2010. In C. Rice & S. Blaho (Eds.) *Modeling Ungrammaticality in Optimality Theory*. Equinox. 237-266.

# THE NEUTRALIZATION APPROACH TO INEFFABILITY IN SYNTAX

**Géraldine Legendre**
**Johns Hopkins University**

## 1. The nature of the debate

Ungrammaticality in syntax comes in two basic flavors which are illustrated in contexts of wh-extraction. In cases of ungrammatical extraction out of a complement clause in (1)-(2), a minor structural repair (at least on the surface) produces a better alternative without altering its interpretation. For example, the complementizer *that* must be dropped when the subject of the complement clause is extracted in English; in a similar context in French the complementizer *que* [kə] changes to *qui* [ki]. As shown in (3a-b), the counterparts to (1a-2a) are grammatical in Italian and Bulgarian. Patterns of ungrammaticality are language-dependent, and so are the repair strategies.[1]

(1) a. *[$_{CP}$Who$_i$ do [$_{IP}$you believe [$_{CP}$that [$_{IP}$t$_i$ came]]]]?
b. [$_{CP}$Who$_i$ do [$_{IP}$you believe [$_{IP}$t$_i$ came]]]?

(2) Fr a. *[$_{CP}$ Qui$_i$ crois$_j$ [$_{IP}$-tu t$_j$ [$_{CP}$ qu(e) [$_{IP}$ t$_i$ est venu]]]]?
b. [$_{CP}$ Qui$_i$ crois$_j$ [$_{IP}$-tu t$_j$ [$_{CP}$ **qui** [$_{IP}$ t$_i$ est venu]]]]?
'Who do you believe that came'

(3) It a. Chi$_i$ credi che t$_i$ sia venuto?
Bu b. Koj$_i$ misliš če t$_i$ e došul ?
Who believe-2sg that is come
'Who do you believe that came?'

In other cases, ungrammaticality cannot be repaired within the same construction because the grammar of a particular language does not provide a better alternative via a minor structural repair. This is known as absolute ungrammaticality or **ineffability**. A well-known case is that of multiple wh-questions in some languages (another is passive in Hungarian and other languages). Compare single wh-questions in (4) with multiple questions in (5). Both types of wh-questions are subject to requirements on the position and argumental status of wh-phrases. In English single wh-questions, the wh-phrase must appear in clause-initial position, regardless of its argumental status. In multiple wh-questions there is a Superiority effect whereby the highest available wh-phrase on the argument-adjunct scale (*who* > *what* > *where, when* > *how* > *why*) must appear in clause-initial position; the other wh-phrase must remain in situ (its d-structure position). Multiple wh-questions like (5a) have a multiple-pair interpretation. Adequate answers are of the form: "John ate pizza, Mary ice cream, etc."

(4) a. Who came?
b. What did John eat?
c. Why did John come?

(5) a. Who ate what?
b. *What did who eat?

Ineffability arises in English when *who* is paired with adjunct wh-phrases *why* or *how,* regardless of which wh-phrase is fronted (6a-b). In other languages requiring wh-fronting in single wh-questions, including Italian for some speakers (Calabrese 1984), multiple wh-questions are all ungrammatical, regardless of the position or argumental status of the wh-phrases involved.

(6)En a. *Who came why?
b. *Why did who come?

It c. *Chi ha mangiato che cosa?
Who has eaten which thing?
'Who ate what?'

For Pesetsky (1997) the existence of ineffability is strong evidence that syntax does not operate on the basis of an optimality-theoretic architecture. Rather, ineffability supports the "clash and crash" model embodied in Principles & Parameters/Minimalist theories. On that view, OT constrains spelling out the abstract syntactic structure only, i.e. its pronunciation.

Pesetsky's conclusion is not the only logical and valid conclusion however. Independent evidence for an OT model of syntax comes from the undeniable existence of significant cross-linguistic variation and economy-based generalizations (which require additional formal constructs in Principles & Parameters or Minimalist models of syntax). Moreover, ineffability is merely one aspect of the mismatch between interpretation and form which manifests itself elsewhere as ambiguity, optionality, etc. (Beaver & Lee 2004). Within OT, the question of ineffability reduces to the characterization of the optimal output in an optimization that does not obviously yield a grammatical output.

Various approaches to handling ineffability in syntax have been developed in the OT literature. In Standard OT (Prince & Smolensky [1993]/2004) it is clear that language-particular absolute ungrammaticality cannot be located in *Gen,* the component which precedes *H-eval*; if so, it would predict universal ineffability. This leaves two options. One is to locate absolute ungrammaticality in a component (Interpretation/LF) that follows *H-eval.* On this view, all members of the candidate set share the same LF and the optimal candidate output of a syntactic optimization is simply uninterpretable (Grimshaw 1997). Note however that all competitors have a valid interpretation under LF equivalency. To say that the Italian winner [who ate what?] crashes at interpretation, while its English counterpart does not, entails that the competitors are not all interpretively equivalent after all.

The alternative is that absolute ungrammaticality is in fact located in *H-eval.* Whether this is viewed as problematic or not depends on one's view of the role of Input-Ouput faithfulness in syntax. For example, Heck et al. (2002) consider that syntax is an information-preserving system with richly structured output candidates, whereas phonology is a system that loses information, so that reference to an underlying input is necessary in phonological constraints. They argue that I-O faithfulness constraints in syntax can easily be reformulated as output constraints provided output representations are enriched (this begs the question of whether the output-oriented approach is just a notational variant of the I-O faithfulness approach). Note also that the very existence of wh- or scope ineffability is a direct challenge to the view that syntax is an information-preserving system.

Proponents of I-O faithfulness in syntax however disagree on how to handle ineffability in standard OT. Based on the original proposal by Prince & Smolensky ([1993]/2004), Ackema & Neeleman (2000) propose a **null parse** approach: The optimal candidate is empty (it has no syntactic structure): "Sometimes it's better to say nothing". Ackema and Neeleman assume that the null parse candidate is not fed into the interpretational component (because it has no syntactic structure at all). Since it has no interpretation they further assume that the null parse candidate doesn't violate the requirement that all candidates be semantically equivalent.

(7) The Null Parse approach

| Italian | Input | Optimal Ouput |
|---|---|---|
| | Q (single wh-question) | Q |
| | QQ (multiple wh-question) | Ø |

By definition the null parse candidate does not contain any of the information contained in the input. Hence, it does not minimally violate I-O faithfulness, contrary to the axiom that OT constraints are violable but **minimally** so. Furthermore, the null parse approach incorporates significant redundancy in the form of relying on both an LF specification in the input and a separate evaluation of the optimal output to check it. Moreover, underparsing is still needed elsewhere in syntax where semantics are not affected (e.g. pro-drop). Finally, for the cases of ineffability at hand, the null parse approach has proven unworkable for a very general reason discussed in Section 3.3 below.

Legendre et al. (1995, 1998) introduce an alternative **neutralization** approach to ineffability, whereby different inputs (interpretations) neutralize to one and the same optimal output because specific input features ([wh]; operator scope) may be underparsed. The optimal candidate is close to the input interpretation but not identical: "Sometimes it's best to say something else".

(8) The Neutralization approach

| Italian | Input | Optimal output |
|---|---|---|
| | Q (single wh-question) | Q |
| | QQ (multiple wh-question) | Q |

The neutralization approach differs from the null parse approach in a number of ways. First, the neutralization approach rests on minimal violations of I-O faithfulness; unlike the null parse approach it does not require relaxing a defining property of OT. Second, all candidates are syntactic structures combined with an LF interpretation in the neutralization approach. In contrast, the null parse approach relies on an odd type of candidate (no structure, no interpretation). Third, the neutralization approach is economical: the input contains a target interpretation; the candidates all have a structure and an interpretation; faithfulness constraints are needed elsewhere in syntax to handle, for example, word order tied to discourse status and expletives. There is no additional interpretational component to operate on the output of the syntax.

