To appear in *Papers in Optimality Theory*, Beckman, Urbanczyk, & Walsh, eds.
University of Massachusetts Occassional Papers, 1995

Optimality and *Wh*-Extraction

Géraldine Legendre,[1,2] Colin Wilson,[1,2] Paul Smolensky,[1,2]
Kristin Homer,[2] and William Raymond[2]

[1]Johns Hopkins University & [2]University of Colorado

I. Issues

The study of *wh*-question formation has historically served as the empirical basis for major constructs in Government-Binding (GB) such as the Empty Category Principle (ECP), the existence of Logical Form (LF) as a separate level of representation—motivated in part by the abstract *wh*-movement at LF analysis of *wh*-in-situ in languages like Chinese (Huang, 1982)—and the central but controversial issue of which principles apply at which levels of representation. For example, Huang (1982) argues, based on Chinese, that the ECP applies at S-structure and LF while subjacency and his Condition on Extraction Domain (CED) apply only at S-structure.

Cross-linguistic investigations have revealed that these ideas are actually hard to formalize in a simple and unified fashion and a brief survey reveals a problematic state of affairs. (1) Standard GB offers no unified treatment of *wh*-fronting—languages which observe movement constraints are analyzed as involving movement; languages which do not observe movement constraints are not analyzed as involving movement (e.g Palauan *wh*-fronting involves base generation; Georgopoulos 1985, 1991). (2) No unified treatment of *wh*-in-situ is offered either: languages like Chinese and Japanese which observe (at least some) movement constraints have most recently been analyzed as involving overt movement of a null operator coindexed with an in-situ *wh*-variable at S-structure (Aoun and Li, 1993, Cole and Hermon (1994)); Malay and Ancash Quechua, which do not observe movement constraints, are not analyzed as involving movement, but rather interpretation in situ (Cole and Hermon, 1994). (3) *Wh*-in-situ languages offer contradictory evidence about the level at which the subjacency principle applies: S-structure in Chinese (Huang, 1982) and Japanese (Lasnik and Saito, 1992), LF in Hindi (Srivastav, 1991). (4) Languages with several *wh*-strategies show different degrees of constraining each strategy: for example, in Ancash Quechua, overt movement is more constrained than LF movement (Cole and Hermon, 1994) while the reverse is true of Iraqi Arabic (Wahba, 1991). While parameterization offers ways of resolving the contradictions in (3) and (4), this raises serious issues for a theory of parameters that typically seeks to limit parameters to the lexicon or to functional categories. (5) Last but not least, we observe a proliferation of principles bearing the same name, the ECP, some of which are clearly distinct in content—Chomsky (1986) seeks to reduce the ECP to antecedent-government, Cinque (1990) reduces it to head-government—while others are formally distinct: conjunctive (Rizzi, 1990; Aoun et al 1987) or disjunctive (Chomsky, 1981, 1986; Manzini, 1992; Lasnik and Saito, 1992).

In this paper, we propose to start addressing the issues outlined above from the perspective of Optimality Theory (OT, Prince and Smolensky, 1993) and test the hypothesis that different patterns of extractability across languages result from different rankings of universal constraints realizing a few simple principles. The scope of this paper does not permit us to address all these issues (which are the topic of our on-going research); here we can only set up the basic framework and explore a few of its consequences. In particular we limit ourselves here to languages which observe movement constraints, whether they make use of *wh*-fronting (English, Bulgarian) or in-situ-*wh* (Chinese). The framework we develop is a version of GB which incorporates the two fundamental constructs of OT, *soft* constraints and *ranking* of soft constraints. In content, some of our constraints are reminiscent of similar constraints in the GB literature; they are, however, formally quite different because of their intrinsic violability and because of the way they interact: they can be violated in well-formed structures, and the force of a given constraint is greater in some languages than in others.[1]

Economy plays a well-known role in *wh*-question formation, ranging from the ultimate economical strategy—in-situ—to successive-cyclic movement, which some languages register morphologically (e.g. Irish, Chamorro). Instantiating the minimalist intuition that shorter movements are better than longer ones (Chomsky, 1992), OT allows a precise formalization of economy in the MINLINK constraints defined below.

The paper proceeds as follows: in Section II, we discuss three *wh*-strategies and illustrate the basics of OT by showing how the distinct *wh*-strategies of English, English, and Bulgarian arise from different rankings of a set of three constraints. In Section III, we turn to the main focus of the paper, an OT treatment of what traditionally falls under the ECP and Subjacency. Because of space considerations, we limit our illustration to *wh*-islands and extraction out of the complement of *think*-type verbs.

---

[1]Our account is comparatively simple, by at least two measures of simplicity: a) The constraints are simple: for example, ECP is reduced to head-government; barriers are defined in terms of L-marking b) We don't have: antecedent government, minimality barriers (rigid or relativized), a distinction between inherent barrierhood vs. barrierhood by inheritance, gamma-marking, adjunction to VP in order to void the barrierhood of VP, etc. The violability of constraints renders these mechanisms unnecessary.

II. An OT analysis of *wh*-strategies

Chinese, Bulgarian, and English exemplify the three distinct *wh*-strategies we focus on in this paper: Chinese leaves *wh*-phrases in situ; Bulgarian obligatorily fronts all available *wh*-phrases, despite having relatively free word order elsewhere; English obligatorily fronts one *wh*-phrase, and only one. The relevant facts are illustrated in (4).

(1) a. Lisi *zenmeyang* chuli zhe-jian shi? (Tsai, 1994)
Lisi how handle this matter
"How did Lisi handle this matter?"
(= what is the means or manner x such that Lisi handled this matter by x)

b. *Koj kakvo na kogo* e da t t ? (Rudin, 1988)
who what to whom has given
"Who gave what to whom?"

c. Who gave what to whom?

Our unified approach starts with a representation of scope in terms of *chains*, not movement. Following May (1985) and others, we retain the basic idea of modelling the representation of *wh*-interpretation on quantifier-variable binding, but modify the standard representation in the following way. A *wh*-chain contains an operator as its head (in highest specifier position of a clause), and a variable as its foot; the latter marks the D-stucture position of a questioned element, the former the scope of its interpretation.[2] The *wh*-fronting strategy places the overt *wh*-operator at the head of the chain, with an empty trace t at the foot of the chain; the *wh*-in-situ strategy places an empty operator Q at the head of the chain and the overt *wh*-phrase at its foot. What matters is a) the relationship between Q and the variable it binds, namely the *chain* (Q,x), b) only one of Q or x can be overt for "free" (resumptive pronouns violate a FILL constraint, as discussed in section III.2.4), and c) any empty variables (and intermediary traces) obey constraints on traces. (Following Aoun and Li (1993), we assume that overt question markers such as optional *ne* in Chinese are generated in C position in the presence of a Q operator via the mechanism of Spec-Head agreement.)

In OT, a grammar is a function which maps inputs to outputs. Inputs consist of raw materials from which the candidate outputs are built: skeletal structures containing predicate-argument structure and scope information. These materials are those employed to express the basic meaning (including discourse properties) and provide the basic building blocks (NP, V, P, Complementizer, etc.) and single clausal brackets. For every input, *Gen* generates the set of *candidate outputs* or *parses* with scope relations (roughly a combination of S-Structure and LF). *Gen* generates all

[2]This proposal goes back to Baker (1970) and has been revived in recent work by a number of linguists. Our proposal is close to that of Aoun and Li (1993) in several respects with a crucial difference: they posit overt movement of a null Q operator in Chinese at S-structure. We have a single level of optimization which subsumes D-structure, S-structure and LF, and we have no movement.

relevant brackets in accordance with standard X′-theory. *Gen* marks as overt Q or x (or both, in which case a resumptive pronoun realizes x). Each candidate output for an input contains that input ('containment'). *Gen* also generates unfaithful candidates which fail to parse some element of the input (e.g. the [wh] feature or the scope of Q, as discussed below.) *Gen* is also responsible for placing Q in the highest Spec position of a clause.[3] In OT, the optimal form is grammatical: roughly, it is the form violating fewer and lower-ranked constraints than any of its competitors. The set of candidates entering the competition is universal.

Consider the candidate set for questioning, say, a direct object out of a simple clause: the universal input is schematically shown in (2) and (3) shows a subset of the corresponding universal candidate set to be evaluated by the constraints.

