When OT Meets LLF: Multiple Wh-questions and Optimality Theory

Huba Bartos

0 Introduction

In this paper I propose a radical revision of the application of Optimality Theory (henceforth: OT) to syntax, as put forth by Grimshaw (1993). Her model will be shown to suffer from a number of conceptual and empirical problems which are demonstrably not resolvable within her framework. Therefore I suggest that a different conception of what serves as the input to the two fundamental modules of OT is necessary. Instead of taking the GEN module to operate on lexical projections and produce a set of possible extended projections (s-structures) corresponding to it, I adopt the representation referred to as Lexico-Logical Form by Brody (1994), and the EVAL module chooses the most harmonic one among the alternative LLF-representations. In arguing for these modifications I will analyse English multiple wh-questions within matrix clauses, as well as similar examples from other language types, in order to show that, in the spirit of some of the definitive properties of OT, this model, but not Grimshaw's one, is capable of accounting for cross-linguistic differences.

1 OT in Syntax — up to now

1.1 OT in general

OT, as a complete theory of grammar and languages, was originally defined by Prince & Smolensky (1993) (P&S), but this book focused mainly on issues and data in phonology, and many parts of the theory, as implemented there, were specific to that branch, including the formulation of certain constraints and definitions, as well as the notion of what constitutes legitimate input and output.

The central concept of OT is *harmony*, or *optimality*, *i.e.* the degree to which a certain linguistic form satisfies an ordered set of well-formedness constraints. The competing linguistic forms arise as different analyses belonging to a given input string (or structure), as produced by the *generator*

(GEN) function (or module). "GEN contains information about the representational primitives and their universally irrevocable relations [...]", as P&S put it (1993 : 4), and generates from some (lexical?) input a set of *candidates*, in compliance with certain entirely universal, and maximally general principles. The other function/module, (H)-EVAL, ranks the candidates with respect to their degree of harmony, as in the above sense, and marks the candidate which is highest ranked as the optimal, hence the only grammatical, output, to be paired with the input of GEN. GEN has certain governing principles, two of which have core importance:

(i) *Freedom of analysis*: Any amount of structure can be posited.

(ii) *Containment*: No element may be literally removed from the input form.

These amount to saying that GEN can randomly add any structural element, but the input must be a substructure of each and every GEN-output.

The EVAL module evaluates the harmony of the candidates by checking them against an array of constraints that are in a "strict dominance hierarchy". The constraints are stated as universal and general requirements, which may be in conflict with each other, and need not all be satisfied by even the most harmonic candidate; what matters is that the candidates strive to satisfy the highest possible constraints, and for any one of them to win out, it has to do no worse on a given constraint than any of the others do, unless for each of the other ones there exists a higher-ranking constraint such that the other one does worse on it. To put it in a simpler and more mechanical fashion: checking of the candidates begins with the highest ranking constraint, and any candidate that satisfies it to a lesser extent than some other one will be eliminated, and in checking further constraints it will not be relevant any more. This cycle is repeated for the constraints one by one, until there remains only one candidate that survives, and that one will be declared the most harmonic.

In this system, the sole source of cross-linguistic variation is the possibly different ranking of the elements of the same set of constraints, that is, the constraints themselves do not vary from language to language, but they have different priority in different languages. Some constraints seem to have no effect whatsoever in a certain language, which is then the result of their being ranked so low that they will fail to ever be operative, since the higher-ranking ones will already be decisive, except, maybe, in a few marginal cases.

Another important feature of OT is the fact that there is no derivation in any sense at all. The input-output pairs are the merely the result of

functions, and the constraints cannot refer to derivational notions, either. OT, therefore, is a purely *representational* model.[1]

1.2 Syntactic applications

So far, only two papers dealing with the relation between syntax and OT have gained publicity,[2] and even these two view this relation quite differently. The earlier of the two is Legendre *et al.* (1993), primarily interested in typological issues hoped to be explained in an OT framework. Its aim is to derive the known Case and voice systems in the world's languages, by rankings of principles governing the mapping from semantic roles to morphosyntactic categories. What Legendre *et al.* (1993 : 2) take to be the input of OT is "simply a clause or a predicate/argument complex", where "each argument [...] is labeled with its thematic role". GEN, then, assigns values for abstract Case to the arguments. Clearly, the authors accept that the syntactic strings are formed prior to the application of OT, but say nothing about these affairs. While in phonological implementations of OT the input material could be claimed to be formed prephonologically, *i.e.* in syntax, or directly as an item in the lexicon, the same is not viable in syntax. Consequently, we have to conclude that Legendre *et al.* do not look upon their OT as *the* syntactic module of grammar; it is at most a link in a chain which may or may not be built up of a whole sequence of independent OT-applications, each working on the output of a previous one, or treating the same input in a parallel fashion. Crucially, OT obviously does not (in fact cannot) replace some syntactic module, but aids it by taking over particular tasks.

