To appear in *Proceedings of the Workshop on Optimality in Syntax*
MIT Press & MIT Working Papers in Linguistics

When is Less More? Faithfulness and Minimal Links in *wh*-Chains[*]

Géraldine Legendre, Paul Smolensky, & Colin Wilson
Department of Cognitive Science, Johns Hopkins University

1. Introduction

It is an attractive prospect to explain and unify a broad spectrum of locality effects in syntax through a principle requiring that movements, or chain links, or dependencies, be as short as possible. But a formalization of such a principle faces great challenges, the most fundamental of which is a proper treatment of the comparative computation inherent in the qualifier 'as possible.' A grammatical framework for formalizing such comparative principles is provided by Optimality Theory ('OT'; Prince & Smolensky 1991, 1993), and our goal in this work is to explore the explanatory power of general comparative constraints in syntax like 'Shortest Link' (e.g., Chomsky 1992) when they are formalized within OT. More specifically, our goal in this paper is a formalization of 'Shortest Link', which we will call MINLINK, which is simultaneously *flexible* enough to accomodate the broad cross-linguistic variation in extraction patterns, and *strong* enough to do the primary work in a wide variety of explanations, including the following general types[1]:

(1) a. *S (sentence S is ungrammatical) because a chain C_i in S is longer than the corresponding chain $C_i{}'$ in a competing structure S′ with the same LF representation (e.g., super-raising).
b. *S because a link of a chain C_i in S is longer than any in a corresponding chain $C_i{}'$ in a competing structure S′ with a *different* LF representation (e.g., wh-islands).
c. *S because a chain C_i in S is longer than the chain of a *different* element $C_j{}'$ in a competitor S′ (e.g., superiority).
d. *S because S has an A′-chain C_i while a competing structure S′ *has no A′-chain* for that element. (e.g., strong islands).

---

[*]Our thanks to Kristin Homer and William Raymond, members of our research group at Colorado; to Joseph Aoun, Luigi Burzio, Robert Frank, Jane Grimshaw, David Pesetsky and Alan Prince, for extremely helpful conversations; to audiences at Arizona, Brown, Cornell, Delaware, Georgetown, Hopkins, Maryland, UCLA, USC, and the MIT conference, for stimulating ideas and questions. For partial financial support, we gratefully acknowledge NSF grant BS-9209265, and then NSF grant IRI-9213894, both to Smolensky and Legendre; and the Center for Language and Speech Processing at Johns Hopkins.

[1]A chain C is 'longer than' chain C′ if the longest link in C is 'longer than' the longest link in C′, in a precise sense defined below. In all the examples in (1), MINLINK interacts crucially with other constraints which also figure in the complete explanations; these are developed below.

Our main claims are these: (a) OT allows such a strong but flexible MINLINK to be constructed directly from general principles, and (b) the resulting MINLINK explains a broad range of cross-linguistic (overt and covert) extraction facts[2]. In addition, we argue for certain general solutions to fundamental questions for comparison-based syntactic formalisms: What structures compete? What is the 'input'? How is language-particular ineffability possible? How are marked outputs possible? (See Section 2.)

The basic ideas underlying our analysis can be summarized as follows. MINLINK is a *universal subhierarchy of constraints* establishing a scale on which longer links are less harmonic (more marked) than shorter links. Inputs to the OT grammar specify *target scopes* for operators; FAITHFULNESS constraints require output chains to realize these targets. (These constraints are violable, of course.) Extraction patterns result from the interleaving of MINLINK and FAITHFULNESS constraints in a language's constraint hierarchy. Cross-linguistic variation in extraction patterns results from the re-ranking of these constraints (and others).

We briefly illustrate these fundamental ideas before beginning our analysis proper. An outline of the paper may be found at the end of this Introduction.

Our example is a fragment of the analysis of Chinese presented more fully below in Section 6. We consider why non-referential adjuncts can be covertly extracted from complements (2a) but not *wh*-islands (2b) (data from Tsai, 1994):

(2) a. Ni renwei [Lisi yinggai zenmeyang chuli zhe-jian shi]
you think L should how handle this-CL matter
"How (manner) do you think that L should handle this matter?"

b. 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?"
"You wonder who handled this matter how"

We claim that the *wh*-island extraction is bad because the (covert) *how*-chain in (2b) is too long, as measured in barriers crossed—despite the fact that the *how*-chain in the [–wh] complement extraction in (2a) is the same length and is grammatical. In our analysis, the relativized minimality effect will be seen to arise as follows (see constraint tableau (3)[3]). The input has wide scope: the target LF has an operator

[2] The large majority of facts we consider involve *wh*-extractions, but we are really concerned more generally with determination of operator scope and of the silent/pronounced status of chain elements, including simple clauses and syntactic operators other than *wh*; we will include all these under the label 'extraction patterns.'

[3] We mark the constraint violations of the optimal candidate with '⊛' rather than '*'. Each '⊛' identifies a constraint violated by grammatical forms; constraints like MINLINK in tableau (3) with both a '*!' and a '⊛' are *active* (Prince & Smolensky 1993:§5) in ruling out candidates ('*!'), yet surface-violated ('⊛').

Q_l marking an extraction of *how*$_l$ out of the embedded clause. (i) The wide-scope chains faithful to the target (3a,a′) compete with narrow-scope chains (3b,b′) in which *how* remains within the embedded clause; these are unfaithful to the operator scope in the input, so they violate the faithfulness constraint PARSESCOPE. (ii) In the *wh*-island case, the matrix verb *wonder* selects a [+wh] complement so this competing, shorter, chain (3b) does not violate selectional restrictions (the constraint SELECT). The shorter chain wins even though it is unfaithful because the PARSESCOPE violation is lower-ranked than the bad MINLINK violation incurred by the long extraction (two barriers crossed, and non-referential, violating the particular MINLINK constraint $\text{BAR}^{2}_{[-\text{ref}]}$ defined below). (iii) In the *think* complement, however, the narrow-scope chain (3b′) yields a [+wh] complement which violates the selectional requirements of *think*; because SELECT is highest-ranked, this rules out the shorter-chain competitor, and the longer chain is optimal, despite its being disfavored by MINLINK.

(3) ***wh*-islands in Chinese**

input: Q_l V_{matrix} [··· how_l ···]	SELECT	MINLINK†	PARSESCOPE
***wh*-island extraction (*)**			
a. Q_l $wonder_{[+wh]}$ [··· how_l ···]		*!	
b. ☞ $wonder_{[+wh]}$ [Q_l ··· how_l ···]			⊛
***think* complement extraction (✓)**			
a′. ☞ Q_l $think_{[-wh]}$ [··· how_l ···]		⊛	
b′. $think_{[-wh]}$ [Q_l ··· how_l ···]	*!		*
Remarks:	No [+wh] complement for *think*	†constraint crucial here: $\text{BAR}^{2\,[-\text{ref}]}$	For an extraction input, extract

Briefly, then, in Chinese, the interleaving of the faithfulness constraints (crucially, PARSESCOPE) and the MINLINK constraints (crucially, $\text{BAR}^{2\,[-\text{ref}]}$) entails that when an input provides a target involving a long non-referential chain link, the output will not be faithful to the wide input scope, provided that a narrow-scope alternative exists which does not violate the selectional restrictions of the matrix verb. This is so with *wonder*, but not with *think*.

In this account, there is no need to stipulate that ‘Shortest Move’ is measured in terms of relativized minimality violations; the relativized minimality effect is a consequence of a MINLINK constraint in which link length is measured simply in terms of maximal projections crossed (barriers, to be precise). None of our constraints require 'relativized' distance measurements, or 'minimality' of any kind: minimality effects arise purely from the constraint *interaction* automatically provided by OT, so the constraints themselves do not refer to 'minimality,' and as we have just seen, relativization effects (e.g., *wh* is harder to extract over *wh*) are also a derived *consequence* of constraint interaction.

The structure of the remainder of this paper is as follows. In Section 2, we provide an informal development of the basic assumptions behind our technical approach; for example, we argue that LF-*in*equivalent structures must compete, and motivate the 'LF-target' aspect of our inputs. In Sections 3 and 4, we give a more formal presentation of our proposed theory. In Section 5, we derive the core typology of question strategies. In Section 6, we develop our analysis of Chinese *wh*-chains. In Section 7, we briefly compare the Chinese analysis to its English counterpart. In Section 8, we begin an exploration of the typology of extraction patterns predicted by the theory. In Section 9, we make a few concluding remarks concerning the commitments of an Optimality Theoretic approach to syntax, comparing them briefly to the Minimalist Program ('MP'; Chomsky 1992, 1995).

2. Informal Preliminaries: Syntactic Competition and Faithfulness

2.1 The Input of the Question, and the Question of the Input

In syntax, the question of the input looms large. As usually conceived, an OT grammar is a device that takes an input and produces an output, the correct structural description or 'parse' of the input. In phonology, an input is canonically taken to be the underlying form of a word, and the output, when phonetically interpreted, the pronunciation of that word.[4] A crucial job of the input is to determine what structures compete: given an input *I*, the competing parses are the set *Gen*(*I*): all possible parses *of I*. That forms with different phonetic interpretation compete is a consequence of the fact that not all the parses of *I* in *Gen*(*I*) are faithful to *I*; the phonetic interpretation of such parses may include material not in *I*, or fail to include material present in *I*. Such unfaithful parses are optimal only when the faithfulness violations they incur are outranked by other violations arising from faithful parsing. Faithfulness violations, visible as 'deep/surface' disparities between the *target phonetics* in an underlying form *I* and the actual phonetic interpretation of *I*'s optimal parse, are readily apparent in phonology via surface alternations, where the same morpheme undergoes faithful parsing in one environment and unfaithful parsing in another.

Much of this structure is not readily apparent in syntax, however; the question of the input, and of what structures compete, and of the role of faithfulness, all seem

[4] For a comparative, violable-constraint approach to phonology with no such concept of 'input,' see Burzio 1994.

rather more opaque. In phonology, the question of the surface form of a given morpheme provides a sound theoretical base, supporting a view of the grammar as a device for mapping a particular underlying form deriving from a lexicon into its correct structural description. In syntax, we believe, this is the wrong view of the grammar. The alternative view, however, was also developed in phonology in Prince & Smolensky (1991, 1993), for example, in the analysis of basic CV syllable structure (§6). There, the question of interest is not, 'what is the structural description of a *particular* input (say, /VCVC/)?'; the question is more global: 'what is the *inventory of all possible* output syllable shapes, as the input is allowed to range over all possible input strings of C's and V's?' Correspondingly, the question of interest in *wh*-theory, we claim, is not, 'given some *particular* input structure, what is its particular output structure?' Rather, it is: 'what is the *inventory of all possible* questions in a given language, deduced by considering all possible inputs?' So, in particular, in a language which allows no extractions from *wh*-complements, our concern is to show that, given our proposed OT grammar, no matter what input we consider, the optimal output structure never contains an extraction from a *wh*-complement. It is not of central concern what the output happens to be given a *particular* input, e.g., one which *would* produce a *wh*-complement extraction in another language the question inventory of which includes such extractions.