(2) Universal input for questioning a direct object out of a simple clause:
$[Q_j\ [\ldots t_j \ldots]]$

(3) Universal candidate set for (2):

a. $[Q_j\ [\ldots wh_j \ldots]]$	faithful parse
b. $[wh_j\ [\ldots t_j \ldots]]$	faithful parse
c. $\langle Q_j \rangle\ [\ldots DP/\langle wh_j \rangle \ldots]$	unfaithful parse

(3a) represents *wh*-in-situ: it wins in Chinese; (3b) represents fronting of a *wh*-phrase to a scope position—the winning strategy in English and Bulgarian. In candidate output (3c), the [wh] feature is not parsed, unfaithful to the semantics of the input. We admit unfaithful parses in our candidate set because in some instances, there is no grammatical way of expressing some input. A language for which candidate (3c) is optimal is a language in which direct objects cannot be directly extracted out of simple clauses, as is the case in Bahasia Indonesia (Saddy, 1991), Kwakwala (Anderson, 1984), Malagasy (Keenan, 1976), Tagalog (Guilfoyle et al, 1992), and other languages. The structure corresponding to a failure to parse [wh] in (3c) is not a question but rather a statement with a [–wh] DP in lieu of a [+wh] DP (See further discussion of failure to parse [wh] in section III.2.1).

We now turn to constraints and rankings. Chinese, as shown in (1a), obviously has a relatively strong constraint against movement of *wh*-phrases which we state as maximally general: *t or 'no traces' (This is the constraint STAY of Grimshaw's 1993, in press OT analysis; cf. Chomsky's 1991 Economy of Derivation).[4]

---

[3]In principle, *Gen* also generates all word orders, with candidates evaluated against word order constraints that are ranked along with other constraints. In the absence of an OT theory of word order, we artificially limit the candidate set in this paper to candidates observing the surface word order of the language.

[4]Because all constraints are violable, this is not equivalent to saying that Chinese never permits movement of any category. In fact, Chinese has productive topicalizations, involving chains whose head is in an adjoined position and a t in D-structure position. See discussion below in section III.2.4.

We suppose that Bulgarian (as shown in (1b))shows the effects of a constraint which has the opposite effect, i.e. forcing *wh*-phrases to front: *Q: 'no empty Q-operators'. English seems to have both constraints at work simultaneously: a constraint against leaving *wh*-phrases in situ that prevails for one element and one element only and a constraint against fronting that prevails for all remaining *wh*-phrases, as shown in (1c). OT resolves such conflicts by claiming that the three languages obey the same constraints, but that the constraints are ranked differently: in Chinese, *t ranks higher than *Q. The situation is exactly reversed in Bulgarian, with *Q ranked higher than *t.[5] That only one *wh*-phrase moves in English results from a third constraint that ranks higher than *t in Bulgarian but lowest in English. Drawing on a standard feature of the analysis of English multiple *wh*-questions with the same scope (Higginbotham and May, 1981), we propose that this constraint is one against absorption of Qs, *ABSORB. In English, *ABSORB ranks low with the result that English allows two operators to convert into one operator marking the scope of two variables. In Bulgarian, *ABSORB ranks relatively high; in effect, Bulgarian does not allow two Qs to be combined.

The constraints are summarized in (4) and the language-particular rankings are given in the standard OT tableau format in (5).

(4) Constraints

*t	"No traces"
*Q	"No empty Q-operators"
*ABSORB	"No absorption of Q-operators"

(5) a. Chinese multiple *wh*

[Q_i Q_j [x_i V x_j]]	*ABSORB	*t	*Q
a. ☞ [Q_i+Q_j [wh_i V wh_j]]			⊛ ⊛
b. [wh_i+wh_j [t_i V t_j]]		*! *	
c. [$wh_{i[j]}$ [t_i V wh_j]]	*	*!	

Tableau (5a) represents Chinese.[6] The input (with scope information) is given above the candidate parses (here it is an instance of multiple *wh*). In the tableaux throughout this paper, **we adopt the convention** of identifying a subject by subscript *i*, a direct object by subscript *j*, a referential adjunct by subscript *k*, and a non-referential adjunct by subscript *l*. Candidates *a, b, c* incur marks or stars '*' for the

[5]The effect of *Q is very close to Grimshaw's (1993, in press) OPSPEC 'Operators must be in Spec position.' Our representations are quite different, however; *Q is violated by a Q—an empty operator—which heads a *wh*-chain in our structures; such an element (and chain) is simply not present in Grimshaw's structures. In conjunction with our representation of scope, *Q essentially replaces the *Wh*-criterion (May, 1985). A language which ranks *Q higher than *t is one in which [wh] functions as a "strong feature" in the sense of Chomsky 1992.

[6]The '+' in multiple-*wh* candidates indicates adjunction to spec of CP.

constraints they violate. Circled stars '⊛' are special marks for constraints violated by the optimal candidate, identified by '☞'. A star followed by an exclamation mark '*!' identifies an *active* constraint—the particular violation identified by '*!' is fatal; it causes the corresponding parse to lose to another candidate parse (e.g. *b* loses to *a* in Chinese by virtue of the mark '*!' in column *t.) Given that, in Chinese, candidate *a* has to defeat candidate *b*, and given that *a* violates *Q , a constraint not violated by the losing candidate *b*, constraint *Q must be ranked lower than a constraint violated by the losing candidate *b*, i.e. *t. We thus obtain the basic Chinese ranking given in (5a)—with higher ranked constraints conventionally placed to the left of lower ranked constraints. The relative ranking of *ABSORB doesn't matter here; absorption is irrelevant in Chinese (it is a factor only when *wh* moves, in our terms when *Q ranks higher than *t; then it decides between *b* and *c*). This situation is suggested by dotted lines separating the *ABSORB column from the others.

The constraint tableau for Bulgarian is given next.

(5) b. Bulgarian multiple *wh*

[Q_i Q_j [x_i V x_j]]	*ABSORB	*Q	*t
a. [Q_i+Q_j [wh_i V wh_j]]		*! *	
b. ☞[wh_i+wh_j [t_i V t_j]]			⊛ ⊛
c. [$wh_{i[j]}$ [t_i V wh_j]]	*!		*

For multiple fronting of *wh* to win, i.e. candidate *b*, some constraint violated by each of the sub-optimal candidates *a* and *c* must be ranked higher than the constraints violated by the optimal candidate *b*, namely *t. Sub-optimal candidate *a* violates *Q; comparing the pair *a* and *b* thus shows that *Q ≫ *t. Comparing *b* and *c* shows that *ABSORB must outrank *t: one violation of *t for both candidates cancels; what remains is a violation of *t for the winning parse *b* and a violation of *ABSORB for the sub-optimal parse *c*. Hence *ABSORB outranks *t. The *ABSORB and *Q columns are separated by a dotted line because their relative ranking is yet to be determined.

Finally, the English tableau is given below.

(5) c. English multiple *wh*

[Q_i Q_j [x_i V x_j]]	*Q	*t	*ABSORB
a. [Q_i+Q_j [wh_i V wh_j]]	*! *		
b. [wh_i+wh_j [t_i V t_j]]		* *!	
c. ☞ [$wh_{i[j]}$ [t_i V wh_j]]		*	⊛

In English, the optimal parse is *c*, which violates *ABSORB and *t. Its competitors lose because *b* encurs two marks against *t (against one for candidate *c*) and candidate *a* violates *Q which is ranked highest.

The resulting (partial) rankings for these languages are summarized in (6).[7]

(6) Partial rankings:

Chinese:	*t ≫ *Q	*ABSORB unranked
Bulgarian:	{*Q, *ABSORB} ≫ *t	
English:	*Q ≫ *t ≫ *ABSORB	

III. Long vs. short movement : Government, Locality, and Referentiality

1. Government

Our representation of scope posits chains that includes traces, hence the need to constrain the occurrence of these traces. We follow much traditional work and in particular Rizzi (1990) and Cinque (1990) in positing that traces are constrained by head-government. In English, the cost of violating head-government is high, hence the relevant constraint is high-ranked. We assume with Cinque (1990) that proper head-government is not sensitive to the distinction between lexical and functional categories.[8] Rather

(7) Gov(t): t must be head-governed by a category non-distinct from [+V]

We only partially adopt Cinque's characterization of which categories are non-distinct from [+V]; namely V [–N, +V], I [+V], and A [+N,+V]. In particular, we reject his assumption that C is a proper governor. Evidence that C is not a proper governor comes from the unexplained absence in (8a) of complementizer stranding in English, which otherwise allows preposition stranding[9] and I stranding (8b).

(8) a. *John left, we think that t.
b. John said that he would leave and leave, he did t.

---

[7]The proposal made here is similar to one independently made in Billings and Rudin (1994). Note that both proposals fail to characterize languages like Italian, Irish, and Quigolani Zapotec which require fronting of one single *wh*-phrase while not allowing any others to remain in-situ. An extension of the present account to these languages is provided in Legendre, Smolensky, and Wilson (to appear).

[8]Grimshaw (in press) proposes constraints requiring that t be head-governed (T-GOV) but distinguishes functional from lexical government (T-LEX-GOV).

[9]The claim that Nouns [+N, -V] are inadequate governors is made in Kayne (1984), Rizzi (1990), and Cinque (1990): cross-clausal NP movement in passive and raising cannot take place within NPs, as in **John's appearance t to be sick*; preposition stranding under NP movement is impossible in NPs: **The new law's vote for t* . We assume that P is [–N] in English, hence a proper governor. P is [–N, –V] in Chinese, French and Bulgarian which do not allow preposition stranding; C is presumably [–V] except when it carries agreement, in which case it is, like I, [+V]. See French *qui* > *que* rule (Rizzi, 1990).

