the class C has a subclass, SC , and that β is a member of this class SC , it means that β is a member of C as well. If we instead call β $\alpha\beta$, the two new environments mentioned above, which the anti-identity constraint must be able to handle, are given in (13) and (14):

- (13) $\alpha\beta \dots \alpha \dots \alpha\beta$
 (14) $* \alpha \dots \alpha\beta \dots \alpha$

The logic is then that in (13) $\alpha\beta$ is a member of both C and SC , which means that it can “choose” to do either C -movement or SC movement. C -movement is ruled out given the C intervener, while SC movement is OK, given that no SC -element intervenes. The configuration in (14), however, can never be good, since α can only do C -movement (since C is the only class to which it belongs), and $\alpha\beta$ is a member of the class C .

Returning to our three island generalizations above, we can conclude that (13) illustrates a successful eWI, in that α does not block movement of $\alpha\beta$ (but cf. 12 above), while (14) shows an SI that blocks all kinds of movement. In other words, (13) shows that a subset of the quantifiers, namely those with additional properties, will be able to escape WI, which is what we mean by eWI. What (14) above and (15) below show is that if something stronger than a WI intervenes, nothing will be able to extract, and the configurations are ruled out (unless a further subclass becomes relevant):

- (15) $* \alpha\beta \dots \alpha\beta \dots \alpha\beta$

Starke takes this to show that one and the same locality principle can be used to derive all three islands. In the following chapters, I provide a brief account of Starke’s chapters 2 through 7, in which he implements this model and this logic using real language data. The reason he spends so much energy on this is that he needs to show that the logic so far developed is on the right track, in order to prepare the ground for what is to come in the two last chapters, i.e. the chapters in which the category Specifier and the operation Move are argued to be redundant and non-existent, respectively.

3 The building up, and the Organisation of, the Syntactic Feature Tree

As already mentioned, the present section contains a brief account of the chapters in which Starke shows and motivates his conception of locality. It is divided into several sub-sections, basically corresponding to the various chapters in Starke’s thesis.

3.1 Unifying eWIs with standard locality [Starke’s ch. 2]

As mentioned above, WIs are weak in the sense that they do not always block movement⁶. When movement is not blocked, although there is a potential blocker intervening, the label eWI is employed. A general informal formulation of the difference between elements moving in eWI configuration and elements unable to move in ordinary WI configurations is that

⁶ If Movement is not part of your favourite syntactic framework, consider the use of Movement in this section to be equivalent to Chain Formation or Internal Merge.

elements that can successfully extract out of WIs have some additional property in relation to those that cannot (i.e. the situation illustrated by 13 above).

Previous research is in agreement that elements capable of extracting out of WI have some additional property, but opinions diverge as to the exact nature of this additional property⁷. Starke's first attempt at a generalization is that we can only extract out of a WI if we have reason to believe that there exists some entity that the interlocutor has in mind as the referent of the *wh*-phrase. To illustrate this, Starke gives us (16) and (17), uttered in the contexts given before the actual examples:

- (16) You are reading a story to Joey. After a while Joey interrupts you and says "I wonder what Belga found! Could it be...?" and stops in the middle of the sentence looking at you starry-eyed. You stop reading and ask:
- a. So? What do you think that Belga discovered?
 - b. So? What do you wonder whether Belga discovered? [Adapted from Starke's (24)]
- (17) I know that you have no clue about what Herbert will cook tonight, and that you are curious about it, so I ask:
- a. What do you hope that Herbert will cook?
 - b. # What do you wonder whether Herbert will cook⁸.

In (16) both the unhindered extraction (16a) and the eWI (16b) are OK, since Joey obviously has a certain referent in mind, as indicated through "Could it be...?". In (17), only the unhindered extraction is OK, since it is obvious from the context that the interlocutor has no referent in mind. Starke concludes that we have good reason to assume that the difference between movement of a Q- element and movement of a Q β -element (where β stands for some additional property) is as stated in (18):

- (18) Q β -movement carries existential presupposition, while Q-movement does not.

This is not, however, the whole truth. According to Starke, existential presupposition *per se* is not enough. In addition to existential presupposition, eWI also requires help from the discourse in terms of either a range or a specific entity. He thus claims that there are two ways out of a WI, one giving rise to 'range' semantics, as in (19), and one giving rise to 'specificity'⁹ semantics, as in (20), both of which are sensitive to the nature of the intervener:

- (19) What is it unclear whether we should repair?
(20) What is it unclear how we should repair?

⁷ For instance, Starke says, Huang (1982), Lasnik & Saito (1984), and Rizzi (1988) take it to be θ -role (i.e. the difference between arguments and non-arguments), Manzini (1992) and Rizzi (2000) consider it to be case or DP-hood, Cinque (1990) and Rizzi (1990) talk about "referentiality" as opposed to pure quantificational readings, and Szabolcsi & Zwarts (1997) take the additional property to be richness of internal semantic structure.

⁸ The # sign indicates that the example in question is pragmatically odd in the given context, even though it is not ungrammatical in itself.

⁹ Starke uses 'specificity' in a particularly restrictive sense. Enç (1991) calls both 'range-based' and 'specificity-based' antecedents 'specific', but he differentiated between 'partitive-specifics' and 'familiar specifics'. Starke uses 'specificity-based *wh*-phrase' to refer to what Enç (1991) called 'familiar-specifics' and 'range-based *wh*-phrase' to refer to Enç's 'partitive-specifics'.

The generalization drawn from this is that eWIs with *whether*-interveners can be used as generic information request questions, given a clear range, while eWIs with *how*-interveners require a stronger context, i.e. that a specific entity is available as a potential antecedent. That (20) requires a stronger context has the effect that (19) is perceived as less degraded than (20) by speakers. An interesting fact to note is that this difference in acceptability disappears in eWI out of infinitive clauses (21 and 22):

- (21) What is it unclear whether to repair?
- (22) What is it unclear how to repair?

We will have reason to come back to this below. Another interesting observation of Starke's is that the contrast between (19) and (20), i.e. the difference between the *whether*-type eWI and the *how*-type eWI, looks like the well-known argument/adjunct asymmetry, illustrated in (23), where argument intervention (23a & b) is less degraded than adjunct intervention (23 c, d & e):

- (23) a. What is it unclear whether we should repair?
b. To whom is it unclear what we should give?
c. What is it unclear when we should repair?
d. What is it unclear how we should repair?
e. What is it unclear why we should repair?

Starke sums up the discussion so far by saying that we now have the following new questions to answer: Why does the distinction between (19) and (20) exist? Why does it disappear in infinitives (21 and 22)? Why is the argument/adjunct nature of the intervener relevant? If range and specificity are two routes out of weak islands, where do their respective classes belong in the syntactic FT?

To start answering these questions, Starke turns to French *wh* in-situ. French *wh* in-situ questions look like English echo questions, even though they are regular questions, for instance (24) (25) and (26)¹⁰:

- (24) Tu crois qu'elle a fait quoi?
you think that she has done what?
- (25) Tu crois qu'elle a dit ça pour inciter Pierrot à séduire qui?
you think that she has said this to incite Pierrot to seduce whom?
- (26) Tu crois qu'ils vont rembourser ceux qui ont voyagé comment?
you think that they will reimburse those who have travelled how ?

Example (24) shows a *wh* in-situ question in which no island is involved, while (25) and (26) in fact contain SIs (an adverbial clause and a relative clause within a noun phrase respectively). This means that French *wh* in-situ, paradoxically enough, survives in a SI context. Inserting a WI, however, leads to sharp degradation, regardless of the presence of a SI and regardless of the position of the WI, e.g. (27) and (28):

- (27) Tu crois qu'elle a pas fait quoi?
you think that she has not done what ?

¹⁰ All glossing throughout the paper is Starke's.

- (28) Tu crois qu'elle a pas dit ça pour inciter Pierrot à séduire qui?
 you think that she has not said this to incite Pierrot to seduce whom?

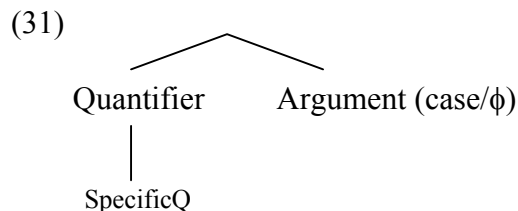
The logical way of interpreting these facts is to say that due to RM, *wh* in-situ (i.e. covert *wh*-movement) is blocked by any intervening Q, but since so-called SIs do not involve Q, they do not block *wh* in-situ. The *wh* in-situ configurations are thus parallel to cases of extraction in which no island is involved. They are even parallel in the sense that (27) and (28), containing WI, are actually grammatical with a slight accent on the *wh* phrase in-situ triggering the same presuppositional interpretation as the eWI cases discussed above¹¹.