To summarize, for syntax we adopt the *inventory* perspective within OT (Prince & Smolensky 1993: §9), rather than the perhaps more familiar *particular input/output mapping* perspective. In any event, we do *not* construe an OT grammar as an account of the 'production' system, responsible to the question, 'given an intention to utter *X*, what does the speaker do?' Such performance models are clearly outside the scope of competence grammars like those of OT.[5] Rather, we take OT to answer the question, 'what is the inventory of grammatical structures in a language?' It is the job of performance theories to explain how speakers manage to deploy the structures made available by the grammar.

2.2 The *Language-Particular Ineffability* and *Marked Output* Problems

In the inventory perspective on OT, the job of the input *I* is to determine what competes: for a given *I*, the structures in the set *Gen*(*I*) compete. Usually, *I* can be thought of as the substructure two candidates must share in order that they be competitors. Thought of this way, we might call *I* the *Index* of its particular candidate set; henceforth, we replace 'input' with 'Index' in order to suppress the 'particular input/output perspective' on OT, and to avoid the baggage which the term 'input' inappropriately brings to syntax.[6]

[5] An OT grammar defines *what* must be computed, not *how* such computation is performed. But OT grammars can often be used directly in parsing algorithms which efficiently compute optimal forms: see Tesar 1994, 1995ab.

[6] The sense of 'index' intended here is exactly the mathematical usage in which each

For studying s-structures, an attractive hypothesis is that the Index consists of a pair [d-structure, LF-structure]; that is, two s-structures compete if and only if they derive from the same d-structure and yield the same LF (Grimshaw, 1993). The interpretive information of d-structure and LF are fixed among competitors; syntax determines the s-structure realization of this interpretive information.[7]

This attractive approach brings us to our first problem, that of *Language-Particular Ineffability*. There are questions which can be realized in some languages but not others, e.g., 'who ate what' is realizable in English, not in Italian. Such a question must be generable by *Gen* since it is realized in some languages, and *Gen* is universal. This question is part of a candidate set which is also universal. It is optimal in English. The problem is: *What in this candidate set is optimal in Italian?* In an OT syntax, the optimal candidate must be *grammatical*. Yet the optimal candidate must mean 'who ate what' since *everything* in the candidate set means 'who ate what', by hypothesis. But in Italian, there is no grammatical form that means 'who ate what'.

So something must give. One option is to abandon OT's equation *optimal = grammatical*, so that in the 'who ate what' candidate set, the optimal structure in Italian would not be grammatical. This might be done by adding language-particular inviolable constraints to an otherwise-OT syntax. Or it might be achieved by sending the optimal parse out of the syntax, but having it 'crash' at interpretation; but since, by hypothesis, all competitors have a valid interpretation, 'who ate what', it's not clear how this would work. To say that the Italian winner crashes at interpretation while the English winner does not would seem to say that the competitors are not all interpretively equivalent after all. Which is the second option: abandon the requirement that competitors must have the same LF interpretation. As we shall see, once this option is adopted, there is no need to appeal to anything beyond a strictly OT syntax. (The Italian case will appear briefly below in (10)).

---

member of a collection is *indexed*—i.e., uniquely labelled—by a member of an 'index set.' Here, a member of the collection in question is a particular candidate set, and its Index uniquely specifies it. Two structures compete—are in the same candidate set—exactly when they are co-Indexed.

[7]An approach within the Minimalist Program that assumes that only LF-equivalent derivations compete (e.g., Fox, this volume) would be viewed from our perspective as having an Index = [numeration, LF-structure]: derivations that compete share the substructure specified by the Index; *Gen*(Index) is the set of derivations using the Index's numeration and having the Index's LF. Faithfulness to the Index is inviolable in such an MP approach (as in OT approaches like that of Grimshaw 1993); for us, such faithfulness is crucially violable. Another difference, that the numeration is an unstructured list while its counterpart in our Index is a predicate-argument structure, amounts to less than it might appear; the LF roles of lexical items and their features (e.g., case) in the numeration together constrain the structural roles of items to a high degree, so it is quite unclear how crucial are the differences between the information in this MP Index and that in our predicate/argument-structure-plus-operator/variable Index. (Another reason we adopt 'Index' in lieu of 'input' is that only part of the Index—the numeration—serves as the *input* to the derivation; the other part—the LF structure—is, however, an equally important part of the Index: the structure which determines which derivations compete.)

Thus the solution we adopt to the first problem, Language-Particular Ineffability, is: competitors need not have the same LF. An LF unrealizable in a language is a structure such that every syntactic output with that LF interpretation is less harmonic in that language than a competitor with a different LF.

To take a concrete example, consider again Chinese adjunct extractions from *wh*-islands, the data for which are schematically summarized in (4)

(4) Chinese adjunct extractions from *wh*-islands (Tsai 1994)

	Scope	
Referentiality	Narrow	Wide
–	✓	*
+	✓	✓

What does the one ungrammatical form, wide-scope [–ref] extraction, lose to? In our analysis, it loses to the *narrow* scope [–ref] structure, which is grammatical. This possibility, open once we let LF-inequivalent structures compete, is a particularly natural one in Chinese, since the winning structure with narrow-scope interpretation and the losing one with wide-scope interpretation are homophonous.

But this immediately raises our second problem: if the wide- and narrow-scope structures compete, how can *both* be optimal in the case of [+ref] extraction? On a level playing field, narrow scope will always win, preferred by 'Shortest Move.' This is a special case of the general problem faced by any optimization-based theory: *How can marked structures ever be optimal—why don't they always lose to unmarked structures? Why doesn't everything surface as the one perfect syllable/word/sentence (perhaps 'ba')?* The general OT answer to this Problem of Marked Outputs is *Faithfulness*. The input, or 'Index' as we're calling it, contains *targets* for the output, and (violable) FAITHFULNESS constraints require the output to hit these targets. What competitors are competing for is the title of optimal compromise between hitting the target, on the one hand, and general well-formedness constraints on the other. Marked outputs arise when FAITHFULNESS constraints outrank those markedness-defining well-formedness constraints which must be violated in order to be faithful to some Index.

Thus our solution to the problem of Marked Outputs, a special case of the general OT solution, has two parts. (1) An Index contains *target scopes* for operators. For example, a schematic input with narrow target *wh*-scope (marked by *wh*-operator Q) might be:

Index$_1$: you wonder [Q_i Q_j x_i ate x_j]

(faithfully rendered in English as 'you wonder who ate what'); a schematic input with wide target scope for subject x_i:

Index$_2$: Q_i you wonder [Q_j x_i ate x_j]

(faithfully realized in 'who [do] you wonder what ate,' not optimal in English). (2) A constraint PARSESCOPE of the FAITHFULNESS family of constraints requires that output structures realize the target scopes of the Index; e.g., 'you wonder who ate what' is an unfaithful parse of Index$_2$: it violates PARSESCOPE for x_i.

In sum: *outputs with different LFs compete, but not on a level playing field:* each Index contains a target LF, and outputs which are not faithful to it suffer constraint violations.[8] Unfaithful outputs are crucial and often optimal: they are essential to our explanation of how sentences can be ungrammatical[9] in the absence of a grammatical alternative realization of the same LF (a situation not explicitly treated in most OT syntax work to date). Note that the LF structure in the optimal *output* determines its interpretation—the target LF in its Index does not (just as in

---

[8] When LF-FAITHFULNESS constraints are undominated, outputs which are not faithful to the LF in the Index cannot be optimal, and their presence in the candidate set has no consequences. Thus we do not predict that it will *always* be crucial to consider LF-inequivalent outputs, only that it will *sometimes* be. On our account, arguments like that of Fox, this volume, show not that LF-inequivalent structures never compete, but rather that the syntactic/semantic constraints relevant to the phenomena under study do not dominate LF-FAITHFULNESS constraints: thus LF-unfaithful competitors all lose, and are irrelevant.

[9] Another approach to ineffability in OT (Prince & Smolensky 1993: 48) employs the *Null Parse* Ø; when this is the optimal candidate in *Gen*(*I*), the Index *I* has no realization. Of course, Ø does not share LF with any non-empty structure, so to allow it in the candidate set is already to abandon the principle that only LF-equivalent structures compete. We have found it unworkable to use Ø as the winner for every case of ineffability, for the following general reason. Since Ø occurs in every candidate set, it determines a fixed Harmony threshold for the *entire* language: Ø wins every competition in which the best alternative has lower Harmony. The Harmony of Ø is governed by the ranking of PARSE, in the simplest analysis: this is the constraint violated by Ø. PARSE must be ranked so that *every* parse of every ineffable Index violates a constraint higher than PARSE (and loses to Ø); PARSE must also be ranked so that *some* parse of every effable Index violates no constraint higher than PARSE (and bests Ø). This has not proved possible: it is imperative that the relative Harmonies of Index-specific faithful and unfaithful parses be decisive. Our solution, in essence, is to use *mini* 'Null Parses' to *selectively* unparse just the most problematic aspects of an Index. Rather than a single unfaithful parse (Ø) in all candidate sets, we have multiple Index-specific unfaithful parses in which just an operator scope, or just a [wh] feature, is not parsed. (Of course, this fully parallels the phonological case.)

The use of Ø for ineffability brings the effect of inviolable constraint into OT: every constraint out-ranking PARSE cannot be violated in optimal forms. Perhaps attractive in principle, we have found 'virtually inviolable' constraints unworkably rigid in practice.

Our interpretation of an LF-unfaithful winner is that it surfaces as a grammatical sentence, but with a different LF interpretation. But it seems that little or nothing would change if instead we gave LF-unfaithful winners no interpretation at all, like Ø: no realization of the Index; this, because every LF-*un*faithful winner can also be obtained as an *LF-faithful* winner for another Index. What is crucial is that LF-inequivalent structures compete, that LF-targets be present in the Index, and that LF-unfaithful candidates win for ineffable LFs; it does not seem crucial whether these unfaithful structures are given their natural interpretation, or none at all.

phonology, where the material in the output parse determines its phonetic interpretation, not the material in the 'phonetic target': the input/underlying form).

3. The Basic Setup

Having motivated our overall OT approach to syntax, we now lay out the basics.

The Index (input). An index is a *target* predicate-argument structure, with scopes indicated for variables[10]; operators mark scope. For example, we will write

Q_i you wonder [Q_j x_i ate x_j] ~

'for which x_i do you wonder for which x_j did x_i eat x_j'

or equivalently, using the notation *predicate*(subject, object; adjunct):

Q_i *wonder*(you, Q_j V(x_i, x_j))

The primary operator of concern here is Q ≡ *wh*-operator, although our Chinese analysis also treats TOp ≡ syntactic operator for topicalization. Arguments in an Index are marked with syntactically-relevant features such as [±wh], [±referential], etc.; indeed, they are best viewed as simply *being* bundles of such features.

The *Gen* function. The generator of competing structures is defined as follows. *Gen* produces licit X′ trees which may contain co-indexed chains of elements. Given an Index *I*, the candidate set *Gen*(*I*) contains two types of outputs: faithful parses and unfaithful parses. These will be described separately.