We further depart from many standard assumptions about which elements are properly head-governed. In addition to direct objects (governed by V), subjects internal to VP and adjuncts (adjoined to VP) are governed by a functional head (I for the sake of simplicity). Subjects in spec of IP are typically not properly head-governed but under certain conditions they may be governed by a matrix V (see discussion of bridge verbs below).[10] The main reason for assuming that adjuncts are properly governed comes from the cross-linguistic observation that extraction of adjuncts is possible: to mention only two cases, adjuncts are extractable out of the complement of *think* and in multiple questions in English. In Chinese, wide scope interpretation of most adjuncts is possible out of complements of *think* as well as weak and strong islands (Tsai, 1994).

We assume with Rizzi (1990) that strict c-command (rather than m-command) underlies head-government and that a head properly governs its complement and the specifier of its complement (or adjoined position if a phrase is adjoined to the complement). Partial evidence for this definition of head-government comes from *that*-t effects in English:

(9) a. *Who do you think $[_{CP}$ that $[_{IP}$ t left early?
b. Who do you think $[_{IP}$ t left early?

In (9a) t is not properly head-governed—C is not a proper head-governor; in (9b) t is head-governed by the matrix verb *think*.

2. Locality

A well-known generalization about *wh*-questions is that extraction is basically *local*: movement cannot place the moved element too far from its originating site. Consider, for example, the extraction of adjuncts in English.

(10) a. How$_1$ did [he [fix it] t$_1$]?
b. How$_1$ do [you [think ⟦ e$_1$ that [he [fixed it] t$_1$]]]]?
c. *How$_1$ does [she [wonder ⟦ what$_j$ [John [fixed t$_j$] t$_1$]]]]?

The standard analysis of the contrast between (10b) and (10c) involves successive cyclic movement, exploiting the escape hatch in the specifier position of the embedded CP. In (10c), the specifier position of the embedded CP is filled with a *wh*-phrase, with the result that movement of *how* is long, in violation of subjacency. Any attempt to precisely characterize the intuition that shorter movement is better than longer one must wrestle with the issue of counting units of length.[11] In OT, we can naturally distribute locality over a family of constraints, MINLINK, which refers

[10] In this paper we assume that subjects are generated under VP but nothing hinges on making this assumption rather than generating them adjoined to VP. We further assume that subjects move to spec of IP in English and Chinese (prior to extraction) while they don't have to in Bulgarian which doesn't show *that*-t effects (See further discussion below). To maximize the readibility of English and Chinese tableaux, we systematically omit the properly head-governed t in spec of VP.

[11] For a proposal within the Minimalist Program, see Collins (1994).

to the length of *chain links*, using Chomsky's 1986 notion of *barrier*: a single link of a chain violates BAR1 when it crosses 1 barrier; it violates BAR2 when it crosses 2 barriers, etc.[12] The constraints BAR1, BAR2, BAR3 are universally ranked, as shown in (12).[13] Like all other constraints in OT, the BAR constraints are violable and are violated in well-formed structures. A preliminary list of the MINLINK constraints is given in (11):

(11) MINLINK family of constraints (incomplete):
BAR1: a single link must not cross one barrier
BAR2: a single link must not cross two barriers
BAR3: a single link must not cross three barriers

(12) Ranking (universal): BAR3 $\gg$ BAR2 $\gg$ BAR1

Assuming the representational equivalent of successive cyclicity, (10b) contains a chain of two short links (how_1,e_1) and (e_1, t_1), which cross 2 barriers and 1 barrier, respectively. In contrast, (10c) contains one long link (how_1, t_1) which crosses 3 barriers. The full OT account requires considering the optimal output which competes with that represented in (10c), as we will see shortly. We comment first on the status of intermediate traces in our account. Each link in (10b) contains an intermediate trace e in the specifier of the lower CP (called e to differentiate it from t in D-structure position). e is a by-product of short links and is itself constrained by GOV(t). Evidence that e is subject to GOV(t) comes from contexts known as strong islands out of which extraction in many languages, including English, is impossible: sentential subjects (14a), adjunct clauses (14b), and complex NPs (14c).

(13) a. [That *he* left early] was obvious to everyone.
b. He got upset [after he saw *Mary*].
c. I found [a man [that would fix it *fast*]].

(14) a. *Who_i [[e_i that [t_i left early]]] was obvious to everyone?
b. *$Who(m)_j$ did he get upset [e_j after [he [saw t_j]]]?
c. *How_1 did you find ⟦ a man [e_1 that [would fix it] t_1]]?

Under standard assumptions, the bracketed clauses contain an escape hatch in spec of CP. Crucially, any intermediate e resulting from an attempt at forming a successive-cyclic chain fails to be properly governed. In (14a) e is not properly head-governed because there is no potential head-governor for e; in (14b), e is not governed because it is not in the government domain of I (assuming CP is a sister of V′); in (14c), e is governed by the head of the relative clause, a [–V] category in violation of the requirement that it be governed by a category non distinct from [+V].

---

[12]In the tableaux we use plain brackets ([) to represent barriers and hollow brackets (⟦) for non-barriers, i.e. L-marked maximal projections (Chomsky, 1986).

[13]As developed in Legendre, Smolensky, and Wilson (to appear), the MINLINK constraints and their universal ranking are a consequence of a general OT mechanism of constraint interaction, Local Conjunction.

The fact that extraction out of these strong islands results in ill-formedness is further evidence for the earlier claim that GOV(t) ranks high in English.

GOV(t) does an important part of the work of Huang's CED (Huang, 1982) but it is a different constraint in at least four respects: a) it is a *general* constraint (on all traces) b) it is a constraint on traces, not a constraint on some extraction domain c) there is no issue of the level at which GOV(t) applies, given our single level of optimization (in Huang 1982, the CED crucially does not apply at LF), and d) GOV(t) takes the place in our account of the ECP as well as the CED.

Having introduced the OT versions of the main constraints we need to handle extraction facts, we now turn to the interaction of these constraints and language-particular rankings. We discuss English first.

2.1. English

Consider the basic case of direct object extraction out of a simple clause. To recover the relative ranking of the constraints we have proposed so far, we apply the standard OT method of systematically comparing two candidates at a time, the optimal parse and a sub-optimal competitor. We compare the marks incurred by each pair of candidates, as shown in (15): candidate *a* violates BAR2 and *t while candidate *b* violates BAR2 and *Q. The violations of BAR2 cancel. Of the two remaining constraints that are violated, *Q must outrank *t, given that *t is violated by the optimal candidate.

(15) Simple questions: direct object extraction

Input: [Q_j [... x_j ...]]	Violated	Comments
a. ☞ [$what_j$ did [he [fix t_j]]]	BAR2 *t	optimal
b. [Q_j [he [fixed $what_j$]]]	BAR2 *Q	*Q ≫ *t
c. [⟨Q_j⟩ [he [fixed NP/⟨wh_j⟩]]]	PARSE(wh)	PARSE(wh) ≫ {BAR2, *t}

The same method is applied to the comparison of candidate *a* and candidate *c*, with the result that PARSE (wh) must outrank BAR2 and *t, the two constraints violated by the optimal candidate. We thus obtain an expanded partial ranking for English:

{*Q, PARSE (wh)} ≫ {*t, BAR2} ≫ *ABSORB.

A more complex case is presented by extraction out of tensed *wh*-islands:

(16) a. ?* What/which dish does she wonder who ate t?
b. * Who/which person does she wonder what t ate?
c. * How/with what speed does she wonder who ate meat t ?

Note first that because verbs like *wonder* subcategorize for a *wh*-complement, they provide a special competitor, one with a narrow scope interpretation, which constitutes a violation of PARSESCOPE whenever the input requires a wide scope interpretation. Consider the fact that direct objects cannot be extracted out of a tensed *wh*-island in English. We examine an input which has wide scope. Tableau (17) displays the competition between (faithful) wide scope interpretation (candidate *a*) and (unfaithful) narrow scope interpretation, which we take to be the optimal parse (candidate *b*).[14] Note that to simplify the discussion we only consider the best candidates in each tableau (it can easily be shown that all other candidates violate additional constraints and hence are suboptimal).

(17) English: Extraction of direct object out of *wh*-island

[Q_j [wonder [Q_i [x_i x_j]]]	*Q	Gv	**P** wh	**B** 3	**P** Sc	**B** 2	1	*t	*AB
a. what$_j$ do [you [wonder ⟦ who$_i$ [bought t$_j$				*$_j$!				*	
b. ☞ you [wonder ⟦ who$_{i[j]}$ [bought what$_j$					⊛				⊛

Here and in other tableaux where space does not permit use of full constraint names, we use the abbreviations:

(18) Abbreviations:
P = PARSE; SC = SCOPE; **B** = BAR; GV = GOV(t); *AB = *ABSORB; **F** = FILL

Tableau (17) shows that the optimal candidate *b* violates PARSESCOPE; suboptimal candidate *a*, which contains a long link, violates BAR3. Hence BAR3 must outrank PARSESCOPE in English.

