The Scrambling Paradox

John Frederick Bailyn

It has been commonly observed that scrambling and wh-movement share sensitivity to strong movement constraints (Webelhuth 1989, Saito 1992, Bailyn 1995). At the same time, the two processes clearly differ in certain other respects, such as wh-island sensitivity, a finding that has inspired a range of analyses of scrambling as entirely distinct from better-understood movement processes (Müller and Sternefeld 1993, Bošković and Takahashi 1998, among many others). Careful comparison of A-scrambling and overt wh-movement in a language that shows both (Russian) reveals that this seemingly paradoxical behavior can be captured effectively in a probe-goal theory of scrambling that obeys a form of Relativized Minimality defined across feature classes, following Rizzi 2004. The resulting analysis exposes the distinct nature of strong and weak islands, with consequences for our understanding of the core architecture of syntactic movement.

Keywords: scrambling, wh-movement, constraints, Relativized Minimality, weak and strong islands

1 Introduction

A troublesome paradox emerges when comparing wh-movement and A-scrambling in Russian and various other scrambling languages. On the one hand, it is well-documented (Webelhuth 1989, Saito 1992, Bailyn 1995) that there are many similarities between the two processes in terms of parallel adherence to strong constraints on movement, such as the Complex NP Constraint, violated in (1) for both A-scrambling and wh-movement.¹

(1) a. *Kogo ty pozvonil [agentu [kotoryj ljubit ___ ]]?
   whoACC you called [spyDAT [who loves ___ ]]
   *‘Who did you call a spy who loves?’
   (*wh-mvt)
b. *Borisa ty pozvonil [agentu [kotoryj ljubit ___ ]].
   BorisACC you called [spyDAT [who loves ___ ]]
   *‘Boris you called a spy who loves.’
   (*scr)

At the same time, there are also significant differences in behavior between A-scrambling and wh-movement (Müller and Sternefeld 1993).

Many thanks for discussion to the students in my Fall 2016 Scrambling seminar at Stony Brook, as well as Andrei Antonenko, Svitlana Antonyuk, Željko Bošković, Richard Larson, Ekaterina Lyutikova, Nerea Madariaga, Andrew Nevins, Asya Perelstvaig, Sergei Tatevosov, Susi Wurmbrand, and (at least) two extremely helpful anonymous reviewers, as well as audiences at UConn, Moscow State, FDSL, FASL, SinFoniJa, and NYI St. Petersburg. All mistakes remain my own.

¹ Dislocated elements are shown in bold; original positions are indicated with a blank. Potential intermediate landing spots are not indicated unless relevant.

Linguistic Inquiry, Volume 51, Number 4, Fall 2020
635–669
© 2020 by the Massachusetts Institute of Technology
https://doi.org/10.1162/ling_a_00361

Downloaded from http://www.mitpressjournals.org/doi/pdfplus/10.1162/ling_a_00361 by guest on 22 May 2021
(2) a. *Kto ty videl [kogda [ ___ pod’ezżal ]]?
   whoNOM you saw [when [ ___ was.arriving]]
   ‘Who did you see when (he) was arriving?’
   (Müller and Sternefeld 1993:468)

   b. Ty doktor videl [kogda [ ___ pod’ezżal ]]? (✓ scr)
   you doctorNOM saw [when [ ___ was.arriving]]
   ‘Did you see when the doctor was arriving?'

The contrast in (2) relies on examples such as (2b), which are originally taken from Zemskaya 1973 and which I will refer to as Zemskaya-type sentences. Such sentences have led to influential claims about the syntax of scrambling, such as those of Müller and Sternefeld (1993), and have also motivated radical departures from derivational accounts of flexible constituent order patterns (Bošković and Takahashi 1998, Van Gelderen 2003). However, a careful examination of the nature of Zemskaya-type sentences has not been undertaken within the generative literature, and there has been no comprehensive attempt to describe the entirety of the comparative data across wh-movement and scrambling, let alone explain why wh-movement and scrambling behave differently with respect to some constraints and identically with respect to others.2

In this article, I offer a resolution of the Scrambling Paradox within the general framework of the Minimalist Program (Chomsky 1995). I show that Zemskaya-type sentences are nothing more than a well-behaved subcase of Ā-scrambling, in which an element undergoes successive-cyclic Ā-movement to a left-peripheral position in the main clause.3 Crucially, I argue that an articulated probe-goal theory, such as that found in Chomsky 2000, can account for the restrictions on wh-movement, as well as the weaker set of restrictions on Ā-scrambling, in terms of feature classes and Minimality, under a revised version of Rizzi 2004. The account allows us to maintain a derivational analysis of both processes and to avoid complex changes to movement theory in order to explain the facts at hand. I examine the interactions among subtypes of Ā-movement (i.e., whether scrambling blocks wh-movement and vice versa) and show that a proper characterization of feature classes and their interplay predicts the observed distribution of blocking effects among the various Ā-movement processes.4

---

2 One reason for the intractability of this problem is that many languages with Ā-scrambling do not also have overt wh-movement (Japanese, Hindi/Urdu, Persian, Turkish, etc.). Conversely, many overt wh-movement languages do not have Ā-scrambling (English, French, etc.). Here, I look at a language that shows both (Russian).

3 In Zemskaya’s examples, the dislocated element typically follows a discourse-anaphoric pronoun (the ty ‘you’ in (2b)), although this is not always the case. See Scott 2012 for an analysis of the unique main-clause far-left Topic position in Russian, where elements such as ty ‘you’ in (2b) might be located.

4 Throughout the article, I use Zemskaya’s (1973) corpus-attested examples wherever possible and contrast them to maximally similar examples involving wh-movement, which also can target the immediately posttopic position. Therefore, most wh-movement examples will involve the same word order pattern in the left periphery, although that particular landing site of long-distance Ā-movement is not essential. In the English translations, I use the far-left position in an attempt to render them as natural as possible, since a posttopic position for wh-movement is not available.
The article is organized as follows. In sections 2 and 3, I present the main similarities and differences between *wh*-movement and (A¯)-scrambling. In section 4, I provide an analysis of this distribution of facts in terms of classes of features and Minimality. In section 5, I argue against the alternative approach of base-generation. In section 6, I conclude by discussing the consequences for movement theory.

2 Scrambling/Wh-Movement Parallels

The existence of parallels between scrambling and *wh*-movement across languages has been discussed quite broadly in the syntactic literature (e.g., Webelhuth 1989, Bailyn 1995, 2001, Karimi 2005, Glushan 2006). Indeed, Glushan (2006:89) concludes her extensive comparison by noting that “[s]crambling in Russian parallels to *wh*-movement . . . in many respects: It is subject to . . . [the] adjunct constraint, [the] complex NP constraint and can apply successive cyclically.”

2.1 Complex NP Constraint

As we saw in (1), and as shown again in (3), the Complex NP Constraint (CNPC) constrains both *wh*-movement and scrambling. Additional evidence is given in (4)–(5) (note the similarity between the ungrammatical (4a) and the Zemskaya-type sentence (2b)).

(3) a. *Ty kogo pozvonil [agentu [kotoryj ljubit ___]]? (*wh-mvt)
   you whoACC called [spyDAT [who loves ___]]
   *‘Who did you call a spy who loves?’
   b. *Ty Borisa pozvonil [agentu [kotoryj ljubit ___]].
      you BorisACC called [spyDAT [who loves ___]]
      *‘Boris you called a spy who loves.’

(4) a. *Ty doktor znaeš’ [ljudej [kotoryx volnuet]]? (*scr)
   you doctorNOM know [people [whoACC worries]]
   *‘The doctor do you know people who worries?’
   b. *Ty kto znaeš’ [ljudej [kotoryx volnuet]]? (*wh-mvt)
      you whoNOM know [people [whoACC worries]]
      *‘Who do you know people who worries?’

(5) a. *Vy [posylku videli [mašinku, [kotoraja zapakovyvaet ___]]]? (*scr)
   you [parcelACC saw [machineACC [which wraps ___]]]
   *‘The parcel, you saw the machine that wraps?’

5 In even considering the similarities between *wh*-movement and scrambling, one is of course operating on the assumption that the two kinds of movement are in some sense distinct. This assumption is not universally shared, however. Thus, Strahov (2001) argues that Russian has no *wh*-movement, and that what appears to be *wh*-movement is in fact simply an instance of (A¯)-scrambling. However, because the former is obligatory (in nonecho contexts) and the latter optional, I will follow the more common assumption that at the very least the driving forces behind the two processes are distinct. This assumption is supported by the differences in behavior shown by the two processes, discussed in section 3, which pose a significant obstacle for claims such as Strahov’s that the two processes are identical.

6 The ungrammaticality of (4a) shows that any base-generated option for such word order variation is also subject to syntactic constraints on movement, an issue to which I return in section 5.
b. *Vy čto videli [mašinku, [kotoraja zapakovyvaet ___]]? (*wh-mvt)  
you what\textsubscript{ACC} saw [machine\textsubscript{ACC} [which wraps ___]]  
*‘What did you see a machine that wraps?’

Other well-known movement constraints restrict both processes, including the Coordinate Structure Constraint (CSC) (Ross 1967), the Constraint on Extraction Domains (CED) (Huang 1985), and the Proper Binding Condition (PBC) (Saito 1992), as I show in the next sections.

2.2 The Coordinate Structure Constraint

Displacement out of coordinate structures is impossible in Russian, as shown in (6)–(7) for both scrambling and wh-movement.\footnote{Note, as is well-known, that the left branch extraction attempted in (7) is otherwise fully acceptable in Russian; the ungrammaticality here is caused entirely by the CSC violation.}

\begin{align*}
(6) & \text{a. *Ty kogo videl [Mašu i ___]?} \hspace{2cm} (*wh-mvt) 
you who\textsubscript{ACC} saw [Masha\textsubscript{ACC} and ___] 
*‘Who did you see Masha and?’ \\
& \text{b. *My Borisa videli [Mašu i ___].} \hspace{2cm} (*scr) 
we Boris\textsubscript{ACC} saw [Masha\textsubscript{ACC} and ___] 
*‘Boris we saw Masha and.’ \\

(7) & \text{a. *Ty kakuju videl [ ___ kvartiru i [Mašin dom ]]?} \hspace{2cm} (*wh-mvt) 
you which\textsubscript{ACC} saw [ ___ apartment\textsubscript{ACC} and [Masha’s house]\textsubscript{ACC}] 
*‘Which did you see apartment and Masha’s house?’ \\
& \text{b. *My novuju videli [ ___ kvartiru i [Mašin dom ]].} \hspace{2cm} (*scr) 
we new\textsubscript{ACC} saw [ ___ apartment\textsubscript{ACC} and [Masha’s house]\textsubscript{ACC}] 
*‘NEW we saw apartment and Masha’s house.’
\end{align*}

Thus, we have good reason to maintain that the CSC constrains Zemskaya-type sentences exactly as it does wh-movement.