In faithful parses, (i) the *foot* of the chain containing any given element is in a position consistent with the input predicate-argument structure (subjects are extracted from their case position [Spec, IP]; adjuncts, from a position adjoined to VP); (ii) the *head* of a variable chain is consistent with the scope of the variable in the input, in the highest position of the appropriate clause; (iii) when multiple *wh*-operators have the same scope, one type of faithful parse has the multiple Q's adjoined to the highest Spec in the clause—this violates ***ADJOIN**; in the other type of faithful parse, Q_i can absorb Q_j, denoted $Q_{i[j]}$, in which case there is no j-chain, the scope of x_j being marked by $Q_{i[j]}$—this violates ***ABSORB**; (iv) one position in a chain contains the *overt* lexical material of the corresponding element of the Index; the other chain positions are empty elements (traces, or, at chain head, an empty operator). E.g., a faithful in situ parse of the preceding Index is:

[$_{\mathbf{CP}}$ Q_i [$_{\mathbf{IP}}$ you [$_{\mathbf{VP}}$ wonder [$_{\mathbf{CP}}$ Q_j [$_{\mathbf{IP}}$ who_i [$_{\mathbf{VP}}$ ate what_j]]]]]]

Given an Index *I*, the unfaithful parses in *Gen*(*I*) are exactly those which are gotten from the faithful parses in *Gen*(*I*) by performing any number of the following three types of operations: (i) failing to respect Index operator scope: the operator is in the highest Spec of the 'wrong' clause—violating **PARSESCOPE**; (ii) failing to 'parse the feature [+wh] on the variable x_i': no x_i chain appears in the output, interpreted as

[10] We adopt a convention in which subject, object, referential adjunct, and non-referential adjunct chains are respectively indexed by i, j, k, l.

an unspecified, non-variable XP (the phonological shape of this XP is of no concern here; we write 'UNSP')—this violates **PARSE(wh)**;[11] (iii) inserting overt (expletive) material in chain elements (resumptive elements)—this violates **FILL** constraints which are relevant, but not of central interest, here.[12]

This concludes our definition of *Gen*. Several points are worth noting: (a) An output is a unitary representation which is simultaneously the interface to semantic and morphophonological interpretation: for the former, it provides the necessary LF information and chain structures; for the latter, linear order and constituency information, and the distinction between pronounced and silent material. (b) The correct unitary structure is selected by a single optimization in which LF-, pronounciation- and chain-based constraints simultaneously conspire to define the optimal structure.(c) The location of the optimal chains in the analyses below cannot in general be determined independently of (e.g., prior to) the determination of the pronunciation of the chain (unlike the approach of Pesetsky, this volume; see footnote 18).

The FAITHFULNESS constraints. The constraints we need have already been introduced. Crucial are **PARSE(F)**, where F = wh, top, or Scope; PARSE(wh) is violated when the Index contains an operator-variable *wh*-chain [Q_i, x_i] (with x_i carrying the feature [wh]) but the output has a non-variable XP corresponding to x_i. PARSE(top) is the analogous constraint for the syntactic operator TOp and the feature it binds, [top]. PARSESCOPE is violated when the chain [Op_i, x_i] is parsed, but with a scope different from that of the Index. Along with the PARSE constraints comes **FILL**, violated by overt traces.

Selection. Pending development of a proper OT theory of selection, we make the following provisional assumptions: (a) verbs like *wonder* select a complement

[11] Ultimately, *Gen* will presumably have to produce candidates that violate other PARSE constraints; a particularly difficult situation (pointed out to us by David Pesetsky at the conference) involves embedding under *wonder* an ineffable question, e.g, **I wonder* [*whom you wonder who loves*]. On our account of *wh*-islands, **whom do you wonder who loves* loses to *you wonder who loves whom*, which is not a direct question and thus violates SELECT when embedded under matrix *wonder*. For this particularly problematic (*wonder*-under-*wonder*) Index, the winner may need to be *you wonder who loves whom*, with PARSE violations for failing to realize the matrix predication, or, once a proper theory of selection is incorporated into the account, something akin to *I **think** you wonder who loves whom*, with a selectional feature of the matrix verb, like [select +wh], unparsed. As emphasized in Section 2.1, what is important is not what the output of this particular Index is; rather, it is accounting for the fact that the inventory of English questions does not include LFs like that of **I wonder* [*whom you wonder who loves*]; either approach could accomplish that.

[12] FILL is the general class of OT constraints requiring that output positions be 'filled' with input material (Prince & Smolensky 1991, 1993). Grimshaw's (1993) FULL-INTERPRETATION is closely related to what we intend by FILL.

(either CP or IP: minimal projection, Grimshaw 1993) which is [+wh], viz., has a Q operator (empty or overt=*wh*-phrase) in highest Spec; (b) verbs like *think* have two forms: $think_{IP}$ and $think_{CP}$, which respectively select an IP and CP complement; (c) verbs like *groan* do not select for a clausal complement, but can take an (optional) clausal adjunct (Doherty 1993); (d) non-matrix clauses which are unselected are CPs.

Head Movement. We implicitly assume the head-movement analysis of Grimshaw 1993; we will not explicitly treat head movement here. In English candidates we will sometimes insert *do* appropriately without comment.

4. Constraints on *wh*-Chains

Having set out in the previous section the constraints that provide the backdrop to our analysis, we pass now to those in our focus.

Operators. A crucial constraint is:

***Q**: No empty question operators.

That is: the scope-marking heads of operator chains must be overt. This is one member of a family of constraints *Op for various syntactic operators; the other we shall have occasion to use is ***TOp**: no empty topicalization operators. *Q favors *wh*-fronting, but conflicts with the next constraint.

Traces. The basic constraint is simply:

***t**: No traces (= STAY Grimshaw 1993)

This requires that a chain's theta-receiving d-structure position, the foot, must be overtly marked; it favors *wh* in situ. This constraint establishes that 'overt movement' is marked, and must be forced by other constraints; it does much of the same work that Procrastinate does in the Minimalist Program. Also, we assume:

GOV(t): t_{XP} must be head-governed by a category non-distinct from [+V] (i.e. V, I, A).

This is our minimal ECP (see also t-GOV, Grimshaw forthcoming). Note that for us, C is not a proper governor, contra Cinque (1990). We assume (loosely following Rizzi, 1990) that a head governs its sister and its sister's highest position (Spec or adjoined). Under this definition, a direct object is head-governed by V; a subject is head-governed by I in its VP-internal position, but *ungoverned* in Spec of IP, except possibly by a matrix V if it takes the IP as a complement. Adjuncts are head-governed by I (when adjoined toVP). These assumptions are briefly argued and explained in Legendre, Wilson, Smolensky, Homer, & Raymond 1995.

Links. Now to the central issue: formalizing 'Shortest Link'. We start with:

BAR: A chain link may not cross a barrier.

We use 'barrier'in the sense of Chomsky 1986: A maximal projection (XP) which is not L-marked. (In our notation, a ‘[’ will denote a barrier, while ‘⟦’ will denote an XP which is not a barrier.)

We come now to our third problem: by itself, BAR is too weak. For example, *ceteris paribus*, chains which are cyclic should be more harmonic; it is locality constraints which require cyclic movement, and a proper characterization of such locality must disfavor non-cyclic chains. But BAR does not.

(5) Cyclic *vs.* non-cyclic chains: equally marked, according to BAR (β = barrier)

		BAR
a. Cyclic, 2 links	☞ $[X_i \cdots \beta \cdots \beta \cdots t_i \cdots \beta \cdots Y_i]$	** *
b. Non-cyclic, 1 link	*☞ $[X_i \cdots \beta \cdots \beta \cdots \cdots \beta \cdots Y_i]$	***

A (‘cyclic’) chain consisting of two shorter links (5a) will violate BAR to the same degree as a (‘non-cyclic’) chain with only one link (5b); this will not do, given contrasts between environments offering intermediate landing sites *vs.* those without, the former broadly tending to better afford extraction.

Power Hierarchies in OT: MINLINK. The solution to this problem will be provided by a general OT mechanism, which directly yields our formalization of ‘Shortest Move’; we now derive the *Local Conjunction Power Hierarchy* of constraints, constructed from the single fundamental constraint, BAR.

Given two constraints $\mathbb{C}_1$ and $\mathbb{C}_2$, their *Local Conjunction* (w.r.t. a domain type *D*), $\mathbb{C}_1 \&_l \mathbb{C}_2$, is a new constraint which is violated when two distinct violations of $\mathbb{C}_1$ and $\mathbb{C}_2$ occur *within a single domain of type D*. (This was proposed for phonological applications in Smolensky 1993.) As a further part of the definition of Local Conjunction, we have the *universal* rankings:

$$\mathbb{C}_1 \&_l \mathbb{C}_2 \gg \{\mathbb{C}_1, \mathbb{C}_2\}$$

(Intuitively: a violation of the conjunction of two constraints is universally worse than violating either one.)

The Local Conjunction of BAR with itself (with domain *D* = link) is a new constraint, **BAR $\&_l$ BAR** $\equiv$ **BAR^2**, which is violated when a link has two distinct violations of BAR, i.e., when it crosses two barriers:

BAR^2: A single link must not cross two barriers

By definition of the Local Conjunction operation, *universally*:

$$\text{BAR}^2 \gg \text{BAR}$$

By recursion, we get the universal **BAR Power Hierarchy** ≡ **MINLINK**:

$$\cdots \gg \text{BAR}^3 \gg \text{BAR}^2 \gg \text{BAR}^1$$

BAR*k*: A chain link must not cross *k* barriers.[13]

Now note that, unlike BAR alone, the BAR Power Hierarchy, our MINLINK, correctly favors cyclic chains:

(6) Cyclic ≻ non-cyclic chains, according to BAR Power Hierarchy: MINLINK

	MINLINK			BAR
	BAR^3	BAR^2	BAR^1	
a. Cyclic ☞ [X_i ··· β ··· β ··· t_i ··· β ··· Y_i]		*	*	** *
b. Non-cyclic [X_i ··· β ··· β ··· ··· β ··· Y_i]	*!			***

The MINLINK constraint hierarchy establishes a harmony scale of chains which can be loosely summarized: ‘A chain is as weak as its longest link,’ where link length is measured in barriers. More precisely: (a) If the longest link of chain C is longer than that of chain C′, then C ≺ C′ (‘≺’ ≡ ‘less harmonic than’). (b) If the longest links of C and C′ have the same length, but C has *more* longest links, then C ≺ C′. (c) If C and C′ have equal-length longest links and the same number of them, so they *tie* w.r.t. the longest links, the decision of which chain is less harmonic is recursively determined by ignoring the longest links and comparing the chains based on the remaining links.

This evaluative structure is provided directly by OT. Universal constraint sub-hierarchies like MINLINK were developed in Prince & Smolensky 1993; their importance for UG was argued on the basis of the roles of the sonority hierarchy in syllabification and Coronal unmarkedness in segmental phonology (the latter argument was extended in Smolensky 1993). That phonological constraint hierarchies, including those involving sonority, can be derived as Power Hierarchies via Local Conjunction was shown in Smolensky 1995.

---

[13]Does a link which crosses three barriers violate BAR^2? Yes, but:

> *Theorem*. If each violation of a higher-ranked constraint logically entails a separable set of violations of lower-ranked constraints, the lower-ranked violations cannot affect the results of optimality computation.

Therefore we may ignore violations of BAR^2 by links crossing more than two barriers, and take BAR^k to be violated only by links crossing exactly *k* barriers.

Why construct a Power Hierarchy from BAR but not from the other constraints? Consider the *Local* Conjunction of other constraints with themselves: e.g., GOV(t)^2 = (a trace must be head governed)^2; this = GOV(t) because *locally*—at the same trace—there can't be multiple distinct violations of head government; the entire Power Hierarchy thus collapses to just GOV(t) itself. For *t, the same story holds: $*t^2$ = (no trace)^2 = *t; the Power Hierarchy generated by *t is just *t itself, because there can't be multiple violations of 'no trace' at a single trace.