Tableau (19) displays the competition for subjects:

(19) English: Extraction of subject out of *wh*-island

[Q_i [wonder [Q_j [x_i x_j]]]	*Q	Gv	**P** wh	**B** 3	**P** Sc	**B** 2	1	*t	*AB
a. ☞ [you [wonder ⟦who$_{i[j]}$ [bought what$_j$					⊛				⊛
b. who$_i$ do[you [wonder ⟦what$_j$ [t$_i$ [bought t$_j$		*!		*$_i$		*$_j$		**	

[14] In OT, one cannot simply say that a particular structure is ungrammatical. For each input and set of candidate outputs, there must be a candidate which is optimal. This optimal output is grammatical though it may not match the input (or intended question) perfectly. It may turn out to be an indirect question or a statement rather than a direct question if that's the best the grammar can do for a given input.

In (19), the optimal candidate *a* violates PARSESCOPE and *ABSORB. Competitor *b* loses to *a* because the subject trace t_i violates GOV(t) and because the subject chain violates BAR3.[15]

The next two tableaux summarize our account of *that*-t effects in English. Under our account, structures with and without *that* compete with one another, but matrix verbs are assumed to select either an IP or a CP complement (contra Grimshaw 1993, in press). Not all verbs allow the *that*/∅ alternation: manner-of-speaking verbs like *grieve, gloat, squeal*, etc. require the complementizer *that*; we assume they select for only one type of complement: CP. Under this view, verbs like *think* have two subcategorization frames, $think_{CP}$ and $think_{IP}$ [16]. This means that they correspond to different inputs. They are not distinguished in terms of government, contra Aoun et al. (1987). We assume that outputs meet subcategorization requirements embodied in a SUBCAT constraint (which for present purposes we take to be undominated).

Consider first extraction of a subject out of the complement of $think_{CP}$:

(20) English: Extraction of subject out of complement of $think_{CP}$

$[Q_i\ [\ think_{CP}\ [\ x_i\ ]]]$	*Q	Gv	**P** wh	**B** 3	**P** SC	**B** 2	**B** 1	*t	*AB
a. who_i do [you [think ⟦ that [t_i [left		*!		*				*	
b. ☞ ⟨Q_i⟩ you [think ⟦that [NP/⟨wh_i⟩ [left			⊛						

In standard English, extraction of a subject out of the complement of $think_{CP}$ is ungrammatical. From an OT perspective, the first question is: what does it lose to? We propose that it loses to an unfaithful parse of the input, namely a failure to parse the [wh] feature of the [+wh]DP. The result is candidate *b*, a grammatical structure, whose surface realization is a statement containing a [–wh]DP, something like *you think that someone/a person left*. (Henceforth, we adopt an abbreviation in tableaux according to which, for example, candidate *b* is denoted simply '⟨wh_i⟩'.) In other words, there is no well-formed pure information question with a *wh*-word for this particular input.[17] The winner cannot be a violation of PARSESCOPE, resulting in a narrow scope interpretation of *who*, because in the case of *think*, which doesn't allow an indirect question as a complement, a narrow scope interpretation amounts to a

[15]We assume with Grimshaw (1993) that *who* remains in specifier of IP position.

[16] Some additional evidence for this move comes from the fact that they are (albeit subtly) different with respect to evidentiality.

[17] As is well known, echo questions like "you think that WHO left" are not requests for new information; they presuppose that the answer is already known; hence their interpretation depends on a restricted set of values for the *wh*-variable, reminiscent of D(iscourse)-linking (see section III.3 for discussion). In our terms, they correspond to an input marked with a feature like D-linking and are the optimal output of a candidate set different from the one under discussion in (20).

violation of SUBCAT, which we assume to be undominated. Candidate *a* loses to *b* because it violates higher ranked GOV(t).

Extraction of a subject out of the complement of *think*$_{\mathbf{IP}}$ is possible because it does not result in a violation of GOV(t). This is shown in (21).

(21) English: Extraction of subject out of complement of *think*$_{IP}$

$[Q_i$ [think$_{IP}$ [x_i]]]	*Q	Gv	**P** wh	**B** 3	**P** Sc	**B** 2	**B** 1	*t	*AB
a. ☞ who$_i$ do [you [think ⟦ t_i [left						⊛		⊛	
b. ⟨wh$_i$⟩			*!						

Optimal candidate *a* violates BAR2 and *t. The subject trace is properly governed by *think*. Its competitor *b* is suboptimal because it violates a higher ranked constraint: PARSE(wh). Note how violability of constraints works in OT: a single constraint, PARSE(wh), is violated in the optimal parse in tableau (20) while it causes a competitor to lose in tableau (21).

Extraction of direct objects and adjuncts out of complement of *think*$_{CP}$ and *think*$_{IP}$ is possible because they don't involve GOV(t) violations. The OT analysis is displayed in the following four tableaux. Note that the adjunct extraction tableau involves constraints that are only fully introduced in section III.3: adjuncts like *how* are non-referential (in the sense of Rizzi, 1990 and Cinque, 1990); chains involving *how* are evaluated against the non-referential counterparts of the BARN constraints, represented as BARN[–ref]. Universally, BARN[–ref] >> BARN.

In the remaining tableaux, some constraints are added while others are dropped, due to demands of space. Abbreviations: $\mathbf{B}^{-r}$ = BAR[–ref]; SBC = SUBCAT; OBH = OBLIGATORYHEADS, Grimshaw 1993: heads of projections must be filled.

(22) English: Extraction out of complement of *think*$_{CP}$

a. Direct object extraction

[Q$_j$ [think$_{CP}$ [x$_j$]]]	SBC OBH GV	B^{-r} 3	P wh	B^{-r} 2	B^{-r} 1	B 4	B 3	B 2	*t
a. what$_j$ do [you [think ⟦he [said t$_j$	*SBC!						*		*
b. ☞what$_j$ do [you [think ⟦e$_j$ that [he [said t$_j$								⊛⊛	⊛⊛
c. what$_j$ do [you [think ⟦ that [he [said t$_j$						*!			*
d. ⟨wh$_j$⟩			*!						

b. Adjunct extraction

[Q$_l$ [think$_{CP}$ [x$_l$]]]	SBC OBH GV	B^{-r} 3	P wh	B^{-r} 2	B^{-r} 1	B 4	B 3	B 2	*t
e. how$_l$ do [you [think ⟦he [left] t$_l$	*SBC!			*					*
f. ☞ how$_l$ do[you [think ⟦e$_l$ that [he [left]t$_l$				⊛	⊛				⊛⊛
g. how$_l$ do [you [think ⟦ that [he [left] t$_l$		*!							*
h. ⟨wh$_k$⟩			*!						

In the case of *think*$_{CP}$ (22a-b), the optimal parses *b* (direct object) and *f* (adjunct) take advantage of successive cyclicity, resulting in shorter links violating BAR2 and BAR2[–ref] (ignoring violations of lower constraints); parses *c* and *g* lose to *b* and *f* respectively because they do not take advantage of successive cyclicity, resulting in a longer link violating BAR4 and BAR3[–ref], respectively; parses *a* and *e* lose because they do not have a CP bracket in violation of the SUBCAT constraint on *think*$_{CP}$. Candidates *d* and *h* also lose, which shows that PARSE(wh) must outrank BAR2[–ref] and BAR2.

As shown in (23a-b) below, *think*$_{IP}$ does not provide the successive cyclicity option (that is, without violating SUBCAT, as shown in candidates *b* and *e*). The optimal parses *a* for direct object and *d* for adjunct extraction violate higher BARN constraints. The latter are outranked by PARSE(wh) violated in suboptimal candidates *c* and *f*.

(23) English: Extraction out of $think_{IP}$

a. Direct object extraction

$[Q_j\ [\ think_{IP}\ [\ x_j\ ]]]$	SBC OBH GV	B^{-r}	P wh	B^{-r}		B			*t
		3		2	1	4	3	2	
a. ☞ what$_j$ do [you [think ⟦he [said t$_j$							⊛		⊛
b. what$_j$ do [you [think ⟦e$_j$ [he [said t$_j$	*SBC! *OBH							**	**
c. ⟨wh$_j$⟩			*!						

b. Adjunct extraction

$[Q_l\ [\ think_{IP}\ [\ x_l\ ]]]$	SBC OBH GV	B^{-r}	P wh	B^{-r}		B			*t
		3		2	1	4	3	2	
d. ☞ how$_l$ do [you [think ⟦he [left] t$_l$				⊛					⊛
e. how$_l$ do [you [think ⟦e$_l$ [he [left] t$_l$	*SBC! *OBH			*	*				**
f. ⟨wh$_l$⟩			*!						