Both approaches ultimately rely on two alternative conceptions of the candidate set. Ackema & Neeleman (2000) assume that the candidate set may only include candidates that share the same interpretation plus the null parse candidate which doesn't. Legendre et al. (1998) hold that the candidate set includes candidate outputs that do not have the same interpretation although they share **the same target interpretation**. In other words, the interpretation of a candidate may deviate from, while remaining as close as possible to, the intended interpretation specified in the input. This relaxing of an LF-equivalency requirement on the candidate set is often held against the neutralization approach. However, a fair comparison must include all other relevant aspects of the proposal, including the fact that the neutralization approach is a very economical solution to the problem of ineffability, which the null parse approach is not. The neutralization approach is explored in detail in Section 2.[2]

Any solution to ineffability is concerned with the nature of the input to syntax and how much structure in the input is desirable. On the one hand, there is general agreement that the input to syntax must at least include argument structure specification in addition to lexical items. On the other hand, some object to the treatment of ineffability in multiple wh-questions in terms of I-O faithfulness on the basis that it requires positing additional structure in the input, including operator scope relations. For Legendre et al. (1998) this is no different from including discourse features like [contrastive focus], [topic], etc. to account for so-called optionality in word order (e.g. Choi 1996; Costa 2001; Legendre 2001a,b; Samek-Lodovici 2001).

More generally, the I-O faithfulness approach advocated here on the basis of Legendre et al. (1995, 1998, 2006) takes the view that the main function of the input in OT syntax is to define the competition in the context of an **inventory** view of the grammar familiar from pre-existing constraint-based approaches to syntax as well as the Basic CV Syllable Structure Theory of Prince & Smolensky ([1993]/2004). The inventory view implies that the main question to ask is "What is the inventory of all possible syllable shapes or questions in a given language, deduced by considering all possible inputs?" rather than "What is the input-output mapping, given a particular input ?" characteristic of some phonology theorizing (Prince & Smolensky [1993]/2004, Ch. 9). On the inventory view, the input to syntax is more appropriately characterized as the **Index** in the mathematical sense of the term, in which each member of a collection is indexed or uniquely labeled by a member of an 'index set'. A member of the collection in question is a particular candidate set, and its Index uniquely specifies it. Henceforth the term Input/Index is used to reflect the mathematical property in question.

The remaining discussion is structured as follows: Section 2 argues for a neutralization approach to ineffability in multiple wh-questions. Section 3 develops an argument against the null parse approach grounded in the need to recognize multiple I-O faithfulness constraints to handle contrasts in wh-extraction out of wh-clauses and *that*-clauses in Chinese and English. Section 4 presents additional evidence for the neutralization approach that is independent of wh-extraction. Section 5 addresses the question of the surface realization of unparsed input features. Section 6 concludes the discussion of neutralization and faithfulness from the perspective of the general architecture of OT.

## 2. The neutralization approach

### 2.1. Cross-linguistic variation in multiple wh-questions

In many languages it is possible to extract – overtly (Bulgarian), covertly (Chinese), or a mixture of both (English) – two or more wh-phrases, subject only to positional requirements associated with overt vs. covert extraction and possible restrictions on which wh-phrases may appear first, second, etc.; see (9). The latter superiority effects need not concern us here.

(9) Bu a. ***Koj kakvo na kogo*** e da ? (Rudin 1988)
who what to whom has given
"Who gave what to whom?"

Ch b. Lisi zhidao ***shenme shenme shihou***?
Lisi know what what time
'What did Lisi know when?'

En c. ***What*** did Congress know ***when***?

However, multiple wh-questions are not universally grammatical. In Irish (McCloskey 1979), Quiegolani Zapotec (Black 2000), and Italian for some speakers (Calabrese 1984), all combinations of argument and adjunct wh-phrases are ungrammatical regardless of wh-phrase order. That is, these languages permit only the fronting of one wh-phrase per question. Inputs/indexes with multiple wh are simply ineffable.

(10) It a. *Chi ha mangiato che cosa?
Who ate which thing?
'Who ate what?'

Ir b. *Cé aL rinne ciadé?
who that did what
'Who did what?'

Za c. *Pa go r-laa de lo txu?
what thing do you face who
'What are you doing to who?'

The patterns in (10) have received a lot of attention in OT syntax for the obvious reason that every optimization in standard OT must yield an optimal output which by assumption is grammatical. What are then the grammatical counterparts to (10a-c) in languages with ineffable multiple wh inputs/indexes? More generally, what consequences does the existence of ineffability have for extending OT to syntax?

**2.2. Theoretical constructs**

The approach to ineffability developed in Legendre et al. (1995, 1998) starts with a structured Input/Index which includes syntactic categories, clausal boundaries, predicate-argument structure plus the **target interpretation** (e.g. [+wh]), thus setting up the prerequisite for Input-Output Faithfulness constraints on interpretation to operate in *H-eval*.

The universal Input/Index to a wh-question depicted in (11) incorporates standard assumptions about the semantics of wh-questions as operator-variable constructions with an abstract operator Q marking scope and a variable x that it binds (May 1985).

(11) Universal Input/Index for questioning a direct object out of a simple clause: [$Q_j$ [...$x_j$...]]

*Gen* generates a universal set of candidates with all relevant brackets (in accordance with standard X'-theory) and is also responsible for placing Q in highest Spec position, typically SpecCP. *Gen* marks as overt Q or x (or both). A candidate with an overt Q looks like English *What did he say t?* where the wh-phrase is fronted to clause-initial position; x is the trace of the overtly moved wh-phrase. A candidate with an overt x looks like an English echo question *He said what*?[3] This in-situ strategy is used in Chinese to express standard information questions, as shown in (9b) above.

*Gen* also generates candidates that fail to parse some element of the input (e.g. the [wh] feature). An unfaithful parse like (12c) is not interpreted as a question but rather a statement with a [-wh] or unspecified DP in lieu of a [+wh] DP. See further discussion in Section 2.3.

(12) Candidate set for a single Q in the input

a. [$Q_j$ [...$wh_j$...]] faithful parse
b. [$wh_j$ [...$t_j$...]] faithful parse
c. +$Q_j$, [...DP/+$wh_j$ ,...] unfaithful parse

Inputs to multiple wh involve 2 Qs which receive a pair-list interpretation. However, English allows only one wh-phrase to be fronted, the other remains in situ; see (9c). Following Higginbotham & May (1981) English involves a process of **absorption** whereby two wh-operators convert into a single operator in SpecCP marking the scope of two variables ($wh_{i[j]}$; candidate c in tableaux T1-T5). This is reinterpreted as a violation of *ABSORB. In contrast, Bulgarian allows multiple wh-phrases to be fronted because it tolerates violations of *ADJOIN thereby allowing two operators to adjoin to SpecCP (Rudin 1988): $wh_i$+$wh_j$; candidate b in T2 below. See (9a). Chinese (9b) uses covert adjunction ($Q_i$+$Q_k$) under the analysis proposed in Legendre et al. (1998) (although an absorption analysis may also be possible).

Summing up, candidates of interest involve two Qs universally, which may be realized as

(13) Best candidates:

a) Both wh-phrases in situ: Chinese (adjunction of empty Qs in SpecCP)
b) Both wh-phrases fronted: Bulgarian (adjunction of overt wh-phrases in SpecCP)
c) Only one wh-phrase fronted, the other in situ: English (absorption of two variables by one overt wh-phrase)
d) One fronted, the other unparsed: Italian, Irish, Zapotec
e) One in situ, the other unparsed: see discussion in section 2.3.

*Con* incorporates a general constraint on economy of movement (*t, equivalent to Grimshaw's STAY) and a number of constraints which build on various technical proposals arising from extensive GB studies of wh-questions in the 1980's and 90's.

(14) Constraints:

| | | |
|---|---|---|
| *t | "No traces" | (general economy of movement) |
| *Q | "No empty Q-operators" | (forces wh-phrases to front) |
| *ABSORB | "No absorption of Q-operators" | (only relevant if wh-phrase is fronted; penalizes the absence of a 1-1 correspondence between wh-operators and variables) |
| *ADJOIN | "No adjunction of Q-operators" | (violated by two fronted wh-phrases) |
| PARSEQ | "[Q] feature must be parsed" | (violated by unfaithful candidates) |

**2.3. A typology of multiple wh-questions**

I adopt the following convention initiated in Legendre et al. (1995): A subject is identified by subscript *i*, a direct object by subscript *j*, a referential adjunct (e.g. *when, where*) by subscript *k*, and a non-referential adjunct (e.g. *how, why*) by subscript *l*. Following Rizzi (1990) **referentiality** is understood as a short name for the universal scale ranging from arguments to adjuncts with core arguments (subjects and direct objects) being most argumental and *how*, *why* least argumental. While respecting this scale, languages typically impose a cut-off point somewhere along the scale which establishes a binary distinction distinguishing 'arguments' from 'adjuncts' in a number of syntactic contexts.