Grimshaw's paper pictures OT as one entire syntax-machine, which does the job of outputting a grammatical sentence in 'one fell swoop'. Accordingly, GEN has to be more wide-scale than in the Legendre *et al.* conception: instead of just specifying a category-value it has to transform the sentential strings in all conceivable ways. The input fed into GEN is a lexical projection, to which GEN assigns all the extended projections (in the sense of Grimshaw (1991)) 'derivable' from it that have non-distinct LF representations. EVAL then filters out the optimal one. If we were to phrase all this sorcery in 'traditional' Principles & Parameters terminology, the input to OT could be seen as something like a d-structure; GEN would comprise the effects of X′-theory (at least as regards the functional projections, since the lexical-projection part is assembled on X′-theoretic grounds before being fed into OT), and Affect-α. EVAL, on the other hand, applies representational filtering conditions, to sift out the ungrammatical extended projections. The

output of OT therefore is the optimal one of all the s-structures stemming from a common d-structure and converging in a (mostly) unique LF.

Several problems arise immediately from such a picture. Firstly, as with Legendre *et al.*'s model, nothing is known about the birth of what serves as input to OT, namely the lexical projections, although, as opposed to the other model, here it might be argued that such assembly can take place within the lexicon, on the ground that lexical projections contain all and only the information specified for the lexical entries in the lexicon. (However, such intralexical manipulation presupposes the presyntactic 'knowledge' of X′-theory.) Secondly, though decisive reference is made to LF, it is unclear what (if any) status LF representation has in this theory. It is obviously not created as a representation in any phase. Thirdly, it is not very appealing to use both LF and d-structure (or lexical projection), especially in the light of recent proposals to treat d-structure as a mere abstraction on LF (*cf.* Brody 1993, 1994). Consider the case where a d-structure uniquely corresponds to an LF. Now one of them should be eliminated from an optimal theory of grammar as redundant. Since d-structure is a (proper) substructure of LF, the former need not be retained, only the latter. On the other hand, suppose that there is no unambiguous mapping from d-structure to LF. Now Grimshaw's theory is at a loss, because the candidate set is defined assuming the unicity of LF to the set of competing s-structures, otherwise we would not know which s-structures can meaningfully compete.

2 Grimshaw and the wh-questions

2.1 Simple wh-questions

Let us now consider Grimshaw's analysis of simple English wh-questions in her OT framework. Her basic claim is that to account for all major cases of subject-auxiliary inversion in English, positing the following five constraints is enough:

PROJ-P – 'selected complements must be the same at d-structure and s-structure'
OP-SPEC – 'syntactic operators must be in specifier position'
OB-HD – 'heads are obligatorily filled at s-structure'
*LEX-FUNC – 'no lexical heads in functional projections'
MIN-PROJ – 'a functional projection must be functionally interpreted'

PROJ-P is a generalized version of Chomsky's (1981) Projection Principle, to the effect that head movement to selected positions is prohibited. Since in this paper we mainly discuss matrix wh-questions, and this principle is mostly used for distinguishing between matrix and embedded questions, I will simply omit PROJ-P unless explicitly needed.

OP-SPEC draws on Rizzi's (1991) analysis of wh-phenomena, in terms of his Wh-Criterion. The OP-SPEC constraint is its Grimshavian reincarnation, with its conceptual problem carried over: a syntactic operator is defined as "a scope-bearing expression which takes its scope by virtue of its syntactic (as opposed to LF) position" (Grimshaw 1993 : 1). Firstly, this definition, and the OP-SPEC constraint rely on each other in a circular way, as was observed by Brody (1994) for Rizzi's formulation of the corresponding notions, already. Secondly, it is dubious to treat 'syntactic' (s-structure) and 'semantic' (LF) operators as distinct, and attribute different behavior to them. The concept of operators has significance only with respect to interpretation, *i.e.* semantics and LF (*cf.* Brody 1993, Haegeman 1994).

OB-HD requires that head slots be filled at s-structure, either by a lexical element, or by a trace, or ϕ-features. Its primary function is to trigger I-to-C movement, in traditional terms. *LEX-FUNC is needed to force *do*-support for questions in the absence of any other auxiliary. Finally, MIN-PROJ is used to block adding functional projections above minimum necessity. As regards the ordering of these constraints, both OB-HD and *LEX-FUNC have to outrank MIN-PROJ, in order to get the correct results in *do*-support, but they are unordered with respect to each other. Consider now a simple object wh-question, as in (1):

(1) *What did you see?*

The lexical projection underlying (1) is [$_{IP}$ you PAST [$_{VP}$ see what]], or possibly [$_{IP}$ you I^0 [$_{VP}$ saw what]]. The alternative extended projections, and their performance against the constraints is shown in (2):

(2) **Simple matrix object wh-questions**

	OP-SPEC	OB-HD	*LEX-FUNC	MIN-PROJ
⇒ what did [you see *t*	√	√	√	⋆
what *e* [you saw *t*	√	⋆!	√	⋆
[you saw what	⋆!	√	√	√
did [you see what	⋆!	√	√	⋆
what saw [you *t* *t*	√	√	⋆!	⋆

(√ means 'pass', ⋆ means 'violate', ! means 'be eliminated',
⇒ marks the optimal candidate)

Obviously, for the correct result to arise, OP-SPEC has to be ranked higher than MIN-PROJ. Furthermore, embedded object wh-questions show that OP-SPEC must also outrank OB-HD. Consider the case of (3), as shown in (4).