Consider the issue of competition under *think* again, say for extraction of direct object or adjunct. Suppose, as in Grimshaw 1993, in press, structures with and without *that* arise from the same input (no selection of IP or CP) via tying for optimality. Then,the analysis should result in two winning candidates per set (given that the presence or absence of *that* does not affect extractability). However, the number of barriers crossed in the presence or absence of *that* differs by one, with the consequence that the candidate with *that* would always lose to the candidate without *that*. Once length of chains enters the picture, it seems that grammatical structures with and without *that* cannot arise from the same candidate set.

2.3. A cross-linguistic prediction

Our account in terms of the interaction between a violable head-government constraint and violable faithfulness constraints makes a cross-linguistic prediction: If GOV(t) were ranked lower than PARSE(wh) and BAR2 were ranked higher than PARSE(wh), we would have the reverse of the well-known pattern: subjects would be extractable, direct objects would not. Such languages in fact exist, among them

Bahasa Indonesia (Saddy, 1991), the Wakashan language Kwakwala (Anderson, 1984), Malagasy (Keenan, 1976), and Tagalog (Guilfoyle et al., 1992). The pattern in information questions in these languages is essentially the same: subjects are extractable, direct objects are not directly extractable; they must be passivized (or topicalized) first. While a systematic analysis of these languages is beyond the scope of this paper, we can still sketch out the following initial proposal for, say, Tagalog. We assume Guilfoyle et al.'s 1992 analysis of Tagalog subjects or topics in spec of IP position. In our terms, object extraction is more costly than subject extraction with respect to barriers: object extraction crosses one more barrier than subject extraction.

Tagalog subject extraction:	constraints violated
☞ subject	BAR1 GOV(t) *t
⟨wh⟩	PARSE(wh)

Comparing the two candidates for subject extraction and their violations, we derive the fact that PARSE(wh) must outrank all other constraints:

PARSE(wh) ≫ {BAR1, GOV(t), *t}

Tagalog direct object extraction:	constraints violated
* direct object	BAR2 *t
☞ ⟨wh⟩	PARSE(wh)

Comparing the two candidates for object extraction and their violations, we can see that BAR2 must outrank PARSE(wh). The proposed ranking for Tagalog is thus:

BAR2 ≫ PARSE(wh) ≫ {BAR1, GOV(t), *t}.

2.4. Chinese

While Chinese has covert *wh*-extraction, it has overt topicalization; according to Huang (1982), the generalization is that covert extraction is generally possible out of complement of *think*, strong and weak islands (except for non-referential adjuncts), while topicalization is more constrained: it is possible out of simple clauses, complements of *think*, and *wh*-islands but impossible out of strong islands. Legendre, Smolensky, and Wilson (forthcoming) claim that resumptive pronouns in Chinese must appear to avoid an ungoverned t. Partial evidence comes from the asymmetry between subjects and direct objects with respect to some topicalizations: the t of a topicalized subject in a simple clause is obligatorily filled with a resumptive pronoun; our informants report that the t of a topicalized direct object is only optionally filled with a resumptive pronoun. This is shown in the contrast between (24) and (25).

(24) a. Zhangsan$_i$, ta$_i$ xihuan kansu.
Z. he like reading
Zhangsan, he likes reading.

b.*Zhangsan$_i$, t$_i$ xihuan kansu.
Z. like reading
Zhangsan, likes reading.

(25) a. Lisi$_j$, Zhangsan hen xihuan ta$_j$.
L. Z. very likes him
Lisi, Zhangsan likes him very much.
b. Lisi$_j$, Zhangsan hen xihuan t$_j$.
L Z very likes
Lisi, Zhangsan likes very much.

The pattern in (25) raises the issue of optionality in an optimization framework. How can we have two optimal parses? One scenario, exemplified here, results from two equally ranked constraints,[18] *t and FILL (a faithfulness constraint which prohibits 'epenthesis' of elements not present in the input, Prince and Smolensky 1993). Consider tableau (26) which displays the competition between (25a) and (25b). We assume that topicalization involves an operator TOp and adjunction to IP.

(26) Chinese: Topicalization of direct object out of simple clause

[Top$_j$, [x$_j$]]	GOV (t)	...†	PARSE (top)	BAR 2	BAR 1	*TOp	FILL	*t	*Q
a. ☞ NP$_j$, ⟦$_{IP}$ NP [$_{VP}$ V t$_j$					⊛			⊛	
b. ☞ NP$_j$, ⟦ NP [V res$_j$					⊛		⊛		
c. TOp$_j$, ⟦ NP [V NP$_j$					*	*!			

Here and in the next several tableau, '...†' abbreviates a contiguous sub-hierarchy of constraints, none of which are violated by the candidates under examination:

(27) Chinese sub-hierarchy†
BAR3[–ref] ≫ PARSE(wh) ≫ BAR2[–ref] ≫ BAR1[–ref]
≫ PARSESCOPE ≫ BAR4 ≫ BAR3

(The BARN[–ref] constraints will be discussed in Section 3.)

In tableau (26), candidates *a* and *b* incur the same pattern of constraint violations and are equally optimal: they each violate BAR1 and FILL/*t. Candidate

[18]We think that a similar scenario obtains in Bulgarian multiple *wh*-questions which display restrictions on the order of multiple *wh*-phrases. According to Rudin (1985, personal communication), subject *koj* must precede object *kakvo* "what" and *kogo* "whom". Other pairs like subject *kakvo* and object *kogo* are unordered; so are pairs involving subject *kakvo* and adjuncts *koga* "when" and *kŭde* "where". We propose that these patterns result from two equally ranked constraints dubbed SUBJECT FIRST and HUMAN FIRST which seem to be cross-linguistically justifiable. A very similar proposal is independently made in Billings and Rudin (1994).

c, topicalization in situ, with an empty TOp operator, is suboptimal because *TOp ≫ {FILL, *t}. (Other candidates not shown in tableau (26), violate higher ranked constraints such as PARSE(top).)

The same ranking, in conjunction with higher ranked GOV(t), yields the contrast between (24a) and (24b). This is shown in tableau (28).

(28) Chinese: Topicalization of subject out of simple clause

[Top$_i$, [x$_i$]]	GOV (t)	…†	PARSE (top)	BAR		*TOp	FILL	*t	*Q
				2	1				
a. NP$_i$, ⟦ $_{IP}$ t$_i$ [$_{VP}$ V NP]]	*!							*	
b. ☞ NP$_i$, ⟦ res$_i$ [V NP]]							⊛		

In tableau (28), candidate *a* loses to *b* because it contains an ungoverned subject t.

When topicalization occurs out of the complement of a bridge verb like *renwei* "think", the pattern shows an intriguing variant: in both subject and object position, t and resumptive pronoun alternate. Why should this be the case? Consider the data in (29) and the competition displayed in tableau (30).

(29) a. Zhangsan$_i$, wo renwei t$_i$/ta$_i$ hen congming.
Z I think (he) very clever
"Zhangsan, I think (he) is very clever".
b. Zhangsan$_j$, wo zhidao ni hen xihuan t$_j$/ta$_j$.
Z I know you very like (him)
"Zhangsan, I know you like (him) very much"

Chinese does not have overt complementizers. Does *renwei* take an IP or a CP (with a null complementizer or in violation of Grimshaw's ObH constraint)? The answer emerges out of the competition displayed in tableau (30).

(30) Chinese: Topicalization of subject out of complement of *renwei* "think"

[Top$_i$, *renwei* [x$_i$]]	GOV (t)	…†	PARSE (top)	BAR		*TOp	FILL	*t	*Q
				2	1				
a. ☞ NP$_i$, ⟦$_{IP}$ [$_{VP}$V ⟦$_{IP}$ res$_i$					⊛		⊛		
b. NP$_i$, ⟦ [V ⟦$_{CP}$ e$_i$ [$_{IP}$ res$_i$					**!		*	*	
c. ☞ NP$_i$, ⟦ [V ⟦ t$_i$					⊛			⊛	
d. NP$_i$, ⟦ [V ⟦ e$_i$ [t$_i$	*!				**		*	**	

Candidates *a* and *c* have IP complements, *b* and *d* have CP complements. The optimal outputs are *a* and *c* because they incur the least number of identical marks (BAR1 and FILL/*t). Candidate *b* loses to both *a* and *c* because it incurs both FILL and *t violations, in addition to BAR1. Candidate *d* fares much worse because t_i is not properly governed, causing a violation of high ranked GOV(t).[19]

Having established that complements of *renwei* are IPs, we turn to the covert *wh*-extraction facts. Note that the candidate set includes *wh*-fronting with resumptive pronoun in-situ, the counterpart of the strategy for subject topicalization. Why does the latter lose to in-situ *wh* in Chinese? Consider tableau (31): candidate *a* represents in-situ *wh*, candidate *b* represents *wh*-fronting with a resumptive pronoun in-situ, candidate *c* involves *wh*-fronting with a t.