2.3 The Constraint on Extraction Domains

The traditional CED consists of two parts (Ross 1967, Huang 1985): one constraining movement out of subjects; the other, movement out of adjuncts.\footnote{Huang (1985) unified these under one common term (CED), although it has been more recently argued (see, e.g., Stepanov 2007) that the two are distinct effects. I will not take a stand on the nature of the two parts of this constraint and the degree to which they are related. For present purposes, it is enough to note that both are active in Russian and both constrain wh-movement and scrambling equally strictly.}

\begin{align*}
(8) & \text{a. *O kom [novaja kniga ___] udivila publiku?} \hspace{2cm} (*wh-mvt) 
about who\textsubscript{PREP} [new book ___] surprised public 
*‘Who did a new book about surprise the public?’
\end{align*}
b. *O Borise [novaja kniga ____] udivila publiku. (*scr)
   about Boris[prep] [new book ____] surprised public
   *‘About Boris a new book surprised the public.’

Similarly, extraction out of sentential adjuncts is impossible.⁹

(9) a. *Ty kogo ušel domoj, [potomu čto Maša ljubit ____]? (*wh-mvt)
   you who[acc] left to.home [because Masha loves ____]
   *‘Who did you go home because Masha loves?’

b. *Ja Borisa ušel domoj, [potomu čto Maša ljubit ____].
   I Boris[acc] left to.home [because Masha loves ____]
   *‘Boris I went home because Masha loves.’

Any theory of flexible constituent order needs to account for these parallel restrictions. Therefore, theories that propose movement to derive noncanonical word orders (i.e., scrambling) do not bear any new burden, whereas nonmovement theories do.

### 2.4 The Proper Binding Condition

The Government-Binding-era PBC captures asymmetries showing that traces must be bound, hence c-commanded, at surface structure. An English violation of the PBC is given in (10).

(10) a. *Who do you wonder [which pictures of ____]k John likes tk?
   b. *[Which pictures of ____]k do you wonder who[ realised] John likes tk?

In (10a), a smaller element is extracted by wh-movement from an already wh-moved larger element. Though the result is marginal, all traces are bound by their antecedents. However, (10b) shows the converse: here, a smaller element is extracted from a larger element, after which that larger element, containing the (now unbound) trace of who, moves to a higher position.

The PBC is known to apply to Japanese scrambling as well (Saito 1992).

(11) *[[CP Mary-ga ____1 katta to ]2 [John-ga [CP sono hon-o1 [TP [Bill-ga ____2
   [[CP MaryNom ____1 bought that]2 [JohnNom [CP that bookAcc [TP [BillNom ____2
   itta ] to ] omotteiru]].
   said]] that] think ]]
   *‘[That Mary bought ____1]2, John [that book]1 thinks that, Bill said ____2.’

In Russian, both wh-movement and scrambling are subject to the PBC. Consider (12)–(13).

(12) a. Ja znaju, čto knigu Boris xočet, [čtoby Maša pročitala ____]. (✓DP scr)
   I know that book Boris wants [that Masha read ____]
   ‘I know that the book, Boris wants Masha to read.’

---

⁹ There is some confusion about the situation with extraction out of PP adjuncts within NP, so I will avoid those here; see Rappaport 2000 for discussion.
b. [Čtoby Maša pročitala [knigu] ] j ja znaju, [(čto) Boris xočet ___].  
   [that Masha read [book ]] I know [(that) Boris wants ___ ]

   ?‘That Masha read the book, I know that Boris wants.’

   (13) a. Knigu, (zavtra) [čtoby Maša pročitala ___], j ja znaju,  
        book (tomorrow) [that Masha read ___ ] I know  
        [(čto) Boris xočet ___].

        [(that) Boris wants ___ ]

   ?‘The book, that Masha read tomorrow, I know Boris wants.’

   b. **[Čtoby Maša pročitala ___], j ja znaju, [(čto) knigu]  
      [that Masha read ___ ] I know [(that) book  
      Boris xočet ___].

      Boris wants ___ ]

   **‘That Masha read, I know that the book Boris wants.’

In (12), we see that long-distance scrambling is fine for either DP (12a) or CP (12b) arguments.10 Scrambling the larger CP first and then a DP contained within it is also fine, as in (13a) (the intervening adverb shows that we are dealing with two distinct displacement operations). However, as the PBC predicts and as (13b) shows, first moving the contained DP, and then the CP containing the trace of that DP, is strongly ungrammatical. The same is true for wh-movement.

   (14) a. ?[O čem], tebe interesno [kakie knigi ___], Maša kupila ___?  
         [about what] you interesting [which books ___ ] Masha bought ___

         *‘What did you wonder which books about Masha bought?’

   b. **[Kakie knigi ___], tebe interesno [o čem], Maša kupila ___?  
      [which books ___ ] you interesting [about what] Masha bought ___

      **‘Which books do you wonder about what Masha bought?’

In (14a), we see that subextraction of a PP from inside a larger DP wh-phrase is mildly awkward, whereas (14b) shows that extracting first the PP and later the larger containing DP is strongly ungrammatical, as the PBC predicts. This parallel is expected under a movement account of both phenomena.

2.5 Reconstruction and Antireconstruction

In standard instances, both wh-movement and ġ-scrambling show typical reconstruction effects (interpretation of the moved phrase at the base position). This can be seen through Principle C violations, which are typically not alleviated by these ġ-movement processes.

---

10 Note that for the extractions to be entirely fine, the indicative complementizer (čto) must be dropped. The marginality of extraction out of čto-indicatives is discussed in detail in section 3.3. Here, I compare the (a) and (b) sentences without the interference of the čto-indicative effect, which is clearest when čto is dropped.
(15) a. *Ona_{i} oprovergla [novye sluxi o Maše_{i}]. (Principle C violation)
    she refuted [new rumors about Masha]
    *‘She refuted the new rumors about Masha_{i}.’
b. *[Kakie sluxi o Maše_{i}], ty xočeš’, čtoby ona_{i} oprovergla ___?
    [which rumors about Masha] you want that she refuted ___?
    ‘Which rumors about Masha_{i} do you want her i to refute?’
c. *Ty [novye sluxi o Maše_{i}] xočeš’, čtoby ona_{i} oprovergla ___?
    you [new rumors about Masha] want that she refuted ___?
    ‘The new rumors about Masha_{i} do you want her i to refute?’
(based on Heycock 1995)

In (15b–c), the violation found before movement in (15a) remains even when the offending R-
expression in the surface order is no longer c-commanded by the pronoun. On movement accounts,
obligatory reconstruction (or interpretation of the lower copy) derives this effect (Heycock 1995,
Barss 2001).

However, as discussed in Lebeaux 1991 and Heycock 1995 for English and Bailyn 2001 for
Russian, there are also instances where this effect is alleviated (antireconstruction). An English
example is given in (16b).

(16) a. *[How proud of John_{i}]_{k} do you think he_{i} is t_{k}? (reconstruction)
    (Heycock 1995:549)
b. [Which question that Gore_{i} hadn’t prepared for]_{k} do you
    think he_{i} messed up on t_{k} the most? (antireconstruction)
    (Heycock 1995:561)

In Russian, wh-movement and scrambling also both show antireconstruction effects, and in the
same contexts (Bailyn 2001).\textsuperscript{11} Compare (15) with (17).

(17) a. *Ja xoču, čtoby on_{i} srazu zabyl [te voprosy, (Principle C violation)
    I want that he immediately forgot [the questions
    kotorye zadali Goru_{i}]_{k}. which_{ACC} asked_{3PL} Gore_{DAT}]
    *‘I want him to immediately forget the questions for Gore.’
b. Ty [kakie voprosy, kotorye zadali Goru_{i}]_{k} (\textsuperscript{\check{w}h-mvt antireconstruction)
    you [which questions which_{ACC} asked_{3PL} Gore_{DAT}]
    xočeš’, čtoby on_{i} srazu zabyl ___?]
    want [that he immediately forgot ___]
    ‘Which questions for Gore do you want him to immediately forget?’
\textsuperscript{11} I do not take a stand here on what the exact conditions allowing antireconstruction effects are. See Lebeaux 1991,
Huang 1993, Heycock 1995, and Barss 2001 for relevant discussion. What matters for present purposes is the parallel
behavior of wh-movement and Ā-scrambling.
c. Ja [nekotorye voprosy, kotorye zadali Goru_{3}^{k}] (\checkmark \text{scr antireconstruction})
I [some questions which_{3}^{ACC} asked_{3PL} Gore_{3}^{DAT}]
xoču, [čtoby on_{1} srazu zabyl \_\_\_\_\_\_\_].
want [that he immediately forgot \_\_\_\_\_\_\_].

‘Some questions for Gore, I want him to immediately forget.’

(17b) and (17c) behave in parallel fashion, implying a parallel derivational history.

To summarize, we have seen parallel constrained behavior of scrambling and $wh$-movement with regard to a series of known movement constraints (the CNPC, CSC, CED (subject and adjunct domains), and PBC) and that both processes undergo reconstruction and allow antireconstruction in the same contexts. Movement accounts of both processes have traditionally been supported by such parallel behavior (e.g., Bailyn 1995, 2001). However, in the next section we will see important instances where the two processes diverge, creating the core of the Scrambling Paradox.