(3) *I wonder what you saw.*

(4) **Embedded object wh-question [wrong]**

	OB-HD	OP-SPEC	...
I wonder [what *e* [you saw	⋆	√	
I wonder [you saw what	√	⋆	

Clearly, if OB-HD were higher than OP-SPEC, the correct form for (3) would die on OB-HD immediately, while the other s-structure form could proceed.[3]

A few remarks may be in order about (2). First, the candidates examined constitute only a subset of the entire candidate set; further candidates, however, would only be superstructures (projectional extensions) of the examined ones, hence they would be ruled out either by OB-HD, if they contained extra empty heads, or by MIN-PROJ. Secondly, two of the examined candidates are apparently acceptable sentential strings: *'you saw what'* could be a proper echo-question, but in that case the wh-word is not an operator, whereby both its d-structure and its LF is distinct from that of (1), and that of an intended non-echo reading, so the echo variant is not to compete within the same candidate set. Similarly, *'what saw you'* could be interpreted as a *subject* wh-question, with an entirely different interpretation, *i.e.* an entirely different d-structure and LF, so once again, the subject-wh reading will not compete here, while the object-wh reading is ungrammatical, as correctly predicted by Grimshaw's analysis.

Now consider simple matrix subject wh-questions, such as (5):

(5) *Who saw you?*

In the standard P&P framework, such wh-questions involve CP-projections with empty C^0 heads, and the wh-expression in SpecCP. Under this state of affairs, the grammatical structure $[_{CP}$ who$_x$ *e* $[_{IP}$ t_x $[_{VP}$ saw you]]] would violate OB-HD, by the unfilled C^0 head, whereas its *do*-support counterpart $[_{CP}$ who$_x$ did $[_{IP}$ t_x $[_{VP}$ see you]]] would not, thus an ungrammatical form would win out in EVAL over the grammatical one. However, as Grimshaw points out, if simple matrix subject wh-questions are just IPs, with the wh-expression in situ, they pass on each relevant constraint, whereas the *do*-support variant will turn out to be suboptimal, on the assumption that

do is a verb, heading its own projection, so it invariably introduces an extra phrasal projection of its own, in violation of MIN-PROJ.

2.2 Multiple wh-questions

Grimshaw demonstrates that her model is capable of explaining a wide range of phenomena in English, with the same small machinery. But she fails to attempt to examine how multiple wh-questions behave under her OT. Let us take a look now at such data, to see how well it fares with them. Consider the multiple wh-question in (6):

(6) *Who saw what?*

We expect that (a) the example being a subject wh-question in part, it need not (in fact, cannot) project to CP-level, and (b) the same constraint hierarchy (notably: OP-SPEC $\gg$ {OB-HD; *LEX-FUNC} $\gg$ MIN-PROJ) will yield the right results. The evaluation is shown in (7):

(7) **Multiple (subject-object) wh-questions**

	OP-SPEC	OB-HD	*LEX-FUNC	MIN-PROJ
who e [t saw what	√⋆	⋆		
who saw [t t what	√⋆	√	⋆	
[who saw what	√⋆	√	√	√
what e [who saw	√√	⋆		
what e [who e [t saw t	√√	⋆⋆		
who did [t see what	√⋆	√	√	⋆
what did [who see t	√√	√	√	⋆
what e [who did see t	√√	⋆		

There are now two wh-expressions per clause. In Grimshaw's original conception, only one of them is a syntactic operator, since only one is in operator position, but they are required by OP-SPEC to be in operator position if they are operators. This circularity was pointed out above. If it is to be resolved, we must give up the notion 'syntactic operator' (*cf.* above). But then two operators per clause have to be considered for OP-SPEC, and the chart reflects these facts.

What becomes problematic is that the expected winner does worse on OP-SPEC than those ones where both operators are in specifier positions. At the same time it is worth noting that if this problem could be overcome, the rest would fall out as expected. The problem, however, is that no matter

what repair strategy we follow, it will not leave us satisfied. Consider the possibilities:

a. *lower* OP-SPEC — Then it has to be lowered under MIN-PROJ, otherwise *'what did who see'* would still win out, but by doing so we lose the explanation for simple matrix object wh-questions, as it crucially hinges on the assumption that OP-SPEC is higher-ranked than MIN-PROJ.

b. *relax* OP-SPEC — This, in effect, amounts to going back to Grimshaw's original proposal, which distinguished syntactic, *i.e.* 'moving', and non-syntactic, *i.e.* 'in situ' operators; the conceptual problems will then return. Relaxing the constraint in any other way voids it of any meaningful content.

c. *kill the 'renegades'* — One could try to devise a constraint which eliminates the candidates that would otherwise outlive the correct one; this should then dominate OP-SPEC in the hierarchy. The unappealing property of such a solution stems from the fact that the problematic candidates are a heterogeneous group, so the new constraint might have a flavor of *ad hoc*ness.