(31) Chinese: Covert extraction of subject out of complement of *renwei* "think"

[Q_i [*renwei*$_{IP}$ [x_i]]]	GOV (t)	...†	PARSE (top)	BAR 2	BAR 1	*TOp	FILL	*t	*Q
a. ☞ Q_i [$_{IP}$ [$_{VP}$ V [$_{IP}$ wh_i				⊛					⊛
b. wh_i [[V [res_i				*			*!		
c. wh_i [[V [t_i				*				*!	

The BAR2 violations cancel with the result that candidates *b* and *c* lose to candidate *a* because their violations (FILL, *t) outrank that of *a* (*Q). The candidate '⟨wh_i⟩' which fails to parse the *wh*-chain incurs *PARSE(wh), which is high-ranked in the segment of the hierarchy abbreviated to '...†' (27); it therefore is sub-optimal. Here and henceforth we omit the '⟨wh⟩' candidates (but see footnote 20).

Chinese permits covert extraction of subjects, direct objects, and certain adjuncts (*where*, *when*, instrumental *how*, purpose *why*) out of wh-islands (Tsai, 1994).

(32) a. Ni xiang-zhidao [shei zai nali gongzuo] ?
you wonder who at where work
"Who do you wonder where works?"
"Where do you wonder who works?"

b. Ni xiang-zhidao [Lisi shenmeshihou mai-le sheme]?
you wonder L when buy-ASP what
"What do you wonder when Lisi bought?"
"When do you wonder what Lisi bought?"

The structure corresponding to the covert extraction in (32b) is given in (33):

[19] Note that this analysis entails a differentiation of two constraints against empty elements: *t and *e. This is because violations of *e by intermediate traces do not license resumptive elements in Chinese: FILL $\gg$ *e. Languages like Irish, Chamorro, Kinande, etc. which morphologically register successive cyclic movement in CP may turn out to be languages in which a violation of *e licenses a resumptive element: relative to FILL, *e is equally- or higher-ranked.

(33) [Q_j [$_{IP}$ni [$_{VP}$xiang-zhidao ⟦$_{CP}Q_k$ [$_{IP}$ Lisi [$_{VP}$ shenmeshihou$_k$ mai-le sheme$_j$]]]]]]

Consider direct object extraction, shown in tableau (34).

(34) Chinese: Covert extraction of direct object out of *wh*-island

[Q_j [*xiang-zhidao* [Q_i [x_i x_j]]]]	GOV (t)	…‡	BAR 4	BAR 3	P (top)	BAR 2	BAR 1	*TOp	**F**	*t	*Q
a. ☞ Q_j [[V ⟦Q_i+e_j [wh_i [wh_j						⊛$_j$⊛$_j$	⊛$_i$			⊛	⊛⊛
b. Q_j [[V ⟦ Q_i [wh_i [wh_j			*$_j$!				*$_i$				**
c. Q_j [[V ⟦wh_i [wh_j				*$_j$!							*

(The omitted segment of the hierarchy here is:
‡ BAR3[–ref] ≫ PARSE(wh) ≫ BAR2[–ref] ≫ BAR1[–ref] ≫ PARSESCOPE)

The optimal parse *a* violates BAR2, BAR1, *t, *Q. The fact that indirect questioning of *shei* "who" is covert is instantiated by an empty Q_i in spec of embedded CP, with the result that the chain corresponding to the wide scope interpretation of the direct object "what" (Q_j, wh_j) can be successive cyclic, i.e. composed of short links; parse *b* (instantiating a non-successive cyclic chain) loses to candidate *a* because the chain (Q_j, wh_j) violates BAR4, universally higher ranked than BAR2. Candidate *c* differs from *b* in lacking 'vacuous covert movement' of the subject. Recall that the chains produced by *Gen* have a Q operator in the highest spec of a clause containing the corresponding variable; here, we have a variable in spec of IP in a clause with no CP, so the variable is already in a legitimate scope position for its Q. Thus the minimal licit chain contains a single element which is simultaneously Q_i and x_i. This single element is overtly realized as *wh*$_i$ in candidate *c*.

Extraction of a subject out of a *wh*-island receives essentially the same account, since in the absence of traces, GOV(t) is irrelevant. The only thing that matters is the number and type of BAR constraints violated. Where GOV(t) is irrelevant, shorter links (instantiated by successive cyclicity) always win over longer links. The tableau for subject extraction is given in (35).

(35) Chinese: Covert extraction of subject out of *wh*-island

$[Q_i$ [*xiang-zhidao* $[Q_j [x_i\ x_j]]]]$	GOV (t)	…‡	BAR		P (top)	BAR		*TOp	F	*t	*Q
			4	3		2	1				
a. ☞ Q_i [[V ⟦Q_j+e_i [wh_i [wh_j						$⊛_i⊛_j$	$⊛_i$			⊛	⊛⊛
b. Q_i [[V ⟦ Q_j [wh_i [wh_j				$*_i!$		$*_j$					**

The one complication in Chinese has to do with adjuncts: the counterparts to *when, where*, instrumental *how*, and purpose *why* behave like direct objects in being covertly extractable out of all islands while manner *how* and reason *why* are covertly extractable out of simple clauses, complements of bridge verbs, but not out of complements of non-bridge verbs and islands (Tsai, 1994). Covert manner and reason *wh*-extractions are exemplified in simple clauses (36a), complements of *renwei* "think" (36b), *wh*-islands (36c), and sentential subjects (36d).

(36) a. Lisi zenmeyang chuli zhe-jian shi?
L how handle this-CL matter
By what means/in what manner did Lisi handle this matter?
b. Ni renwei [Lisi yinggai zenmeyang chuli zhe-jian shi]
you think L should how handle this-CL matter
"How (means/manner) do you think that L should handle this matter?"
c. Ni xiang-zhidao [shei zenmeyang chuli zhe-jian shi]
you wonder who how handle this-CL matter
*"How (manner) do you wonder who handled this matter?"
d. *[Women weishenme chuli zhe-jian shi] bijiao hao
we why handle this-CL matter more appropriate
"Why (reason) is it more appropriate for us to handle this matter?"

Assuming that the distinction is one in referentiality, as has been proposed in much recent literature (e.g. Rizzi, 1990, Cinque 1990, Tsai 1994), we show in section 3 how referentiality can be incorporated into the MINLINK family of constraints.

3. Referentiality

Under 'referentiality,' many linguists have subsumed distinct properties which affect extractability in similar ways. The two most common sources of "referentiality" discussed in the literature are theta-roles and discourse-linking. Rizzi (1990) and Cinque (1990) distinguish "referential" theta-roles from "non-referential" ones, which we may interpret on a hierarchy of participants to an event with central participants (agent, patient) at one end and exterior conditions at the other (manner, reason). The grammar of a particular language selects a cut-off point on the hierarchy which dichotomizes it into two parts which syntax treats differently. In Chinese, for example, the cut-off point is between adjuncts: locative, temporal, means, and purpose *wh*-phrases pattern with agent and patient *wh*-phrases—they allow wide scope interpretation out of all context islands—while manner and reason *wh*-phrases (homophonous to means and purpose *wh*-phrases respectively) allow wide scope interpretation out of a restricted set of contexts (see (36a-d) above). In English, as discussed in Rizzi (1990), the distinction shows up for example in extraction out of *wh*-islands with ambiguous verbs like *weigh*.

(37) ?What did John wonder how to weigh t?

Rizzi comments that (37) can only be properly answered with a patient-phrase like *apples* and not with a measure phrase like *200 lbs*. The former is characterized as a referential theta-role, the latter as a quasi-argumental nonreferential theta-role (Rizzi, 1990:86). Rizzi's analysis in terms of a binding relationship is by definition sensitive to this distinction.

Discourse (D)-linking is invoked by Pesetsky (1987) to handle the absence of expected superiority effects in English, i.e. the contrast exemplified in (38).

(38) a. *Mary asked $what_j$ who read t_j ?
b. Mary asked which $book_j$ which man read t_j ?

In (38a), there is no presupposition that either speaker or hearer has a particular set of objects and readers in mind: *what* and *who* are non-D-linked; in (38b) the range of answers is limited by the particular set of books and readers both speaker and hearer have in mind: *which book* and *which man* are D-linked. Rudin (1988) discusses contrasts in Bulgarian extractions out of wh-islands that seem to fall squarely under D-linking (see (43) below). Comorovski (1989) discusses additional examples in Bulgarian. In the absence of a fully specified OT theory of referentiality, we simply adopt the distinctions proposed in the literature and integrate them to our account.

If our family of MINLINK constraints includes constraints on the length of chains containing non-referential elements, then we can account for these patterns in a straightforward fashion. A proper treatment of the referentiality hierarchy, and typological variation in the cut-off point, should be naturally handled by OT (much as is the sonority hierarchy in Prince and Smolensky 1993). But for the purposes of this paper, we assume that what counts as referential is specific to a language and that the constraints refer only to [–ref].