The next question Starke considers is whether this presuppositionality effect concerns 'range' or 'specificity'. Example (29) illustrates that *wh* in-situ configurations are ungrammatical with intervening specific NPs, while they are grammatical within NPs explicitly indicating a range (30):

- (29) * Tu amerais avoir cette/ma photo de qui?
 you would-like to-have this/my picture of whom ?
 (30) Tu amerais avoir une des photos de qui?
 you would-like to-have one of-the pictures of whom ?

This indicates that the additional property β that makes extraction out of WI possible is 'specificity' not range, since quantifier movement can chose to involve a 'specific quantifier' when Q intervenes, while extraction is impossible when a 'specific Q' intervenes.

In conclusion, covert movement is a pure Q-class movement, and SpecificQ is a subclass of Q. Covert movement can thus ignore SIs, since SIs do not involve Q, but they can only extract out of WI by taking the SpecificQ route. Overt movement, on the other hand, has access to the range-based route, and therefore it does not entail specificity. All in all, this means that (31) illustrates the FT argued for so far:



The discussion so far leaves Starke with two main questions to be answered: (i) What is the nature of the difference between overt and cover *wh*-movement, such that SIs block overt but not covert *wh*-movement? (ii) What is the role of the range-based route out of WIs?

3.2 Strong Islands (\emptyset across \emptyset) [Starke's ch. 3]

In the previous sub-section we learnt that overt *wh*-movement cannot be movement of Q only (since it is blocked by SIs), but must involve something else as well. In the present section, Starke's answer to this question is discussed. Starke argues, on the basis of English and French data, that *wh*-phrases can move only if they are directly connected to the predicate, i.e. they cannot move if they are related to the entire proposition. When related to the entire proposition, they are directly merged in the left periphery. This means that (32) is ungrammatical if we consider *how* to have moved from the lower clause and thus to be related to the predicate *cost*.

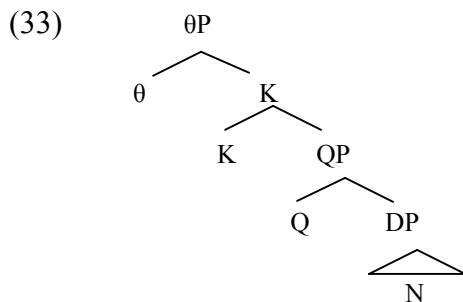
¹¹ Note that both the interpretation and the intonation are distinct from the echo reading.

- (32) How do you think that this can cost so much now? It was still affordable yesterday.

In order to explain these facts, Starke introduces some new terminology, and some changes to the scope of old terminology. Firstly, he informs us that he will use the term ‘theta role’ to refer to all relations between a predicate and its modifiers. In this terminology, adverbials of manner, time, reason, etc. all have theta-roles, regardless of their argumental status. Secondly, Starke tells us that the term ‘outcast’ will be employed to denote elements lacking theta-roles, and that the term ‘insider’ will be used to denote those that do. Starke can now restate the generalization above and say that ‘ θ -less *wh*-phrases cannot move (regardless of islands)’ or ‘outcasts cannot move’. The question is then why this is so, and the underlying question is why theta-roles should matter to Q-movement in the first place. Could it be that the reason SIs do not block covert movement, but block overt movement of Q, is that overt Q-movement involves some additional property, namely θ ? Unfortunately, Starke says, this matter cannot yet be settled.

To decide whether it is one and the same property, theta role, that makes overt *wh*-movement sensitive to SI and makes outcasts immobile, Starke turns to SIs themselves. He claims that ‘everything is a SI, except DPs with structural case’. To support this claim he argues that structural case is nothing but a geometrical relationship, in which no theta role is involved, while inherent case is the spell-out of a theta role. Given this, Starke concludes that (i) only theta-marked phrases can be (overtly) *wh*-moved, (ii) θ -marked phrases are SIs for overt *wh*-movement, and (iii) covert Q-movement does not respect SIs, which means that overt *wh*-movement involves movement of θ , i.e. θ -movement, and that what Starke has said so far makes logical sense.

For θ to be visible to RM it has to be structurally represented. Starke assumes there to be a θ P on top of every functional sequence or ‘Extended Projection’, so that, for instance, a nominal sequence looks like the one in (33)¹²:



When the nominal sequence lacks inherent case, the nominal θ P is empty or, rather, missing, so that TP (through a low verbal theta-assigning projection also called θ P) establishes a relation with KP instead of with θ P.

The next thing we have to understand is why overt Q-movement triggers θ -movement, while covert Q-movement does not. The reason, in the present framework, is, quite simply, that we cannot move parts of θ e.g. QP, since Q is a bound morpheme (in English and most other languages) and we then split a morphological unit, giving rise to something unpronounceable¹³. Overt *wh*-movement must thus, according to Starke, be pied-piping of a *wh*-phrase by some constituent containing it, i.e. overt *wh*-movement is movement of a θ P in

¹² In the same way, a clause is a CP dominated by a θ P.

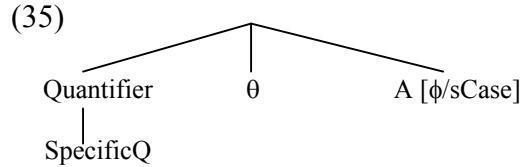
¹³ Note that the obvious prediction is that in languages expressing the *wh*-feature as a free morpheme, it should be possible to escape SIs, which, according to Starke, is also the case.

order to pied-pipe the Q-feature. Pied-piping is thus two movements in one (in this case movement of Q and movement of θ), which means that it is subject to double locality restrictions. So what SIs reveal is the following constraint on pied-piping:

- (34) If α -movement involves pied-piping of α by γ , then the locality restrictions of both α and γ must be respected.

In other words, every θ -marked syntagm is a SI.

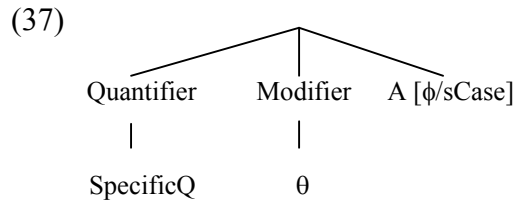
In terms of Starke's FT, the only possible position for θ is a sister to Q and A(rgument), i.e. (35):



It is tempting here, Starke points out, to view θ , the way we have defined it here, as identical to Rizzi's (2001) 'modifier' class, a class claimed to be relevant to RM (along with 'quantifiers' and 'arguments'). Starke shows, however, that modifier-movement and θ -movement cannot be collapsed, since intervening modifiers do not block *wh*-movement (36):

- (36) How do you usually sleep?

A more plausible solution is, according to Starke, to view his θ as a subset of Rizzi's modifiers, i.e. 'predicate-modifiers' is a subset of all modifiers in the clause. This explains the grammaticality of (36), in that θ , being a subclass of M(odifiers), can move over M by the same logic that lets SQ move over Q, i.e. (37):



3.3 θ -movement as 'Long Scrambling' [Starke's ch. 4]

As already mentioned, Starke claims there to be two routes out of a WI, one specificity-based (38) and one range-based (39):

- (38) **what**_{SQ}...V[CP how_Q...[θ P **t**_{what} [θ P **t**_{how}...[VP...**t**_{what} **t**_{how}

- (39) **what**_Q...[θ P **t**_{what}...V[CP how_Q...[θ P **t**_{what} [θ P **t**_{how}...

That (38) exists has already been shown in connection with, for instance, *wh* in-situ. The next thing to show is thus that (39) exists as well.

To begin with, Starke points out that there must be an intermediate landing site used in eWI configuration, and shows that this position must be between the subject and the predicate of the matrix clause. For reasons related to reconstruction, this position must be a θ -position,

not an SQ-position, since we would otherwise have a situation in which Q-elements could move passed SQ, which is impossible given the FT in (37), which shows that successive-cyclic θ -movement (39) is a way to escape WI (given that 37, and the reasoning behind it, is on the right track).

Similarly, French so-called long *tout*-movement data shows that *tout* can move over Q and A interveners, at the same time as Q- and A-movement can take place over *tout*. In addition, *tout*, just like the θ -step of eWI discussed above, (i) lands in low positions, (ii) is able to scramble across clauses, (iii) is able to jump over adverbs ('modifiers'), and (iv) is unable to reconstruct for scope.

Another interesting fact revealed through the examination of long *tout*-movement is that it carries 'range'-based presuppositional semantics, i.e. (40), with long *tout-movement*, can only be used if there is general knowledge of a range of objects directly threatened by the Nile, were it to flood:

- (40) Les crues du Nil ont tout commencé à emporter sur leur passage.
the floods of-the Nile have everything started to remove on their path

This 'range' interpretation is achieved regardless of scope. It can thus be concluded that θ -movement triggers 'range'-interpretations instead of specificity interpretations, which would then suggest that whether we have θ -movement or SQ-movement decides whether we get a 'range' or a specificity based interpretation of eWI.