Finally, mention must be made of the fact that there is one type of multiple wh-questions which is problematic not only for Grimshaw's model of OT, but for any implementation of OT. Consider (8):

(8) **Who left why? / *Why did who leave?*

No multiple wh-question whatsoever can be grammatically formed in English where the two wh-expressions are subject and adjunct. Standard P&P theory accounted for this fact by the ECP: both adjunct and subject wh-traces have to be antecedent-governed, but there is just one available SpecCP per clause for an antecedent-governor, so in cases like (8) one of the wh-traces at LF remains without being properly governed, no matter how things are arranged. In OT, once a lexical projection corresponding to (8) is taken as an input, GEN will generate a number of candidates, of which there will inevitably arise a most harmonic one, which then has to be marked as grammatical. So such cases have to be taken care of before OT, *i.e.* before syntax.[4] Notice also, that the equivalents of (8) are perfect in several other languages, including Chinese, or Hungarian. Therefore trying to constrain GEN to rule (8) out is not an option, GEN being universal in all respects. Unless we are willing to allow inviolable constraints in EVAL (constraints whose violation incurs a fatal mark on the candidates, regardless

whether there will be at least one of them that passes), the solution must be pushed back to whatever comes before OT—the lexicon, in all likelihood.

3 OT meets Lexico-Logical Form

3.1 Lexico-Logical Form

Let us adopt the notion of Lexico-Logical Form (LLF) from Brody (1994). His theory of grammar is a non-derivational, single-level one; LLF is the only level of representation, which is assembled directly from lexical material, and to which semantic interpretation has access to. It has all the properties of the LF of the traditional P&P framework, but contentive elements occupy their PF-appropriate positions, so while traditional overt movement is represented here as a chain whose contentive is in its head position, what is referred to as LF-movement in much of the literature corresponds to chains headed by scope-markers in LLF, with the contentives in such chains in a lower slot. The input of PF, *i.e.* s-structure, can simply be read off LLF, and what Brody calls the *d-set*, the set of chain root positions, where lexical, thematic requirements hold, and 'projection' takes place, is just an abstraction on (L)LF. Building LLF into OT can rid us immediately from the conceptual weakness pointed out above, that the simultaneous use of d-structure and LF leads to redundancy.

3.2 The model of OT proposed

I now take the input of OT to consist of the set of lexical items constituting the given sentence, which is much like what Chomsky (1994) terms as a 'numeration'. GEN builds structures from the input; this operation basically comprises chain formation and X′ projection, and conforms to some principle of *Respect* (Newson 1994), which ensures that the lexical requirements of the input are satisfied.[5]

The candidate set will then consist of the alternative LLF representations mapped from the input; their differences may affect the order and number of chain links, or the linear order of the contentive elements. The LLF that comes out as optimal can be fed into PF.

As follows from the way LLF is organized, 'moved' wh-expressions will be in scope-taking positions themselves, while for in-situ wh-expressions there will be empty scope markers (SMs) as links of the given wh-chain, in the position corresponding to the scopal properties of the wh-element. This shows already, what crucial importance is attributed to chains in this model.[6] Trying now to view the problem of multiple wh-questions in general,

let us attempt to devise a version of OT that will allow us to subsume other language types under the same system. This is in fact one of the primary aims of OT: different rankings of the same constraints should yield different languages, and this is taken to be the sole source of cross-linguistic variation within the grammar (and apart from the lexicon); and as Tesar & Smolensky (1993) note, the role of constraint-ranking is analogous to that of parameters in the P&P framework.

English can be a representative of the type of languages where one wh-operator has to move overtly to its scope-position, and further wh-expressions of the same clause, having the same scope, can remain in their 'd-structure' sites. In our LLF model this translates as the requirement that one wh-chain must be headed by its contentive element, and in the other wh-chains there need only be a SM in the chain-head. Chinese is a prototypical example for languages where no wh-operator moves overtly, that is, in our terms, all wh-chains can be headed by SMs. On the other hand, in languages like Hungarian, each wh-operator must move overtly into a functional projection,[7] so in LLF there can be no SM-headed wh-chains in such a language. Intuitively, then, in the Chinese type some principle to the same effect as Chomsky's (1992) *Procrastinate* has to prevail, while the Hungarian type is characterizable by an *Earliness* (Pesetsky 1987) or *Transparency* (Brody 1994) principle.[8] In the case of English, Procrastinate may be operative (thus blocking movement of a second wh-operator), but some higher-ranked constraint must provide for the overt movement of the first one. I propose the following set of constraints:

OP-SCOPE – 'operator-chains must have a link corresponding to their scope'
LIC-SM – 'scope markers must be licensed by a (PF-)adjacent wh-element'
PROCR – 'Procrastinate, *i.e.* contentive elements must be in chain roots'
FILL – 'no expletives or empty categories'

Grimshaw's OP-SPEC has been generalized to OP-SCOPE, and formulated to fit an LLF-type representation. (It renders adjunction to be an available option, *contra* Grimshaw.) LIC-SM will force at least one wh-chain per clause to have a contentive head; it draws on the insight of Brody (1994), who treats secondary wh-chains as parasitic on primary ones. PROCR is simply the chain-theoretic equivalent of Chomsky's (1992) derivational lateness principle with the same name; eventually, FILL corresponds to Grimshaw's MIN-PROJ, mostly, as regards its effects.

3.3 The case of English in LLF OT

Let us turn our attention now to linguistic data. Examine first English matrix multiple (subject–object) wh-questions, as in (6), repeated here as (9):

(9) *Who saw what?*

The results are shown in (10).

(10) **Multiple (subject-object) wh-questions**

	OP-SCOPE	LIC-SM	PROCR	FILL
[who saw what	⋆			
who [t saw what	⋆			
⇒ SM_x [who saw what_x	√	√	√	⋆
what [who saw t	√	√	⋆	
SM_y [who_y saw what	⋆			
SM_x SM_y [who_y saw what_x	√	?	√	⋆⋆
what who [t saw t	√	√	⋆⋆	
who did [t see what	⋆			
what did [who see t	⋆			
SM_x did [who see what_x	⋆			
SM_y did [who_y see what	⋆			
SM_x SM_y did [who_y see what_x	√	⋆		
what who did [t see t	√	√	⋆⋆	
⇒ SM_{xy} [who_y saw what_x	√	√	√	⋆
SM_{xy} did [who_y see what_x	√	⋆		

What we find is that under this dominance hierarchy there are two 'winners', so far as only these constraints are concerned. It is a theoretical question, which of the two we really want to come out as the optimal one. Favoring the first one is in keeping with Grimshaw's assumption that a wh-subject need not be associated with SpecCP, *i.e.* that simple matrix subject wh-questions can just be IPs.

It is clear from (10) that OP-SCOPE must be ranked above FILL. The ? mark in line 6 is due to the fact that we have not taken sides as to whether SMs count as appropriate licensing wh-elements for each other or not. Not much hinges upon it, however, since FILL turns the candidate in question suboptimal. Let us analyse simple wh-questions now, to demonstrate that LLF OT is able to reproduce the effects of Grimshaw's model:

(11) **Simple matrix subject wh-questions**

	OP-SCOPE	LIC-SM	PROCR	FILL
⇒ [who saw Joe	√	√	√	√
SM_x [who_x saw Joe	√	√	√	⋆
who [t saw Joe	√	√	⋆	
who did [t see Joe	√	√	⋆	
SM_x did [who_x see Joe	√	⋆		
saw [who t Joe	⋆			
SM_x saw [who_x t Joe	√	⋆		

The result here suggests in (10) the first 'winner' should be the truly optimal candidate (*e.g.* by blocking a category from having two or more distinct indices), as analogous to this case: the subject wh-expression is in its scope-appropriate position in situ, so there is no need to redundantly relate it to SpecCP. Consider now object wh-questions:

(12) **Simple matrix object wh-questions**

	OP-SCOPE	LIC-SM	PROCR	FILL
[Joe saw what	⋆			
what [Joe saw t	√	√	⋆	⋆
what did [Joe see t	√	√	⋆	⋆⋆
SM_x [Joe saw $what_x$	√	⋆		
SM_x did [Joe see $what_x$	√	⋆		
saw [Joe t what	⋆			
SM_x saw [Joe t $what_x$	√	⋆		
what saw [Joe t t	√	√	⋆⋆	

The analysis has run into a serious problem here. No winner has been marked since the candidate that OT would assign this title to is not the real grammatical one. What has happened is that the candidate without the supportive *do*, that is, the one that would fit embedded contexts, has done better on FILL. Remember that this difference between matrix and embedded questions was taken care of by PROJ-P in Grimshaw's theory, together with its being ranked above OB-HD. Since the closest equivalent of PROJ-P in the present system is the *respect* principle governing GEN's workings, it will invariably precede any EVAL principle — in fact, it can never be violated by any candidate, since such candidates simply cannot be generated. What remains for us to do is reintroduce OB-HD (which can work

properly only if adjunction is once again prohibited). Such a step necessitates reconsidering the previous cases, too, to see if any trouble is caused there. At first sight the only condition for keeping the correct results seems to be that OB-HD cannot outrank OP-SCOPE. If so, no significant change can arise with the multiple questions, since all the suboptimal candidates either fail OP-SCOPE, or have two (contentive or empty) operators in front of the IP-domain, but they will multiply violate OB-HD (and FILL) by this, unless the heads are *do*-supported; this latter type is either in violation of LIC-SM (if SMs are separated by *do*), or of PROCR (if overt wh-expressions are in the pre-IP specifiers). Therefore the consequence is that OB-HD must be ranked below LIC-SM and PROCR. Simple subject wh-questions are not affected, because there the winner will pass OB-HD anyway.