(39) Family of MINLINK constraints (expanded)
BAR1: a single link must not cross one barrier
BAR2: a single link must not cross two barriers
BAR3: a single link must not cross three barriers
BAR1[–ref]: a single [–ref] link must not cross one barrier
BAR2[–ref]: a single [–ref] link must not cross two barriers
BAR3[–ref]: a single [–ref] link must not cross three barriers

We propose that BARN[–ref] universally outranks BARN because of the additional markedness a link has in virtue of being non-referential. (This issue is addressed in detail in Legendre, Smolensky, and Wilson, to appear). For readability of the tableaux, we keep the name 'BAR1,' but henceforth intend it to be interpreted as BAR1[+ref].

We now turn to a MINLINK analysis of Chinese and Bulgarian referentiality contrasts in *wh*-islands.

3.1. Chinese

Examples like (40)—with a non–referential adjunct—are possible in Chinese but they are not interpreted as having wide scope. Rather they have a narrow scope interpretation, as pointed out in Aoun et al. (1989).

(40) Ni xiang-zhidao shei weisheme mai-le shu? (Aoun et al., 1989)
you wonder who why buy-ASP book
"you wonder who bought books why".

Tableau (41) displays the competition between the two interpretations and the role played by constraints on non-referential links.

(41) Chinese Extraction of adjunct [–ref] out of *wh*-island:

$[Q_l$ [*xiang-zhidao* $[Q_i$ $[x_i$ $x_l]]]]$	Gv	**B**$^{-r}$	**P** wh	BAR [–ref]		**P** Sc	BAR		**P** top	BAR		*T Op	**F**	*t	*Q
		3		2	1		4	3		2	1				
a. Q_l $[_{IP}[_{VP}V$ ⟦$_{CP}Q_i+e_l$ $[_{IP}$whi wh_l				$*_l$!	$*_l$						$*_i$			*	
b. ☞ V ⟦ Q_l+Q_i [wh_i wh_l					⊛$_l$	⊛$_l$					⊛$_i$				

The optimal parse *b* includes a non-referential chain (Q_l, wh_l) which violates PARSESCOPE: both wh-forms are paired in the embedded clause. We can ignore the referential chain (Q_i, wh_i) as its violations in *a* and *b* cancel. In parse *a,* one link in the non-referential chain (Q_l, e_l) violates BAR2[–ref] while parse *b*'s only MINLINK violation is of BAR1[–ref].

It is worth examining the interplay of violations: a middle length link of type BAR2 [+ref] is better than a failure to parse scope, but a middle-length link of type BAR2[–ref] is worse than a failure to parse scope.[20] It is not the case that Chinese

[20]And a longer link violating Bar3[–ref] is worse than a failure to parse [wh]. Recall that extraction of a non-referential adjunct out of a sentential subject is ungrammatical (36d). The successive-cyclic parse corresponding to (36d) violates Gov(t); its non-successive-cyclic counterpart violates Bar3[–ref]. For that input, failure to parse [wh] is optimal.

equally disprefers middle length links across the board. This shows that not all instances of locality can be analyzed in terms of one single constraint stating that shorter chains are better than longer chains. We need two families of contraints that are interrupted by unrelated constraints—like PARSESCOPE—on the hierarchy.

3.2. Bulgarian

We finally turn to Bulgarian, where extraction of *koj* (subject) and *kakvo* (direct object) out of a *wh*-island are equally bad, as shown below[21]:

(42) a. *Koj se čudiš kude e otišul?
who refl wonder-2s where aux gone
"Who are you wondering where has gone?"
b. *Kakvo pitaš koj e čel?
what ask-2s who aux read
"What are you asking who has read?"

When *kakvo* is replaced by a D-linked *wh*-phrase like "which of these books", the result is basically fine. The same contrast is observed with *koj* "who."

(43) a. Koja kniga pitaš koj e čel?
Which of these books ask-2s who aux read
"Which of these books are you asking who has read?"
b. ?Koj student se čudiš kakvo e napisal?
which student refl wonder-2s what aux written
"Which student are you wondering what has written?"

In our terms, (42a) and (43a) are not in competition with each other since they have different discourse properties and hence are outputs of different inputs. For an input containing a non-D-linked *wh*-phrase, the optimal ouput does not faithfully parse the input as a direct question (there is no grammatical way of directly asking (42b)), rather it fails to parse the intended scope of the input. The optimal output is thus a narrow scope (rather than wide scope) interpretation of the input.

[21]Our informant, Tzvetelina Ganeva, who generously provided the examples in (42) and (43) informs us that (42a) is acceptable under an echo reading. Our discussion here and below pertains only to neutral, non-echo readings.

(44) Bulgarian: Extraction of a non-D-linked direct object $x_j^{[-dl]}$ out of *wh*-island

$[Q_j\ [\ V\ [Q_i\ [\ x_i\ \ x_j^{[-dl]}\ ]]]]$	SBCGV *AB	B^−r^	P	B^−r^		P	BAR			B^±r^	*t
		4	wh	3	2	Sc	4	3	2	1	
a. $wh_j^{[-dl]}[_{IP}\ [_{VP}\ ⟦_{CP}\ wh_i{+}e_j[_{IP}\ [_{VP}\ t_i\ t_j$					$*_j!*_j$				$*_i$		***
b. ☞ $[\ [\ ⟦\ wh_i{+}wh_j^{[-dl]}\ [\ [\ \ t_i\ \ t_j$						⊛					⊛⊛

The optimal parse *b* violates PARSESCOPE. Focusing strictly on the wh_j chain, competitor *a* loses to *b* because it violates BAR2[–ref] twice.[22]

This situation contrasts with tableau (45) for an input containing a D-linked direct object *wh*-phrase.

(45) Bulgarian: Extraction of a D-linked direct object $x_j^{[+dl]}$ out of wh-island

$[Q_j\ [\ V\ [Q_i\ [\ x_i\ \ x_j^{[+dl]}\ ]]]]$	SBC GV *AB	B^−r^	P	B^−r^		P	BAR			B^±r^	*t
		4	wh	3	2	Sc	4	3	2	1	
a. ☞ $wh_j^{[+dl]}[_{IP}[_{VP}⟦_{CP}wh_i{+}e_j[_{IP}[_{VP}\ t_it_j$									⊛$_j$⊛$_j$⊛$_i$		⊛⊛⊛
b. $[\ [\ ⟦\ wh_i{+}wh_j^{[+dl]}\ [\ [\ \ t_i\ \ t_j$						*!			$*_j\ *_i$		**

The optimal parse *a* violates BAR2 which is lower ranked than the highest constraint violated by competitor *b*, PARSESCOPE. The Bulgarian pattern shows that the BAR constraints are interrupted by an unrelated constraint, PARSESCOPE. Locality is distributed over the hierarchy of constraints: longer chains are better than not parsing intended scope if they are referential, but worse than not parsing intended scope if they are non-referential. Note again the violability of constraints at work here: PARSESCOPE is violated in an optimal parse (44) while it is active (i.e., its violation is fatal) in a sub-optimal parse in (45).

Turning to subjects, we note first that Bulgarian shows no *that*-t effects in non-*wh* complements, despite the presence of an obligatory complementizer *če:*

(46) Koj misliš če e došŭl? (Rudin, 1985)
who think-2s that has come
Who do you think that came?

[22]Rudin (1985:84-5) gives data showing that *wh*-phrases in spec of IP or in post-verbal position can only have an echo interpretation: *Kazvat če KOJ e došŭl?* "They say that WHO came?" *Kazvat če e došŭl KOJ?* "They say that WHO came?". We follow Rudin in interpreting this as indicating that neutral interpretations of *koj* correspond to location in spec of CP.

The relatively free word order of Bulgarian discussed in Rudin (1985) suggests that the subject is free to remain under VP; the absence of *that*-t effects supports this hypothesis[23] because VP internal subject traces are, in our terms, properly head-governed (by I). Extraction of subjects then looks like extraction of direct objects; no violation of GOV(t) occurs. Tableau (47) represents extraction of a (non-D-linked) subject out of a *če* "that" complement.

(47) Bulgarian: Extraction of subject out of *če*-complement

[Q$_i$ [V [*če* [x$_i^{[-dl]}$]]]]	SBC GV *AB	**B**$^{-r}$	**P** wh	**B**$^{-r}$		**P** SC	BAR			**B**$^{\pm r}$	*t
		4		3	2		4	3	2	1	
a. wh$_i$ [$_{IP}$ [$_{VP}$V ⟦ $_{CP}$ *če* [$_{IP}$ [$_{VP}$ t$_i$		*!									*
b. ☞ wh$_i$ [[V ⟦ e$_i$ *če* [[t$_i$					⊛⊛						⊛⊛
c. ⟨wh$_i$⟩			*!								

The optimal output in (47) is *b* because, compared to its competitors, it incurs the least costly violation (BAR2[–ref]); it involves two successive cyclic links.