3 Scrambling/$Wh$-Movement Differences

Despite the fairly well-known $\hat{A}$-movement similarities illustrated above, it has also been shown that scrambling and $wh$-movement differ in important ways. First, in overt $wh$-movement languages that allow scrambling (German, Russian), the scrambling of $wh$-elements is disallowed (Müller and Sternefeld 1993).\(^{12}\) Second, German and Dutch exhibit locality distinctions (Müller and Sternefeld 1993, Neeleman 1994, Grewendorf and Sabel 1999): long-distance $wh$-movement is possible, but long-distance scrambling is not.\(^{13}\) Third, scrambling is insensitive to some constraints, notably $wh$-islands. Early movement approaches (e.g., Saito 1992, Bailyn 1995) do not account for this. After an in-depth discussion of Russian scrambling, Glushan (2006:87) concludes that “$wh$-movement is more sensitive to $wh$-islands than scrambling” and decides to “leave this issue for future research.” That is, a central characteristic of this kind of movement, the core of the Scrambling Paradox, is left unanalyzed.\(^{14}\) In what follows, I describe three syntactic contexts where scrambling fares far better than $wh$-movement: interrogative $wh$-islands (section 3.1), perception verb complements introduced by the $wh$-elements kogda ‘when’ and kak ‘how’ (section 3.2), and indicative complements introduced by the complementizer čto (section 3.3). In each case, we will see clear distinctions in acceptability between $wh$-movement and $\hat{A}$-scrambling.

\(^{12}\) Whereas in in-situ $wh$-movement languages such as Japanese and Hindi/Urdu, $wh$-scrambling is well-attested (Takahashi 1993, Kidwai 2000).

\(^{13}\) Neeleman (1994) argues that long-distance scrambling in Dutch and German is available when the scrambled element serves as a Focus.

\(^{14}\) One theory that does maintain movement and attempt to account for the distinctions presented here is Müller and Sternefeld’s (1993). Following Yadroff (1991), Müller and Sternefeld describe “a surprising asymmetry between $wh$-movement and scrambling, which . . . calls for a sophisticated theory of improper movement” (p. 468). Their particular solution, involving distinct landing sites for scrambling and $wh$-movement, and the Principle of Unambiguous Binding, is not available in a Minimalist framework, but shares various characteristics with the account given here, as a movement analysis that distinguishes between $wh$-movement and scrambling. The two accounts are thus similar in spirit, though the account given here explains the Scrambling Paradox in a way that Müller and Sternefeld’s account cannot. See section 4 for more discussion.
3.1 Interrogative Wh-Islands

In Zemskaya-type sentences, dislocation out of \textit{wh}-islands is essentially unconstrained (although adjuncts, as in (18c), are somewhat degraded, for reasons that are not immediately clear).

(18) a. Ty \textit{musor} sprosil, [kogda uvozjat \_\_\_]? (\checkmark obj scr)
you trash\textsubscript{ACC} asked [when take.away \_\_\_]
‘Did you ask when they take the trash away?’
Literally: ‘The trash, did you ask when they take [it] away?’
b. Ty \textit{doktor} sprosil, [kogda ___ pod’edet ]? (\checkmark subj scr)
you doctor\textsubscript{NOM} ask [when ___ will.arrive]
‘Did you ask when the doctor would arrive?’
Literally: ‘The doctor, did you ask when [he] will arrive?’
c. ??Ty \textit{molotkom} sprosil, [kogda [Mitja \textit{\_\_\_\_\_} inil ma...]
you hammer\textsubscript{INSTR} ask [when [Mitya fixed car ___ ]]?
‘Did you ask when Mitya fixed the car with a hammer?’

On the other hand, \textit{wh}-movement out of interrogative \textit{wh}-islands yields a typical weak-island effect, whereby object movement, though degraded, is far better than subject or adjunct movement, as documented extensively for English \textit{wh}-movement (e.g., Rizzi 1990).

(19) a. ??\textit{Čto} ty sprosil, [kogda uvozjat \_\_\_]? (??obj \textit{wh-mvt})
what\textsubscript{ACC} you asked [when take.away \_\_\_]
??‘What did you ask when they take away?’
b. *Ty \textit{kto} sprosil, [kogda ___ pod’ezžaet]? (*subj \textit{wh-mvt})
you who\textsubscript{NOM} asked [when ___ is.arriving]
*‘Who did you ask when was arriving?’
c. *Ty \textit{čem} sprosil, [kogda [Mitja činil mašinu ___ ]]? (*adj \textit{wh-mvt})
you what\textsubscript{INSTR} asked [when [Mitya fixed car ___ ]]?
*‘With what did you ask when Mitya fixed the car?’

Thus, we see weak-island behavior with \textit{wh}-movement, which is as expected. What is unexpected is the acceptability of \~A-scrambling in similar contexts.

3.2 Perception Verb Complements

Zemskaya (1973) provides a multitude of examples of displacement out of perception verb complement clauses introduced by the \textit{wh}-phrases \textit{kogda} ‘when’ and \textit{kak} ‘how’.

\footnote{The examples in (21) do not adhere to the usual Zemskaya-type pattern whereby the displaced element follows the topical pronoun. However, the contrasts illustrated are not dependent on this choice; with the posttopic position, the same distinctions hold. I therefore use the original examples verbatim.}
a. Ty musor slyšala, [kogda uvozili ___ ]?
you trash_{ACC} heard [when took.away ___ ]
‘Did you hear them taking the trash away?’
(Zemskaya 1973:399)

b. Ty doktor videl, [kogda ___ pod’ezžal ]?
you doctor_{NOM} saw [when ___ was.arriving]
‘Did you see the doctor arriving?’
(Zemskaya 1973:399)

(21) a. Mitja ty slyšal, [kak ___ čitaet ]?
Mitya you heard [how ___ reads ]
Not exactly: ‘Did you hear Mitya reading?’
Not exactly: ‘That Mitya! Did you hear how he was reading??’

b. Vasja ty pomniš’, [kak ___ rasskazyval]?
Vasya you remember [how ___ told.stories ]
Not exactly: ‘Do you remember Vasya telling stories?’
Not exactly: ‘That Vasya! Do you remember how he was telling stories??’

Parallel examples with _wh_-movement are highly degraded.16

(22) a. ???Ty čto slyšala, [kogda uvozili ___ ]?
you what_{ACC} heard [when took.away ___ ]
‘What did you hear them taking away?’

b. *Ty kto videl, [kogda ___ pod’ezžal ]?
you who_{NOM} saw [when ___ was.arriving]
*‘Who did you see when arrived?’

(23) a. *Kto ty slyšal, [kak ___ čitaet]?
who_{NOM} you heard [how ___ reads ]
Intended: ‘Who did you hear reading?’
Literally: *‘Who did you hear how reads?’

b. *Kto ty pomniš’, [kak ___ rasskazyval]?
who_{NOM} you remember [how ___ told.stories ]
Intended: ‘Who do you remember telling stories?’
Literally: *‘Who do you remember how told stories?’

Zemskaya does not provide adjunct examples parallel to those given here; however, given that subject extraction, typically bad in such contexts with _wh_-movement, is fine here, it is not surprising to find that adjunct extraction is also fine.

16 Note that for some speakers the object _wh_-displacement cases are markedly better than subject displacement. This is as expected in a weak-island environment.
(24) a. Ty doktor videl, [kogda pod’ezžal ]?  
    you doctorNOM saw [when was.arriving]  
    ‘Did you see the doctor arriving?’ (same as (20b))  

b. Vy posylku videli, [kak zapakovali [ ]]]?  
    you parcelACC saw [how [packed.up [ ]]]  
    ‘Did you see how they did up the parcel?’  

(Müller and Sternefeld 1993:467)  

c. Ty molotkom videl, [kak [Mitja činil mašinu [ ]]]?  
    you hammerINSTR saw [how [Mitya fixed car [ ]]]  
    Not exactly: ‘Did you see Mitya fixing the car with a hammer?’  
    Not exactly: ‘A hammer! Did you see Mitya fixing the car with one?’  

Again, as we have already seen in (22a–b), similar constructions with wh-movement are ill-formed.

(25) a. ??Vy čto videli [kak zapakovali [ ]]?  
    you whatACC saw [how [packed.up [ ]]]  
    ‘Did you see how they did up the parcel?’  
    (based on Müller and Sternefeld 1993:468)  
    Intended: ‘What did you see them pack up?’  
    Literally: ‘What did you see how they packed up?’  

b. *Ty kak videl [kak Mitja činil mašinu [ ]]?  
    you how saw [how [Mitya fixed car [ ]]]  
    Intended: ‘How did you see Mitya fixing the car?’  
    Literally: ‘How did you see how Mitya fixed the car?’

Here again, it is the wh-movement violations in (25) that are expected (due to the island formed by the kak-phrase) and the well-formedness of the non-wh displacements that is unexpected.

3.3 Movement out of Čto-Indicatives

Zemskaya (1973:398–405) gives quite a few examples of displacement out of indicatives introduced by the complementizer čto, shown in (26), of which the last three are quoted in Müller and Sternefeld 1993:467.

17 A reviewer points out a problem with examples like (25a–b), namely, that there are two kinds of kak-phrases: true how-adject islands and declarative clauses embedded under perception verbs. And indeed, Glushan (2006) observes that semantically, kak-phrases are ambiguous between these two readings. However, syntactically these are wh-islands in either interpretation, as shown by the degraded nature of (25a–b), an effect Glushan does not discuss. Note that with a change in matrix verb, the latter reading is excluded and the ungrammaticality of the resulting attempted question confirms that the source of the ungrammaticality is the wh-island.

(i) *Čto vy sprosili, [kak zapakovali [ ]]?  
    what you asked [how [packed.up [ ]]]  
    ‘What did you ask how they packed up [ ]?’
(26) a. Ogurcov žal’, [če to malo ___].
   pickles too.bad [that there are few ___]
   ‘Pickles, it’s too bad that there are so few of [them].’

b. Plašč mne ne nравится, [če to pridetsja s soboj taščit’ ___].
   raincoat me NEG likes [that have to with self bring ___]
   ‘The raincoat, I don’t like (the fact) that I have to bring with me.’

c. Konfety on skazal, [če ___ vkusnye].
   candies he said [that ___ tasty ___]
   ‘These candies he said are tasty.’

d. Vot bumagi mne neprijatno, [če vy ne kupili ___].
   here paper me unpleasant [that you NEG bought ___]
   ‘The paper, it’s unpleasant that you didn’t buy [it] for me.’

e. Buločki skazali, [če ___ čerstvye].
   rolls said [that ___ stale ___]
   ‘The rolls they said are stale.’

f. Ego sestra govorjat, [če ___ priexala].
   his sister say [that ___ arrived ___]
   ‘His sister they say arrived.’

g. On skazal, čto noski [on rad [če kupil ___]].
   he said that socks [he glad [that bought ___]]
   ‘He said that the socks he is glad that he bought.’

h. Mne Katju kažetsja, čto [otpustit’ ___ odnu tak pozdno] bylo by me
   KatyaACC seems [to let go ___ alone so late ___] would be
   ‘It seems to me that it would be insane to allow Katya out alone so late.’

i. . . . čto Petrov stranno, [če [___ nam pomogal]].
   that PetrovNOM odd [that [___ us helped ___]]
   ‘. . . that Petrov it is odd (*that) helped us.’