There is one more observation about *tout*-movement that we have to be aware of. Firstly, *tout* can relate to subjects, direct objects, and indirect objects, which means that there can be more than one instance of *tout* in the same sentence, which all move, granted that they obey the order [nominative > dative > accusative], i.e. the C-command relations between the moved constituents must be obeyed throughout the derivation. Similar facts have been observed by Haegeman (1993) and Watanabe (1992) concerning scrambling and *wh*-movement. According to Starke, it all boils down to a 'Round Robin Constraint' (henceforth RRC) (41) which tells us that a chain *j* can never cross a chain *i*¹⁴:

- (41) $*\alpha_j \alpha_i \dots \alpha_i \alpha_j$

Starke claims that the standard RM-configuration in (42) is a special case of (41), in which the intervening chain is a degenerate one with only one member:

- (42) $*\alpha_j \dots \alpha_i \dots \alpha_j$

This reasoning requires a theory in which RM is a constraint on chains, not on individual landing sites of various elements that move.

We have just seen that more than one instance of the universal quantifier *tout* can appear in the same sentence, as long as they obey a certain order. This order can actually marginally be violated, Starke says, if the highest quantifier is specific and focused as in (43):

- (43) ??Il a TOUT toutes voulu qu'elles admettent.
he has everything everything_{fem} wanted that they_{fem} admit

¹⁴ It has been pointed out to me that Starke's reasoning here might be problematic in the light of Richard's (1998) Principle of Minimal Compliance. Due primarily to the limited space at my disposal, I will not discuss to what extent this objection is valid.

Starke takes this to show that the route out of eWI given in (38) above has to be slightly modified. Either (44), (45), or (46) represents the true state-of-affairs, Starke says, before he concludes that (44) is ruled out since we have a RM violation in the form of a Q-over-SQ-movement:

- (44) ***what**...[SqP *t_{what}*...[CP how_Q...[SqP *t_{what}* [θP *t_{what}* [θP *t_{how}*
 (45) **what**...[SqP *t_{what}*...[θP *t_{what}*... [CP how_Q...[θP *t_{what}* [θP *t_{how}*
 (46) **what**...[SqP *t_{what}*...[CP how_Q...[θP *t_{what}* [θP *t_{how}*...

Whether (45) or (46) is correct is further discussed below.

All in all, successful θ-movement (i) can move over intervening A, Q, and M, (ii) does not block A, Q, or M-movement, (iii) can give rise to Q- and SQ-movement, (iv) reconstructs for binding, but not for scope, (v) give rise to 'range'-presuppositions, and (vi) respects the RRC. Properties (i) through (iii) follow logically from the FT in (37), (iv) and (v) show that every type of chain has its own interpretative characteristics, while (iv) will later make it possible to reformulate the principle of locality.

3.4 The 'Argument/Adjunct' Asymmetry [Starke's ch. 5]

The Argument/Adjunct asymmetry refers to the contrast between (47) and (48) ever since Huang (1982):

- (47) a. What do you wonder whether to cook today?
 b. *How do you wonder whether to cook today?

The first question is whether the relevant factor is actually 'argument or adjunct', or if the true nature of the contrast is to be sought elsewhere. The second question is in what way the contrast relates to presuppositionality. If presuppositionality is the relevant factor, as argued in 2.1, why should there be an additional 'argument/adjunct' asymmetry?

Starke argues that the relevant difference is not between arguments and adjuncts, but between cased and uncased elements. Concerning *wh*-phrases, there is a cut-off point between those that pattern with (47a) and those that pattern with (47b), in that cased DPs pattern with (47a) and uncased adverbials pattern with (47b) as illustrated in (48)¹⁵:

- (48) a. Who do you wonder whether I should invite?
 b. What do you wonder whether I will cook?
 c. (?) To whom do you wonder whether I will tell how to cook this?
 d. (?) Where do you wonder whether they cook this kind of things?
 e. ?* When do you wonder whether they cook this kind of things?
 f. * How do you wonder whether they cook this kind of things?
 g. * Why do you wonder whether they cook this kind of things <why>?

In conclusion, Starke argues, cased *wh*-phrases can extract out of WI, while uncased ones cannot. This is what underlies Huang's asymmetry.

Starke goes on to discuss various kinds of modifiers. Since temporal and manner adjuncts all have relations to the predicate, they bear a θ-role in Starke's terminology, but since these

¹⁵ An essential part of Starke's argument here is the observation that natural language has locative, but not temporal, case.

roles are never associated with inherent case, only a proper subset of θ -roles is associated with inherent case. Starke introduces the term $i\theta$ for those θ -roles that correspond to (inherent) case, and concludes that the correct description of Huang's so-called argument/adjunct asymmetry is that it is easier to extract *wh*-phrases with inherent case from WIs. This means that we now seem to have two unrelated conditions on eWI, namely (i) 'only cased *wh*-phrases can extract' and (ii) 'only presuppositional *wh*-phrases can extract'. Obviously, Starke's next step is to show that the two conditions are only apparently unrelated to each other.

Enç (1991) discusses the observation that morphological presence of accusative case correlates with specificity, in that all and only case-marked accusative NPs are specific. Getting rid of Enç's and others' misleading terminology, Starke argues, the correct description of the situation discussed by Enç (1991) is that accusative morphology correlates with presuppositionality (since Enç uses 'specific' in the same sense as Starke uses 'presuppositional'), or, rather, that accusative inherent case correlates with presuppositionality.

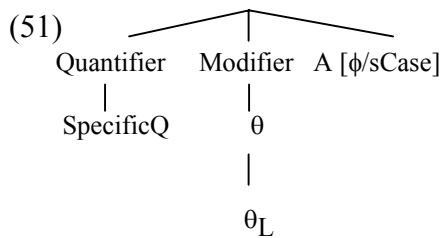
It is, however, the case that presuppositionality, but not inherent case, is necessary for eWI. If presuppositionality is controlled for, it can be observed that adjuncts are actually extractable when presuppositional, as illustrated in (49):

- (49) How didn't he want to eat the dish: with a fork or with Chinese sticks?

The descriptive generalisation behind the 'argument/adjunct' asymmetry is thus that inherent case makes it easy to build the relevant presuppositions, i.e. (i) eWI requires presuppositionality on the *wh*-phrase, (ii) inherent case facilitates presuppositionality, but (iii) presuppositionality can be forced by an adequate discourse setting, independently of case. From this it follows that uncased adverbials need help from the discourse context to extract out of WI, while cased DPs do not, hence the 'argument/adjunct' asymmetry.

Referring to a licensed θ as θ_L , what Starke has been arguing for in this section can be summarized as in (50), which gives us the FT in (51), since licensed θ is a subset of θ :

- (50) a. eWI requires θ_L
b. θ_L is licensed either by inherent case or by discourse



According to Starke, this tree entails, among other things, (i) that WI-escaping θ -movement should be rephrased as θ_L -movement, which means that SQ-movement and θ_L -movement allow eWI while A-movement and θ -movement are clause-bound, (ii) that SIs are environments in which inherent case intervenes, (iii) that overt *wh*-movement is pied-piping of the *wh*-feature by θ , and therefore θ_L when θ is licensed, (iv) that *wh* in-situ is only sensitive to WI because no θ -feature is involved, (v) that outcasts cannot move overtly, since no θ is present, which means that no pied-piping is possible, (vi) that so-called high *tout*-movement is successive-cyclic θ_L -movement, (vii) that WIs are cases of Q-across-Q, (viii)

that eWI is either θ -across-Q or SQ-across-Q, and, finally (ix) that the so-called definiteness island is an SQ-across-SQ violation.

3.5 Clause-internal θ -movement [Starke's ch. 6]

In his chapter 6, Starke sets out to explore what he terms the lower θ -step of *wh*-movement. He makes the following initial stipulations (i) θ -movement and θ_L -movement land in the same type of position, namely θP (i.e. there is no $\theta_L P$), (ii) round robin-configurations of multiple θ_L or θ phrases must be homogeneous, i.e. either they are all θP or they are all $\theta_L P$ ¹⁶ (we could temporarily think of the multiple θ -phrases as multiple specifiers of a single head, but see section 5 for a “Starkean” version of that idea), and (iii) movement of the θ -type must be uniform in that it cannot start by doing θ_L -movement and then continue by doing θ -movement, and vice versa.

One prediction based on these assumptions and the machinery developed above is that intervening subjects cannot be θ -phrases. In order to intervene legally, a subject must be a θ -less DP, since any intervening θ blocks the lower θ -step of *wh*-movement. According to Starke, this prediction matches the asymmetry that pre-verbal subjects in *wh*-movement configurations cannot be non-specific¹⁷ and indefinite, i.e. (52) is bad, while (53) is OK:

- (52) *Who did a policeman see?
(53) Who did this policeman see?

If θ -movement maps onto non-specificity just like SQ-movement maps onto specificity, the empirical observation that subjects cannot be non-specific in *wh*-movement configurations is the same fact as the theoretical observation that subjects in such configurations cannot be θ -phrases. Given that θ -movement maps onto non-specificity, it is tempting, Starke says, to conclude that A-movement maps onto specificity. If this is indeed the case, each type of movement seems to be associated with a particular semantics, in that θ -movement is non-specific, θ_L -movement is associated with range, SQ-movement is a specific referent for a quantifier, A-movement is a specific DP, and Q-movement is pure quantification.