T1-T5 display the optimizations in the four languages under discussion, each of which exemplifies a different optimal output. Only one possible ranking yielding a given optimal pattern is considered -- with the understanding that other rankings exist which may yield the same results (in each case, further consideration of question patterns in each language is needed to possibly narrow down alternative constraint rankings). Finally, the position of V in T1-T5 candidates reflects word order properties of the target languages with no consequence for the main issue under investigation.

Chinese is an in-situ-wh language: abstract Q operators in SpecCP bind wh-phrases in situ. Chinese permits adjunction of Q operators in violation of *ADJOIN. Given that candidate *a* has to defeat candidate *b* and that *a* violates *Q, a constraint violated by candidate *b* must outrank *Q, namely *t. (Note: The relative ranking of * ABSORB is a factor only when wh is fronted, i.e. when *Q ranks higher than *t; then it decides between candidates *b* and *c*.)

T1. Chinese

| [ $Q_i$ $Q_k$ [V $x_i$ $x_k$ ]] | *t | *ABSORB | PARSEQ | *Q | *ADJOIN |
|---|---|---|---|---|---|
| *a*. Λ[ $Q_i$+$Q_k$ [$wh_i$ $wh_k$ V ]] | | | | ** | * |
| *b*. [ $wh_i$+$wh_k$ [$t_i$ $t_k$ V ]] | *! * | | | | * |
| *c*. [ $wh_{i[k]}$ [$t_i$ $wh_k$ V ]] | *! | * | | | |
| *d*. [ $wh_i$ +$Q_k$, [$t_i$ DP/+$wh_k$, V ]] | *! | | * | | |
| *e*. [ $Q_i$ +$Q_k$, [$wh_i$ DP/+$wh_k$, V ]] | | | *! | * | |

Bulgarian exemplifies the overt counterpart to Chinese, i.e. adjunction of multiple wh-phrases in SpecCP binding wh-traces in situ. The optimal candidate is *b.* Overt movement of two wh-phrases results in two violations of *t and one violation of *ADJOIN. PARSEQ, *Q, AND *ABSORB must outrank *t AND *ADJOIN in Bulgarian.

T2. Bulgarian

| [ $Q_i$ $Q_j$ [V $x_i$ $x_j$]] | *ABSORB | *Q | PARSEQ | *t | *ADJOIN |
|---|---|---|---|---|---|
| *a.* [ $Q_i$+$Q_j$ [V $wh_i$ $wh_j$ ]] | | *! * | | | * |
| *b.* Λ [ $wh_i$+$wh_j$ [V $t_i$ $t_j$ ]] | | | | * * | * |
| *c.* [ $wh_{i[j]}$ [V $t_i$ $wh_j$ ]] | *! | | | * | |
| *d.* [ $wh_i$ +$Q_j$, [V $t_i$ DP/+$wh_j$, ]] | | | *! | * | |
| *e.* [ $Q_i$ +$Q_j$, [V $wh_i$ DP/+$wh_j$, ]] | | *! | * | | |

English exemplifies the mixed strategy of fronting a subject wh-phrase while leaving an object one in situ. For an input/index containing for example *who* and *what*, the winner is candidate *c*, a single overt wh-phrase in SpecCP co-indexed with both variables (a wh-phrase and a trace in situ). This means that *ABSORB and *t must be outranked by *Q, PARSEQ, and *ADJOIN in English. The ungrammaticality of **Who came why?* will be addressed in Section 3.2. At that point it will become clear why PARSEQ must be outranked by *Q and *ADJOIN in English.

T3. English

| [ $Q_i$ $Q_j$ [V $x_i$ $x_j$]] | *Q | *ADJOIN | PARSEQ | *t | *ABSORB |
|---|---|---|---|---|---|
| *a.* [ $Q_i$+$Q_j$ [$wh_i$ V $wh_j$ ]] | *! * | * | | | |
| *b.* [ $wh_i$+$wh_j$ [$t_i$ V $t_j$ ]] | | *! | | * * | |
| *c.* Λ [ $wh_{i[j]}$ [$t_i$ V $wh_j$ ]] | | | | * | * |
| *d.* [ $wh_i$ +$Q_j$, [$t_i$ V DP/+$wh_j$,]] | | | *! | * | |
| *e.* [ $Q_i$ +$Q_j$, [$wh_i$ V DP/+$wh_j$,]] | *! | | * | | |

The two basic strategies of multiple wh-questions – fronting vs. in-situ – yield two alternative candidates that fail to parse one input Q feature: *d* and *e*. In Italian the optimal output is *d*: one Q-feature results in a fronted wh-phrase, the other Q feature is unparsed. One possible constraint ranking is the one displayed in T4: The constraints violated by the optimal candidate *d* (*t and PARSEQ) are outranked by *Q, * ABSORB, and *ADJOIN.

T4. Italian

| [ $Q_i$ $Q_j$ [V $x_i$ $x_j$]] | *Q | *ABSORB | *ADJOIN | PARSEQ | *t |
|---|---|---|---|---|---|
| *a*. [ $Q_i$+$Q_j$ [V $wh_i$ $wh_j$ ]] | *! * | | * | | |
| *b*. [ $wh_i$+$wh_j$ [V $t_i$ $t_j$ ]] | | | *! | | * * |
| *c*. [ $wh_{i[j]}$ [V $t_i$ $wh_j$ ]] | | *! | | | * |
| *d*. Λ [ $wh_i$ +$Q_j$, [V $t_i$ DP/+$wh_j$, ]] | | | | * | * |
| *e*. [ $Q_i$ +$Q_j$, [V $wh_i$ DP/+$wh_j$, ]] | *! | | | * | |

Under the ranking (*t >> *ADJOIN >> *Q >> *ABSORB >> PARSEQ), candidate *e* is optimal. See T5. The outcome is an in-situ-wh language which only allows a single Q feature to be realized. No such language has been discussed in the literature. Yet, a query by Ralf Vogel (LinguistList, 7/3/2001) revealed that Omaha-Ponca (Siouan) may turn out to exemplify this pattern, courtesy of Catherine Rudin who has conducted fieldwork on this language. In particular she reports that Omaha-Ponca does not seem to have multiple questions of the form “who what likes”, “you what where did?”. Native speakers either drop the second wh-phrase or give “who likes all these things?”.

T5. Predicted possible: one wh-phrase in situ only

| [ $Q_i$ $Q_j$ [V $x_i$ $x_j$]] | *t | *ADJOIN | *Q | *ABSORB | PARSEQ |
|---|---|---|---|---|---|
| *a*. [ $Q_i$+$Q_j$ [$wh_i$ $wh_j$ V ]] | | *! | * * | | |
| *b*. [ $wh_i$+$wh_j$ [$t_i$ $t_j$ V ]] | *! * | * | | | |
| *c*. [ $wh_{i[j]}$ [$t_i$ $wh_j$ V]] | *! | | | * | |
| *d*. [ $wh_i$ +$Q_j$, [$t_i$ DP/+$wh_j$, V ]] | *! | | | | * |
| *e*. Λ [ $Q_i$ +$Q_j$, [$wh_i$ DP/+$wh_j$, V ]] | | | * | | * |

The reader may verify that the present analysis of multiple wh-questions makes the right predictions for single wh-questions in each language. The constraints pertaining to possible combinations of operators, *ADJOIN and *ABSORB, are vacuously satisfied and the ranking of PARSEQ relative to *Q and *t correctly yields wh-fronting in Bulgarian, English, and Italian vs. wh-in-situ in Chinese (and Omaha-Ponca).

## 3. Why the null parse approach is unworkable

The neutralization approach is grounded in the interplay of markedness and faithfulness constraints. One objection raised against the approach concerns the overall validity of I-O faithfulness constraints in syntax. This section presents evidence that I-O faithfulness plays a role in contexts other than multiple questions, in particular where it interacts with **locality restrictions** on movement: extraction out of [-wh] complement clauses vs. [+wh] complement clauses.[4]

Another objection to the neutralization approach comes from the existence of an alternative approach – the null parse approach – which is often held to have the advantage of not relaxing the LF-equivalency constraint on the candidate set that is required by the neutralization approach. This section discusses evidence that the null parse approach cannot provide an account of the sort we propose below to account for the relative difficulty of extracting out of a selected wh-clause compared with a selected *that*-clause.