---

18 A reviewer correctly points out that one might expect this example to be judged ill-formed, given that it involves displacement out of a sentential subject (a strong island) and hence violates the CED, which as we have seen constrains A-scrambling as strongly as wh-movement. However, Stepanov (2007) has shown that languages vary with regard to the severity of subject island effects and that Russian, in particular, does not show a severe extraction effect in infinitival subject constructions exactly like this one. Thus, the acceptability of (26h) is not an exception to Russian’s adherence to strong islands (such as adjunct islands), illustrated above. (The same holds for (27h).)
In all of these examples, a non-*wh* element is displaced from within an embedded čto-indicative clause and the results are essentially fine. Crucially, *wh*-extraction from the same contexts is degraded.

(27) a. *Čego žal’, [čto malo ___ ]?
   what too.bad [that there.are.few ___ ]
   ‘What is it too bad that there are so few of?’

b. ??Kakie vešči tebe ne nравится, [čto priedetsja s soboj taščit’ ___ ]?
   what things you NEG likes [that have.to with self bring ___ ]
   ‘What things don’t you like (the fact) that you have to bring with you?’

c. ??Kakie konfety on skazali, [čto ___ vkusnej]?
   which candies he said [that ___ tasty ]
   ‘Which candies did he say [that] were tasty?’

d. ??Čto neprijatno, [čto vy ne kupili ___ ]?
   what unpleasant [that you NEG bought ___ ]
   ‘What is it unpleasant that you didn’t buy?’

e. ??Kakie buločki skazali, [čto ___ čerstvyej]?
   which rolls said [that ___ stale ]
   ‘Which rolls did they say were stale?’

f. *Kto govorjat, [čto ___ priexal]?
   who NOM say [that ___ arrived]
   ‘Who did they say arrived?’

g. ??Čto on skazal, [čto [on rad [čto kupil ___ ]]]?
   what he said [that [he glad [that bought ___ ]]]
   ‘What did he say that he is glad that he bought?’

h. ??Kogo kažetsja, [čto [optusit’ ___ odnogo tak pozdno]] bylo by
   who ACC seems [that [to.let.go ___ alone so late ]] would be
   insality
   ‘Who does it seem that it would be insane to allow out alone so late?’

i. *Kto stranno, [čto [ ___ nam pomogal]]?
   who NOM odd [that [ ___ us helped ]]
   ‘Who is odd (*.that) helped us?’

19 Note that examples (26a,b,d,g,h) involve underlying objects (either of N or of V), whereas examples (26c,e,f,i) involve subjects. For objects, nothing here is unexpected—it is normal for objects to be successfully fronted out of indicative embedded contexts, as the English translations of (26b) shows. Instances of successful subject displacement are somewhat unexpected, since those often trigger *that*-effects when moved over overt complementizers (e.g., Rizzi 1990). However, the *that*-effect is also obviated when the element does not move from Spec,TP (Rizzi 2007, Stepanov 2007), which is possible with unaccusatives and subjects of copular predicates, as in (26c,e,f). It thus appears that (26i) is degraded because a true Spec,T subject is fronted across an overt complementizer. However, I leave the proper characterization of Russian *that*-violations to future research. See Pesetsky 1982 for discussion.

20 A similar contrast is found in Japanese (see Saito 1992, Bošković and Takahashi 1998).
In cases of extraction of *wh*-phrases from such contexts, note that there exist significant subcontrasts in acceptability. Direct object extraction in (27b,g,h) results in a mild violation, whereas subject extraction is worse. Of course, this is highly reminiscent of weak-island effects (Rizzi 1990), for which what is relevant is the degradation in all cases—the amelioration with objects is related to independent factors (Rizzi 1990, 2004). Thus, all cases are instances of the Scrambling Paradox, given that the scrambling counterparts, as shown above, are essentially fine.

When adjunct phrases are added to the picture, the puzzle becomes more familiar, in that adjunct *wh*-extraction out of indicative čto-clauses is worse than object *wh*-extraction, while scrambling is fine.

(28) a. Včera govorjat, [čto ego sestra priexala ____].
   ‘Yesterday they say that his sister arrived.’

b. *Ty kogda dumaeš’, [čto ego sestra priexala ____]?
   ‘When do you think that his sister arrived?’

Given this contrast, the proper characterization of čto-clauses appears to be that they are weak islands for *wh*-movement—far worse for adjuncts than objects. Indeed, the argument/adjunct asymmetry is quite strong; compare (29) and (30).

(29) ??[Kakuju knigu] ty dumaeš’, [čto Petr pročital ____]? (obj extraction)
   ‘Which book do you believe that Peter read?’
   (Müller and Sternefeld 1993:466; judgment based on my consultants)

(30) *Kak ty dumaeš’, [čto Petr počinil mašinu ____]? (adj extraction)
   ‘How do you believe that Peter fixed the car?’

Thus, we can conclude that čto-clauses behave as *wh*-islands do: they induce the adjunct/argument asymmetry for *wh*-movement but do not constrain scrambling. Descriptively, I refer to this as the Čto-Clause Weak-Island Restriction.

(31) The Čto-Clause Weak-Island Restriction
Čto-clauses are weak islands for *wh*-movement.

---

21 The well-known amelioration effect of *0*-marked objects moving out of weak islands (Rizzi 1990, 2004) is assumed to have a source distinct from the blocking effects found under Relativized Minimality. See Rizzi 1990, 2004 and Bailyn 2018 for discussion of the nature of that amelioration. For present purposes, these cases are consistent with the cases given in (27g–i). See also footnote 29.

22 In (27a), *wh*-extraction from within a quantificational nominal appears to be independently further degraded (Bailyn 2012), although, as a reviewer points out, extraction of adnominal elements is not universally ill-formed. But (27a) is clearly worse than cases without the subextraction.

23 A similar effect is found in Polish and other Slavic languages (see Lubańska 2005, Orszulak 2010).
Any successful theory of displacement should of course address both the weak-island-like nature of Slavic long-distance \(wh\)-movement and the acceptability of non-\(wh\) displacement in similar contexts. In section 4, I propose an approach to resolve this issue.

### 3.4 Summary of Displacement Data

Table 1 summarizes the data any theory of displacement facts must cover. Lines (a)–(e) show we are dealing with movement: \(wh\)-movement and scrambling both violate the CNPC, CSC, PBC, and CED, and reconstruction feeds a Principle C violation in parallel instances. On the other hand, lines (f)–(h) tell us that although \(wh\)-movement of arguments out of interrogative \(wh\)-islands, \(kak\)-phrases, and \(čto\) (indicative) clauses is mildly degraded and \(wh\)-movement of adjuncts is highly degraded, in a manner similar to \(wh\)-movement out of English weak islands, scrambling in these contexts is fine.

This separation of constraint behavior is very revealing: so long as (31) (the weak-island behavior of \(čto\)-clauses) can be accounted for independently, the Scrambling Paradox reflects a distinction in sensitivity to \(wh\)-islands. In particular, it can be reduced to two questions, one global and one language-specific: (a) why is scrambling not sensitive to \(wh\)-islands? and (b) why do Russian \(čto\)-clauses pattern with (traditional) \(wh\)-islands? The remainder of this article is an attempt to answer these two questions and thus situate the Scrambling Paradox within a broader picture of movement typology.

### 4 A Feature Class Account of the Scrambling Paradox

#### 4.1 Movement Features and Relativized Minimality

As we have seen, indicative \(čto\)-clauses behave like weak \(wh\)-islands. Without providing a full account of why this is so, I will proceed under this assumption. If \(čto\)-clauses are a kind of \(wh\)-island, then the differences between \(wh\) and non-\(wh\) displacement can be reduced to this question: why is scrambling not sensitive to \(wh\)-islands?
The answer follows from a specific application of the more articulated theory of Relativized Minimality proposed in Rizzi 2004, the kind of relativized Relativized Minimality that Rizzi (2018) refers to as “Featural Relativized Minimality.” On this approach, A-blocking of the kind familiar from Rizzi’s (1990) Relativized Minimality is further relativized with respect to the feature classes of the elements involved in the movement and the potential blocking (“[s]ome finer typology is then needed” (Rizzi 2004:229)). Crucially, I assume that A-scrambling is feature-driven movement, in the spirit of Grewendorf and Sabel (1999) and Kawamura (2004), who argue for the scrambling feature [+Σ], which is attracted to the left periphery. In what follows, I will assume that [+Σ] drives scrambling, while the full set of [−Q] elements includes [−Mod] for base-generated adjuncts and [+Top] for topicalized elements, as in Rizzi 2004. (32a–cii) present Rizzi’s (2004) classification, with the addition in (32ceiii) of the feature that drives scrambling.

(32) Classes of features
   a. Argumental: person, number, gender, case
   b. [+Q] Quantificational: Wh, Neg, Measure, Focus...
   c. [−Q] Nonquantificational:
      i. [−Mod] Modifiers: evaluative, epistemic, Neg, frequentative, measure, manner, ...
      ii. [+Top] Topic
      iii. [+Σ] Scrambling

Thus, the predicted state of affairs for extraction is summarized in (33).

24 Describing certain properties of Russian scrambling under Rizzi’s (2004, 2018) Featural Relativized Minimality approach is not an entirely novel idea; something similar was proposed for Russian case “confusion” constructions in Yadroff 1992 and for adverb scrambling in Shields 2012. Here, I present a system that generalizes ideas in those works.

25 Two observations are in order here. First, I do not assume any necessary Information Structure component for scrambling. Thus, as we will see, [+Σ] elements can carry [+Foc], but need not. There is controversy around the question of whether or not scrambling always reflects some change in information structure; like Miyagawa (2006), I assumed in Bailyn 2001 that it does, but much of the literature on Russian syntax and intonation, such as Lyutikova 2009, does not. Therefore, I will assume that [+Σ] and [+Top] are distinct features. The interaction among these [−Q] features is discussed below. Thanks to an anonymous reviewer for discussion of this issue.