Consider, finally, multiple wh-questions where question is formed at direct and indirect object, as *e.g.* in (13):

(13) a. *What did you give to whom?*
b. *?Who did you give what to?*

The EVAL results are shown in (14).

(14) **Multiple matrix wh-questions (direct & indirect object)**

	OP-SCOPE	LIC-SM	PROCR	OB-HD
what did [you give t to whom	$\star$			
what [you gave t to whom	$\star$			
[you gave what to whom	$\star\star$			
whom did [you give what to t	$\star$			
whom [you gave what to t	$\star$			
what whom did [you give t to t	$\surd$	$\surd$	$\star\star$	
whom what did [you give t to t	$\surd$	$\surd$	$\star\star$	
whom what [you gave t to t	$\surd$	$\surd$	$\star\star$	
SM_x SM_y [you gave what_y to whom_x	$\surd$	$\star$		
SM_x SM_y did [you give what_y to whom_x	$\surd$	$\star$		
SM_x what did [you give t to whom_x	$\surd$	$\surd$	$\star$	$\star$
SM_y whom did [you give what_y to t	$\surd$	$\surd$	$\star$	$\star$
SM_{xy} did [you give what_y to whom_x	$\surd$	$\star$		
SM_x what [you gave t to whom_x	$\surd$	$\surd$	$\star$	$\star\star$
SM_y whom [you gave what_y to t	$\surd$	$\surd$	$\star$	$\star\star$

The situation is much the same as with simple object wh-questions: the matrix *vs.* embedded context contrast is resolved by OB-HD (and the prior effect of the Projection Principle). However, there appears an additional problem here: we still have two 'winners', which differ with respect to the

choice of the overtly 'moved' wh-expression. This problem falls under the 'traditional' label of *Superiority*, *i.e.* the condition that of two possible source categories movement targets the one that is 'higher' in the tree in some sense (that is, the one that asymmetrically c-commands the other). It is interesting to note that with multiple wh-questions involving the subject this effect was the result of PROCR not allowing the 'movement' of the object wh-phrase, since the subject did not have to move at all, thus offering a less costly 'derivation'. Here we face a different case: either the direct or the indirect object wh-phrase has to 'move', PROCR thus makes no difference. The issue is further complicated by the fact that double object constructions have long been a matter of debate: Are they sisters, or does either of them asymmetrically c-command the other? (see *e.g.* Larson (1988) for a discussion) If the grammaticality difference in (13), marked by *?*, really holds, it suggests that a direct object is superior to an indirect one at least when the latter is embedded in a *to*-PP. To account for this possible distinction, some further constraint needs to be introduced, but I will not pursue this question here.[9]

4 Typological issues

4.1 Chinese

In the remaining part of the paper I will briefly discuss how the model devised here is capable of treating wh-questions in languages with characteristically different wh-strategies than that of English. Take Chinese first. As is widely known, there is no overt movement of wh-phrases in this language to any functional position within the sentences. All wh-phrases (in fact, all operator phrases) are in the position corresponding to their grammatical function (or subcategorization): subject wh-phrases are in subject position, object wh-phrases are in object position, etc. In LLF terms, then, all operator chains are headed by SMs, and the contentives stay in chain-root position. Scope must be marked, though, by a chain link, since there are signs of an existing relationship between the scopal positions and the operator phrases, as has been shown by Huang (1982a, b). OP-SCOPE must therefore remain active (*i.e.* high enough in the hierarchy), but clearly, PROCR must be promoted to a higher position to take effect over primary wh-relations, too. This can be achieved by lowering LIC-SM under PROCR; once PROCR gains more importance, primary wh-chain contentives are exempt from the requirement that their empty SMs be licensed by a PF-adjacent overt wh-operator. The relative order of OP-SCOPE and PROCR is immaterial.

4.2 Hungarian

Hungarian constitutes a more complicated case. For one thing, wh-phrases move not to SpecCP, in terms of the P&P-type analysis, but to the Spec of what is usually referred to as *Focus Projection (FP)*, accompanied by movement of V to the head position of FP. For another, in multiple wh-questions all wh-phrases must move to the front of the V in F^0—but there are a few exceptional cases, as well, with some wh-phrases staying in situ.