### 3.1. Extraction out of wh-clauses

Since the early 80's much discussion in generative syntax has been devoted to locality or clause-boundedness effects in syntax despite some surface evidence to the contrary. For example, both English and Chinese permit extraction of direct object *what* (15) and a non-referential *how* adjunct (16) from a [-wh] complement clause selected by the matrix verb *think*, although the extraction is covert in Chinese. In both languages, extraction of a wh-phrase out of a [-wh] complement clause embedded under *think* involves long-distance movement (or movement which spans more than a clause). The present discussion assumes that *think* is lexically marked to select an IP complement in both languages, for two reasons: a) Chinese does not have a complementizer, b) independent evidence for an IP analysis is provided in Legendre et al. (1995).

(15) En a. [$_{CP}$What$_j$ do [$_{IP}$ you think [$_{IP}$ John fixed t$_j$ ]]]]?
Ch b. [$_{CP}$ Q$_j$ [$_{IP}$ Ni renwei [$_{IP}$ Lisi yinggai chuli shenme$_j$]]]?
you think L should handle what
"What do you think (that) L should handle t?"

(16) En a. [$_{CP}$ How$_l$ do [$_{IP}$ you think [$_{IP}$ John fixed it t$_l$ ]]]]?
Ch b. [$_{CP}$ Q$_l$ [$_{IP}$ Ni renwei [$_{IP}$ Lisi yinggai zenmeyang$_l$ chuli zhe-jian shi]]]?
you think L should how handle this-CL matter
"How (manner) do you think (that) L should handle this matter?"

According to Tsai (1994), Chinese wh-islands display a pattern of covert extraction (as indicated by scope interpretation) which is sensitive to the universal referentiality scale introduced earlier (*who* > *what* > *where, when* > *how* > *why*). Covert extraction of a referential wh-phrase (*who*) out of a wh-clause is possible, yielding two alternative direct questions or wide scope readings as shown in (17a). In contrast, the target wide scope reading is impossible when covert extraction of a non-referential wh-phrase (*how*) takes place, as shown in (17b). Yet, (17b) is grammatical under the narrow scope reading of an indirect question, as shown in (17b'). A similar situation obtains with (overt) extraction of referential *who* out of a wh-clause in English, as shown in (17c)-(17d).

(17) Ch a. [$_{CP}$ Q$_i$ +Q$_l$ [$_{IP}$ Ni xiang-zhidao [$_{IP}$ shei$_i$ zai nali$_l$ gongzuo]]]
you wonder who at where work
"Who do you wonder where works?" (wide scope)

Ch b. * [$_{CP}$ Q$_i$ +Q$_l$ [$_{IP}$ Ni xiang-zhidao [$_{IP}$ shei$_i$ zenmeyang$_l$ chuli zhe-jian shi]]]
you wonder who how handle this-CL matter
"How (manner) do you wonder who handled this matter?" (wide scope)

Ch b'. [$_{IP}$ Ni xiang-zhidao [$_{CP}$ Q$_i$ +Q$_l$ [$_{IP}$ shei$_i$ zenmeyang$_l$ chuli zhe-jian shi]]]
you wonder who how handle this-CL matter
"You wonder who handled this matter how" (narrow scope)

En c. *[$_{CP}$ Who$_i$ do [$_{IP}$ you wonder [$_{CP}$ what$_j$ [$_{IP}$ t$_i$ bought t$_j$ ]]]]? (wide scope)
En d. [$_{IP}$ you wonder [$_{IP}$ who$_{i[j]}$ bought what$_j$ ]]? (narrow scope)

In a nutshell, Legendre et al. (1995, 1998) provide the following account of the Chinese patterns in (16) and (17). The input to wh-island extractions includes a target wide scope specification, e.g. $Q_i$ you wonder [$Q_j$ $x_i$ ate $x_j$]. In (17b) the extraction of the adjunct wh-phrase *how* out of a wh-clause in Chinese is ungrammatical because the chain [$Q_l$, *how*] is non-referential and too long, as measured in barriers crossed (to be made precise below as the MINLINK power hierarchy) – despite the fact that the non-referential *how*-chain in the [-wh] complement extraction in (16b) is the same length and is grammatical. The selectional restrictions on the matrix verbs *wonder* vs. *think* provide part of the answer. *Wonder* selects for a wh-clause with the result that two candidates with different scope compete: a narrow scope interpretation associated with a shorter chain and a wide scope interpretation associated with a longer chain. The shorter chain may win even though it is unfaithful to the target wide scope specification in the input if PARSESCOPE is lower-ranked than the particular constraint violated by the long extraction, as must be the case in (17b) and (17c). *Think* does not select for a wh-clause, the shorter chain with narrow scope violates the selectional restrictions on *think* (SELECT is high-ranked) and the longer chain with wide scope is optimal despite being disfavored by MINLINK. The relevant optimizations are sketched out for Chinese in T6 (and further elaborated upon in T10 and T12).

T6. Chinese: Comparative covert extraction of a non-referential wh-phrase

| input: $Q_l$ $V_{matrix}$ [ ⌡ $how_l$ ⌡ ] | SELECT | MINLINKH | PARSESCOPE |
|---|---|---|---|
| ***wh*-island extraction (*)** | | | |
| *a.* $Q_l$ $wonder_{[+wh]}$ [ ⌡ $how_l$ ⌡ ] | | *! | |
| *b.* Λ $wonder_{[+wh]}$ [$Q_l$ ⌡ $how_l$ ⌡ ] | | | * |
| ***think* complement extraction (T)** | | | |
| *aN.* Λ $Q_l$ $think_{[Bwh]}$ [ ⌡ $how_l$ ⌡ ] | | * | |
| *bN.* $think_{[Bwh]}$ [$Q_l$ ⌡ $how_l$ ⌡ ] | *! | | * |
| Remarks: | No [+wh] complement for *think* | Hconstraint crucial here: $BAR^{2\ [Bref]}$ | For an extraction input, extract |

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 the case with *wonder*, but not with *think*.

### 3.2. Locality vs. I-O faithfulness

The complete analysis of extraction out of complement vs. wh-clauses involves an I-O faithfulness constraint modeled on PARSEQ (Section 2), namely PARSESCOPE, as in (18).

(18) PARSESCOPE: Target scopes of the Input/Index must be realized.

We also need a family of **barrier** constraints, MINLINK, establishing a scale on which longer wh-links are less harmonic (more marked) than shorter ones. MINLINK exploits both the concept of barrier to movement from Chomsky (1986) and a formal technique of OT, Local Conjunction of constraints (Smolensky 1995, 1997), plus one of its consequences, universal Power Hierarchies.

A barrier is a maximal projection (XP) which is not theta-governed (Chomsky 1986). For the extraction cases discussed here, VP as well as IP preceded by a complementizer count as barriers; they are governed by a functional category, Infl or C, which by definition does not theta-mark its complement. CP and IP in the absence of a complementizer are governed and theta-marked by the matrix verb (*wonder* or *think*), hence they do not constitute barriers.

(19) BAR: A chain link may not cross a barrier.

A question regarding BAR immediately arises. Is a single constraint sufficient or do we need a family of BAR constraints? The comparison in T7 provides the answer. A single BAR constraint cannot differentiate local from non-local movement (each candidate incurs 3 violations of BAR); it therefore fails to characterize one fundamental property of syntactic operations.

T7. Cyclic *vs.* non-cyclic chains: equally marked, according to BAR (β = barrier)

| | | BAR |
|---|---|---|
| *a.* Cyclic, 2 links | Λ [$X_i$ ⌡ β ⌡ β ⌡ $t_i$ ⌡ β ⌡ $Y_i$] | ** * |
| *b.* Non-cyclic, 1 link | *Λ [$X_i$ ⌡ β ⌡ β ⌡ ⌡ β ⌡ $Y_i$] | *** |

This is easily remedied by **locally conjoining** BAR with itself (with domain *D* = link); we thus obtain BAR $\&_l$ BAR / $\text{BAR}^2$. By recursion we then obtain a universal BAR Power Hierarchy. T8 demonstrates that a cyclic chain is universally preferred to a non-cyclic one.