26 (32a) contains “argumental” features, which I ignore below, because they do not interact directly with the A-system (Rizzi 1990). Ideally, given this more articulated theory of features, one could do away with the A/A¯ distinction entirely, deriving the original A vs. A¯ asymmetries in blocking the same way that wh-movement and scrambling will be distinguished here. Other A vs. A differences might also be handled under an articulated theory of features (see Bailyn 2002 and Shields 2012 for such an approach to A- vs. A-scrambling).

27 Abels (2012) argues that the A-features in (32) (types (b)–(c)) form a set of nested subset/superset feature dependencies, in which subset features block others of the same type as well as those that are in their superset. In section 4, I consider what kind of subclassifications of features the Russian data lead us to posit for the A-system. For now, I will simply work with [±Q] and [Σ], adding [−Mod] and [−Top] to the mix directly below.

28 I assume that “blocking” happens as follows: In a probe-goal system, following Chomsky (2000), attracting features require a match in feature class to establish an Agree relation, which then leads to movement, if required by a “strong” or generalized EPP feature. Blockers prevent the establishment of the Agree relation necessary to begin this process, and the constraint effects appear. For partially successful extraction of objects out of weak islands, see footnote 29.
(33) Feature-class Relativized Minimality predictions for extraction
   b. [−Q] elements do not block [+Q] elements.

(I return in section 4.3 to the issue of blocking of [−Q] elements by other [−Q] elements.) Thus, the derivation of a *wh*-question and an instance of Ā-scrambling will be as follows:

(34) Derivation of simple Ā-relations
   a. Wh-movement
   \[
   C_{[+Q(wh)]} \left[ \ldots XP_{[+Q(wh)]} \ldots \right]
   \]
   (feature match)
   b. Ā-scrambling
   \[
   F_{[+Σ]} \left[ \ldots XP_{[+Σ]} \ldots \right]
   \]
   (feature match)

(35) Relativized Minimality effects (e.g., *wh*-islands)
   a. Wh-island
   \[
   C_{[+Q(wh)]} \left[ YP_{[+Q(wh)]} \left[ \ldots XP_{[+Q(wh)]} \ldots \right] \right]
   \]
   (feature block)
   b. Scrambling out of *wh*-island
   \[
   F_{[+Σ]} \ldots \left[ YP_{[+Q(wh)]} \left[ \ldots XP_{[+Σ]} \ldots \right] \right]
   \]
   (no feature block)
   (feature match)

The [+Q] feature can establish an Agree relation with the lower [+Q] feature in (34a), and the same goes for [+Σ] in (34b). In (35a), the relationship is impeded by the intervening YP bearing the same feature. Crucially, however, [+Q] does not block [+Σ] in (35b).

Thus far, the account rides on Russian being sensitive to the distinction between [+Q] and [−Q] (and in particular [+Σ]) features. The core prediction is that distinct feature types will not block elements bearing features of other classes. That the feature-class system can go a long way in accounting for the patterns observed with Russian scrambling and *wh*-movement supports a movement account of the relevant phenomena. However, some adjustments to the system will be necessary. Next, we examine closely the interactions among classes of Ā-features with regard to movement and blocking.

4.2 [+Q] Interveners

In Rizzi 2004, all [+Q] features involve some form of quantification (Rizzi lists [wh], [Neg], [Measure], and [Focus]). Here, I survey relevant interactions among some of these features and show that Rizzi’s generalization holds more broadly in Russian: [+Q] elements block other [+Q] elements, but [+Q] and [+Σ] elements do not block each other.
4.2.1 [+wh] Blocks [+wh], but Does Not Block [+Σ] With respect to [+Q] features, we saw in (20) and (21) that [+wh] blocks [+wh], but does not block [+Σ]. The former is the standard case of wh-islands, familiar from many languages. The relevant examples, (22a) and (20a), are repeated here.

(36) a. ???Ty čto+[wh] slyšala, [kogda+[wh]] uvozili ___ ]?
‘What did you hear them taking away?’

b. Ty musor+[Σ] slyšala, [kogda+[wh]] uvozili ___ ]?
‘Did you hear them taking the trash away?’

(36a–b) are cases of split wh-extraction from the specifier of a DP across a potentially blocking adverbial element. Russian shows a similar contrast.30

4.2.2 [+Q] Blocks [+wh]; [−Q] Does Not Rizzi (2004) shows for French that the quantificational adverb ([+Q]) beaucoup ‘many’ blocks wh-movement (a finding also reported in Rizzi 1990), whereas purely modificational adverbs (Rizzi’s (2004) [+Mod] elements) such as attentivement ‘attentively’ do not.

(37) a. *Combien+[wh] a-t-il beaucoup+[Q] consulté [ ___ de livres]?
‘How many books has he consulted a lot of?’

b. Combien+[wh] a-t-il attentivement+[Mod] consulté [ ___ de livres]?
‘How many books has he consulted attentively?’

(37a–b) are cases of split wh-extraction from the specifier of a DP across a potentially blocking adverbial element. Russian shows a similar contrast.30

(38) a. *Kakogo+[wh] on mnogo+[Q] citiroval [ ___ avtora]?
‘Which author did he cite a lot?’

29 Note that Rizzi (1990, 2004) does not consider the well-known “weak” effects of wh-islands, whereby objects are less severely constrained than moved subjects and adjuncts, to undermine the notion of Relativized Minimality as a constraint on movement. Rather, the amelioration that is found with argument extraction results from independent factors such as referentiality (Rizzi 1990). As Rizzi (2004:231) puts it, “[T]he only wh-elements successfully extractable from an indirect question are arguments with special interpretive properties (specific, presupposed, D-linked).” I do not attempt to account for variation of adjunct/argument asymmetries with different kinds of blockers (if such exist) in this article, but see Bailyn 2018 for a possible account.

30 Thanks to a reviewer for helpful adjustment of these examples. Note that I do not use the posttopic position for the dislocated elements here because the adjacency between the displaced element and the quantificational adverb causes an additional confound.
b.  **Kakogo** [+wh] on **včera** [+Mod] citiroval [ __ avtora]? (✓wh-mvt over včera)
   whichACC he yesterday cited [ __ author]ACC
   ‘Which author did he cite yesterday?’

In (38a), the [+Q] feature on *mnogo* ‘much’ blocks the establishment of an Agree relation with the lower [+wh] feature, rendering movement impossible. So (38a) cannot be derived. However, the nonquantificational adverb *včera* ‘yesterday’ does not block [+wh] in (38b).

4.2.3 [+Q] Does Not Block [+Σ] As now expected, scrambling out of contexts similar to those just illustrated is fine.

   (39) **Étogo** [+Σ] on **mnogo** [+Q] citiroval [ __ avtora].
       thisACC he much cited [ __ author]ACC
       ‘This author he cited a lot.’

4.2.4 [+Foc] Blocks [+wh], but Does Not Block [+Σ] Focus adverbs carry a [+Q] feature (see Rizzi 1997 for extensive argumentation that Focus is quantificational) and therefore should block (at least long-distance) *wh*-movement.31

   (40) a.  Ty **komu** [+wh] xočeš’, [ čtoby ona pozvonila ___ ]? (✓wh-mvt)
           you whoDAT want [that she called ___ ]
           ‘Who do you want her to call?’
   b.  *Ty **komu** [+wh] tol’ko [+Foc] xočeš’, [ čtoby ona pozvonila ___ ]? (✓wh-mvt over only)32
           you whoDAT only want [that she called ___ ]
           ‘Who do you only want her to call?’ [* on reading where ‘only’ restricts ‘who’]

However, ‘only’ does not block Ā-scrambling.

   (41) a.  Ty **Ivanu** [+Σ] xočeš’, [ čtoby ona pozvonila ___ ]? (✓scr)
           you IvanDAT want [that she called ___ ]
           ‘Ivan, do you want her to call?’
   b.  Ty **Ivanu** [+Σ] tol’ko [+Foc] xočeš’, [ čtoby ona pozvonila ___ ]? (✓scr over only)
           you IvanDAT only want [that she called ___ ]
           ‘Ivan, do you only want her to call?’

31 It is important to limit this discussion to long-distance *wh*-movement because long-distance *wh*-movement is necessarily Ā-movement, driven by exactly the features we are testing for. More local movements (traditional Ā-movements) might first involve movement of [+Arg] (A) features, not subject to blocking by any of the Ā-features, which might void the relevant blocking effect, in a manner similar to Ā-movement not being subject to Weak Crossover effects.32 A reviewer points out that the Focus-marking adverb *tol’ko* ‘only’ might itself not be an intervener; rather, it might determine a Focus domain that is opaque, so we might be seeing a case of dominance blocking rather than c-command blocking. Regardless of such a possible implementation, the scrambling vs. *wh*-movement contrasts remain and can be accounted for as proposed here.
4.2.5 [+wh] Blocks [+Foc] The degree to which Russian scrambling can serve as an example of contrastive Focus movement is controversial (see Neeleman and Titov 2009 for discussion). However, a survey of 15 native speakers shows that most feel that if strong contrastive Focus intonation is applied on the element scrambled out of wh-islands (shown here in capitals), the results are significantly worse than the ostensibly similar scrambling examples. This is shown in (42) (compare with typically successful Zemskaya-type sentences such as (2a)).33


   DOCTORNOM students asked [when ___ arrived]
   ‘[It’s] the DOCTOR the students asked when [he] arrived.’


   MASHANOM students asked [when ___ arrived ]
   ‘[It’s] MASHA the students asked when [she] arrived.’

In section 4.5, I return to syntactically marked instances of Focus fronting, which show similar results.

4.2.6 Scrambling of [+Q] Elements If the feature typology proposed here is on the right track, we might expect lexically quantified nominals (každyj mal’čik ‘every boy’, vse pesni ‘all the songs’, etc.) that scramble to be blocked by [+Q] elements.34 That is, we might expect them to be sensitive to wh-islands in a way that other scrambled nominals are not (recall that we have seen that scrambling is typically fine out of wh-islands). In fact, however, these cases seem to pattern with scrambling and not with wh-movement. Thus, if we take the sentences from (20), repeated here as (43), and use quantified elements as the scrambled items, as in (44), the derivations are not degraded in any significant way.