As regards simple wh-questions, the only significant difference from English (leaving the problems related to topics preceding the wh-operators in Hungarian aside) is the lack of *do*-support; instead of inserting *do* in the head position of the projection hosting the wh-phrase, the verb itself turns up there. In our model, this option was excluded in English by PROCR: the V-chain {do, V} complies with it, whereas a chain like {V, t_V} does not. In Hungarian, no corresponding supportive elements exist in the lexicon, and this yields the solution already. There is an additional consideration, however: matrix and embedded wh-questions are not different in Hungarian, so nothing disallows ranking OB-HD above PROCR, to prevent the V from remaining within VP. Relevant examples are given in (15), and the EVAL results in (16).

(15) a. *Kit* *látott* *Mari?*
who-ACC see-3sg-PAST-indef_obj Mary-NOM
'Who did Mary see?'

b. *Ki* *látta* *Marit?*
who-NOM see-3sg-PAST-def_obj Mary-ACC
'Who saw Mary?'

(16) **Simple wh-questions in Hungarian**

	OP-SCOPE	LIC-SM	OB-HD	FILL, PROCR
⇒ Kit_x $látott_v$ Mari t_v t_x	√	√	√	
Kit Mari látott t	√	√	⋆	
Mari látott kit	⋆			
$Látott_v$ kit_x Mari t_v t_x	⋆			
SM_x Mari látott kit_x	√	⋆		
⇒ Ki látta Marit	√	√	√	
SM_x $látta_v$ ki_x t_v Marit	√	⋆		
SM_x $látta_v$ $Marit_y$ ki_x t_v t_y	√	⋆		
Ki $Marit_x$ látta t_x	√	√	⋆	

Consider now multiple questions. (17) is an appropriate example:

(17) *Ki* *kit* *látott?*
who-NOM who-ACC see-3sg-PAST-indef_obj
'Who saw whom?' $(\forall x ?y\ [\text{person}(x) \wedge \text{person}(y) \wedge \text{saw}(x, y)])$

At first sight, the situation is relatively clear: if there are more than one wh-chains in a clause, each must have its contentive phrase in front of the V (which is in F, arguably). For this, there are two viable analyses. Either SMs are not available in Hungarian (a condition implementable by a constraint straightforwardly disallowing them), or some *Earliness* principle is operative (as mentioned above), which conflicts with, but outranks, PROCR. It would confine contentive elements to chain-head positions. This latter version is illustrated in (18):

(18) **Multiple wh-questions in Hungarian**

	OP-SCOPE	LIC-SM	EARLY	OB-HD	...
⇒ Ki kit_x látott t_x	√	√	√	⋆	
Ki látott kit	⋆				
SM_x SM_y látott_v ki_x t_v kit_y	√	⋆			
SM_x ki látott kit_x	√	√	⋆	√	
SM_x ki kit_x látott t_x	√	√	⋆	⋆	
?⇒ kit_x ki látott t_x	√	√	√	⋆	

Apparently, there are two winners here, again. Their case could probably handled by superiority, whatever the way to build it into our system. But is it really the correct solution? If we reconsider (17), more notably: its logical interpretation, it turns out that the two winners will not be attributed the same LF. The first of the two wh-phrases in (17) translates as a universal quantifier in Hungarian, as pointed out and explicated by É. Kiss (1993). But if the two winners do not share their LFs, then they are not meant to compete, presumably because the chain of the first *ki*-phrase is just not a [+wh]-chain, but a quantifier-chain. Thus {*ki*-[+Q]; *kit*-[+wh]; *látott*} and {*ki*-[+wh]; *kit*-[+Q]; *látott*} are two distinct numerations, or OT-inputs. Arguments in support of this approach can come from the fact that non-referential wh-phrases (like *hogyan* 'how', *miért* 'why') can never precede the true question-phrases. Incidentally, the quantifying pseudo-wh-expressions occupy the proper quantifier position, which is after topics and before focus in Hungarian s-structures.

There is, however, another type of wh-questions, illustrated in (19):

(19) *Ki* *látott* *kit?*
who-NOM see-3sg-PAST-indef_obj who-ACC
'Who saw whom?' $(?x ?y\ [\text{person}(x) \wedge \text{person}(y) \wedge \text{saw}(x, y)])$

This is the true multiple question in Hungarian, with both wh-expressions bearing focal stress, though they are far apart. Actually, these questions allow the insertion of *és* 'and' before the second wh-phrase, in effect splitting the clause into two, with the second conjunct as an elliptical clause whose remnant is just its question focus:

(20) *Ki látott (és) kit?*

It may be the case then that even in (19) we may posit conjoined clauses, each of which conforms to the EVAL constraints. Alternatively, the second wh-phrase may be assumed to be in a second focus projection in a right-dislocated position (É. Kiss 1993). Under this second assumption, chains can be taken to behave in the way required by the constraints. Thus this problem with Hungarian multiple wh-questions proves to be resolvable.