Referentiality. More referential elements tend to be more extractable; but the notion of 'referentiality' at work here is complex. In the ultimate OT theory, a family of constraints pertaining to different aspects of referentiality must be developed, and cross-linguistic variation in referentiality effects explained by re-ranking of these constraints relative to others. (For an OT analysis of related effects, see Baković, this volume.) Here, however, we make the provisional assumption that a language makes a binary [±ref] distinction based on either:

Event structure peripherality: Argument ··· adjunct [Here: English, Chinese] (Rizzi 1990, Cinque 1990), or:

D[iscourse]-linking [In our analysis: Bulgarian; in reality, English also] (Pesetsky 1987, Cinque 1990).

Our basic constraint, then, is simply:

REF: Chains are referential.

Cross-linguistically, the interaction between REF and MINLINK has a general character nicely illustrated by Chinese:

a. Non-referential chains are good, if short
b. Long chains are good, if referential
c. Chains violating *both* MINLINK *and* REF are bad.

This chain inventory 'bans the worst of the worst' (Prince & Smolensky 1993:180): generally, this implicates Local Conjunction (Smolensky 1993), so we posit in UG:

$\textbf{MINLINK \&}_{\textbf{l}}\ \textbf{REF} \equiv \textbf{MINLINK}^{\textbf{[–ref]}}$.

This is a universal sub-hierarchy built of the following constraints[14]:

$\textbf{BAR}^{k}\ \textbf{\&}_{\textbf{l}}\ \textbf{REF} \equiv \textbf{BAR}^{k\,\textbf{[–ref]}}$: A link in a non-referential chain must not cross *k* barriers.

The universal constraint sub-hierarchy $\textbf{MINLINK}^{\textbf{[–ref]}}$ derived by $\&_l$ is thus:

$\cdots \gg \text{BAR}^{3\,[–ref]} \gg \text{BAR}^{2\,[–ref]} \gg \text{BAR}^{1\,[–ref]}$

The definition of Local Conjunction also entails the universal relative rankings:

$\text{BAR}^{k\,[–ref]} \gg \text{BAR}^{k}$

[14] Does a [–ref] chain violate the more general BAR^k constraints, in addition to $\text{BAR}^{k\,[–ref]}$? The same analysis as for BAR^3 and BAR^2 above (footnote 13) applies. The theorem given there ensures that we can ignore violations of the lower-ranked, more general constraint BAR^k which are incurred by non-referential chains, recording only violations of $\text{BAR}^{k\,[–ref]}$. Thus BAR^k can be treated as though it were $\text{BAR}^{k\,[+ref]}$.

Summary. The following schematically summarizes our account:

(7) Summary of proposed constraints on *wh*-chains

	BAR, REF		two general constraints on chains
+	Local Conjunction	+	a general, independently proposed OT mechanism
=	MINLINK $^{([-ref])}$	=	'Shortest Move'
+	*t, GOV(t), *Q	+	three general constraints on traces, operators
=	our proposed theory of *wh*-chains		

5. Core Typology of Question Strategies

5.1 Single-variable Questions

The basic typology of single *wh*-chain strategies arises from re-ranking the constraints *Q, *t, and FILL, which respectively disfavor in situ, fronted, and resumptive chains. This is discussed in Legendre et al. 1995; here, in the interest of space, we just summarize with the table in (8).[15]

(8) Basic typology of single *wh*-chain strategies

Winning Ranking	Strategy	Chain	*Q	*t	FILL
a. {*t, FILL} ≫ *Q	in situ	$[Q_i \cdots wh_i]$	*		
b. {*Q, FILL} ≫ *t	fronted	$[wh_i \cdots t_i]$		*	
c. {*Q, *t} ≫ FILL	resumptive	$[wh_i \cdots res_i]$			*

To make a central point about the role of the lexicon in cross linguistic variation, and illustrate one approach to optionality within OT,[16] we consider the distribution of resumptives in Chinese topicalization (overt movement). For our informants, the resumptive res is optional when the chain foot is governed; otherwise, res is mandatory. We treat the optionality by allowing *t and FILL to have equal ranking (i.e., they share a *stratum* in the sense of Tesar & Smolensky 1993, Tesar, this volume). Higher-ranking GOV(t) forces res in ungoverned position:

[15] In some languages (e.g. French), in situ and fronted strategies co-exist, and seem to involve two different rankings. This is analogous to a now rather standard treatment in OT phonology of different constraint domains in the lexicon.

[16] For an important approach to (apparent) optionality in OT, complementary to that of the text, see Grimshaw & Samek-Lodovici 1995, this volume, and Legendre, Raymond & Smolensky 1993: 'free variation' in outputs actually results from input differences.

(9) Chinese Optional Resumptives in Topicalization

	GOV(t)	*TOp	*t	FILL
Governed position, topicalization from t				
a. [TOp_j ··· DP_j]		*!		
b. ☞ [DP_j ··· t_j]			⊛	
c. ☞ [DP_j ··· res_j]				⊛
Ungoverned, topicalization from t position				
d. [TOp_i ··· DP_i]		*!		
e. [DP_i ··· t_i]	*!		*	
f. ☞ [DP_i ··· res_i]				⊛

This little example illustrates an important thread in our OT approach to syntax: *the functional lexicon is slave to the syntax.* Thus, whether a language has resumptives is not an arbitrary fact about its lexicon. It is a *consequence of its grammar.* Specifically, whether a language has a resumptive pronoun is *dictated* by the ranking of FILL. If this ranking is such that the grammar declares that res surfaces in optimal forms, the lexicon must provide res a phonological shape. If the grammar declares that res never surfaces, there is nothing the lexicon can do about it. (For 'grammar-determines-lexicon', see also Grimshaw forthcoming [for *do*], and Grimshaw & Samek-Lodovici, this volume and Samek-Lodovici 1995 [for *pro*, expletive subjects, and agreement morphology].)

Crucially, the constraint determining *whether* resumptives appear, FILL, is independently needed to determine *where* they appear, as illustrated in (9). That is, simply by positing that UG contains a constraint FILL (part of the FAITHFULNESS family that is an inseparable part of OT) we *simultaneously* derive:

i. the *absence* of resumptives in some languages (FILL undominated); and

ii. the *possible distributions* of resumptives in languages that have them (FILL dominated by conflicting constraints)[17]

Constraint re-ranking in the grammar is the one locus of cross-linguistic variation in

[17] Analogously, whether [wh] behaves as a 'strong feature' is *dictated* by the grammar: it depends on the ranking of *Q (relative to *t, primarily); in the absence of resumptives, for example:

i. *Q ≫ *t ⇔ [wh] strong

ii. *t ≫ *Q ⇔ [wh] weak

OT: the lexicon does not provide another independent source of variation. This is the syntactic case of OT's general theory of inventories/lexicon developed by Prince & Smolensky 1993 and applied to inventories of basic syllable types (§6); sonority-based inventories of possible onset, nucleus, coda segments (§8); and segmental inventories (§9).

5.2 Multiple Questions

The basic typology of multiple *wh*-chain strategies arising from the constraints above was also developed in Legendre et al. 1995. We simply summarize here:

(10) Basic typology of multiple *wh*-chain strategies

Winning Ranking	Strategy	Chain Structure	*ADJOIN ↓	*ABSORB	PARSE(wh) ↓
a.{*ABSORB, PARSE(wh)} ≫ *ADJOIN	adjunction Bulgarian	[$\underline{Q_i\ (Q_j}$ ··· t_i] ··· t_j)	*		
b. {*ADJOIN, PARSE(wh)} ≫ *ABSORB	absorption English	[$Q_{i[j]}$ ··· t_i] ··· $wh_{[j]}$		*	
c. {*ADJOIN, *ABSORB} ≫ PARSE(wh)	unparsing Italian	[Q_i ··· t_i]··· DP/⟨wh⟩			*

A few remarks: the notation $\underline{Q_i\ Q_j}$ in candidate *a* denotes two operators adjoined in [Spec, CP]; the ranking that gives the adjunction strategy also permits cyclic movement via adjunction: $\underline{e_i\ Q_j}$; Chinese uses adjunction under our analysis (although an absorption analysis may also be possible). A subset of this OT typology was independently proposed in Billings & Rudin 1994 (for another OT analysis of a closely related typology, see Ackema & Neeleman, this volume.)

The candidate set is essentially a Cartesian product of faithfulness and *wh*-strategy; to over-simplify somewhat, the candidates determine, for each chain:

[faithful or unparse scope or unparse *wh*?] × [*in situ* or front?] × [res or t?]

That is, for each chain in the input, there can be a corresponding chain in the output with the same scope, or different scope, or no chain at all; if there is a chain, the overt *wh*-phrase can occupy the chain foot or chain head; the other chain element can be empty (Q or t) or overt (res); and if two chains have the same scope, the two chain heads can be adjoined or one can be absorbed in the other. (Other possibilities, like a *wh*-phrase at an intermediate position—partial *wh*-fronting—await future research.)

In our English, Chinese, and Bulgarian rankings, if there is a *wh* chain, the fronting strategy is independent of any properties of the chain, so in the analysis of a particular language of this sort we can (and will) ignore those candidates that do not conform to the language's *wh* strategy: they are always sub-optimal. Whether there *is* a chain in the optimal form can however depend on the language's *wh* strategy: we cannot first determine the location of the chains, then determine which chain elements are pronounced.[18]

6. Language Study I: Chinese

We have used the constraints above to analyze in detail the extraction patterns of Chinese, English and Bulgarian. Here, we present a portion of the Chinese analysis; in Section 7 we very briefly summarize the English analysis; space does not permit discussion of Bulgarian (but see Legendre, et al. 1995, Legendre & Smolensky in progress). For Chinese, we first deduce the ranking (6.1), then indicate the coverage provided by this this ranking (6.2), and finally *encapsulate* the ranking (Prince & Smolensky 1993:§8.4) to give better insight into its predictions.

6.1 Deducing the Ranking

For starters, we have:

(11) Undominated constraints: SELECT; GOV(t)

GOV(t) is relevant to intermediate traces (responsible for CED-like effects, Huang 1982), and to the traces at the foot of topicalization chains. Next, following (8):

(12) In situ vs. fronting
  a. *wh* in situ: *t ≫ *Q
  b. fronting in topicalization: *TOp ≫ *t

(13) Resumptive pronouns
  a. res required when t is ungoverned: GOV(t) ≫ FILL
  b. res optional when t is governed: FILL ≈ *t (equal ranking)

A simplified tableau showing how these rankings work was given in (9).

For the remaining rankings, we use 'winner/loser pairs,' exploited in Tesar & Smolensky 1993, 1994, for OT ranking/learning algorithms. The first 'winner' is the (covert) subject extraction from a *think* complement (14*a*); an informative 'loser' is the competitor (14*b*) in which the *wh*-chain of the Index is not parsed. The result of

[18] E.g., for our Chinese ranking illustrated in (9), an Index with a topicalized subject results in a chain with res (9f); with FILL (and GOV(t)) undominated (a language without res) this Index results in no chain at all (a PARSE(top) violation is optimal). Whether a chain exists at all is determined by the same optimization that determines pronunciation (e.g., t vs. res).

the comparison is that PARSE(wh) $\gg$ BAR^3; the columns for these constraints are in white, while the others, irrelevant for *a* vs. *b*, are in grey. (The ranking shown is one of those consistent with the Chinese facts; the constraints in white cells at the top are ranked by the examples in (14).) 'UNSP' in candidate *b* denotes a non-quantificational, non-referring DP whose phonological shape is not important here: only the syntactically relevant features are.