(43) a. Ty musor[+Σ] slyšala, [kogda[+wh] uvozili ___ ]? (✓[-Q] scr over [+wh])

   you trashACC heard [when took.away ___ ]
   ‘Did you hear them taking the trash away?’
   (Zemskaya 1973:399)

   b. Ty doktor[+Σ] videl, [kogda[+wh] ___ pod’ezžal ]? (✓[-Q] scr over [+wh])

   you doctorNOM saw [when ___ was.arriving]
   ‘Did you see the doctor arriving?’
   (Zemskaya 1973:399)

(44) a. Ty [vsex studentov][+Q], [+Σ] slyšala, [kogda[+wh] ___ ]

   you [all students ]ACC heard [when ___ ]

33 A reviewer asks whether it is significant that the focused element in these examples precedes the Topic (if so, that would require additional discussion). However, it appears that given the proper intonation pattern that accompanies focused constructions such as these, the relative order with respect to the Topic is not crucial to the contrasts at hand, just as it is not crucial above. The survey conducted used this order, so I will leave these sentences as judged.

34 Thanks to an anonymous reviewer for suggesting that I examine these cases.
uvozili took.away ‘Did you hear them taking all the students away?’

b. Ty [každyj doktor][+Q], [+Σ] videl, [kogda][+wh] (✓ [+Q] scr over [+wh])
you [every doctor]NOM saw [when] was.arriving
‘Did you see every doctor arriving?’

The fact that (44a–b) are essentially no worse than (43a–b) indicates that the lexically [+Q] nature of the scrambled elements does not trigger any significant blocking.

Similarly, scrambling quantified nominals out of our other [+Q] contexts is also fine: across a quantificational adverb (compare (45a) with (45b)) and across a Focus-marking adverbial (compare (41b), repeated as (46a), with (46b)).

(45) a. Étogo[-Q] on mnogo[+Q] citiroval [ ___ avtora],
thisACC he much cited [ ___ author]ACC
‘This author he cited a lot.’

everyACC he much cited [ ___ author]ACC
‘Every author, he cited a lot.’

you IvanDAT only want [that she called ___ ]
‘Ivan, do you only want her to call?’

you [each student ]DAT only want [that she called ___ ]
‘Every student, do you only want her to call ___ ?’

I assume that the acceptability of (44), (45b), and (46b) is related to the fact that the feature triggering movement ([+Σ]) is [−Q], and the [+Q] element carried lexically by the quantified nominal is not involved in the feature-matching process that initiates movement. Thus, we can conclude that only the feature type that is being probed for in a certain derivation is subject to blocking; hence, scrambled quantifiers behave as [−Q] elements, despite the presence of a [+Q] feature within the feature bundle that is not part of the Agree relation. This motivates a reorganization of the feature bundle associated with a lexical item that we can call marking for scrambling, of the following form:
(47) Marking for scrambling

a. Features involved in marking for scrambling
   i. \([+\Sigma]\) + ii. \([+D]\) (or \([+Q]\), etc.) →

b. Feature bundles after marking for scrambling

\[
\begin{align*}
[+\Sigma] \\
\{[+D]/[+Q]\}
\end{align*}
\]

Essentially, (47) shows how a DP (or CP or other element) to be scrambled is marked, via the creation of a feature bundle whose only visible feature is \([+\Sigma]\), that is, a phrase that must be the goal of a higher scrambling probe.\(^{35}\) Thus, we now have an explicit claim about the nature of scrambled elements: they behave as syntactic \(\Sigma\)Ps, and if the DP that \([\Sigma]\) embeds had a lexical \([+Q]\) feature, that feature is not reflected in the visible features of the phrase. Hence, scrambled quantifiers are not subject to \([+Q]\) blocking.\(^{36}\)

4.2.7 The \([+Rel]\) Feature  Relative clauses themselves are strong islands, and this holds in Russian, as we have seen, constraining \(wh\)-movement and scrambling equally severely. This is unsurprising, and is well-known from other languages (e.g., Rizzi 1990). However, as Lyutikova (2009) has shown, following Testelets (2001), the Russian relative pronoun \(kotoryj\) ‘which’ behaves at times more like a scrambled \([\sim Q]\) element than a \([+Q]\) element, in that Russian corpora provide multiple instances of successful relativization out of \(wh\)-islands. One of Lyutikova’s many examples is given here:

(48) I tut pojavljaetsja novyj mir, v kotorom ja ne znaju [kak žit’ ___ ].

‘And here a new world appears in which I don’t know how to live.’

(Lyutikova 2009:472)

\(^{35}\) I express this in tree format to clarify the prominence of the scrambling feature in such instances. This is an explicit statement of what is implied in Kawamura 2004; there, Kawamura argues for the \([+\Sigma]\) feature for scrambling exactly so that no minimality requirement forces the probe to prefer non-\(\Sigma\)-marked DPs higher in the structure (so that objects can scramble over subjects, etc.).

\(^{36}\) Clearly, there are significant theoretical consequences that follow from this conclusion about how elements are marked for scrambling (and potentially other processes)—in particular, that all features of an element are not fused into a single feature set with all features equally active in any Agree process. Rather, particular features are probed in particular derivations and other features of the goal in question are not relevant to blocking configurations. One especially interesting consequence involves scrambling of \(wh\)-elements themselves, which should also be free of \([+Q]\) blocking (which is mostly the case, as shown in Takahashi 1993). Russian does not allow \(wh\)-scrambling (Bailyn 2012), a property that is related to its overt \(wh\)-movement status. The theory of features given here would exactly predict that overt \(wh\)-movement languages should not allow syntactic marking for scrambling on \(wh\)-elements, since the \([+wh]\) feature would no longer be visible for later true \(wh\)-movement. The fact that Japanese \(wh\)-scrambling can have the side effect of \(wh\)-scope marking (Takahashi 1993) needs to be reexamined in this light. I leave further discussion of these consequences for future research.
Clearly, [+Rel] does not behave as a [+Q] feature in such derivations.\footnote{An anonymous reviewer asks about the compatibility of this account with the clearly quantificational semantic behavior of relative operators at least in restrictive relative clauses. This is an important question, but it is clear that syntactically, kotoryj ‘which’ behaves as a [−Q] element. The exact connection between the semantic representation and the syntactic derivation is beyond the scope of this article.} Note that this is also not unique to Russian. In fact, English shows a similar contrast between wh-movement and relativization out of wh-islands.

(49) a. the book which the journalist doesn’t remember who wrote ___      (relativization)
b. *What do you remember who wrote ___ ?                          (wh-mvt)

The same is true in Italian. For example:

(50) Tuo fratello, a cui mi domando [che storie abbiano raccontato ___ ], era molto preoccupato.

‘Your brother, to whom I wonder which stories they told, was very troubled.’

(Abels 2012:236, quoting from Rizzi 1982:50)

The theory given here provides a solution for this otherwise unexpected contrast: in instances of relativization, such as (48) and (49a), there is a feature match between two elements bearing [+Rel], which, despite having the form of a [+wh] element, behaves as a [−Q] element, being essentially modificational. This seems incompatible with the relative pronoun being a [+wh] element, but is in fact consistent with the fact that these elements are not actually quantificational in the same way that [+Q] questions are. Abels (2012) identifies a similar asymmetry and concludes, following Starke (2001), that features are organized into “subclasses and superclasses” and that

\[\text{the construction of movement dependencies and the application of Relativized Minimality can then be understood in terms of the elsewhere or Pāṇini principle: the application of a more specific process preempts the application of a less specific one. Thus, an element that belongs only to a superclass will always move as a member of that superclass and this movement will be blocked by any intervener from that superclass. An element that belongs to a subclass, however, will be able to undergo the more specific rule of moving elements in that subclass and be able to circumvent blocking by elements in the superclass. Itself, it will block elements in the superclass and in the subclass. (Abels 2012:248)}\]

In the case of relativization, we are dealing with a [+wh] element being used in a modificational context. Abels (2012) represents this as follows:

\[
\begin{array}{c}
\text{Modifier} \\
\text{Q}
\end{array}
\]

\[(+Q) \text{ elements that are also modificational} \] (based on Abels 2012:248, (36))
This then instantiates Abels’s superset condition and explains why the attracted feature behaves as a [−Q] element with respect to potential [+Q] blockers such as \(wh\)-islands. That is, we end up with the satisfying result that the insensitivity of relativization to \(wh\)-islands has the same explanation as successful scrambling out of \(wh\)-islands. This result in turn strengthens the proposed account of the Scrambling Paradox.\(^{38}\)

4.3 [−Q] Interveners

In this section, I address the issue of whether [−Q] adverbials serve as blockers for both [+Q] and [−Q] potential goals.

4.3.1 Base-Generated [+Mod] Does Not Block [+\(wh\)]

As documented in Shields 2005, certain base-generated adverbs block certain instances of adverb movement (as shown in (52a)). However, the same adverbs do not block \(wh\)-movement (as shown in (52b)).

(52) a. ??Ja \textbf{bystro}\(_{\mathsf{+[\times]}}\) xoču, [čtoby ona \textbf{často}\(_{\mathsf{+[Mod]}}\) ___ exala].
   \(\text{I quickly want [that she often ___ went]}
   \text{‘I want her to often go quickly.’}\
   \) (Shields 2005:156, my diacritics)\(^{39}\)

b. \textbf{Gde}\(_{\mathsf{+[\mathsf{wh}]}\})\) ty xočes’, [čtoby ona \textbf{často}\(_{\mathsf{+[Mod]}}\) obedala ____].
   \(\text{where you want [that she often dined ___]}
   \text{‘Where do you want her to often eat?’}\

(52b) is as expected; [+\(\mathsf{Mod}\)] elements do not block [+\(Q\)] elements. See section 4.3.3 for discussion of blocking interactions among [−Q] elements.

\(^{38}\) Note that in Abels’s (2012) full account, based on Starke 2001, features are related to each other in various cross-cutting ways, as shown in (i).

(i) \textit{Abels’s (2012:249) feature space}

\text{\begin{center}
\begin{tikzpicture}[level distance=2cm,
  sibling distance=3cm,]


\node {F}
  child {node {Argumental}}
  child {node {Mod}}
  child {node {Top}};

\node {Op}
  child {node {Rel}};
\end{tikzpicture}
\end{center}}\)

(i) has important consequences for the theory of features and may or may not be fully compatible with the approach proposed here, whereby syntactic marking (such as marking for scrambling) supersedes other (lexical) feature components of the element in question. A full characterization of the interaction of feature spaces such as (i) and the syntactic approach to marking for scrambling shown in (47) remains outside the scope of this article.