(20) a. $\text{BAR}^2$: A single link must not cross two barriers
b. By definition of the Local Conjunction operation, *universally*: $\text{BAR}^2 \gg \text{BAR}$
c. By recursion: MINLINK (universal BAR Power Hierarchy)…$\gg \text{BAR}^3 \gg \text{BAR}^2 \gg \text{BAR}^1$

T8. Universal BAR Power Hierarchy

| | MINLINK | | | BAR |
|---|---|---|---|---|
| | $\text{BAR}^3$ | $\text{BAR}^2$ | $\text{BAR}^1$ | |
| *a.* Cyclic Λ [$X_i$ ⌡ β ⌡ β ⌡ $t_i$ ⌡ β ⌡ $Y_i$] | | * | * | ** * |
| *b.* Non-cyclic [$X_i$ ⌡ β ⌡ β ⌡ ⌡ β ⌡ $Y_i$] | * ! | | | *** |

One additional application of Local Conjunction is needed because not all cyclic chains are equal. Referentiality matters because it interacts with locality as demonstrated by the Chinese patterns in (17). Non-referential chains are good, if short. Long chains are good, if referential. That is, chains violating *both* MINLINK *and* REF are bad.

(21) REF: Chains are referential.

Using Local Conjunction and recursion we obtain (22).

(22) a. $\text{BAR}^k$ $\&_l$ REF /$\text{BAR}^{k\,[\text{Bref}]}$: A link in a non-referential chain must not cross *k* barriers.
b. $\text{MINLINK}^{[\text{Bref}]}$ : ⌡ $\gg \text{BAR}^{3\,[\text{Bref}]} \gg \text{BAR}^{2\,[\text{Bref}]} \gg \text{BAR}^{1\,[\text{Bref}]}$
c. $\text{BAR}^{k\,[\text{Bref}]} \gg \text{BAR}^k$

We are now in a position to examine the specific optimizations underlying the English and Chinese patterns in (17). Extracting a referential wh-phrase (*who*) out of a wh-clause (i.e. assigning it a wide scope interpretation) is impossible in English: in T11 the narrow scope candidate *c* wins. In Chinese, a wide scope interpretation is fine in the same context (candidate *a* wins). In the interest of simplicity, only a subset of candidates informed by earlier optimizations (T1 and T3) are considered in T9-T10. Full brackets ([) represent barriers and hollow brackets (v) represent non-barriers. Violations are annotated with relevant indices corresponding to argument and adjunct type to facilitate readability. SELECT is undominated and satisfied by all candidates; for space considerations it is omitted in T9-10.

In an English matrix wh-extraction the best faithful candidate involves fronting of one wh-phrase and absorption (see T3). The equivalent in a long-distance extraction is candidate *a* in T9. However the extraction results in a long link violating BAR3 (B3). The latter can be obviated by failing to parse Q, as in candidate *b*; this too is suboptimal (see T3). The optimal candidate *c* violates PARSESCOPE instead and incurs a minimal MINLINK violation: BAR1 (for *what*).

T9. English: Extraction of subject *who* out of wh-clause

| $[Q_i$ [ *wonder* [ $Q_j$ [ V $x_i$ $x_j$ ]]] | *Q | PQ | B3 | PSC | B2 | B1 | *t | *ABS |
|---|---|---|---|---|---|---|---|---|
| *a.* $[_{CP}who_{i[j]}$ do$[_{IP}$you $[_{VP}$wonder $\vee_{CP}what_j$ $[_{IP}t_i$ $[_{VP}$V $t_j$ ]]]]]] | | | $*_i$! | | $*_j$ $*_j$ | | ** | * |
| *b.*$[_{CP}who_i$ $+Q_j$, do$[_{IP}$you $[_{VP}$wonder $\vee_{IP}t_i$ $[_{VP}$V DP/+$wh_j$ ,]]]]] | | $*_j$! | | | * | | * | |
| *c.* Λ $[_{IP}$you $[_{VP}$wonder $\vee_{IP}who_{i[j]}$ $[_{VP}$V $what_j$ ]]]] | | | | * | | * | | * |

In Chinese the best faithful candidate involves adjunction of two Q operators in violation of *ADJOIN (see T1). In T10, the most faithful candidate with such adjunction is candidate *a* which incurs violations of BAR3[-REF] (for *how*) and BAR2 (for *who*). Failing to parse one Q is suboptimal, as in candidate *b*. The optimal candidate *c* violates PARSESCOPE instead and comparatively incurs minimal MINLINK violations: BAR2[-REF] (for *how*) and BAR1 (for *who*). The outcome is the same as in English except for the absorption vs. adjunction strategy (and fronting vs. in-situ, of course). To save space in T10 BAR2[-REF] and BAR1[-REF] are conflated into one constraint, as are BAR3 and BAR2, without altering the outcome of the optimization.

T10. Chinese: Extraction of *how* out of wh-clause

| [Q$_l$ [*wonder* [Q$_i$ [ V x$_i$ x$_l$]]]] | *t | PQ | B3 -ref | B2-1 -ref | P SC | B3-2 | B 1 | *Q | *ADJ |
|---|---|---|---|---|---|---|---|---|---|
| *a*.[$_{CP}$Q$_{l+}$Q$_i$[$_{IP}$you [$_{VP}$wonder ∨$_{IP}$wh$_i$ [$_{VP}$ wh$_l$ V NP ]]]]]] | | | $*_l$! | | | $*_i$ | | ** | * |
| *b*. [$_{CP}$Q$_l$ +Q$_i$,[$_{IP}$you [$_{VP}$wonder ∨$_{IP}$ DP/+wh$_j$, [$_{VP}$ how$_l$ V NP]]]]]] | | $*_i$! | | | | | | | |
| *c*. Λ [$_{IP}$you [$_{VP}$ wonder ∨$_{CP}$ Q$_l$+Q$_i$ [$_{IP}$ wh$_i$ [$_{VP}$ how$_l$ V NP ]]]]] | | | | $*_l$ | * | | $*_i$ | | * |

Extrapolating from T10, extraction of a direct object wh-phrase (referential *what*) instead of non-referential *how* out of a wh-clause in Chinese would violate BAR3 (instead of BAR3[-REF]) and result in wide-scope candidate *a* (rather than narrow scope candidate *c*) being optimal.

With respect to extraction out of the complement of *think*$_{IP}$ the narrow scope candidate *c* in T11 and T12 is not a viable output for either language because it violates the selectional restriction (SELECT/SEL) on the matrix verb: *think* is lexically marked to select a [-wh] complement with the consequence that an operator (Q or wh) cannot appear in the Spec position of the immediate complement of *think* without incurring a fatal violation.[5] A failure to parse Q resulting in a declarative statement rather than an information question is also suboptimal; see candidate *b*. A long link violating BAR3 in English (for *what*) or BAR3[-REF] in Chinese (for *how*) – candidate *a* – is optimal in both languages.

T11. English: Extraction of *what* out of the complement of *think*$_{IP}$

| [Q$_j$ [ *think*$_{IP}$ [ V x$_j$ ]]] | SEL | *Q | PQ | B3 | PSC | B2 | B1 | *t |
|---|---|---|---|---|---|---|---|---|
| *a*.Λ[$_{CP}$what$_j$ do[$_{IP}$you [$_{VP}$think ∨ $_{IP}$John [$_{VP}$fixed t$_j$ ]]]]] | | | | * | | | | * |
| *b*. [$_{CP}$ DP/+wh$_j$, do[$_{IP}$you [$_{VP}$think ∨$_{IP}$ John [$_{VP}$fixed +t$_j$,]]]]] | | | *! | | | | | |
| *c*. [$_{IP}$you [$_{VP}$think ∨$_{CP}$what$_j$ [$_{IP}$John [$_{VP}$fixed t$_j$ ]]]]] | *! | | | | * | * | | * |

T12. Chinese: Extraction of *how* out of the complement of *think*$_{IP}$

| [Q$_j$ [ *think* [ V x$_j$ ]]] | SEL | *t | PQ | B3 -ref | B 2-1 -ref | PSC | B 3-1 | *Q |
|---|---|---|---|---|---|---|---|---|
| *a*.Λ[$_{CP}$Q$_l$ [$_{IP}$you [$_{VP}$think ∨$_{IP}$Lisi [$_{VP}$how$_l$ handle this matter ]]]]] | | | | * | | | | * |
| *b*.[$_{CP}$+Q$_l$, [$_{IP}$you [$_{VP}$ think ∨$_{IP}$Lisi [$_{VP}$DP/+how$_l$, handle this matter]]]]]] | | | *! | | | | | |
| *c*.[$_{IP}$you [$_{VP}$think ∨$_{CP}$Q$_l$ [$_{IP}$Lisi [$_{VP}$how$_l$ handle this matter]]]]] | *! | | | | | * | | * |

In sum, it is clear that not all instances of locality effects in syntax can be analyzed in terms of a single constraint stating that shorter links are better than longer links.[6] In particular, it is not the case that Chinese disprefers long links across the board. Rather, a particular type of link in a particular context is dispreferred: A long link of type Bar3[+ref] is better than a failure to parse input/index scope but a long link of type Bar3[-ref] is worse than a failure to parse scope. However, a long link of type Bar3[-ref] is tolerated if the alternative is to violate selectional restrictions.