Another mystery that Starke argues is solved by a proper recognition of the RRC is that *wh*-movement of an object is impossible if dative-shift has taken place, i.e. (54) is bad¹⁸:

- (54) *What did you give Paul <what>?

The idea is that a dative-shifted phrase undergoes θ -movement, thereby blocking further *wh*-movement.

It seems to be a problem for Starke that the Swedish correspondent of (55) is fully grammatical:

¹⁶ Unintentionally, Starke here seems to introduce a contradiction, in that he, after stating that they do not exist, talks about obviously existing θ_L phrases. One way out of this paradox is to say that there are no θ_L phrases in the sense that there are no such projection in the hierarchy of functional projections, but that the term $\theta_L P$ can be used to denote phrases, in a more traditional sense, that carry the feature θ_L .

¹⁷ Note that providing a ‘range’ is not enough.

¹⁸ It has been pointed out to me that (54) (and 56 below) are not so bad that they deserve to be starred. Perhaps two question marks would have been enough. In either case, the English examples are much worse than their Swedish counterparts.

- (55) Vad gav du Paul?
what gave you Paul [my example]

There is one obvious difference between the two questions (54) and (55) in that Swedish, being a V2 language, instead of making use of *do*-insertion, has the main verb in second position. To show that this difference does not seem to be relevant here, I provide you with (56) and (57), which should be structurally identical. Still, the English (56) is bad, while the Swedish (57) is fine:

- (56) *What have you given Paul? [my example]
(57) Vad har du givit Paul?
what have you given Paul [my example]

I do not know how to account for this apparent inconsistency in Starke's system. It would seem that further research is required.

In section 3.1 above no explanation of the difference in acceptability between (19) and (20) (repeated here as 58 and 59) was provided:

- (58) What is it unclear whether we should repair?
(59) What is it unclear how we should repair?

The extraction in (59) is considered much more degraded than the one in (58), while still being grammatical in the appropriate context. This difference in grammaticality disappears with infinitival complements, another observation not accounted for in 2.1 (the relevant example sentences are repeated as 60 and 61):

- (60) What is it unclear whether to repair?
(61) What is it unclear how to repair?

The assumption above on the homogeneity of θ -movement once again proves to be very useful. Since the intervener in (59) is an uncased adverbial, its first step is θ -movement, not θ_L -movement. This means that the extracted *what* cannot do θ_L -movement, because of the RRC. Neither can it do θ -movement, Starke points out, since θ -movement is clause-bound. This means that the only way to extract is to do SQ-movement, with strong requirements on the context. This is why (59) is unacceptable in an out-of-the-blue context. The reason this difference disappears with infinitival complements is, according to Starke, that infinitivals allow θ -movement out of them.

Concluding what has been said on the issue of locality in chapters 2 to 6, Starke states that he thinks that we have reason to believe that we only need one principle of locality, namely RM. Having such a unified notion of locality, Starke feels prepared to go on to ask deeper questions about how locality (or, rather, the operations 'move' or 'chain') interacts with and relates to "other tools in the syntactic toolbox". This is what will be discussed in sections 4 and 5 of this paper.

4 How the syntactic operations Merge and Move are unified [Starke's ch. 8]

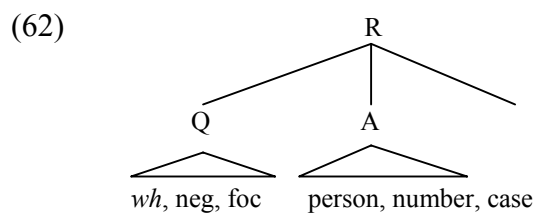
Given what we now know about the locality of 'movement' or 'chains', the time has come to unify Merge and Move. Since Move requires Merge, but not vice versa, Starke intends to show that what has been called Move is actually a kind of Merge. Traditionally, Merge and

Move (in their various guises) are considered to be two different, fundamental, syntactic processes. Starke focuses instead on their similarities, the most important of which being that both processes create an abstract unit by grouping together two units. In the case of Merge, the two units grouped together differ from each other, but in the case of Move/Chain, the two units are identical. The next step is to derive this apparent difference between Merge and Move from the ways in which the elements they are applied on differ from each other.

One reason why viewing Merge and Move as separate operations has traditionally been considered unproblematic is that it has been generally claimed that the most important independent property of Move is locality, while Merge has no such property, and that the most important property of Merge is labelling, or projection, while Move has no interesting labelling properties. Starke argues that this is a misunderstanding, i.e. Move/Chain has labelling, and the locality of Merge is a special case of the locality of Move. Eventually, this will lead to a unification of Merge and Move in which Move dissolves into Merge.

4.1 Locality and 'Labels' [Starke's section 8.1]

We always merge sisters in a tree. The reason, Starke says, is, once again, RM. RM is normally thought of as a constraint telling us that it is impossible to establish a relation between two elements belonging to the class X, if another element from the class X intervenes. When we merge two elements, it would make sense to say that they always belong to different classes in the syntactic FT, but this is actually not true, since both belong to the root class R (see 62). Logically, then, Merge-operations thus create R-chains, and any intervening member of R (i.e. any syntactic feature) would give rise to a RM violation, which forces Merge to take place between sisters and nothing else.



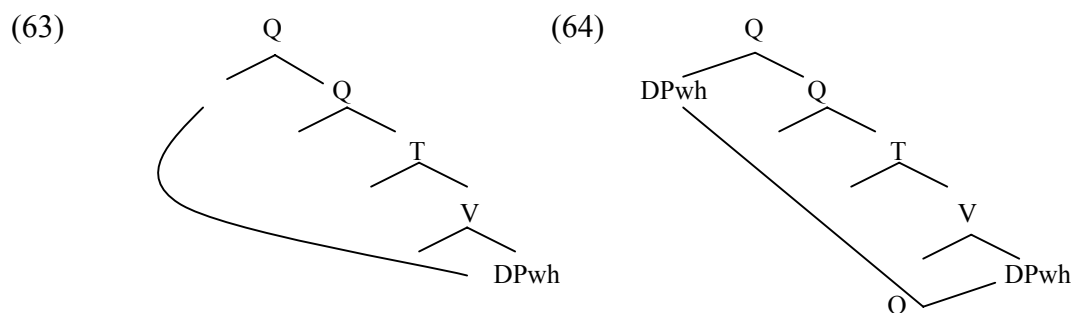
Given this, RM is the only tool we need to explain the locality of both Move and Merge, and the fact that one requires sisterhood, while the other does not, now follows from the fact that they operate on different input.

According to Starke, 'Labels' express three facts: (i) non-terminal nodes are of various types, (ii) the type of a certain non-terminal node is inherited, and (iii) this inheritance is asymmetric. When we understand that labelling is nothing but a name for this collection of properties, it is easy to see, Starke claims, that Move has labelling as well: (i) the locality properties of a given chain are a function of the chain type, i.e. A-chains differ from A'-chains in terms of locality, (ii) the chain type is a function of the links by which it is built up, and (iii) the chain type is determined by the highest, not the lowest, link in the structure. This is taken to indicate that there is nothing in labelling itself that forces us to assume Merge and Move to be different operations. On the contrary, the fact that Merge, just like Move, has been shown to be a grouping operation subject to RM, whose type is asymmetrically determined from its input entities, suggests a unified account of the two.

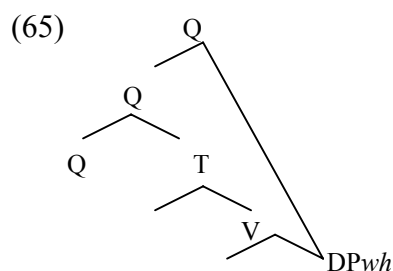
4.2 Move as 'Distant' Rmerge [Starke's section 8.2]

As mentioned above, unifying Merge and Move means getting rid of Move, since Merge is the core of syntax and Move requires Merge, while the reverse is not true. This brings us to a

situation in which Move is Merge ‘at a distance’. If we just mechanically substitute every instance of Move for instances of Merge, and let every step of Move/Chain inherit a label from its landing-site and quit graphically indicating the two relations using different types of connectors, we get (64) instead of (63):



According to Starke, the graph in (64) captures the idea that what we mistook for a Move operation is nothing but a Merge operation operating on two non-adjacent nodes¹⁹. Given this knowledge, the structure in (64) can be further simplified, so that we get (65)²⁰:

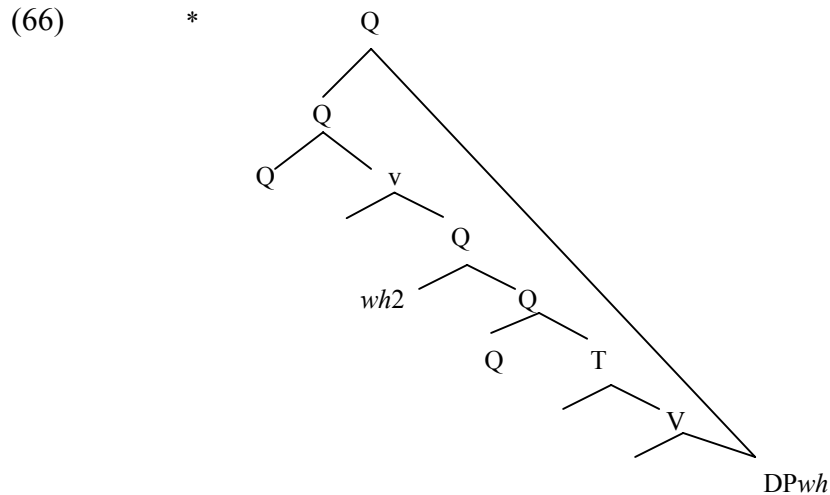


Reading off the graph in (65) enables us to observe that it captures the similarity between Merge and what has been called move, in that the relation between the DP and its long distance sister is the exact same relation as that between a head and what has traditionally been called a specifier. It is important to note that we are allowed to merge things on top of the highest Q in (65), i.e. the node corresponding to what has previously been called Move or Chain. This means that the node in question participates in phrase structure just like any other node. When Move and Merge are one, we no longer need traces or copies, since what used to be Move is now simply a ReMerger of a node already present in the structure. According to Starke, his graphs can easily be translated into traditional trees, and vice versa, which means that no old results are lost in the process.

Looking at locality again, Starke’s graphs express the same relations as did the old trees. For instance, the head-complement relation between Q and T in (65) cannot be a ‘long distance sister’ relation, since it is an R-grouping, and any intervening node would rule it out. The relation between DP_{wh} and Q in (65), the old specifier-head relation, can, however, be a ‘long distance sister’ relation, since it is a Q-Q-relation, only blocked by intervening Q-elements. This means that the RM-logic of eWI developed above is kept intact, and that the intervening spec-QP blocks extraction in (66):

¹⁹ Note that we can just as well draw the line showing the merger of DP_{wh} to the right of the rest of the graph, as in (65).

²⁰ For further details, I refer the reader to Starke (2001: 140-141).



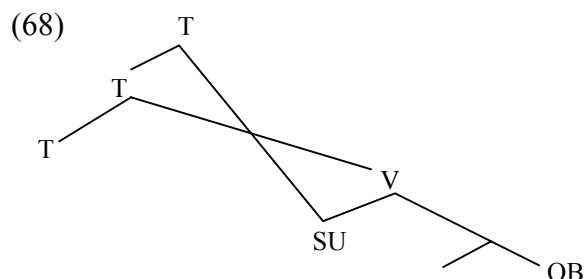
To achieve these results, Starke has not changed much concerning the logic of RM. When formalizing the notion of locality, there is a need to exchange C-command for Dominance, however, since the locality restrictions hold between the highest node in the tree/graph and the in-situ element in question, not between a specifier and its trace. This is a welcome consequence, Starke points out, since Dominance is more basic than C-command. Starke's tentative formalization of the locality principle (to be modified in section 5 to express the asymmetric nature of Merge) is given in (67):

- (67) X-merge (α, β) iff
 $\alpha, \beta \in X$, α dominates β , and $\neg \exists \gamma, \gamma \in X$, α dominates γ and γ closer to α than β is²¹.

4.3 Representations, Derivations and Imperfections [Starke's section 8.3]

Since Move has been shown to be a type of Merge, this means that syntactic theory need not contain much more than Merge. The project of motivating, or accounting for, the imperfection Move has now become obsolete, which, in a sense, means that the core of minimalism, i.e. the attempt to justify imperfections, is perhaps not the right strategy.

In Starke's system an ordinary declarative clause, in which the subject (SU) is merged in VP and remerged in TP, has the simple structure given in (68):



This unification of Merge and Move gives rise to some apparent problems: If there are no traces or copies, and syntagms are simply remerged from their base position, how do the interfaces know where to spell out the phrase? Obviously, this question concerning which

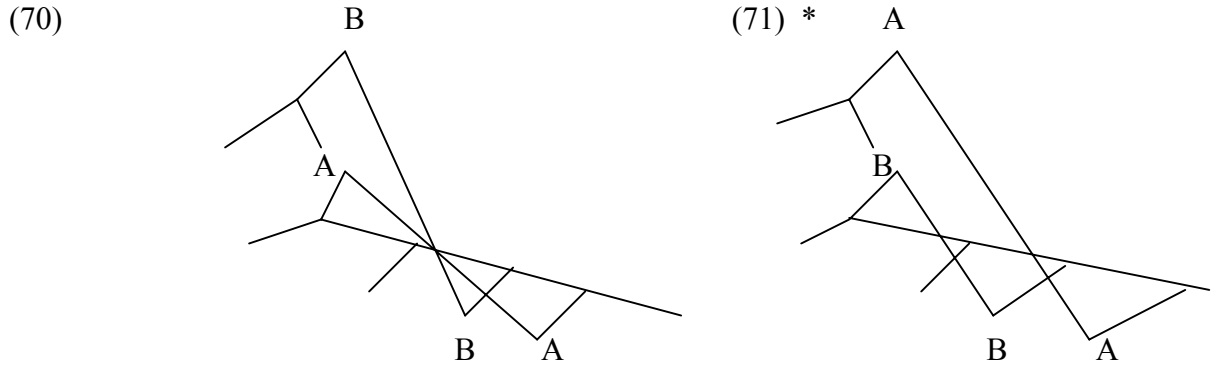
²¹ X-merge (α, β) means 'merge α and β and call the result X'.

merger to spell out is the same problem as the old one concerning which copy to spell out. The same holds for cases of so-called ‘split-reconstruction’, which, instead of a matter concerning multiple traces, becomes a matter concerning multiple mergers. Similarly, so-called reconstruction paradigms are now analysed as spell-out of multiple merger operations, instead of as spell-out of multiple copies. All in all, Starke does not solve these problems, but his theory does not give rise to further problems of that same kind.

For lack of space, I refrain from discussing Starke’s views on representations and derivations in detail. Starke’s main conclusion, after having examined the RRC once more, is that the locality restrictions on Merge apply after the last step of the derivation, which means that our syntactic model must be representational. Consider the RRC again:

$$(69) \quad * \alpha_j \alpha_i \dots \alpha_i \alpha_j$$

A felicitous phrase structure, i.e. one not violating (69) is given in (70) while (71) violates (69) in that the B-chain is contained within the A-chain²²:



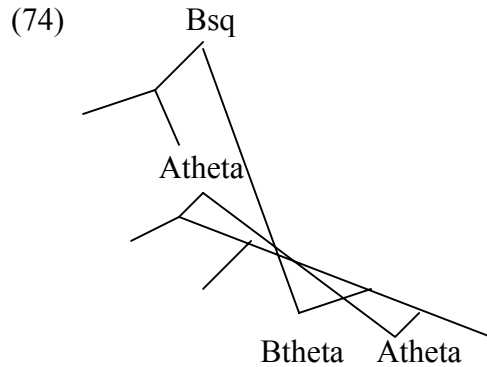
It seems as if the locality principle generating the result that (70) is fine and that (71) is bad can be stated either in terms of domination/inclusion, or in terms of C-command. In terms of domination, the situation in (71) is such that A contains B, since the remerger of A dominates all mergers of B, while this is not the case in (70). In terms of C-command, the lower B in (70) is no intervener, since it C-commands the higher A. Starke wants to formalize the locality principle in terms of inclusion, so he gives the following formalization of “dominates all mergers” above:

$$(72) \quad \alpha \text{ X-includes } \beta \text{ iff } \alpha \text{ dominates all X-mergers of } \beta$$

He then formulates the locality principle in terms of inclusion, as in (73), and his reason for preferring inclusion to C-command is that inclusion, but not C-command, is a relative notion. In other words, a definition based on C-command would not be able to separate the two kinds of B-remergers in (74), while the definition in (73), based on inclusion, can, since A includes all theta-occurrences of B. The definition in (73) thus correctly tells us that (71) and (74) are ungrammatical, while (70) is not:

$$(73) \quad \begin{array}{l} \text{X-merge } (\alpha, \beta) \text{ iff} \\ \alpha, \beta \in X, \text{ and} \\ \neg \exists \gamma, \gamma \in X, \alpha \text{ X-includes } \gamma \text{ and } \gamma \text{ C-commands } \beta \end{array}$$

²² The classical RM-violation is, according to Starke, a subset of (70).



We now see that the reason why Starke's approach requires a representational approach is that in, for instance, (70) it is not until B has remerged that the remerger of A is legal, i.e. the principle of locality does not apply until after the very last step of the derivation. According to Starke, this suggests that the locality component of Merge is a Chomskyan bare output constraint, leaving syntax to be a trivial and unrestricted concatenation engine²³.