In (14*c,d*), the extracted non-referential adjunct wins; the non-quantificational 'UNSP' corresponds to 'for a reason.' Thus PARSE(wh) $\gg$ BAR$^{2\ [-\mathrm{ref}]}$.

Finally (14*e,f*) shows how an LF unrealizable in Chinese (non-referential adjunct extraction from a sentential subject) is treated: the faithful parse *e* loses to unparsing of the *wh*-chain (*f*). This illustrates the general type of MINLINK/FAITHFULNESS interaction, mentioned among the goals of the analysis listed in (1d), in which an overly-long link loses to no chain at all.

At this point, PARSE(wh) has been ranked above BAR^3 and between BAR$^{3\ [-\mathrm{ref}]}$ and BAR$^{2\ [-\mathrm{ref}]}$. To rank PARSESCOPE, we return to *wh*-islands, in which, we claimed in Section 1, unrealizable LFs arise because wide scope chains in the Index are unfaithfully parsed as narrow scope chains in the output, violating PARSESCOPE.

Tableau (15) is an expanded version of the simplified tableau (3). The faithful parse of the wide-scope extraction of a non-referential adjunct (15*a*) loses to the unfaithful narrow-scope parse (15*b*); this yields the ranking BAR$^{2\ [-\mathrm{ref}]}$ $\gg$ PARSESCOPE. (Another unfaithful parse (15*c*) fails to parse the adjunct chain altogether. But we know from (14*c,d*) that the resulting violation of PARSE(wh) is worse than the BAR$^{2\ [-\mathrm{ref}]}$ violation of the faithful parse (15*a*), so this cannot be the unfaithful parse responsible for the ungrammaticality of (15*a*).) In (15*a*), $\underline{e_l\ Q_i}$ are adjoined in [Spec, CP]. The pair (15*a,b*) illustrates another general type of MINLINK competition listed among our goals in (1b): a chain (*l*-chain in *a*) loses to a shorter chain of the same element (*l*-chain in *b*) which has a different LF. (The Chinese structures *a* and *b* have the same pronunciation. Only the structure *b* can correspond to this pronunciation; only its LF is a possible interpretation of this PF.)

The competition for extraction from under *wonder* (15*a–c*) contrasts with the corresponding competition with *think* (15*a′–c′*). Because the unfaithful narrow-scope parse *b′* now violates SELECT, it can no longer win; as with *wonder*, the unfaithful parse *c′* violating PARSE(wh) is less harmonic than the faithful parse a': $c' \prec a'$. Note that the same link that *loses* because of its length (*BAR$^{2\ [-\mathrm{ref}]}$) in a *wh*-island extraction ($a \prec b$) *wins* with *think* ($a' \succ b'$).

Our analysis predicts that universally (for all rankings), *wh*-clauses are harder to extract from than *think*-type complements. The unfaithful competitor violating PARSE(wh) is equally marked in the two cases; but the competitor violating PARSESCOPE is less marked with *wonder* than with *think* since only the latter violates SELECT. Thus *wonder* extractions must beat a better competitor than *think* extractions. Reranking PARSE(wh), PARSESCOPE and even SELECT in (15) moves from languages like Chinese, with extraction from *think* but not *wonder*, to languages with extractions from both, or languages with extractions from neither; but never languages with extractions from *wonder* but not *think*. This, for an extraction of particular type (particular MINLINK violation; here, $\text{BAR}^{2\ [-\text{ref}]}$)[19]. Our analysis entails a number of such implicational universals (e.g., if have *wonder* extractions, then have *think* extractions); see the brief discussion in Section 8.6.

The difficulty of extracting *wh* over *wh* is *derived* in this theory: nowhere is there a constraint that refers to the configuration *wh*-over-*wh*.[20] Rather, we have a true competition effect driven by minimization of link length. It is not literally the configuration *wh*-over-*wh* that figures here; rather it is that the presence of the embedded *wh* is a consequence of the selectional requirements of *wonder*, which license the embedded [Spec, CP] as a *wh* 'landing' site. This licensing makes viable a better competitor to extraction: narrow scope. Note that this effect is independent of any *additional* effect that may arise in a language (which, unlike Chinese on our analysis, has *ADJOIN highly ranked) where a *wh* element blocks cyclic movement through [Spec, CP]. Note that the minimality effect derived in our analysis holds of both *wh* in situ and *wh*-fronting chains.

In the interest of space we omit the ranking arguments for PARSE(top) in Chinese, based on the topicalization facts. As usual, the Chinese facts do not suffice to determine a unique total ranking of the constraints; one ranking consistent with the facts is shown in tableau (15).

---

[19]Any BAR^k ranked such that PARSE(wh) $\gg$ BAR^k $\gg$ PARSESCOPE defines a link length *k* fatal to *wonder* extraction but possible for *think* extraction.

[20]A related but potentially more general approach to Relativized Minimality effects aimed specifically at cases usually handled with the Binding Theory is developed in Wilson 1995. Using primitives somewhat reminiscent of those in "On Binding" (Chomsky 1980), this theory derives 'Specified Subject Conditions' (SSCs) from competition along two dimensions: Locality and (inflectional) Prominence (based on Burzio 1992). Categories compete, not for the semantic role of antecedent, but for membership in an abstract (syntactic, autonomous) dependency, the main privilege of which is the ability to control deletion of another category's phi-features under referential identity (*anaphorization*). Languages differ in how they rank the two fundamental criteria: roughly speaking, those in which Locality dominates Prominence observe strict SSCs, while those in which Prominence dominates Locality allow more categories (crucially, more 'subjects') to intervene between an anaphor and a conspicuous antecedent. This OT theory of coreference 'inventories' bears a strong resemblance to the theory presented in this paper: it unifies a wide range of empirical variation (see Burzio 1992: footnote 11 and references cited) under a small set of principles; and it claims that the possibility of syntactically realizing a 'target semantic' relation of an Index depends on the non-existence of more harmonic, though less faithful, alternatives: longer (faithful) dependencies can be blocked by shorter (unfaithful) ones, even when a shorter dependency does not allow anaphorization.

6.2 Illustrative Consequences of the Ranking

A ranking deduced from the four sentences (winner/loser pairs, more precisely), of §6.1, plus a few more involving topicalization, entails the correct facts for subject, object, referential adjunct, and non-referential adjunct chains in the following:

(16) Chinese extractions correctly predicted
- a. questions out of simple clauses
- b. questions out of [–wh] complements
- c. questions out of *wh*-islands
- d. questions out of sentential subjects
- e. questions out of complex NPs
- f. questions out of adjunct clauses
- g. topicalization out of simple clauses
- h. topicalization out of [–wh] complements
- i. topicalization out of *wh*-islands
- j. topicalization out of sentential subjects

6.3 Encapsulating the Chinese Grammar

A method for analyzing complex grammars containing markedness sub-hierarchies like MinLink was introduced by Prince & Smolensky (1993:§8): *encapsulation* involves abbreviating segments of the sub-hierarchy which are contiguous in a ranking, defining a language-specific constraint which is actually just a label for a ranked group of universal constraints. We now apply this to our Chinese grammar.

Here is the MinLink/Faithfulness portion of our Chinese ranking:

$$\text{Bar}^{3\,[-\text{ref}]} \gg \text{Parse(wh)} \gg \text{Bar}^{2\,[-\text{ref}]} \gg \text{Bar}^{1\,[-\text{ref}]} \gg \text{ParseScope} \gg \text{Bar}^{3} \gg \text{Parse(top)} \gg \text{Bar}^{2} \gg \text{Bar}^{1}$$

Now we separate the MinLink, MinLink$^{[-\text{ref}]}$, and Faithfulness sub-hierarchies:

$\text{Bar}^{3\,[-\text{ref}]} \gg$ $\quad\quad$ $\text{Bar}^{2\,[-\text{ref}]} \gg \text{Bar}^{1\,[-\text{ref}]}$

$\quad\quad$ $\text{Parse(wh)} \gg$ $\quad\quad\quad\quad$ $\gg \text{ParseScope} \gg$ $\quad$ $\text{Parse(top)} \gg$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $\text{Bar}^{3} \gg$ $\quad\quad$ $\text{Bar}^{2} \gg \text{Bar}^{1}$

Next we define the encapsulated constraints:

*LM = ‘No long movement’ $\equiv [\cdots\ \text{Bar}^{4} \gg \text{Bar}^{3}]$

*SM = ‘No short movement’ $\equiv [\text{Bar}^{2} \gg \text{Bar}^{1}]$

Recalling that $\text{Bar}^{k\,[-\text{ref}]} \equiv \text{Bar}^{k}\ \&_{l}\ \text{Ref}$, these definitions entail:

$\text{*LM}\ \&_{l}\ \text{Ref}$ = ‘No long [–ref] movement’ $\equiv [\cdots\ \text{Bar}^{4\,[-\text{ref}]} \gg \text{Bar}^{3\,[-\text{ref}]}\ ]$

$\text{*SM}\ \&_{l}\ \text{Ref}$ = ‘No short [–ref] movement’ $\equiv [\ \text{Bar}^{2\,[-\text{ref}]} \gg \text{Bar}^{1\,[-\text{ref}]}\ ]$

This gives the *encapsulated ranking* (SELECT, GOV(t) are superordinate):

(17) Encapsulated Chinese

*LM $\&_l$ REF ≫ *SM $\&_l$ REF
≫
*LM ≫ *SM

PARSE(wh) PARSESCOPE PARSE(top)

∴ in *wh*-islands, –ref extractions are bad (LM and SM)
∴ in sentential subjects (LM), only –ref is bad
∴ in [–wh] complements, all extractions are good (either SM or +ref)

The encapsulated ranking displays how, in a rough sense, '*REFERENTIALITY* ⋙ *LOCALITY*' in Chinese: the contrast between long and short movement carries a smaller Harmony difference than the contrast between [–ref] and [+ref].

7. Language Study II: English

Space constraints demand that we discuss only a few points from our extended analysis of English *wh*-extractions (see Legendre et al. 1995, Legendre & Smolensky in progress). It is interesting to contrast the English ranking with the Chinese one by comparing the encapsulated rankings; for English:

(18) Encapsulated English

*LM $\&_l$ REF ≫ *SM $\&_l$ REF
≫
*LM ≫ *SM

PARSE(wh) PARSESCOPE

∴ in *wh*-islands: only good extraction is objects (+ref, GOV) from an untensed clause (SM)
∴ in islands (tensed *wh*, sentential subjects, adjunct clauses—LM): no extractions
∴ in CP complements: object, adjunct extr. good (SM, GOV);subject bad (*GOV)
∴ in IP complements: all extractions good (SM, GOV)

This encapsulated ranking is the same as that of Chinese, except *LM is higher. This has the effect that in English, '*LOCALITY* ⋙ *REFERENTIALITY*'; the long/short dimension now dominates the [±ref] dimension. In English, the referentiality cutoff is different than in Chinese: arguments are [+ref] while all adjuncts are [–ref].[21] The abbreviations used in the encapsulation for English differ slightly from Chinese:

> *LM ≡ [··· BAR^4 ≫ BAR^3], as in Chinese, but
> *SM ≡ BAR^2

In English, PARSESCOPE dominates BAR$^{1\ [-ref]}$ (but not BAR$^{2\ [-ref]}$).