It is worth emphasizing that the neutralization account of wh-extraction significantly departs from the traditional one based on locality violations. In our terms, it is harder to extract out of a wh-island not because of locality violations caused by the presence of an intervening wh-phrase but because embedded wh-islands offer a competitor (narrow scope interpretation) that other types of embedded complements do not offer.

### 3.3. A fatal problem for the null parse approach

The comparison between extracting out of a complement of *think* vs. a complement of *wonder* reveals a fatal problem for the null parse approach relying on a single PARSE WH constraint (Ackema & Neeleman 2000). To account for the range of English and Chinese patterns discussed in this paper it is necessary to posit two separate PARSE constraints: PARSEQ and PARSESCOPE. In the wh-extraction contexts discussed in Section 3.2. both constraints are active because two Qs are present in the input. The option of not parsing one Q (on a par with multiple wh-questions discussed in Section 2.2) exists – see candidate *b* in T9-10 – but is dispreferred. What gives instead is wh-scope in complements of *wonder*, an option simply not available for multiple wh-questions in a single clause. It is evident that a single parse constraint cannot have the two different positions in the hierarchy needed for PARSEQ and PARSESCOPE within a language, e.g. PARSEQ >> $\text{BAR}^{3}$ >> PARSESCOPE in English (T9) and PARSEQ >> $\text{BAR}^{3\text{-}1[\text{-REF}]}$ >> PARSESCOPE in Chinese (T10).

More generally, since Ø occurs in every candidate set in the null parse approach, 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. PARSE must be ranked so that every parse of every ineffable Input/Index violates a constraint higher than PARSE (and loses to Ø). PARSE must also be ranked so that some parse of every effable Input/Index violates no constraint higher than PARSE (and bests Ø). However, it is imperative that the relative Harmonies of Input/Index-specific faithful and unfaithful parses be decisive. This can be accomplished by positing multiple Input/Index-specific unfaithful parses in which just an operator scope or just a Q feature is not parsed, depending on the candidate set. This cannot be accomplished by positing a single unfaithful parse (Ø) in all candidate sets.[7]

## 4. Independent evidence for neutralization in syntax

As argued above, neutralization offers a very economical solution to the ineffability problem that takes full advantage of existing OT resources. But neutralization is not a special strategy solely deployed to handle the minimal underparsing of a Q feature or operator scope in syntax. In fact, neutralization is a crucial component of the interpretive optimization in Wilson (2001)'s Bidirectional Optimization approach to cross-linguistic patterns in anaphoric binding of the sort *Tom said that Sue loves self*. Wilson argues that 'relativized minimality' (or locality) effects on anaphor binding are a consequence of neutralization in the interpretive optimization rather than the effect of a relativized minimality constraint per se. In his terms, the conflict is between a rigid locality constraint (a version of Principle A of the Binding Theory) and I-O Faithfulness. The ranking LOCALITY >> FAITH maps non-local binding to local binding, resulting in neutralization of a possible contrast.

T13 Neutralization for relativived minimality in anaphor binding (Wilson 2001)

| Input | Output | |
|---|---|---|
| a. [$T_i$ [ $S_j$ ……self$_j$ ….]] | [$T_i$ [ $S_j$ ……self$_j$ ….]] | faithful output |
| b. [$T_i$ [ $S_j$ ……self$_i$ ….]] | [$T_i$ [ $S_j$ ……self$_j$ ….]] | unfaithful output |

Neutralization also offers a solution to another 'unexpected' pattern under OT assumptions, namely optionality of forms. For example, Baković & Keer (2001) propose that the input to clausal complements of the matrix verb *think* be specified as [+complementizer] or [-complementizer] and derive the two patterns in (23a-b) from distinct inputs.

(23) a. I think [$_{CP}$ that [$_{IP}$ the coat doesn't fit him]].
b. I think [$_{IP}$ the coat doesn't fit him].

(24) a. The coat$_j$ [$_{CP}$ that [$_{IP}$ he always wears $t_j$]] doesn't fit him.
b. The coat$_j$ [$_{IP}$ he always wears $t_j$] doesn't fit him.
c. The coat$_i$ [$_{CP}$ that [$_{IP}$ $t_i$ doesn't fit him]] might fit me.

d. *The coat$_i$ [$_{IP}$ t$_i$ doesn't fit him] might fit me.

Baković & Keer argue that there is optionality (23a-b, 24a-b) precisely where there is no neutralization. When the complementizer is either obligatory or prohibited (24c-d) it is because certain markedness constraints dominate the proposed faithfulness constraints. As a result, markedness constraints, which favor the same output for both inputs, prevail.

In relative clauses with subject extraction (24c-d), the complementizer *that* is obligatory because the requirement that the subject trace (t$_i$) be governed (TGOV, Grimshaw 1997) outranks FAITH[COMP], an I-O faithfulness constraint regulating the input and output value of the input specification [+/- complementizer]. In (24c) t$_i$ is governed by *that*, satisfying TGOV. See T14. In (24d) TGOV is fatally violated because relative clauses, being adjunct structures, by definition are ungoverned. See T15: The optimal counterpart – which violates FAITH[COMP] is candidate *a*, with complementizer *that*. T14 and T15 are reproduced from Baković & Keer (2001:103).[8]

T14. English: [+comp] relative clause with subject extraction

| [+comp] | TGOV | FAITH[COMP] |
|---|---|---|
| *a*. ʌ The coat$_i$ [$_{CP}$ that [$_{IP}$ t$_i$ doesn't fit him]] might fit me | | |
| *b*. The coat$_i$ [$_{IP}$ t$_i$ doesn't fit him] might fit me | *! | * |

T15. English: [– comp] relative clause with subject extraction

| [– comp] | TGOV | FAITH[COMP] |
|---|---|---|
| *a*. ʌ The coat$_i$ [$_{CP}$ that [$_{IP}$ t$_i$ doesn't fit him]] might fit me | | * |
| *b*. The coat$_i$ [$_{IP}$ t$_i$ doesn't fit him] might fit me | *! | |

In contrast, no neutralization takes place in relative clauses with object extraction (24a-b). By definition, the object position is governed by V. The faithful candidate wins for both [+comp] and [-comp] inputs. The result is optionality of the complementizer.

In sum, the Baković & Keer analysis extends the neutralization analysis beyond the realm of interpretational properties – the focus of Legendre et al. (1995, 1998) – to regulating formal properties of syntactic structures in individual languages.

## 5. The surface realization issue

For many critics of the underparsing approach advocated here the main question is: If the optimal output is a minimally unparsed structure (as opposed to a null structure Ø), what is its surface realization? A more substantial question in our view is: What is the interpretation of the optimal candidate for ineffable inputs? Is a separate component, subject to optimization or not, needed to interpret the optimal candidate? In the neutralization approach discussed above, there is no additional component. Candidates have LF in them; a single optimization of the usual sort is all that is needed to yield a solution to the ineffability problem.

It is important to keep in mind that a candidate set in syntax is a set of possible abstract realizations of an input with many surface properties irrelevant to the optimization at hand, in particular their pronunciation. This means that the job of determining the pronunciation of an underparsed structure falls to another optimization subject to lexical, phonological, and pragmatic constraints.

However, the possible patterns of interpretational repairs in ineffable multiple wh-questions cross-linguistically display properties worth taking a look at here. In some languages (e.g. Chinese), wh-phrases share a lexical form with indefinite quantifiers, resulting in a single form interpreted as *what* in wh-contexts but *something* elsewhere. This led Legendre et al (1995, 1998) to suggest that ineffable multiple wh-questions can be repaired by a single wh-phrase plus an (in)definite quantified phrase (the [-wh] counterpart of wh-phrases). The question to be entertained here is the extent to which languages like English and Italian avail themselves of a similar strategy.