5 The Redundancy of Specifiers, the Nature of the Functional Sequence, and the formalising of Merge and Locality [Starkes's ch. 9]

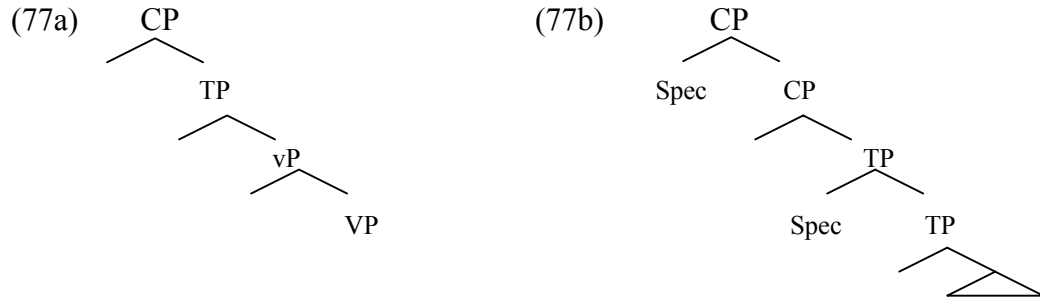
When the theory of syntax develops, Starke points out, certain theoretical tools and concepts are no longer needed and consequently disappear, while others are still being used. This keeping of old tools and concepts is not a problem, when the new version of the theory builds naturally on the old one. In other cases, however, tools and concepts seem to be retained for no good reason at all. According to Starke, X-bar theory has no place in modern syntactic theory.

Starke claims that all theories of functional projection have the following cornerstones:

- (75) There is a concatenation engine (Merge)
 (76) There is a functional sequence (FS) such that what (75) produces must respect the FS.

In (75) we have the core of syntax, and (76) expresses the implicit consensus that CP is above TP and not vice versa. On the basis of the cornerstones, we would expect the structure in (77a), but present-day theories instead postulate structures in which each instance in the FS is entitled to a second merge, i.e. a specifier (77b):

²³ It has been pointed out to me that when reading off Starke's graphs, it is not clear how they are supposed to be linearized (note that I am not referring to the fact that it is not clear which merger to spell out, c.f. section 4.3 above). The way I understand Starke, the potential linearization problem, for instance whether or not to pronounce B before A in (70) in a situation where the higher merger of B is to be spelt out, is no problem, given the discussion in section 5.3 below, since the FS tells us whether A is a legal continuation of B or vice versa.



The principle added to (75) and (76), giving rise to (77b) instead of (77a) is the assumption in (78) which is taken to be a reflex of the deeper tacit assumption in (79):

- (78) Every XP within the FS (i.e. every specifier or adjunct) must be locally associated with a corresponding head.
- (79) Asymmetric Projection (AP): a feature f in an XP node cannot legitimate its mother (i.e. it cannot project), but the same f in an X^0 node can legitimate a maximal projection (i.e. it can project).

The obvious question is why we have to add (78) and (79). In other words, what do we gain by postulating a null head accompanying each specifier in the FS? Starke's expected answer is that we gain nothing. On the contrary, adding (78) and (79) to our cornerstones gives rise to a number of inconveniencies, as discussed below.

5.1 Doing away with specifiers [Starke's section 9.1]

Within present-day theory, Starke claims, (78) and (79) are only used to trigger Merge or 'Move', since every insertion of an XP creates a spec-head configuration, making it trivial to equate the cause of the insertion with the need to create that configuration, which is what current theories do. If there should be another trigger for insertion, (78) and (79) would thus be obsolete. There is such a trigger, namely (76). Quite simply, having done away with (78) and (79), each feature is represented only once, as in (77), and therefore every insertion will alter the FS. Since (76) regulates possible FS²⁴, (76) will regulate what can (and must) be inserted, triggering Merge and ReMerge. To give an example: given the FS $CP > TP > vP > VP$, the subject has to remerge and project T (or an expletive has to be merged) to avoid the illegitimate FS $CP > vP > VP$.

When (78) and (79) have been shown to be obsolete, this means that X-bar is obsolete, and, consequently, that the difference between heads and specifiers has become obsolete. The only structural configuration we need is the head-complement relation, and the 'head' of a projection can be either a terminal node (80), or a non-terminal, i.e. a phrase (81):



The configurations in (80) and (81) are identical, in that they turn a β -terminated sequence into a α -terminated sequence. Using natural language examples, (80) illustrates an

²⁴ Given that we know what the (universal) FS looks like.

interrogative CP in which the head is, for instance, *if*, and (81) illustrates an interrogative CP in which the head is a *wh*-phrase, the important point being that a DP that is [+Q] can project a QP directly, since we have made obsolete the assumption that XPs cannot project.

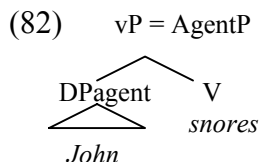
5.2 The downsides of Spec-Head

Before elaborating on his FS-based approach to Merge and Rmerge, Starke mentions a number of downsides with the spec-head/X-bar approach. One obvious problem is that spec-head agreement, if at all existent, does not seem to extend beyond ϕ -features. In other words, it could be the case, Starke says, that the idea of the importance of spec-head rests on the idea that all features behave as if they were ϕ -features, which is not case. To the contrary, Starke argues, not even “obvious” cases of spec-head agreement such as the English CP, French negation, and Russian yes/no questions need be actual examples of spec-head agreement. Given that no clear cases in which spec-head agreement concerning features other than ϕ -features occur, Starke argues, it is a welcome fact that our cornerstones in themselves make X-bar theory obsolete, i.e. we do not need the notion of agreement to trigger insertion.

Getting rid of X-bar has the following welcome consequences, according to Starke: (i) Since we no longer need to postulate empty heads accompanying every XP, an approach to clause structure such as that advocated by Cinque (1999) becomes much more attractive, even though postulating empty heads is not necessarily bad in itself, (ii) We no longer need to duplicate every feature, i.e. a feature present in an XP need no longer be borne by a head too, (iii) We no longer need to think of phrase structure as split into two layers, one in which X-bar units are created, and one in which these units are assembled into a FS, and, finally (iv) when we get rid of X-bar, we automatically get rid of the problem that even though the relation between the specifier and the head looks the same, given that locality regulates identity, the two relations differ from each other in terms of locality (the spec-head relation being radically more stringent than the relation between, say, a *wh*-phrase and its trace).

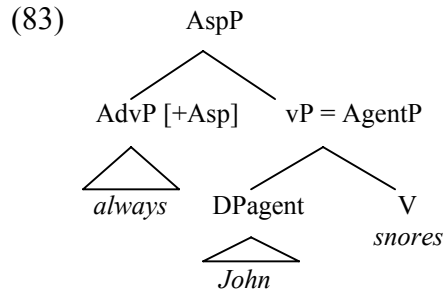
5.3 Formalising Merge and Locality [Starke's section 9.2]

The argumentation summarized in the preceding sections results in a situation in which the checking-relation (or the spec-head relation) has been made redundant, that is, in more general terms, we have dispensed with specifiers. In 5.1 above, it was shown, on a theoretical level, how the need to comply with the FS drives Merge and Rmerge. Deriving a simple declarative main clause such as “John always snores” can now be described as follows: Given the FS $\langle \text{Agr}, \text{Asp}, \text{v/Agent}, \text{V} \rangle$ ²⁵ and that the DP has already been formed, the first step is to group together the verb and the DP, so that the DP projects its Agent feature, in order to get the FS $\langle \text{v/Agent}, \text{V} \rangle$ (82):

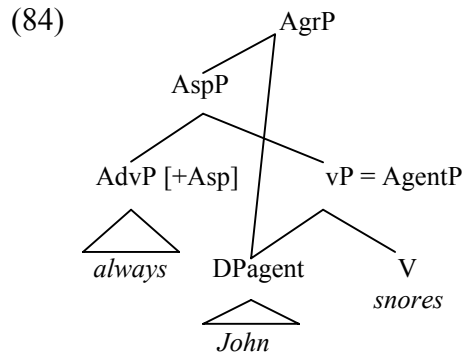


²⁵ Which phrases belong in the FS in which order remains an empirical question.

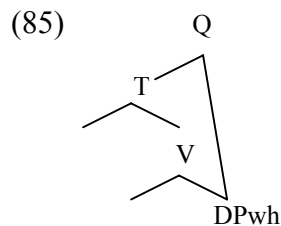
In order to adhere to the FS we merge the aspectual adverb *always* in the next step (83):



The final step is to add Agr (or T, I, or EPP), and the only way to do that is to remerge the subject and project its Agr features (84):



This might seem fairly simple and straightforward, Starke says, but the reformulation has important consequences for locality; if a given feature only occurs once, locality cannot be a constraint on identity. In other words we never merge (or remerge) when identity obtains, which means that instead of (65) above, we get (85), i.e. instead of the grouping at a distance *Merge* (*Q*, *Q*) we get the grouping at a distance *Merge* (*Q*, *T*):



According to Starke, the obvious generalisation in terms of locality is that when merging T with Q, only the closest Q is available, as formalised in (86):

- (86) X-merge (α , β) iff
 $\beta \in X$, and
 $\neg \exists \gamma, \gamma \in X, \alpha$ includes γ and γ c-commands β .