The other point of interest in the English analysis which we take up here is the illustration, via simplified tableaux, of the two remaining general types of competition listed in our goals (1). The first is exemplified by super-raising:

(19) English super-raising

seems(likely(win(John)))	BAR^2	BAR^1	*t
a. ☞ 𝕀t seems ⟦ that [John$_i$ is [likely ⟦ t_i to [win		⊛	⊛
b. John$_i$ seems ⟦ that [𝕚t is [likely ⟦ t_i to [win	*!		*

Here a chain (*b*) loses to a corresponding chain (*a*) which is shorter, where the two structures have the same LF: this illustrates the simplest general type of MINLINK competition mentioned in (1a). To illustrate the final case (1c), we consider English superiority (20), where MINLINK causes a chain (j-chain in *b*) to lose to a chain involving a different DP (k-chain in *a*).

(20) English superiority

Q_j Q_k fix(he, wh$_j$; wh$_k$)	BAR^2	BAR$^{1\ [-ref]}$	*t	*ABSORB
a. ☞ how$_{k[j]}$ did [he [fix what$_{[j]}$] t_k		⊛$_k$	⊛$_k$	⊛$_k$
b. what$_{j[k]}$ did [he [fix t_j] how$_{[k]}$	*$_j$!		*$_j$	*$_j$

8. Exploring the Typology

In OT, an analysis of *wh*-chains in one language, say Chinese, is necessarily implicitly a theory of *wh*-chains in *all* languages: the factorial typology generated by re-ranking the proposed constraints in all possible ways. While the wide-ranging exploration of this typology for the constraints proposed here is a long-term project, both empirically and theoretically (computing the predictions), it seems intuitively clear that our theory

[21] Ultimately, cross-linguistic variation in referentiality must be handled by re-ranking relative to a sub-hierarchy of referentiality; see section 4.

predicts quite a broad spectrum of extraction patterms. Here, we simply sample from this spectrum, based on preliminary stages of on-going analysis of the languages mentioned.

8.1. Prediction 1: Broad orthogonality of fronting/in situ and extraction patterns

The particular constraint rankings which determine the fronting/in situ/resumptive strategy (*Q, *t, FILL) are independent of the rankings which mostly determine the extractibility patterns (MINLINK, PARSE). A rather high degree of independence is thus predicted between these two aspects of *wh*-chains. (But not complete independence, however; e.g., GOV(t)'s effects on extractability depend on whether there are traces at the foot of *wh*-chains: the fronting/in situ/resumptive strategy.)

One piece of evidence is provided by Bulgarian (Rudin, 1985, 1988), whose *wh*-strategy is exactly opposite of Chinese's: Bulgarian (SVO) fronts *all wh*-phrases. Yet the basic extraction patterns, stated in our terms, are virtually identical in the two languages; in both languages extractability is sensitive to referentiality, though what counts as referential elements differs in the two languages: in Chinese, these include arguments and some adjuncts (corresponding to *when*, *where*, instrumental *how* and purpose *why;* Tsai, 1994), while in Bulgarian they include D(iscourse)-linked arguments. (See further discussion in Legendre et al, 1995).

Our ranking for Bulgarian is the same as for Chinese, stated over the *encapsulated* constraints. But the necessary encapsulation is slightly different:

(21) a. Chinese: $*\text{LM} \equiv [\cdots \text{BAR}^4 \gg \text{BAR}^3]$ $*\text{SM} \equiv [\text{BAR}^2 \gg \text{BAR}^1]$
b. Bulgarian: $*\text{LM} \equiv [\cdots \text{BAR}^5 \gg \text{BAR}^4]$ $*\text{SM} \equiv [\text{BAR}^3 \gg \text{BAR}^2 \gg \text{BAR}^1]$

Because clausal complements are CPs in Bulgarian (obligatory complementizer) and IPs in Chinese (see Legendre et al. 1995 for evidence), superficially comparable chains differ by one barrier across the two languages.

A second piece of evidence is provided by Iraqi Arabic (SVO). Its in situ *wh* extraction pattern is almost identical to the fronting extraction pattern of English: as shown in Wahba (1991), extraction of arguments and adjuncts is grammatical out of simple clauses, but ungrammatical out of *wh*-islands and complex NPs; extraction of arguments and adjuncts out of [–wh] complements (no complementizer) is grammatical if the complement is non-tensed, ungrammatical if tensed (but the latter can be saved by affixing a Q particle to the main clause verb).

8.2. Prediction 2: Subordination of GOV(t) and reversal of subject-object asymmetry

Unlike object traces, subject traces in [Spec, IP] incur *GOV(t) (unless they are governed by a matrix verb). On the other hand, extractions from [Spec, IP] incur one less BAR violation than extractions from object position; e.g, in simple questions, the subject incurs $*\text{BAR}^1$, the object $*\text{BAR}^2$. Thus if MINLINK ≫ GOV(t), we can get the reverse of the well-known pattern: subjects are directly extractable, direct objects are not. This is indeed the case in Austronesian languages like Tagalog (VOS; Guilfoyle et al. 1992, Nakamura 1994), Bahasa Indonesia (SVO; Saddy, 1991, Guilfoyle et al. 1992), and Malagasy (VOS; Keenan, 1976), and in the Wakashan language Kwakwala (VSO; Anderson, 1984). The pattern is essentially the same in all these languages: informally, a direct object is extractable to the extent that it has undergone passivization or topicalization. Note that the pattern in question is the opposite of that predicted by the ECP, a problem for any approach based on inviolable constraints.

In (22), we present a preliminary analysis of Tagalog which assumes Guilfoyle et al.'s 1992 analysis of Tagalog subjects as occupying either [Spec, VP] or [Spec, IP], the latter when they are what is traditionally termed the 'topic' of the clause. In Tagalog, the 'topic' position may be filled by objects or obliques, given proper verb morphology identifying the theta role of the topic. Subject wh-phrases are in [Spec, CP] position, as shown by the presence of the complementizer *ang*.

(22) Tagalog simple argument extraction

	BAR^2	BAR^1	GOV(t)	*t
a. ☞ $wh_i\ [_{IP}\ V_{AT}\ [_{VP}\ t_i\ NP]\ \ t_i\ ]$		$⊛_i⊛_i$	$⊛_i$	⊛⊛
b. $wh_i\ [_{IP}\ V_{TT}\ [_{VP}\ t_i\ \ t_j\ ]\ NP_j]$	$*_i!$	$*_j$		**
c. $wh_j\ [_{IP}\ V_{AT}\ [_{VP}\ t_i\ \ t_j\ ]\ NP_i]$	$*_j!$	$*_i$		**
d. ☞ $wh_j\ [_{IP}\ V_{TT}\ [_{VP}\ NP\ t_j\ ]\ \ t_j\ ]$		$⊛_j⊛_j$	$⊛_j$	⊛⊛

In candidates *a* and *d*, the topic position ([Spec, IP]) is occupied by a t whose subscript "agrees" with the morphology on the verb: subject t_i with 'agent topic' morphology ('AT'), object t_j with 'theme topic' morphology ('TT'). In candidates *b* and *c*, the topic position is occupied by an object topic and a subject topic, respectively. Candidate *a* represents a subject extraction chain consisting of two links (wh_i, t_i, t_i) as indicated by t_i in both [Spec,VP] and [Spec, IP], incurring two violations of BAR^1; candidate *b* represents a subject extraction chain consisting of a single link (wh_i, t_i) with t_i in [Spec,VP], incurring one violation of BAR^2. Candidates *c* and *d* represent object extraction, with chain (wh_j, t_j) in candidate *c* and (wh_j, t_j, t_j) in candidate *d*.

What matters in Tagalog is the length of the links, not whether a t is governed or not: unlike Chinese, English, and Bulgarian, here GOV(t) is subordinated and surface-violated in optimal candidates *a* and *d* (by t_i and t_j in [Spec, IP]).[22] Tableau (22) predicts that long extraction in Tagalog should be severely restricted. And in fact it is, in a very interesting way: extraction of an embedded subject requires the embedded verb and the matrix verb to carry AT and TT morphology respectively, while extraction of an embedded object requires TT morphology on both matrix and embedded verbs (based on data from Nakamura, 1994). The verbal morphology reveals the successive cyclic character of the extraction (via the topic position), with matrix TT morphology corresponding to the topicalization of the entire embedded clause (bringing it closer to the landing site of the *wh* phrase). Space considerations preclude full presentation of tree structures and tableaux here; yet we are in a position to assert that in the optimal candidates for both subject and object long extraction, the relevant chains each violate BAR^2 once and BAR^1 twice. While impossible in simple questions (22), violation of BAR^2 is possible in Tagalog, in long extractions, where competitors fare worse, i.e. violate BAR^3 or violate BAR^2 more than once.

Another language in which subordination of GOV(t) is proposed is Hindi: Hindi lacks *that*-t effects despite the presence of a complementizer (Mahajan, 1990). In our terms, this shows that GOV(t) is ranked lower than PARSE(wh).

8.3. Prediction 3: Subordination of MINLINK and absence of 'subjacency effects'

Extractibility via short but not long links arises in Chinese, English and Bulgarian because MINLINK (including $\text{MINLINK}^{[-\text{ref}]}$) is interrupted by PARSE(wh) or PARSESCOPE. If MINLINK is not interrupted by PARSE constraints, locality requirements cannot block extraction; e.g., {PARSE(wh), PARSESCOPE} $\gg$ MINLINK. The result is an absence of subjacency effects, observable in the in situ strategy of Ancash Quechua and the fronting strategy of Palauan. In Ancash Quechua (SOV), covert extraction of subjects, objects, and adjuncts is grammatical out of [–wh] complements, complex NPs, and adjunct clauses (no available data on *wh*-islands) (Hermon, 1985; Cole and Hermon, 1994). Extraction out of all kinds of islands (*wh*-islands, adjunct clauses, sentential subjects, and relative clauses) is grammatical in Palauan (VOS), both in its fronting and its in situ *wh* strategies (Georgopoulos 1985, 1991). In previous accounts, these patterns have contributed to the proliferation of mechanisms for handling *wh* strategies cross-linguistically: *wh*-indexing or LF

[22] Nakamura (1994) offers a minimalist account of these facts in terms of minimizing the length of chain links, which in spirit is very similar to the account offered here. However, his account of other extractions in Tagalog such as extraction of subject and objects in the recent past (which are both allowed without topicalization) and extraction of adjuncts incorporates a number of additional assumptions which we do not need, as discussed in work in progress.

interpretation for Ancash Quechua in situ *wh* questions (Cole and Hermon, 1994) as opposed to LF *wh*-movement in Chinese, and base-generation of *wh* fronting in Palauan (Georgopoulos, 1985, 1991) as opposed to s-structure movement in English. In contrast, re-ranking of constraints like MINLINK and PARSE together with a single level of representation allows for a unified treatment of *wh* strategies and extractability.[23]