\(^{39}\) Some native speakers report finding this sentence entirely unacceptable. Others, including an anonymous reviewer, find it essentially fine, given an appropriate intonation pattern, which implicates the participation of discourse factors that may influence the judgment (and perhaps the derivation). Here, I present the judgment published in Shields 2005. A more fine-grained study of the interaction among various nonquantificational adverbials, both scrambled and base-generated, will have to be set aside for future study.
4.3.2 (Moved) [+Σ] Does Not Block [+wh]  Turning to elements that have undergone movement themselves, we find that in Russian, scrambled elements (with a non-Focus interpretation) do not block wh-movement of either arguments or adjuncts.40

(53) a. Ty komu [+wh] srazu [+Σ] xočeš’, [čtoby ona pozvonila you whoDAT right.away want [that she called __ __ ]?
   __ __ ]
   ‘Who do you want her to call right away?’

b. Ty komu [+wh] iz doma [+Σ] xočeš’, [čtoby ona pozvonila you whoDAT from home want [that she called __ __ ]?
   __ __ ]
   ‘Who do you want her to call from home?’

(54) a. Ty kak [+wh] srazu [+Σ] xočeš’, [čtoby ona počinila mašinu you how right.away want [that she fixed car __ __ ]?
   __ __ ]
   ‘How do you want her to fix the car right away?’

b. Ty kak [+wh] iz doma [+Σ] xočeš’, [čtoby ona počinila you how from home want [that she fixed mašinu __ __ ]?
   car __ __ ]
   ‘How do you want her to fix the car from home?’

4.3.3 Which [−Q] Elements Block [+Σ]?  As illustrated in (52a), repeated as (55), [−Q] movement of an adverbial over a base-generated nonquantificational adverb is degraded.

(55) ??Ja bystro [+Σ] xoču, [čtoby ona často [+Mod] __ exala].
   I quickly want [that she often __ went ]
   ‘I want her to often go quickly.’

(Shields 2005:156, my diacritics)

In (55), with both adverbs present, we see again a Russian case of [+Mod] blocking [+Σ]. Shields (2005:156) generalizes as follows: “Long distance scrambling obeys the RMC [Relativized Mini-

40 I assume now that moved [−Q] elements are [+Σ], since [+Σ] is the feature that drives their movement, according to (47) (marking for scrambling), whereas [+Mod] is the feature they bear lexically, by virtue of being base-generated adverbials (assuming a traditional (noncartographic) approach to the syntax of adverbs). Following (47), this feature is not relevant for blocking. Glushan (2006) makes a distinction between [+Mod] and [+Top] as different kinds of base-generated nonquantificational adverbials. I assume something simpler: nonquantificational adverbs are all base-generated with a [+Mod] feature; if they scramble (excluding Focus movement), the driving feature is [+Σ]. The possibility of more kinds of base-generated nonquantificational adverbs is not relevant here; what matters is the blocking potential of [−Q] adverbials on [+wh] and [+Σ] goals. See also Abels 2012.
mality condition], as expected.” However, as Shields also notes, not all multiple scrambling constructions are subject to such blocking. In particular, two non-Focus-scrambled items do not appear to interfere with each other.

(56) a. Ty Ivanu[+Σ] srazu[+Σ] xočeš’, [čtoby ona pozvonila you IvanDAT right.away want [that she called 
   ___ ___ ]?
   ___ ___ ]
   ‘Ivan, do you want her to call right away?’

b. Ty Ivanu[+Σ] iz doma[+Σ] xočeš’, [čtoby ona pozvonila you IvanDAT from home want [that she called 
   ___ ___ ]?
   ___ ___ ]
   ‘Ivan, do you want her to call from home?’

It now appears that we have encountered a crucial distinction among [−Q] elements. In particular, base-generated [+Mod] appears to block [+Σ], but [+Σ], the dominant feature when an adverbial itself has been scrambled, does not. This is the converse of the core distinction above, and supports the overall system.

4.4 Summary of Blocking Data

To summarize, we have seen evidence that [+Q] elements block [+Q] movements, but do not block [−Q] movements (such as [+Σ]-driven scrambling). Scrambled elements, even if lexically [+Q], are not blocked by [+Q] elements. [−Q] elements, on the other hand, do not block [+Q], accounting for the core of the Scrambling Paradox. Further, base-generated [+Mod] elements block [+Σ], though they never block [+Q] movement, while moved [+Σ] elements do not block other instances of [+Σ].

We can now describe the overall situation in elegant fashion: standard movement constraints (the CNPC, CSC, CED, PBC, etc.) apply equally to scrambling and wh-movement, exactly because these are not blocking effects. They involve strong islands, opaque domains, impenetrable phases, none of which should be sensitive to the type of element moving. Relativized Minimality–style blocking effects, described in terms of feature classes, explain the rest of the paradigm. In particular, the illicit wh-movement cases are all instances of [+Q] blocking of wh-movement, whereas scrambled [+Σ] elements can freely move over [+Q] elements. This explains the primary component of the Scrambling Paradox—namely, the fact that scrambling is not blocked by wh-islands. Table 2 summarizes the interactions, which can now be understood in terms of feature classes and Relativized Minimality.

4.5 The Scrambling Paradox Revisited

We now have a basic resolution of the Scrambling Paradox. To provide additional support for the account, I will rely on the claim that the [+Σ] feature of a scrambled element does not interact with the [+wh] feature of the embedded clause, and no blocking takes place. The account therefore
THE SCRAMBLING PARADOX 661

predicts that if we manipulate the features of the displaced element, the blocking effect should return. The data confirm this. Recall the primary contrasting pair, (20b) vs. (22b), repeated here.

(57) a. *Ty kto videl, [kogda ___ pod’ez’al ]?
   you whoNOM saw [when ___ was.arriving]
   ‘Who did you see when arrived?’

b. Ty doktor [+Σ] videl [kogda [+wh] [ ___ pod’ez’al ] ]?
   you doctorNOM saw [when [ ___ was.arriving]]
   ‘Did you see when the doctor was arriving?’

We now have the explanation for this basic distinction: (57a) contains a [+wh] feature that blocks movement of a [+wh] element, whereas movement in (57b) does not involve a [+Q] feature, and so the intervening [+wh] feature has no effect.

In starting to test whether changes in the feature makeup of the moved element influence acceptability, let us turn our attention to the discourse status of the element displaced in Zemskaya-type sentences. (57b), a typical Zemskaya-type sentence, can be paraphrased as ‘The doctor came by. Did you or did you not see that happen?’ ‘The doctor’ is strongly presupposed information (consistent with the yes/no question). As we have seen, the element in question carries a [+Q] feature, which is not sensitive to blocking by true [+wh] islands. When the element bearing a [+Σ] feature is changed to one bearing a contrastive [+Foc] feature, as in (42), repeated here, the result is notably worse.

   DOCTORNOM students asked [when ___ arrived]
   ‘[It’s] the DOCTOR the students asked when he arrived.’

   MASHANOM students asked [when ___ arrived ]
   ‘[It’s] MASHA the students asked when she arrived.’

Crucially, this can now no longer be considered an instance of scrambling (if it were, the element would be “marked for scrambling,” with [+Σ] subsuming any other features). Instead, this is an instance of focalization, which is driven by a [+Foc] feature distinct from [+Σ], as expected in
Rizzi’s typology of features. In a strong Focus reading, extraction from perception clauses of the Zemskaya type should also be worse. And it is:

(59) ??Studenty [tol’ko DOKTOR$_{[+\text{Foc}]}$ [videli [kogda$_{[+\text{wh}]}$ [ ____ pod’ezžal ]]]].

students [only DOCTOR$\text{NOM}$ [saw [when [ ____ was.arriving]]]]

‘The students saw when only A DOCTOR came.’

This example can be paraphrased as ‘We know the students saw a set of people arriving. In fact, all they saw was that it was a doctor who arrived’. The ungrammaticality is caused by [+wh] blocking [+Foc].

However, as pointed out by a reviewer, it is unreliable to use intonation alone to distinguish a truly [+Foc] interpretation from a [−Q] ([+Σ]) reading. (Perhaps this is why the degradation is not as severe as with full wh-movement.) Additional confirming evidence is found in the attempt to use the èto-cleft construction here. The èto-cleft, which has a strong Focus effect, as does its English equivalent, is perfectly acceptable on internal arguments that, when combined with èto, undergo movement.

(60) a. Èto IVANU$_{[+\text{Foc}]}$ ja pozvonil ____.
   it’s IVAN$_{\text{DAT}}$ I called ____
   ‘It’s IVAN I called.’

b. Èto [S IVANOM]$_{[+\text{Foc}]}$ ja igral v šaxmaty ____.
   it’s [WITH IVAN ] I played at chess ____
   ‘It’s WITH IVAN I played chess.’

Such clefts are degraded when formed with arguments from within traditional wh-islands and sharply ill-formed when adjuncts are involved, thus implicating the effects of a weak island, as predicted.

(61) a. ??Èto IVANU$_{[+\text{Foc}]}$ ja sprosil, [kogda$_{[+\text{wh}]}$ pozvonili ____].
   it’s IVAN$_{\text{DAT}}$ I asked [when called ____ ]
   ‘It’s IVAN I asked when they called.’

b. *Èto [S IVANOM]$_{[+\text{Foc}]}$ ja sprosil, [gde$_{[+\text{wh}]}$ ty žil ____ ].
   it’s [WITH IVAN ] I asked [where you lived ____ ]
   *‘It’s WITH IVAN I asked where you lived.’

The account makes the further prediction that Focus clefting out of an indicative èto-clause should also be degraded, and Focus clefting an adjunct extracted from the same context should be sharply degraded. This appears to be exactly the distribution.

(62) a. ??Èto DOKTORA$_{[+\text{Foc}]}$ ja nadejus’, [čto oni nanjali ____ ].
   it’s DOCTOR$_{\text{ACC}}$ I hope [that they hired ____ ]
   ‘It’s A/THE DOCTOR I hope that they hired.’