English is particularly relevant because multiple wh-questions display a pattern of ineffability that is sensitive to referentiality. Multiple wh-questions are good, if referential (25a). Non-referential extraction is good, if it involves a single wh-phrase (25b). But multiple wh-questions involving at least one non-referential wh-phrase (e.g. *why*) are bad (25c).

(25) a. Who ate what?
b. Why did John come?
c. *Who came why?

(25a) corresponds to optimal absorption candidate *c* which incurs violations of *t and *ABSORB in T3. Candidate *c*, in particular, beats candidate *d* which fails to parse Q. At first glance, this appears to be problematic for (25c). According to the present analysis the optimal output of ineffable multiple Qs is a failure to parse. How can a failure to parse Q beat the absorption candidate if ParseQ is high-ranked in English? The solution lies in the conjunctive effect of referentiality and *ABSORB. On the model of $\text{BAR}^k$ $\&_l$ REF /$\text{BAR}^{k\ [\text{Bref}]}$ in (22a) a Local Conjunction of *ABSORB and REF yields $\text{*ABSORB}^{[\text{Bref}]}$ which outranks both constraints in isolation. The Local Conjunction captures the generalization that chains violating *both* *ABSORB *and* REF are bad in English.

T16. English: ineffable multiple wh-questions

| [ $Q_i$ $Q_l$[V $x_i$ $x_l$]] | *Q | *ABSORB$^{-REF}$ | *ADJOIN | PARSEQ | *t | *ABSORB |
|---|---|---|---|---|---|---|
| *a*. [ $Q_i$+$Q_l$[$wh_i$ V $wh_l$ ]] | *! * | | * | | | |
| *b*. [ $wh_i$+$wh_l$ [$t_i$ V $t_l$ ]] | | | *! | | * * | |
| *c*. [ $wh_{i[l]}$ [$t_i$ V $wh_l$]] | | *! | | | * | * |
| *d*. Λ [ $wh_i$ +$Q_l$, [$t_i$ V DP/+$wh_l$,]] | | | | * | * | |
| *e*. [ $Q_i$ +$Q_l$, [$wh_i$ V DP/+$wh_l$, ]] | *! | | | * | | |

In contrast to Chinese, English wh-phrases *who, what*, are lexically distinct from indefinite quantifiers (e.g. *something*, *anything*). At first glance, a number of alternative wh-questions (26b-d) come to mind as possible repairs for ineffable inputs such as (26a).

(26) a. *Who came why?
b. Why did each person come?
c. *?Who came for each reason?
d. Who came for any reason?
e. Who came, and why?

In (26b-c) the unparsed Q feature is realized as a definite quantified expression (rather than an indefinite one, as in Chinese). *Each person* and *each reason* are d(iscourse)-linked expressions (Pesetsky 1987). Both presuppose a pre-existing limited set of referents. However, (26c) is pragmatically marked, compared to (26b). While it is easy to conceive of a limited list of people enrolled in a class for example (by consulting a class roster) it is much harder to conceive of a similar list of reasons. This may be the reason why native speakers in fact never volunteer (26c) as a repair for ungrammatical (26a). There is in fact a grammatical and pragmatically natural alternative to (26c), namely (26d), with a characteristic but unfaithful single wh reading.

Note that substituting *each person* for *who* (as in 26b) allows the retention of the multiple-pair reading characteristic of multiple wh-questions. If (26b) were uttered by a professor on the first day of classes then Johnny would give his reason, Martha hers, etc. If so, then English manages to unparse Q while preserving the characteristic multiple pair reading of multiple wh-questions.

(26e) represents a distinct strategy sometimes offered by native speakers when confronted with (26a). Both Q features are in fact parsed but separately in two single wh-questions joined by a coordinating conjunction itself preceded by an intonational break. What most obviously gives here is the target syntactic structure specified in the input, a simple sentential structure, as well as the target prosodic structure. Native speakers consulted also find that the target multiple-pair interpretation is not the preferred interpretation of (26e). Rather, the most natural answer to *Who came, and why?* associates a single reason for *why* with a set of referents for *who,* pointing to a distinct input.

The English strategies largely hold for Italian repairs of ineffable (27a). Like English (26b), Italian (27b) retains a multiple-pair reading while failing to parse one Q feature in all multiple wh-questions. According to native speakers, the alternative (27c) has a distinctive yes/no pair list interpretation 'Who ate something? Gianni, yes; Monica no, etc.' (27d) in turn is comparable to English (26e) and involves an unfaithful coordinate structure.

(27) It a. *Chi ha mangiato che cosa?
who has eaten which thing
'Who ate what?'
b. Che cosa ha mangiato ciascuna persona?
which thing has eaten each person
'What did each person eat?'
c. Chi ha mangiato qualcosa?
Who ate something?
d. Chi ha mangiato e che cosa ha mangiato?
Who ate and what did (they) eat?

In sum, English and Italian are languages in which a Q unparsing repair strategy may result in a surface structure which retains the multiple-pair reading associated with multiple wh-questions. Others strategies likely to lose competition for favored repair involve unparsing of Q with concomitant unparsing of multiple-pair reading on the one hand and unparsing of input syntactic and prosodic specifications without unparsing of multiple-pair reading on the other. The input to wh-questions must therefore contain the target multiple-pair interpretation ( left implicit in Section 2) in addition to Q features. The two properties are independent: A Q feature may be unparsed while the target multiple-pair interpretation is preserved. All other things being equal, minimal LF (rather than structural) unparsing is preferred. This in turn confirms a basic assumption of the neutralization approach: the input includes multiple specifications; it is structured.

## 6. Conclusion

Ineffability is one of the input-output mismatches providing strong evidence for a decisive role of Input-Output Faithfulness in syntax. Structural optionality of the kind discussed in Baković and Keer (2001) is another, which similarly calls for a neutralization approach.

Neutralizing the input-output mapping is a comparatively simple and elegant solution to the challenge of handling ineffable inputs/indexes in an output-optimizing system. At the syntax/semantics interface of wh-questions two separate instances of two inputs (e.g. multiple vs. single wh, wide vs. narrow scope) were shown to be mapped to the same output each, thus eliminating universally possible contrasts from the languages under consideration.

The neutralization approach requires abandoning the assumption that competitors have the same LF interpretation – though they have the same target interpretation. On the neutralization view, an LF unrealizable in a given language is a structure such that every syntactic output with that LF interpretation is less harmonic in that language than a competitor with a (minimally) different LF.

The neutralization approach is grounded in a 'traditional' view of OT. The original concept of input is retained. Much work is done by I-O faithfulness. No additional component (such as an interpretational one) is needed to operate on the output of the syntax. No additional constraint on the candidate set is imposed either.

The debate surrounding ineffability in syntax ultimately bears on one fundamental single question: What defines the candidate set in an OT system? The answer is: the structured input.

## References


Ackema, P. & A. Neeleman. 2000. Absolute Ungrammaticality. In J. Dekkers, F. van der Leeuw, J. van de Weijer (Eds.) *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford University Press. 279-301.

Bakovi, Eric and Edward Keer. 2001. Optionality and Ineffability.In G. Legendre, J. Grimshaw, and S. Vikner (eds). *Optimality Theoretic Syntax*. MIT Press.

Beaver, D. & H. Lee. 2004. Input-Output Mismatches in Optimality Theory. In R. Blutner & H. Zeevat (Eds.), *Optimality Theory and Pragmatics*. Palgrave Macmillan. 112-153

Black, C. 2000. *Quiegolani Zapotec Syntax: A Principles and Parameters Account*. SIL International and University of Texas at Arlington. Publications in Linguistics 136.

Calabrese, A. 1984. Multiple Questions and Focus in Italian. In W. de Gest and T. Putseys (Eds.) *Sentential Complementation*. Dordrecht: Foris. 67-74.

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

Chomsky, N. 1993. A Minimalist Program for Linguistic Theory. In K. Hale & S.J. Keyser (Eds.), The View from Building 20: Essays in Honor of Sylvain Bromberger. Cambridge, Mass: MIT Press. 1-52.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, Mass: MIT Press

Choi, Hye-Won. 1996. *Optimizing Structure in Context: Scrambling and Information Structure*. PhD Dissertation. Stanford University.

Choi, Hye-Won. 2001. Binding and Discourse Prominence: Reconstruction in =Focus= Scrambling. In G. Legendre, J. Grimshaw, and S. Vikner (eds). *Optimality Theoretic Syntax*. Cambridge, Mass: MIT Press MIT Press. 143-169.