8.4. Prediction 4: Subordination of PARSESCOPE to FILL and distribution of res

Given a ranking PARSE(wh) $\gg \mathbb{C} \gg$ PARSESCOPE, a constraint $\mathbb{C}$ cannot be violated by extraction from *wh*-islands, but $\mathbb{C}$ can be violated by extraction from [–wh] complements. In Chinese and English, $\mathbb{C}$ is a MINLINK constraint, as shown in tableau (14) for Chinese. (For English, see (18) and Legendre et al. 1995). In our typology, the contraint $\mathbb{C}$ can be another faithfulness constraint: FILL. If PARSESCOPE is subordinated to FILL, extractions from *wh*-islands cannot involve a resumptive element, while extractions from [–wh] complements can. Such a pattern is found in the (SVO) Kru language Vata (Koopman and Sportiche, 1986): subjects can be extracted from a [–wh] complement if the gap is filled with a resumptive pronoun but they cannot be extracted out of a *wh*-island, with or without a resumptive pronoun. Objects can be extracted out of both contexts but forbid the resumptive pronoun. That is, subjects can only be extracted from [–wh] complements (with a resumptive pronoun in situ) while objects can be extracted both out of [–wh] and [+wh] complements (without a resumptive pronoun in situ). The difficulty for Koopman and Sportiche is to explain why a resumptive pronoun fails to save subject extractions out of [+ wh] complements, given that one offsets an ECP violation in [–wh] complements. They end up stipulating a condition on long extraction which restricts long extraction sites to theta-positions, merely reflecting the empirical generalization given above. Our OT analysis handles the Vata pattern in exactly the same fashion as other languages discussed above. As shown in tableau (23), the pattern follows from the ranking:

{GOV(t), PARSE(wh)} $\gg$ FILL $\gg$ PARSESCOPE $\gg$ *t .

(We assume high-ranked *Q: fronting.)

[23] We thank Peter Cole for clarification on Quechua.

(23) Extraction and resumptives in Vata

		GOV(t)	PARSE(wh)	FILL	PARSESCOPE	*t
Extraction from [–wh] complements (*na* = complementizer)						
subj	*a.* ☞ wh$_i$ V ⟦ *na* [res$_i$ V			⊛		
	b. wh$_i$ V ⟦ *na* [t$_i$ V	*!				*
	c. ⟨wh$_i$⟩		*!			
obj	*d.* wh$_j$ V ⟦ *na* [NP V res$_j$			*!		
	e. ☞ wh$_j$ V ⟦ *na* [NP V t$_j$					⊛
	f. ⟨wh$_j$⟩		*!			
Extraction from [+wh] complements						
subj	*a′.* wh$_i$ V ⟦ wh [res$_i$ V			*!		
	b′. wh$_i$ V ⟦ wh [t$_i$ V	*!				*
	c′. ☞ V ⟦ wh [wh$_i$ V				⊛	
obj	*d′.* wh$_j$ V ⟦ wh [NP V res$_j$			*!		
	e′. ☞ wh$_j$ V ⟦ wh [NP V t$_j$					⊛
	f′. V ⟦ wh [NP V wh$_j$				*!	

8.5. Prediction 5: Subordination of {*Q, *t} and context-dependent fronting

Important to our analysis is the claim that in situ and fronted chains compete in every language. The respective rankings *Q ≫ *t, *t ≫ *Q give fronting, in situ uniformly across a given language—if one of {*Q, *t} is undominated.

If both *Q, *t are dominated, however, the choice of fronting/in situ can be modulated by the dominating constraints. The Nilo-Saharan language Lango and the Bantu language Kikuyu provide some interesting examples of such rankings, with a slight difference between the two languages. In Lango, as described in Noonan, 1992, fronting and in situ stategies are mirror images of each other in *wh* questions and topicalizations:[24]

i. *wh*-questions: subjects are fronted; non-subjects appear in situ
ii. Topicalization: subjects are in-situ; non-subjects are fronted

According to Noonan, Lango (SVO) obeys a topic-first principle which results in having old information (topics) in initial position; subject *wh*-phrases which represent new information cannot simply be placed in initial position: they appear focussed via relativization or clefting. We assume that Top is the highest IP position ([Spec, IP] or adjoined to IP) and posit that NPs are marked with features [foc], [top] in the Index, on a par with scope and [wh] features; we also assume that elements with [wh] are also [foc]. It is natural to expect structural constraints on foci and topics of the following sort:

(24) *[foc]-in-Top: foci are not in Top position
[top]-in-Top : topics are in Top position

The ranking {*[foc]-in-Top, [top]-in-Top} ≫ *t ≫*Q governing fronting yields the intriguing Lango pattern of *wh*-questions and topicalizations, as shown in tableau (25); since *t ≫*Q, the default strategy is in situ, but since these constraints are dominated by the constraints on [foc] and [top], fronting can be forced:

(25) Modulation of fronting/in situ by topic and focus in Lango

	Index	Output	*[foc]-in-Top	[top]-in-Top	*t	*Q
object	x_j [wh] [foc]	*a.* ☞ x_j in situ				⊛
		b. x_j in [Spec, CP]			*!	
	x_j [top]	*c.* x_j in situ		*!		
		d. ☞ x_j adjoined to IP			⊛	
subject	x_i [wh] [foc]	*a′.* x_i in situ	*!			*
		b′. ☞ x_i in [Spec, CP]			⊛	
	x_i [top]	*c′.* ☞ x_i in situ				
		d′. x_i in [Spec, CP]		*!	*	

[24] We thank Bill Raymond for uncovering Lango as one language which meets our prediction.

The optimal output of an object extraction is the in situ strategy (candidate *a*) despite a violation of *Q because moving the object (*b*) violates the higher ranked constraint *t: this is the default case. The reverse strategy operates in object topicalizations because fronting the topic (*d*) encurs a less costly violation than leaving it in situ (*c*). Subject extraction (candidates *a′* and *b′*) is the mirror image of object extraction because leaving a subject *wh*-phrase in situ violates the structural constraint on foci. Finally, subject topicalization (*c′* and *d′*) is the mirror image of object topicalization because subjects are in [Spec, IP] and hence are interpreted as topics by default (*c′*). Thus the complex pattern in Lango is the result of the interaction of violable constraints on foci, topics, economy of movement, and scope.[25]

Kikuyu (SVO) offers the added complication of optionality: the in situ vs. fronting alternation in *wh*-questions is essentially the same as in Lango, except that non-subjects can optionally be fronted as well as appear in situ (Clements, Mailing, and Zaenen, 1983). Only a minor change in the ranking proposed above for Lango needs to be made: a tie among the constraints *t and *Q will result in both strategies for extractions of non-subjects only, because only these two constraints are relevant to non-subject extraction (see candidates *a* and *b* in the Lango tableau). Crucially, this tie does not affect topicalizations and subject extractions, because the optimal outputs are determined by the other two constraints (ranked higher than those tied).

8.6 General Restrictive Predictions: Implicational Universals

Sections 8.1–8.5 illustrate some of the range of variation in extraction patterns predicted by our account. On the reverse side, the account also *restricts* cross-linguistic variation in many ways. One way of stating these limits is as implicational universals like the one briefly mentioned in Section 6.1: if a language permits extractions (crossing *k* barriers) from [+wh] complements, then it permits extractions (crossing *k* barriers) from [–wh] complements; other examples are: if a language permits extraction of a [–ref] element in some environment, then it permits extraction of a [+ref] element in that environment; if a language having both *wh* in situ and *wh* fronting allows subject *wh* fronting, then it allows subject *wh* in situ (but the corresponding universal for adjuncts does not hold). A number of such universals are theorems entailed by our account; their theoretical scope and empirical adequacy are the subject of current research.

[25] While it was suppressed in Section 8.2 for simplicity, we assume that in Tagalog too NPs are marked with [top] features in the Index. The analysis of §8.2 goes through because the [top]-in-Top constraint is dominated in Tagalog by MINLINK: it is active, e.g., in simple clauses, where it determines which argument occupies [Spec, IP], since MINLINK does not decide the matter. MINLINK does decide for *wh*-extractions, as discussed.

9. Concluding remarks

9.1 Summary

We have argued that general techniques from Optimality Theory for formally managing the computation and analysis of grammatical competition (especially, markedness sub-hierarchies, local conjunction, and constraint encapsulation) offer a natural formalization of 'Shortest Move,' MINLINK, which is flexible enough to predict, through re-ranking, much cross-linguistic diversity in extraction patterns, yet strong enough to determine optimality, with formal rigor, in competitions between: different chains of a single element in structures with the same LF; different chains of a single element in structures with different LFs; chains of different elements; and chains competing against no chain at all.

This analysis is framed within a general approach to syntax in OT which solves the fundamental problems of Language-Specific Ineffability and Unmarked Outputs by taking syntactic competition to occur among structures weighing the (violable) demands of structural well-formedness against (violable) faithfulness to a target 'Index' consisting of predicate/argument, operator/variable structure. Each competing syntactic structure is a unified representation combining d-structure, LF, and pronunciation information, a structure which simultaneously serves as the interface to both semantic and morphophonological interpretation, a structure the grammaticality of which is determined by a single optimization over well-formedness constraints evaluating syntactic, LF, and pronunciation structure.

It is the interleaving within a language's constraint hierarchy of the constraints in the MINLINK and FAITHFULNESS families that determines its question formation strategy and extraction pattern. Relativized minimality effects arise through the interaction of the MINLINK and FAITHFULNESS constraint families, with no individual constraint referring to either 'relativization' or 'minimality'.

Another central principle of the general approach pursued here is that the sole source of cross-linguistic variation is constraint re-ranking: all differences in the syntactic-structure inventory and functional lexicon are derived from this. We illustrated this in the case of resumptive pronouns: the relative ranking of a FAITHFULNESSconstraint determines whether a language has resumptives, and if so, their distribution. Thus the functional lexicon is determined by the grammar.

9.2 On the relation to the Minimalist Program

In support of the notion that there are meaningful connections between OT approaches to syntax and the Minimalist Program (MP), a notion represented in part by the conference from which this volume derives, we close with a few very general and somewhat speculative observations regarding some of those connections.

While some see a major divide between the derivationally-oriented MP and OT, we do not. Of course, there are likely to be differences of empirical import between the non-derivational, chain-based theory of 'Shortest Move' developed here and a particular derivational MP proposal, but such differences seem comparable to those between different approaches to syntax within OT, or to those between different proposals within MP: they do not seem to follow from some major divide between the OT and MP frameworks. In fact, derivational theories can be naturally formalized within OT, in at least two ways. 'Harmonic serialism' is a derivational version of OT developed in Prince & Smolensky 1993 (pp. 15, 20, 79–80) in which each step of the derivation produces the optimal next representation. Another approach, seemingly needed to formalize MP within OT, has *Gen* produce *derivations;* it is these that are evaluated by the constraints, the optimal derivation being determined via standard OT evaluation. Thus, on our view, while the issue of derivations is an important one, it is largely orthogonal to OT: any evidence there may be *for* derivations is not ipso facto evidence *against* OT.