Finally, consider the analysis of ungrammatical subject extractions out of a *wh*-island (neutral interpretation). Example (42b) is repeated here for convenience:

(48) a. *Koj se čudiš kude e otišul?
who refl wonder-2s where aux gone
"Who are you wondering where has gone?"

(49) Bulgarian: Extraction of subject x$_i^{[-dl]}$ out of *wh*-island

[Q$_i$ [V [Q$_k$ [x$_i^{[-dl]}$ x$_k$]]]]	SBC GV *AB	**B**$^{-r}$	**P** wh	**B**$^{-r}$		**P** SC	BAR			**B**$^{\pm r}$	*t
		4		3	2		4	3	2	1	
a. wh$_i$ [$_{IP}$[$_{VP}$V ⟦$_{CP}$ wh$_k$ [$_{IP}$ [$_{VP}$ t$_i$]t$_k$		*$_i$!								*$_k$	**
b. wh$_i$ [[V ⟦ wh$_k$+e$_i$ [[t$_i$] t$_k$					*$_i$*$_i$!					*$_k$	***
c. ☞ [[V ⟦ wh$_i$+wh$_k$ [[t$_i$] t$_k$					⊛$_i$	⊛				⊛$_k$	⊛⊛
d. ⟨wh$_i$⟩			*$_i$!							*$_k$	

In (49) candidate *a* loses as a combined effect of two factors: 1) the chain (wh$_i$,t$_i$) consists of a long link and 2) *koj* "who" is non-D-linked; hence MINLINK violations involve long non-referential links or BAR4[–ref] violations. Candidate *b* involves

[23] The absence of *that*-t effects is of course consistent with extraction out of a post-verbal position given that subjects can appear post-verbally (Rudin, 1985). In that case, the subject might be extracted out of an adjoined position which still qualifies for proper head-government. We are not aware of any arguments that distinguish the two analyses and hence choose the minimal one.

shorter links, but as they are non-referential, the candidate is still sub-optimal: the optimal output is *c*, which fails to parse the intended scope of the input. Failing to parse [wh], candidate *d*, is a less harmonic unfaithful option.

IV. Summary

When expressed in the violable constraint framework of Optimality Theory, simple government, locality, and referentiality constraints account for a rich set of *wh*-extraction patterns in three typologically distinct languages. That the same constraints are used to build these grammars explains the cross-linguistic commonalities in the extraction patterns; differential rankings of the constraints explain the cross-linguistic contrasts.

The constraints we have proposed are these:

(50) Constraints:

a. Faithfulness family
  - PARSE family: PARSE(wh), PARSE(top), PARSESCOPE
  - FILL

b. MINLINK family:
  - BAR1, BAR2, BAR3, ...
  - BAR1[–ref], BAR2[–ref], BAR3[–ref], ...

c. Government: GOV(t)

d. Operator family: *Q, *TOp, *ABSORB

e. *t

f. SUBCAT

Acknowledgements

For most helpful conversations we thank Luigi Burzio, Robert Frank, Jane Grimshaw, and Alan Prince; for useful comments, audiences at the Society for Philosophy and Psychology, the Association for Computational Linguistics, the University of Toronto, Johns Hopkins University, the University of Delaware, Brown University, and Georgetown University. For help with the Bulgarian facts, we thank Tzvetelina Ganeva and Catherine Rudin; for help with the Chinese facts, Yi-ching Su and Ningsheng Liu. This work was supported in part by NSF grant DBS-9209265.

References

Anderson, S. 1984. Kwakwala Syntax and the Government-Binding Theory. *Syntax and Semantics 16: The Syntax of Native American Languages.*

Aoun, J., N. Hornstein, D. Lightfoot, A. Weinberg. 1987. Two types of locality. *Linguistic Inquiry* 18:4, 537-577.

Aoun, J. and Y. Li. 1993. *Wh*-Elements in situ: Syntax or LF? *Linguistic Inquiry* 24:2, 199-238.

Baker, C.L. 1970. Note on the description of English questions: The role of an abstract question morpheme. *Foundations of Language* 6:197-219.

Billings, L. and C. Rudin. 1994. Optimality and Superiority: A new approach to overt multiple *wh* ordering. To appear in *Proceedings of the Third Symposium on Formal Approaches to Slavic Linguistics.*

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

Chomsky, N. 1986. *Barriers*. MIT Press.

Chomsky, N. 1991. Some notes on the economy of derivation, in Freidin (ed.), *Principles and Parameters in Comparative Grammar*, 417-54.

Chomsky, N. 1992. A Minimalist Program for Linguistic Theory. *MIT Occasional Papers in Linguistics* 1.

Cinque, G. 1990. *Types of A'-Dependencies*. MIT Press.

Cole, P. and G. Hermon. 1994. Is There LF *Wh*-Movement? *Linguistic Inquiry* 25:2:239-262.

Cole, P. and G. Hermon. 1994. Is *Wh*-in-Situ Really in Situ? Evidence from Malay and Chinese. To appear in *WCCFL Proceedings.*

Comorovski, I. 1989. Discourse-Linking and the *wh*-Island Constraint. In *Proceedings of NELS 19*. GLSA, University of Massachusetts, Amherst.

Collins, C. 1994. Economy of Derivation and the Generalized Proper Binding Condition. *Linguistic Inquiry* 25:1, 45-61.

Georgopoulos, C. 1985. Variables in Palauan Syntax. *Natural Language and Linguistic Theory 3*, 59-94.

Georgopoulos, C. 1991. *Syntactic Variables: Resumptive pronouns and A' Binding in Palauan.* Dordrecht: Kluwer.

Grimshaw, J. 1993. Minimal Projection, Heads, and Optimality. Rutgers U. manuscript. Technical Report, Rutgers Center for Cognitive Science.

Grimshaw, J. In press. Projection, Heads, and Optimality. *Linguistic Inquiry*.

Guilfoyle, E. H. Hung, and L. Travis. 1992. Spec of IP and Spec of VP: Two Subjects in Austronesian Languages. Natural Language and Linguistic Theory 10, 375-414.

Higginbotham, J., and R. May. 1981. Questions, Quantifiers, and Crossing. *Linguistic Analysis* 1, 41-79.

Huang, C.-T. J. 1982. *Logical Relations in Chinese and the Theory of Grammar*. MIT Ph.D. Dissertation.

Huang, C.-T.J. and R. May (Eds). 1992. *Logical Structure and Linguistic Structure*. Dordrecht: Kluwer.

Lasnik, H. and M. Saito. 1984. On the Nature of Proper Government. *Linguistic Inquiry* 15: 235-289.

Keenan, E. 1976. Remarkable Subjects in Malagasy. In *Subject and Topic* (C.N. Li, ed.). Academic Press.

Lasnik, H. and M. Saito 1992. *Move alpha: Conditions on its Application and Output*. MIT Press.

Legendre, G., P. Smolensky, and C. Wilson. To appear. When is Less More? Faithfulness and Minimal Links in *wh*-Chains. In *Proceedings of the 1995 MIT Workshop on Optimality in Syntax*. MIT Press.

Legendre, G., P. Smolensky, and C. Wilson. Forthcoming. Optimality in Chinese (C)overt 'Movement.'

Manzini 1992. *Locality*. MIT Press.

May, R. 1985. *Logical Form: Its Structure and Derivation*. MIT Press.

Pesetsky, D. 1987. *Wh*-in-situ: Movement and Unselective Binding, in Reuland and ter Meulen (eds.), *The Representation of (In)definiteness*, 98-129.

Prince, A. and P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, and Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder. To appear in Linguistic Inquiry Monograph Series, MIT Press.

Rizzi, L. 1990. *Relativized Minimality*. MIT Press.

Rudin, C. 1985. *Aspects of Bulgarian Syntax: Complementizers and Wh Constructions*. Slavica Publishers, Inc.

Rudin, C. 1988. On multiple questions and multiple *wh*-fronting. *Natural Language and Linguistic Theory* 6, 445-501.

Saddy, D. 1991. Wh-Scope Mechanisms in Bahasa Indonesia. In *More Papers on Wh-Movement* (L.L.S. Chen and H. Demirdache, eds.). MIT Working Papers in Linguistics, Vol.15.

Srivastav, V. 1991. Subjacency Effects at LF: The Case of Hindi *wh*. *Linguistic Inquiry* 22:3, 762-769.

Tsai, W-T. 1994. On Nominal Islands and LF Extraction in Chinese. *Natural Language and Linguistic Theory* 12:1, 121-175.

Wahba, W.A.B. 1991. LF Movement in Iraqi Arabic. In Huang and May (eds.). 1992. *Logical Structure and Linguistic Structure*. Dordrecht: Kluwer.

Watanabe, A. 1992. *Wh*-in-Situ, Subjacency, and Chain Formation. *MIT Occasional Papers in Linguistics* 2.

Department of Cognitive Science
Johns Hopkins University
Krieger Hall/3400 N. Charles St.
Baltimore, MD 21218-2685

`{legendre, wilson, smolensky}@mail.cog.jhu.edu`