Merging α and β is now asymmetric, in that the FS decides which one of them is allowed to project, i.e. either α is a legal continuation of β , or vice versa, but both options are not available. In other words the locality principle does not regulate the identity between the landing site and the starting point of a given remerge operation. What it regulates is which node is allowed to remerge. The question of the direction of the asymmetry is shown by Starke, through a discussion of successive cyclic merger, to be such that a remerger of Q

above T is blocked by Q-interveners, not by T-interveners, i.e. the projecting element determines the nature of the intervener, not vice versa, which means that (86), not its reverse, can be concluded to be the final formulation of Starke's locality principle, perhaps, Starke says, the only locality principle necessary, given the view of functional projections and the FS advocated here. However, it still remains an empirical question to what extent (86) and the FS actually cover all relevant cases in all languages.

5.4 The Nature of the FS [Starke's section 9.3]

Given the role of the FS in the present system, we need to know not only the identity of the features within the FS, as do all other P&P approaches, but also to what extent the FS allows variation. There are two old questions to be addressed: (i) Is there more than one FS? (ii) Given a certain FS, need all its features always be present, or are subsets also legal? Referring to Starke (1995), Rizzi (1997) and Cinque (1999), Starke points out that there seems to be consensus that there is one and only one universal functional sequence²⁶.

The second question appears to be more problematic as well as more controversial. According to Starke, there are three different answers that are commonly given, namely (i) the 'rigid' approach: everything must always be present (e.g. Starke 1995, Cinque 1999), (ii) the 'peeling' approach: projections can be missing, but only by peeling off from the top (e.g. Rizzi 1994, Platzack 1996, Cardinaletti & Starke 1999), and (iii) the 'laissez-faire' approach: any projection can be missing (e.g. Wexler 1994, Thráinsson 1996). According to Starke, what we need is the 'laissez-faire' approach, so that projections can be omitted if no element in the numeration demands them, in order (i) to explain why there is no NegP/PolP in positive statements and (ii) to be able to allow adverb positions that are not put to use. It thus seems to be the case that we could replace (76) with (87):

- (87) There is one FS – a unique sequence of functional projections – such that the output of (75) must be a subset of FS.

This formulation is inadequate, however, since it (i) allows too many unattested combinations, and (ii) expresses the generalization that any projection can be omitted (which is not the case). What the empirical facts suggest is instead that there is some sort of regularity behind what can, and cannot, be omitted. Starke reaches the generalisation that only unmarked features can be omitted from the FS. What this generalization is intended to capture is the observation that the features [+neg], [+wh], and [+foc] have to be present in the structure to get a semantic interpretation, while their unmarked counterparts [-neg], [-wh], and [-foc] need not (c.f. the approach in Cinque 1999). In other words an FS-element can be left out iff it is recoverable at LF. This actually entails that at the level of interpretation the FS is always complete. What can, and cannot, be left out from the FS follows from the principles (88) and (89):

- (88) Semantic representations must respect FS, i.e. there is a FS – a unique sequence of functional projections – such that the output of (75) must be LF-interpretable as an instance of FS²⁷.
- (89) The interpretative component's reading of syntactic representations is based on recoverability, so that absence from syntactic structure corresponds to unmarked values.

²⁶ Obviously, not all linguists would agree with Starke here.

²⁷ Note that the formulation in (88) means that the FS-requirement is not a constraint of narrow syntax.

What these principles boil down to is the conclusion that the ‘rigid’ approach is correct as regards the interpretative component, while the ‘laissez faire’ approach is correct as regards syntax, as an artefact, Starke says, of how interpretation takes place.

Another interesting argument in favour of the ‘laissez-faire’ approach is that only the marked value of a given feature seems to give rise to remerge, i.e. there is no language in which the unmarked values of [*wh*], [*neg*], and [*foc*] give rise to remerge. If (i) features with marked values need be present, while those with unmarked values do not, and (ii) the FS is only expanded in order to comply with (88), unmarked values of features will never give rise to remerge, since the FS is only expanded when necessary. Marked values, on the other hand, correspond to projections that cannot be omitted, causing expansion of the FS²⁸.

Given this view of FS, the stipulations at the beginning of section 3.5 above, i.e. that within the same RRC, features do not mix, can be translated into FS terms as the observation that the FS $\langle \dots \alpha \dots \beta \dots \gamma \dots \rangle$ can also be instantiated as $\langle \dots \alpha \dots \beta \beta \beta \gamma \dots \rangle$. All instances of β are here interpreted as one single instance, and to function as one single unit of interpretation, they have to have the same value.

5.5 Apparent Problems [Starke’s section 9.4]

Starke discusses some apparent problems for his approach, the most important of which are mentioned here. According to the logic developed in Starke’s thesis, the head-complement relation (including the long distance sister relation) is the only syntactically relevant relation, meaning that syntactic structures are nothing but layers of head-complement relations in which each layer is independent of the other. Given this idea, it is unexpected that we should find agreement relations with ϕ -features and subject-verb inversion in root *wh*-questions (where the insertion of the inverted verb depends on the insertion of the *wh*-phrase).

Regarding the agreement with ϕ -features, Starke not only wants to be able to express the fact that the morphology of the verb co-varies with the morphology of the subject, but also why such agreement only ever happens with ϕ -features. A similar case is simple binding situations, involving co-variation regarding ϕ -features only (“He washes himself”, vs. “She washes herself”). One way of unifying the two phenomena would be to say that the morphology of the verb enters a binding relation with its argument, so that the subject and the verb are in different projections, without a geometrical relation like spec-head being involved (c.f. Rizzi 1982, where it is argued that verb agreement is pronominal, which in Starke’s terms would mean that the pronominal nature of Agr allows it to enter binding relationships, making spec-head agreement redundant).

Regarding inversion, Starke’s view is that the FS-requirement gives us ‘dependent insertion’, i.e. a situation in which the insertion of a higher node is dependent on the insertion of a node immediately below it. As already seen above, unmarked features can be omitted, which means that dependent insertion (e.g. English type inversion) will only occur with marked features. Within Starke’s FS-system, contrary to the standard spec-head system, the insertion of one YP can be dependent on the insertion of a number of elements, and the elements that need be inserted can be either phrases or heads, and, Starke argues, his system seems empirically correct.

Another apparent problem (actually a problem for all approaches) is the EPP, which states that there is one projection in the FS that can never be left out. Starke suggests that the analysis of ϕ -agreement in terms of binding might explain the EPP. If verb agreement enters

²⁸ This logic does not apply to the first merge, Starke points out, since every element in the numeration need enter the structure, regardless of marked and unmarked features.

a binding relation with the argument, this agreement is either anaphoric or pronominal. Hypothesizing that it is anaphoric, it follows that a local antecedent needs to be present. If we assume that languages generally have anaphoric agreement (often null) on the verb, the EPP follows, Starke claims. A further related idea is that languages with poor verbal inflection are anaphoric-inflection languages, in which a subject is needed to bind the inflection. Languages with rich inflection, on the other hand, are pronominal and may thus escape the EPP requirement through the very morphology of verbal agreement.

One potentially serious objection to Starke's reasoning concerns the fact that he seems to run the risk of having to double all the features in the FS, since if, in traditional terms, both the head and the specifier are filled, Starke has to say that they, in fact, project two different features. The underlying question is whether or not the doubly filled comp filter extends to each and every projection in the FS. If it extends to all projections, Starke's approach is rescued. To be able to address this question, Starke distinguishes two types of terminal nodes, particles and verbs, which seem to have different properties, in that (i) particles never move, but verbs appear to move, and (ii) the insertion point of particles has a semantic correlate, wherever the meaning of the particle can be independently determined, while the position of verbs has lost all correlation to meaning or form. Starke calls particles Rigid Heads and verbs Soft Heads. Rigid Heads seem to behave as Starke's approach predicts, so the potential doubling problem only arises with Soft Heads. The question concerning the doubling of features is, according to Starke, a side effect of an independent problem, namely the question why verb movement in a system such as that advocated in Cinque (1999) is, to a very large extent, optional and without correlation with semantics and morphology.

To begin with, Starke argues, not all verb movement is optional; V2 movement and subject-aux inversion in question behave just like Rigid Heads, which means that only the first "default" step of verb movement is actually optional.