5 Conclusion

The above discussion was meant to show that Grimshaw's picture of OT applied to syntax cannot overcome certain conceptual problems, which give rise to empirical problems as well. However wide the range of data accounted for by her model is, its extension to multiple wh-questions fails. Therefore I proposed to alter the structure of the model, combining it with Brody's LLF theory, as well as to modify some of its particulars, while retaining those parts that fit into the LLF–OT model. English simple and multiple wh-questions have been shown to fall out of the system, except for one specific type for which neither Grimshaw's mode, nor the present one is able to account for. Finally, the model has been extended to languages with different wh-question forming strategies, to demonstrate that, in the spirit of the basic tenets of OT, cross-linguistic 'parametrization' can be reproduced with LLF–OT.

NOTES

[1] Practice, at least so far, has not perfectly complied with this, inasmuch as many concrete analyses have tried to encode what used to be seen as derivational history or information into the representations valid in OT. If this turns out to be inevitable, a fundamental tenet of OT will be challenged, and its spirit seriously marred.

[2] This statement is based on the contents of a bibliography file of works in OT, which is available by Internet ftp from `ruccs.rutgers.edu`, in the directory /pub/OT/TEXTS. The file dates from Sep. 16, 1994. It cites only these two works in syntax as papers riper (and more accessible) than a simple manuscript.

[3] Irrelevantly to our discussion, PROJ-P has to outrank OB-HD as well, to exclude **I wonder what did you see.* PROJ-P need not be ordered with respect to OP-SPEC.

[4] This problem does not yield itself to remedy in the model advocated by the present paper, either.

[5] Newson (1994) is hesitant whether *respect* is an output condition, *i.e.* one of the constraints of EVAL, or a constraint on GEN. I opt for the latter view, even though GEN is normally seen as unconstrained, since under the former option *respect* would just be one of the ranked constraints, and there could exist some language that ranked it too low to have effect — which is not in keeping with the widely held view that this part of the theory is universal, with no known exceptions at all.

[6] Note that since both Brody's LLF theory and OT are non-derivational, they must make use of the notion of chains to capture regularities covered by the name 'Move-α' in standard P&P theory.

[7] Apparent counterexamples will be treated below.

[8] Brody (1994) argues against Procrastinate on theoretical grounds, and rules it out as a possible principle of grammar, thereby relying entirely on Earliness. For his theory, two such conflicting principles would be mutually exclusive, but in OT constraint-conflict is allowable, or even desirable.

[9] If, on the other hand, it turned out that (13a) and (13b) are equally acceptable, we would have a case of double winners, *i.e.* one input results in two distinct outputs, both of which are marked by EVAL as grammatical. Then we must conclude that *ties* are a possibility in OT.

REFERENCES

Brody, M. (1993) 'Theta theory and arguments'. *Linguistic Inquiry* **24.** 1–13.

Brody, M. (1994) 'Lexico-Logical Form—A Radically Minimalist Theory'. Ms., to appear.

Chomsky, N. (1981) *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, N. (1992) 'A Minimalist program for Linguistic Theory'. *MIT Occasional Working Papars in Linguistics.*

Chomsky, N. (1994) 'Bare phrase structure'. Ms., MIT.

É. Kiss, K. (1993) 'Wh-movement and specificity'. *Natural Language and Linguistic Theory* **11.** 85–120.

Grimshaw, J. (1991) 'Extended Projection'. Ms., Brandeis University.

Grimshaw, J. (1993) 'Minimal Projection, Heads & Optimality'. *Technical Report #4 of the Rutgers Center for Cognitive Science.* Rutgers University.

Haegeman, L. (1994) *The Syntax of Negation.* Cambridge: CUP.

Huang, C.-T. J. (1982a) 'Move–WH in a language without WH-movement'. *The Linguistic Review* **1:4.** 369–416.

Huang, C.-T. J. (1982b) 'Logical relations in Chinese and the theory of grammar'. PhD diss., MIT.

Larson, R. K. (1988) 'On the double object construction'. *Linguistic Inquiry* **19.** 335–391.

Legendre, G., W. Raymond and P. Smolensky (1993) 'An Optimality-Theoretic typology of case and grammatical voice systems'. *Proceedings of the 19th Meeting of the BLS.*

Newson, M. (1994) 'Negation, double negation and Optimality Theory'. Paper presented at a BuPhoC session, 25 May 1994, at Eötvös Loránd University, Budapest.

Pesetsky, D. (1987) 'WH-in-situ: movement and unselective binding'. In: Reuland, E. & A. ter Meulen, eds. *The Representation of (In)definiteness.* Cambridge, MA: The MIT Press. 98–129.

Prince, A. and P. Smolensky (1993) 'Optimality Theory: constraint interaction in generative grammar'. *Technical Report #2 of the Rutgers Center for Cognitive Science.* Rutgers University.

Rizzi, L. (1991) 'Residual Verb Second and the Wh-Criterion'. *Technical Reports in Formal and Computational Linguistics 2.* Faculté des Letters, Université de Genève.

Tesar, B. and P. Smolensky (1993) 'The learnability of Optimality Theory: an algorithm and some basic complexity results'. Ms., Univ. of Colorado.