Costa, João. 1998. *Word Order Variation: A Constraint-based Approach*. HIL Publications.

Costa, João. This volume. The Emergence of Unmarked Word order. In G. Legendre, J. Grimshaw, and S. Vikner (eds). *Optimality Theoretic Syntax*. Cambridge, Mass: MIT Press MIT Press. 171-203.

Dayal, V. 2006. Multiple wh-questions. In M. Everaert, H. van Riemsdijk, R. Goedelmans, and B. Hollebrandse (Eds.), *The Blackwell Companion to Syntax*. Blackwell Publishing. Vol. III.

Fanselow, G. & C. Féry. 2002. Ineffability in Grammar. In *Resolving Conflicts in Grammar: Optimality Theory in Syntax, Morphology, and Phonology*. Special Issue 11 of *Linguistische Berichte*. 265 -307.

Grimshaw, J. 1997. Projection, Heads, and Optimality. *Linguistic Inquiry* 28,3: 373-422.

Heck, F, G. Müller, R. Vogel, S. Fischer, S. Vikner, and T. Schmidt. 2002. On the Nature of the Input in Optimality Theory. *The Linguistic Review* 19, 345-376.

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

Legendre, Géraldine. 2001. Masked Second-Position Effects and the Linearization of Functional Features. In G. Legendre, J. Grimshaw, and S. Vikner (eds). *Optimality Theoretic Syntax*. MIT Press.

Legendre, G., C. Wilson, P. Smolensky, K. Homer, and W. Raymond. 1995. Optimality and Wh-Extraction. In *Papers in Optimality Theory* (J. Beckman, S. Urbanczyck, and L. Walsh, eds.). UMOP 18: 607-636.

Legendre, G., P. Smolensky, and C. Wilson. 1998. When is Less More? Faithfulness and Minimal Links in Wh-Chains. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky (eds), *Is the Best Good Enough? Optimality and Competition in Syntax*. MIT Press. 249-289.

Legendre, G., C. Wilson, P. Smolensky, K. Homer, and W. Raymond. 2006. Optimality in Syntax II: wh-Questions. In P. Smolensky & G. Legendre, *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. MIT Press. Vol. II: 183-230.

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

McCloskey, J. 1979. *Transformational Syntax and Model Theoretic Semantics: A Case Study in Modern Irish*. Dordrecht: Reidel.

Müller, G. & W. Sternefeld. 2001. The Rise of Competition in Syntax: A Synopsis. In G. Müller & W. Sternefeld (Eds.) *Competition in Syntax*. Berlin: Mouton de Gruyter. 1-68.

Pesetsky, David. 1997. Optimality Theory and Syntax: Movement and Pronounciation. In D. Archangeli and D.T.Langendoen (eds), *Optimality Theory, An Overview*. Blackkwell Publishers. 134-170.

Prince, A. and P. Smolensky. [1993]/2004. *Optimality Theory: Constraint Interaction in Generative Grammar.* Blackell Publishing.

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

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

Samek-Lodovici, Vieri. 2001. Cross-linguistic Typologies in OT. In G. Legendre, J. Grimshaw, and S. Vikner (eds). *Optimality Theoretic Syntax*. MIT Press.

Schmid, T. 2001. OT Accounts of Optionality: A Comparison of Global Ties and Neutralization. In G. Müller & W. Sternefeld (eds), *Competition in Syntax*. Berlin: Mouton de Gruyter. 283-319.

Smolensky, Paul. 1995. On the internal structure of the constraint component *Con* of UG. Paper presented at UCLA.

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction, or random rules in UG. Paper presented at HOT, Baltimore, MD.

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

---

[1] These repair strategies prevent the ECP/TGOV (Empty Category Principle) from being violated. High-ranked in English and French, TGOV requires that traces be head-governed (Legendre et al. 1995). See also Grimshaw (1997). (1a)-(2a) are ungrammatical because the subject trace $t_i$ fails to be properly governed (complementizers *that* and *que* are functional rather than lexical categories, a defining attribute of proper governors). When *that* is dropped as in (1b), $t_i$ is properly governed by the matrix verb. Rizzi (1990) assumes that French *qui* is the 'agreeing' version of *que*, somehow making the complementizer eligible to serve as a proper governor of $t_i$. Finally, the grammaticality of corresponding Italian and Bulgarian extractions in (3a,b) is accounted for by positing that subject extraction is from a post-verbal position (Rizzi 1990), i.e. a position which is by definition properly governed by the embedded verb. This is independently motivated by the fact that both languages freely allow postverbal subjects in non-wh-contexts.

[2] Others (e.g. Müller & Sternefeld 2000:51) object to neutralization on the basis that it creates massive derivational ambiguity characterized as a serious problem in language acquisition and parsing which can be avoided only by positing additional meta-optimization procedures (e.g. lexicon optimization). Such comments confound competence and performance and ignore the fact that OT is explicitly formulated as a theory of competence by its founders.

[3] Echo questions are not requests for new information. They presuppose that the answer is already known; hence, their interpretation depends on a restricted set of values for the wh-variable. They correspond to a distinct input.

[4] Legendre et al. (1995) discuss a pattern of resumptive pronouns under overt DP topicalization which is sensitive to a subject-object asymmetry. In subject position these resumptive elements

offset violations of TGOV at the cost of violating the I-O Faithfulness constraint, FILL, penalizing epenthesis of elements not present in the input.

(i) Zhangsan$_i$, ta$_i$ xihuan kansu.
Z he like reading
'Zhangsan, he likes reading'

(ii) *Zhangsan$_i$, t$_i$ xihuan kansu.
Z like reading
'Zhangsan, likes reading'

[5] Candidate *b* also violates faithfulness to the structural specification of the input by realizing a complement CP structure; see discussion of Baković & Keer (2001) in Section 4.

[6] It is important to appreciate how the MINLINK approach differs from related concepts in the Minimalist Program also grounded in economy such as "Shortest Move" (Chomsky 1993) or the Minimal Link Condition (MLC, Chomsky 1995). In the present account there is no need to stipulate that Shortest Move is measured in terms of relativized minimality violations. Relativized Minimality captures the generalization that locality is not an absolute condition on movement but rather is dependent on each type of intervening element ("Don't move α across a place where α could have landed" where potential landing positions are specified for each type of movement: A-bar-, A-, or head-movement). The result is a complex definition of the relevant conditions which incorporate both each type of intervening element and the concept of barrier (Rizzi 1990). In other words, the comparison inherent to evaluating link length is built into the definition of the MLC as well as Rizzi's antecedent government. In the present OT analysis, the relativized minimality effect is a **consequence** of a general MINLINK constraint in which link length is measured simply in terms of barriers/nodes crossed. None of the 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 relativization effects (e.g. *wh* is harder to extract over *wh*) are also a derived consequence of constraint interaction.

[7] Ackema & Neeleman (2000:297-9) discuss another problem arising from their null parse approach to ineffability in passive . In particular, when the possibility of failing to parse passive and the failure to parse Q are put together under the ranking PARSEPASSIVE >> MARKEDNESS CONSTRAINT >> PARSEWH, strange languages are predicted to exist: multiple wh questions exist except in passive sentences. They save the null parse approach to ineffability by blocking the interaction of multiple PARSE constraints. This is achieved by replacing the standard single procedure of constraint evaluation in OT with a series of evaluation cycles stipulated to involve only one parse constraint at a time. The output of one optimization is taken to be the input for the next, "with the effect that if the null parse is optimal in one evaluation, it will be also be the optimal output of the total procedure" (p. 298). They resort to generic performance considerations to motivate their specific treatment of PARSE constraints "If in general multiple performance of a simple task is preferred over single performance of a more complicated task, the proposed evaluation procedure for PARSE constraints is indeed preferred" (p. 299). It is far from obvious that the additional, unmotivated machinery warrants preserving LF-equivalency of the candidate set.

[8] Unlike relative clauses, complement clauses are lexically governed by the main verb. Subject traces thus satisfy TGOV whether there is a complementizer or not: (i) *[Which coat]$_i$ do you think t$_i$ doesn't fit?* (ii) **[Which coat]$_i$ do you think that t$_i$ doesn't fit?* (ii) is ungrammatical because t$_i$ violates TLEXGOV (Grimshaw 1997) which outranks FAITH[COMP].