Moreover, the behaviour of Agreement, Negation, and verbs is similar. As Cinque (1999) has convincingly argued, it is possible to generate a NegP above every adverb related functional projection, and the same situation, Cinque argues, seems to be true of "DP-related" projections, i.e. AgrP, as well. Given the fact that verbal forms canonically contain an agreement morpheme and that this morpheme is typically the outermost morpheme of the verb, Baker's (1988) Mirror Principle entails that the verb will be displaced into the position of Agr. The random insertion of Agr will thus appear to be the random insertion of V. In terms of the FS, the conclusion must be that the Agr head has a liberal positioning within FS – each language seems to make an arbitrary choice – and the verb ends up in that apparently arbitrary position. We thus need no doubling of the FS at all. All apparent problems discussed by Starke have thus been shown not to be severe threats to his approach, even though much further work needs to be done in several areas only touched upon here, before the final judgement is made.

6 Summary and Critical Remarks

Throughout his thesis, as summarized above, Starke argues quite convincingly for the benefits of a syntactic model in which there is no need for two of the three types of fundamental relationships between entities commonly recognized in syntactic theory, namely the specifier-head relationship and the movement/chain relationship. We are left with (i) the complement-head relation, i.e. the core of the grammar organ, underlying the human capacity for recursive compositionality, (ii) the operation Merge, which is argued to incorporate what has formerly been known as Move, and (iii) the universal Functional Sequence, (implicitly) recognized in standard generative syntax, but seldom made full use of.

To be able to exclude specifiers, and thus X-bar, from the theory, Starke argues successfully (i) that not only heads, but also phrases (specifiers) can project, and (ii) that the operation Merge never merges elements with identical features, and (iii) that the doubly filled comp filter extends to all projections. To be able to unify Merge and Move, so that Move dissolves into Merge, Starke (i) carries out an extensive investigation of the locality of “movement”, showing that all locality principles can be reduced to a slightly modified version of Rizzi’s (1990) Relativized Minimality, operating on hierarchically structured trees of syntactic features, (ii) shows that the locality of Merge is the same as the locality of “Move”, (iii) shows that the labelling properties traditionally associated with Merge are also present in the old “Move” operation, and (iv) shows how Merge can operate at a distance, giving rise to long distance sister relationships.

In the model proposed by Starke, we need no checking, no specifiers, no movement, no traces, no copies and no chains. We only keep the central syntactic operation Merge, which gives rise to sister relations, i.e. groupings of two non-related entities into one compositional unit. Starke’s final comment in the thesis is that “it seems possible to express all the results of current theories by simply iterating head-complement mergers, without indulging into the rich apparatus standardly thought to be necessary to achieve those results” (Starke 2001: 177).

In the title of the paper, I promise the reader critical comments on Starke’s work. I have hitherto mainly been presenting a summary of Starke’s main ideas and arguments. As mentioned in the introduction, and as has become obvious, I am fascinated and impressed by Starke’s work, the logic in his arguments, and the straightforward way in which he presents his thesis. Unlike most theses I have previously read, the present one is a thesis in the proper sense of the word, in that every part of it follows logically from what has been said in the previous one, forming one comprehensive, controversial, and bold argument that appears to build on everything that has previously been claimed within modern syntactic theory at the same time as it turns a large part of it upside down by drawing some central parts of syntax to and beyond their logical consequences, thereby showing the redundancy of some and the true nature of others.

On the negative side I would like to list the following: (i) Even though keeping a thesis short is generally a good virtue, Starke’s thesis would have benefited from further elaboration on certain points. As it stands, too much previous knowledge is taken for granted, (ii) Even though Starke’s use of informal language enhances readability, I sometimes get the impression that his informal language deludes the reader into thinking that his arguments are more obvious and less controversial than is actually the case, (iii) The redefinition of concepts such as ‘theta-role’, ‘specificity’, and ‘cased’, and the introduction of terms like ‘outcast’ tend to confuse the reader, (iv) The thesis unfortunately contains a few spelling mistakes and some spelling/notational inconsistencies, as well as a number of sentences/sequences which could have been formulated differently, all of which is probably due to the time pressure under which the thesis was finalized (Starke p.c.), (v) In my opinion, the introduction provides too precise and sudden information about the overall findings of the thesis, and too little background information concerning the problems to be discussed, and (vi) Even though such areas are mentioned throughout the thesis, the concluding chapter would have benefited from containing a discussion of unsolved matters and suggestions for future research.

Finally, even though Starke’s two main suggestions regarding modern syntactic theory, i.e. that he (i) reduces Move to Merge²⁹, and (ii) argues for the redundancy of specifiers, could,

²⁹ It must be pointed out that Starke’s unification of Merge and Move is very reminiscent of Chomsky’s (2001) use of Internal and External Merge. I remain agnostic as to whether Starke has inspired Chomsky, or vice versa,

and have been (wrongly, especially in the case of Move and Merge), argued to be mere cosmetic and terminological modifications driven by the search for theoretical elegance, no one should be able to refute Starke's unification of locality using such arguments. Whether or not Starke's idea that we only need Relative Minimality and a structured feature tree is on the right track, his proposal is well argued, empirically motivated, and well worth taking seriously in future research.

References

- Baker, M. 1988. *Incorporation. A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- Belletti, A. (ed.). To appear. *Structures and Beyond. The Cartography of Syntactic Structures*, vol.3. Oxford: Oxford UP.
- Cardinaletti, A. & M. Starke. 1999. The Typology of Structural Deficiency: On the Three Grammatical Classes. In *Clitics in the Languages of Europe*, ed. H. Van Riemsdijk, 145-290. Mouton de Gruyter.
- Chomsky, N. 1986. *Barriers*. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2001. Beyond Explanatory Adequacy. *MIT Occasional Papers in Linguistics* 20. Cambridge, Mass.: MIT.
- Cinque, G. 1990. *Types of \bar{A} -Dependencies*. Cambridge, Mass.: MIT Press.
- Cinque, G. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford UP.
- Enç, M. 1991. The Semantics of Specificity. *Linguistic Inquiry* 22: 1-25.
- Haegeman, L. 1993. Object Clitics in West Flemish. *GenGenP* 1: 1-30.
- Huang, C.-T.J. 1982. Move *wh* in a Language without *wh* Movement. *The Linguistic Review* 1:369-416.
- Lasnik, H & M. Saito. 1984. On the Nature of Proper Government. *Linguistic Inquiry* 15.2: 235-289.
- Manzini, R. 1992. *Locality*. Cambridge, Mass.: MIT Press.
- Manzini, R. 1997. Syntactic Dependencies and their Properties: Weak Islands. In *Studies on Universal Grammar and Typological Variation*, ed. A. Alexiadou & A.T. Hall. Amsterdam & Philadelphia: John Benjamins.
- Platzack, C. 1996. The Initial Hypothesis of Syntax. A Minimalist Perspective on Language Acquisition and Attrition. In *Generative Perspectives on Language Acquisition*, ed. H. Clahsen. Amsterdam: John Benjamins.
- Richards, N. 1998. The Principle of Minimal Compliance. *Linguistic Inquiry* 29.4: 599-629.
- Rizzi, L. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, L. 1988. On the status of referential indices. Unpublished manuscript, University of Geneva.
- Rizzi, L. 1990. *Relativized Minimality*. Cambridge, Mass.: MIT Press.
- Rizzi, L. 1994. Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives. *Language Acquisition* 3: 371-393.
- Rizzi, L. 1997. The Fine Structure of the Left Periphery. In *Elements of Grammar: A Handbook of Generative Syntax*, ed. L. Haegeman, 281-337. Kluwer Academic Publishers.

noting that ideas of this kind have been around for some years, and that their exact origin might be difficult to pinpoint.

- Rizzi, L. 2000. Reconstruction, Weak Island Sensitivity, and Agreement. Unpublished manuscript, University of Geneva.
- Rizzi, L. 2001. Relativized Minimality Effects. In *A Handbook of Syntactic Theory*, ed. M. Baltin & C. Collins, 89-110. Oxford: Blackwell.
- Starke, M. 1995. On the Format for Small Clauses. In *Small Clauses*, ed. A. Cardinaletti & T. Guasti, 237-269. New York: Academic Press.
- Starke, M. 2001. *Move Dissolves into Merge: A Theory of Locality*. Unpublished doctoral dissertation.
- Szabolcsi, A. & F. Zwarts. 1997. Weak Islands and an Algebraic Semantics for Taking Scope. In *Ways of scope taking*, ed. A. Szabolcsi, 217-262. Kluwer Academic Publishers.
- Thráinsson, H. 1996. On the (Non-)Universality of Functional Categories. In *Minimal Ideas*, ed. W. Abraham, S.D. Epstein, H. Thráinsson & C.J-W. Zwart. Amsterdam & Philadelphia: John Benjamins.
- Watanabe, A. 1992. Subjacency and S-Structure Movement of *wh*-in-situ. *Journal of East Asian Linguistics* 1.3: 255-291.
- Wexler, K. 1994. Optional Infinitives, Head Movement, and Economy of Derivation. In *Verb Movement*, ed. N. Hornstein & D. Lightfoot, 305-350. Cambridge: Cambridge UP.