This is consistent with Bošković’s (2004) claim that focalization and scrambling are distinct operations, although his claim is more extreme, namely, that the former is movement and the latter is not.
The generalization thus emerges that [+Foc] elements cannot escape from either traditional wh-islands or čto-islands, which is exactly as predicted by the present analysis.

This section has shown that it is indeed the feature makeup of the moved element in Zemskaya-type examples such as (57b) that makes them acceptable. Therefore, we do not need to appeal to a radically different grammar of displacement to account for the scrambling/wh-move-ment contrasts from section 3, and of course we can maintain a principled account of the scrambling/wh-movement similarities presented in section 2.

5 Remarks on Base-Generation

The insensitivity of long-distance scrambling to wh-islands has been taken to motivate radical nonmovement accounts of flexible constituent order (Bošković and Takahashi 1998, Van Gelderen 2003). However, when carefully examined, the facts reveal that this more comprehensive collection of parallels and nonparallels between wh-displacement and non-wh-displacement strengthens arguments against accounts that claim Ā-scrambling is not an instance of standard (upward) movement. In particular, base-generation approaches such as Bošković and Takahashi’s (1998) and Van Gelderen’s (2003) are unable to account for the distribution of these constructions.

There are three primary arguments against treating Zemskaya-type sentences as the result of a base-generation process: (a) sensitivity to islands and other movement constraints, including those discussed in section 2; (b) the possibility of both surface and reconstructed interpretive effects; and (c) what I call “longer displacement” facts. Let us briefly consider each.

42 Base-generated processes may still exist, of course, but the strong constraints on non-wh displacement show that a movement approach is still required, unless standard constraints are to be reworked for non-wh displacement and maintained for wh-displacement, an obviously ad hoc complication of existing movement theory.

43 Polinsky and Potsdam (2014) provide important diagnostics for differentiating base-generated from moved elements in numerical quantificational constructions; they show that both strategies can be observed, dislocated paucals being derived by movement and dislocated genitive plural topics being base-generated. They draw the following distinctions:

(i) a. Movement (paucal complements)

Knigi u menja [dve ___].

books at me [two ___]

‘I have two books.’

b. Base-generation (GenPl numerical topics)

Knig u menja [dve pro].

books at me [two pro]

‘Books, I have two (of them).’

According to Polinsky and Potsdam, the paucal construction involves movement because it obeys islands, triggers Weak Crossover, licenses parasitic gaps, undergoes reconstruction for Principle C, and cannot (easily) appear with a resumptive pronoun. Clearly, this shows more evidence for movement strategies for flexible word order patterns, as well as the availability of base-generated options at least for numerical topics.
5.1 Movement Constraints

The very existence of the strong movement-constraint-obeying parallels shown in section 2 undermines the viability of nonmovement approaches. Thus, the sensitivity of the Zemskaya-type sentences shown in section 2 to the CNPC, CSC, CED, and PBC argues strongly for a movement approach to these constructions.

5.2 Interpretive Effects

Using interpretive effects as a diagnostic to tease apart movement and nonmovement approaches to base-generation is not straightforward, because it depends on the particular theory of base-generation (and theory of reconstruction) at hand. Some assume a low position for the “scrambled” element, and some do not. Thus, a base-generation theory with “obligatory lowering” such as Bošković and Takahashi’s (1998) makes the same predictions as a movement theory with “radical reconstruction” such as Saito’s (1992)—namely, that the lower LF (nonscrambled) position is the only one relevant for interpretation. On this kind of theory, for all relevant constructions, the nonscrambled position should be the position from which all LF scope and binding relations are determined. However, this appears to be too strong, as antireconstruction and surface scope relations are found with such sentences (Bailyn 2006). Thus, (63) shows a preference for surface scope (the only option for some speakers), consistent with the analysis of Russian scope in Antonyuk 2015, which is exactly the opposite of what is claimed to be found with Japanese long-distance scrambling (Saito 1992).

(63) Ty [kakuju-to devušku] videl kak [každyj mal’čik] celoval ___?
you [some girl] ACC saw how [every boy] NOM kissed ___?
‘Did you see when every boy kissed some girl?’

Examples such as (63) demonstrate clearly, then, that the only viable base-generation theory of such constructions would be one similar to Polinsky and Potsdam’s (2014) account of the topical genitive plural constructions illustrated in footnote 43, in which a high base-generated element is discourse-related to a lower pro element in the structure. Such a theory could explain (63) without sacrificing base-generation: the surface left-dislocated element could serve as the locus of semantic interpretation, allowing for the high scope reading. Thus, this diagnostic cannot necessarily distinguish between a viable base-generation theory and a derivational movement account of such constructions.

However, Principle C reconstruction shows that this kind of base-generation also will not account for all Â-scrambling sentences. With Zemskaya-type sentences, we see evidence of required reconstruction, an effect not predicted by base-generation theories that allow the dislocated element to stay in its surface position at LF. Thus, in (64) (repeated from (15)) it must be the lower position that is relevant at LF, forcing the Principle C violation.
Principle C reconstruction violations such as (64b) argue against those base-generation theories that assume the LF position of the dislocated element to be the same as its surface position. Thus, the overall picture of Principle C reconstruction supports a movement account of Ā-scrambling sentences.

5.3 “Longer Displacement”

This conclusion is strengthened by a closer look at another set of Zemskaya-type sentences, those where the displaced element is moved to the edge of an adjunct domain. If this were a base-generated construction, we would not expect difficulty with relocating the displaced element outside the island. However, this is clearly not the case. Thus, consider the pairs in (65) and (66), in which the (a) sentences, attributed to Zemskaya via Yadroff 1991 by Sabel (2002) and Müller (2002), involve a displaced element at the edge of the domain, whereas the (b) sentences show displacement to a minimally higher position, clearly outside of the relevant local CP. CED-type violations emerge, where even direct objects are unable to participate, as shown in (65b) and (66b), violations similar to those found in wh-movement (as in (67)).

(65) a. Vse usnuli [CP groza [CP kogda ___ končilas’]]. (✓CP adjunction) everybody fell.asleep [CP storm [CP when ___ ended ]] ‘Everybody fell asleep when the storm ended.’

b. *Vse groza usnuli [CP kogda ___ končilas’]. (*scr from adjunct CP) everybody storm fell.asleep [CP when ___ ended ]

(66) a. My byli udivenly [CP vodka [CP potomu čto on we were surprised [CP vodkaACC [CP because he prines ___ ]]. brought ___ ]] ‘We were surprised because he brought vodka.’

b. *My vodka byli udivenly [CP potomu čto on prines we vodkaACC were surprised [CP because he brought ___ ]]. ___ ]

(67) a. *Vy čto usnuli [CP kogda ___ končilos’]? (*wh-mvt from adjunct CP) you what fell.asleep [CP when ___ ended ] ‘What did you fall asleep when ended?’
b. *Vy čto byli udivleny [CP potomu čto on prines ___ because he brought ___ ]?

*‘What were you surprised because he brought?’

The lexical array in (65a–b) and (66a–b) is identical across the scrambling/wh-movement pairs, of course, indicating that something about the derivation of the (b) sentences causes their deviance. This shifts the burden of proof to nonmovement accounts to capture what is otherwise coincidental parallel adherence to constraints. Clearly, a movement-based account of the phenomena at hand is preferable.44

6 Consequences for Movement Theory

In this article, I have shown that the Scrambling Paradox can be successfully reduced to the basic question of why wh-islands (more exactly, [+Q] islands) do not constrain (non-Focus) scrambling. The feature-class system of Relativized Minimality accounts for this. And a range of other blocking and nonblocking facts are accounted for as well, on the assumption that scrambling involves a [+Σ] feature, which does not interact with [+Q] features. If the moved element carries a [+Q] feature such as [+Foc], then the expected blocking effects emerge. We thus have a solution that is elegant, is predictably constrained, and requires nothing to be added to existing theories of movement: long-distance scrambling, like wh-movement, is overt, leftward A-bar movement. Wh-movement/scrambling similarities result from similarities in derivation, subject to the usual strong island constraints (CNPC, CSC, CED, PBC, etc.) while wh-movement/scrambling differences all reduce to (featural) Relativized Minimality effects in the A-bar system. Note also that, with respect to the landing site of movement, this approach is fully compatible with either cartographic (Rizzi 1997) or adjunction (Kidwai 2000) approaches.

And there is a further consequence significant for future research—the evidence presented here shows that constraints applying to configurations traditionally labeled “strong islands” have an entirely distinct character from those applying to configurations traditionally labeled “weak islands” (a difference discussed in detail in Boeckx 2012 and elsewhere), and yet both kinds are true grammatical constraints. Strong islands constrain both arguments and adjuncts equally, and they also constrain both scrambling and wh-movement equally, being insensitive to blocking.

44 Müller and Sternefeld (1993) provide a movement-based account of the Zemskaya facts, through the use of what they call “unambiguous binding.” Their idea is that wh-movement and scrambling target distinct landing sites and therefore move through distinct escape hatches. In particular, they claim that scrambling targets an adjunction position, while wh-movement targets Spec,CP. They formalize this as follows:

(i) Principle of Unambiguous Binding

A variable that is α-bound must be β-free in the domain of the head of its chain (where α and β refer to different types of positions). (Müller and Sternefeld 1993:461)

Although within Minimalist theory distinct escape hatches are not possible, current theories of the cartography of the left periphery, following Rizzi (1997), do allow distinct landing sites to be a viable approach to the problem (see also Sabel 2002). What I proposed in section 4 is in many ways based on the spirit of this proposal and shares with it the strong claim that both scrambling and wh-movement are standard upward A-bar-displacement processes.
This provides an additional diagnostic with regard to these apparently disparate phenomena: not only does the traditional distinction between arguments and adjuncts in wh-movement contexts diagnose weak islands, but it is also now the case that any lack of distinction between Ā-scrambling and wh-movement diagnoses strong islands. This then sets the stage for unique analyses of the strong constraints as unrelated to Relativized Minimality.

References

Antonyuk, Svitlana. 2015. Quantifier scope and scope freezing in Russian. Doctoral dissertation, Stony Brook University, Stony Brook, NY.


Scott, Tanya. 2012. Whoever doesn’t HOP must be superior: The left periphery and the emergence of superiority in Russian. Doctoral dissertation, Stony Brook University, Stony Brook, NY.


*Department of Linguistics*
*Stony Brook University*
*Stony Brook, NY 11794*

*john.bailyn@stonybrook.edu*