Another such issue is the feature-checking theory of movement embodied in MP. Any evidence there may be for the feature-checking theory is also evidence for OT theories with constraints governing chains (or movement) such as:

CHECK(F): Spec/Hd F-agreement with overt chain element ≻
Spec/Hd F-agreement with empty chain element ≻
no F-agreement

That is, agreement with an overt element (overt movement to check F) is best; agreement with an empty chain element like our 'Q' (covert movement) is next best; no chain (movement) is worst. (CHECK(wh) is essentially our '*Q'.) CHECK(F) conflicts with *t, essentially PROCRASTINATE. When CHECK(F) ≫ *t, F is a strong feature, and we get overt movement to check F; when *t ≫ CHECK(F), F is a weak feature and we get covert movement. Cross-linguistic variation arises from re-ranking the CHECK(F) features (this is also observed in Grimshaw forthcoming and Legendre et al. 1995). To analyze a language with a ranking in which

$\{$CHECK(F_1), CHECK(F_2), …, CHECK(F_n)$\} \gg$ *t

we can encapsulate the constraints $\{$CHECK(F_1), CHECK(F_2), …, CHECK(F_n)$\}$ into a constraint CHECKSTRONGF whose content is defined by the language-specific set of strong features $\{F_1, F_2, \ldots, F_n\}$.

So neither the use of derivations nor the feature-checking theory of movement conflict with the commitments of OT. Those commitments are: (a) grammatical = optimal; (b) optimal = most harmonic; (c) Harmony is computed from ranked violable constraints as formally specified in OT; (d) cross-linguistic variation = alternative rankings of universal constraints (only: no functional parameters in the lexicon). None of these commitments is obviously inconsistent with MP—even (d), in light of the preceding observation that the strong/weak distinction can be naturally formalized in OT in terms of whether a constraint CHECK(F) is literally stronger or weaker than a constraint amounting to PROCRASTINATE. And as we have argued above, OT's Harmony calculation affords formalizations of economy principles such as 'Shortest Link' which are natural, explanatory, and empirically powerful. Further explanatory purchase arises from violable versions of non-economy constraints such as the ECP, and from Faithfulness, integral to Optimality Theoretic competition.

Obviously these remarks barely begin to scratch the surface of the relationship between MP and OT. They already suggest, however, that this relationship is potentially quite interesting and fruitful to pursue in depth. In our view, the OT approach developed here offers several attractive 'minimizing' features for linguistic theory. For example, it employs a single, unitary syntactic representation that serves both the phonological and interpretative interface. And it opens the door to a grammatical theory in which a single theoretical framework serves both the phonological and syntactic components of grammar. As we have occasionally illustrated in this paper, OT phonology and OT syntax have enjoyed considerable cross-fertilization, and they are now jointly contributing to the development of Optimality Theory's characterization of UG.

References

Ackema, P. and A. Neeleman. This volume. "Optimal Questions."

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, 537–577.

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

Bošković, E. J. This volume. "Optimality and Inversion in Spanish."

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." In *Proceedings of the Third Symposium on Formal Approaches to Slavic Linguistics.*

Burzio, L. 1992. "The Role of the Antecedent in Anaphoric Relations." In R. Freidin, ed., *Current Issues in Comparative Grammar*, Kluwer.

Burzio, L. 1994. *Principles of English Stress.* Cambridge University Press.

Chomsky, N. 1980. "On Binding" *Linguistic Inquiry* 11, 1–46.

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*, 417–454.

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

Chomsky, N. 1995. *The Minimalist Program*. MIT Press.

Clements, G. N., J. McCloskey, J. Maling, and A. Zaenen. 1983. "String-Vacuous Rule Application." *Linguistic Inquiry* 14, 1–18.

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

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

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

Doherty, C. 1993. *Clauses Without that: The Case for Bare Sentential Complementation in English*. Doctoral Dissertation, UC Santa Cruz.

Fox, D. This volume. “Locality in Pronoun Binding.”

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*. Kluwer.

Grimshaw, J. 1993. “Minimal Projection, Heads, and Optimality.” Technical Report, Rutgers Center for Cognitive Science; Rutgers Optimality Archive.

Grimshaw, J. forthcoming. “Projection, Heads, and Optimality.” Revision forthcoming in *Linguistic Inquiry*.

Grimshaw, J & V. Samek-Lodovici. 1995. “Optimal Subjects.” In J. Beckman, S. Urbanczyk, and L. Walsh, eds., *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, GLSA, University of Massachusetts, 589–605.

Grimshaw, J & V. Samek-Lodovici. This volume. “Optimal Subjects and Subject Universals.”

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.

Hermon, G. 1985. *Syntactic Modularity*. Foris.

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

Keenan, E. 1976. “Remarkable Subjects in Malagasy.” In Ch. Li, ed., *Subject and Topic*, Academic Press, 247–301.

Koopman, H., and D. Sportiche. 1986. “A note on Long Extraction in Vata and the ECP.” *Natural Language and Linguistic Theory 4*, 357-374.

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

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

Legendre, G., W. Raymond, P. Smolensky. 1993. “An Optimality-Theoretic Typology of Case and Grammatical Voice Systems.” *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA. February. 464–478.

Legendre, G. and P. Smolensky. In progress. *Optimality Theoretic Syntax: The Case of Wh-Chains*.

Legendre, G., C. Wilson, P. Smolensky, K. Homer, and W. Raymond. 1995. “Optimality in *Wh*-Chains.” In J. Beckman, S. Urbanczyk, and L. Walsh, eds., *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, GLSA, University of Massachusetts, 607–636.

Mahajan, Anoop K.. 1990. *The A/A-Bar Distinction and Movement Theory*. Doctoral Dissertation, MIT. Distributed by MIT Working Papers in Linguistics.

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

Nakamura, Masanori. 1994. “Extraction in Tagalog and Economy of Derivation.” McGill University Manuscript.

Noonan, Michael. 1992. *A Grammar of Lango*. Mouton.

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

Pesetsky, D. This volume. “Optimality Principles of Sentence Pronunciation.”

Prince, A. and P. Smolensky. 1991. "Notes on Connectionism and Harmony Theory in Linguistics." Technical Report, Computer Science Department, University of Colorado at Boulder. [Course notes, LSA Linguistic Institute, UC Santa Cruz.]

Prince, A. and P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report , Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, and Computer Science Department, University of Colorado at Boulder.

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 L.L.S. Cheng and H. Demirdache, eds., *More Papers on Wh-Movement*. MIT Working Papers in Linguistics, Vol. 15.

Samek-Lodovici, V. 1995. *Constraints on Subjects: An Optimality Theoretic Analysis*. Doctoral Dissertation, Rutgers University.

Smolensky, P. 1993. "Harmony, Markedness, and Phonological Activity." Talk presented at the Rutgers Optimality Workshop–1, October 23; handout: Rutgers Optimality Archive.

Smolensky, P. 1995. "On the Internal Structure of *Con*, the Constraint Component of UG." Talk presented at UCLA, April 7; handout: Rutgers Optimality Archive.

Tesar, B. B. 1994. "Parsing in Optimality Theory: A Dynamic Programming Approach." Technical Report, Computer Science Department, University of Colorado at Boulder.

Tesar, B. B. 1995a. "Computing Optimal Forms in Optimality Theory: Basic Syllabification." Technical Report, Computer Science Department, University of Colorado at Boulder.

Tesar, B. B. 1995b. *Computational Optimality Theory*. Doctoral Dissertation, University of Colorado at Boulder.

Tesar, B. B. This volume. "Error-driven Learning in Optimality Theory via the Efficient Computation of Optimal Forms."

Tesar, B. and Smolensky, P. 1993. "The Learnability of Optimality Theory: An Algorithm and Some Basic Complexity Results." Technical Report, Computer Science Department, University of Colorado at Boulder. Revision forthcoming in *Linguistic Inquiry*.

Tesar, B. and Smolensky, P. 1994. "The Learnability of Optimality Theory." *Proceedings of the West Coast Conference on Formal Linguistics XIII*. 122–137.

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

Wahba, W.A.B. 1991. "LF Movement in Iraqi Arabic." In Huang and May, eds., *Logical Structure and Linguistic Structure*. Kluwer.

Wilson, C. 1995. "Optimality-Theoretic Constraints on Pronouns and Reflexive Predicates." Ms., Johns Hopkins.

Last revision: Dec. 30, 1995 Formatted for ROA: Mar. 10, 1996

(14) Deriving the Chinese ranking

	MinLink[–ref]					MinLink			
	Bar[–ref] 3	Parse (wh)	Bar[–ref] 2	Bar[–ref] 1	Parse Scope	Bar 3	Parse (top)	Bar 2	Bar 1
Q_j *think* (s, V(s′, x_j)) Q out of complement of *think*$_{IP}$, direct object (✓)									
a. ☞ Q_j [$_{IP}$ [$_{VP}$ V ⟦$_{IP}$ [$_{VP}$ V wh$_j$ *Q_j you think Z. saw what$_j$*						⊛			
b. [[V ⟦ [V DP/⟨wh⟩ *you think Z. saw UNSP*		*!							
Q_l *think* (s, V(s′, o′; x_l)) Q out of complement of *think*$_{IP}$, adjunct, –ref (manner *how*, reason *why*) (✓)									
c. ☞ Q_l [$_{IP}$ [$_{VP}$ V ⟦$_{IP}$ wh$_l$ V *Q_l you think Z.why left*			⊛						
d. [[V ⟦ XP/⟨wh⟩ V *you think Z. UNSP left*		*!							
Q_l V (V(s′, o′; x_l), o) Q out of sentential subjects, adjunct, –ref (*)									
e. Q_l [$_{IP}$ [$_{CP}$ [$_{IP}$ wh$_l$ V ... *she why left upset him*	*!								
f. ☞ [[[XP⟨wh⟩ V ... *she UNSP left upset him*		⊛							

(15) Relativized Minimality effect in *wh*-island extraction: Chinese Q out of *wonder*- vs. *think*-complement

		MINLINK$^{[-ref]}$					MINLINK			
	SELECT	BAR$^{[-ref]}$ 3	PARSE (wh)	BAR$^{[-ref]}$ 2	BAR$^{[-ref]}$ 1	PARSE Scope	BAR 3	PARSE (top)	BAR 2	BAR 1
Q_l *wonder* (s, Q_i V(x_i ; x_l))	**–ref adjunct over subject (*)**									
a. **Q_l** [$_{IP}$ [$_{VP}$ V ⟦$_{CP}$ **e_l** Q_i [$_{IP}$ wh_i **wh_l** V ***Q_l*** *you wonder Q_i who_i* ***why_l*** *left*				$*_l$!	$*_l$					$*_i$
b. ☞ V ⟦ **Q_l** Q_i [wh_i **wh_l** V *you wonder* ***Q_l*** *Q_i who_i* ***why_l*** *left*					$⊛_l$	$⊛_l$				$⊛_i$
c. V ⟦ Q_i [wh_i **XP**/⟨wh⟩ V *you wonder Q_i who_i UNSP left*			*!							$*_i$
Q_l *think* (s, V(s′, o′; x_l))	**–ref adjunct (✓)**									
a′. ☞ **Q_l** [$_{IP}$ [$_{VP}$ V ⟦$_{IP}$ **wh_l** V ***Q_l*** *you think Z.**why_l** left*				$⊛_l$						
b′. V ⟦ **Q_l** [Z. **wh_l** V *you think* ***Q_l*** *Z.* ***why_l*** *left*	*!				$⊛_l$	$⊛_l$				
c′. [[V ⟦ **XP**/⟨wh⟩ V *you think Z. UNSP left*			$*_l$!							