Splitting Puzzles in South Slavic

Steven Franks

Indiana University

Anita Peti-Stantić

University of Zagreb

0. Introduction. This talk is partly born of an increasing sense of uncertainty about how to do syntax. Time was, we linguists had at our disposal a generally accepted arsenal of analytic devices to deal with a familiar range of constructions and phenomena. However, as the phenomena under study began to proliferate, so did the analyses. The increasingly baroque GB architecture began to collapse under the weight of Japanese, but it was the vagaries of Slavic that surely "brought the house down". Now, one might think that Minimalism—with its appeal to elegance and its purported aim of narrowing the search space for credible hypotheses to those that stray as little as possible from "virtual conceptual necessity"—would at the very least provide us with a general blueprint for how to construct the right sort of analysis for any given problem. Unfortunately, it does not seem to us that we are there yet. The leading ideas are broad, expansive, and on occasion even profound, but actually applying them to real linguistic data can, we have found, be a very frustrating exercise. Now this frustration partly derives from the truism that has always bedeviled formal linguistic efforts: raw data is just that. Raw. It does not come with any instructions on the label. So, given for example the judgment that some string of words is not particularly felicitous, without an explicit model we cannot know in advance where to look for the source of the degradation. We often really have no idea why the string was deemed infelicitous. So, taking syntax crudely to be a formal system relating sound and meaning, we do not know whether the reason for the judgment was purely formal or has instead to do with the need to interface with articulation, on the one hand, or with interpretation, on the other. Moreover, each of these interface exigencies—the so called "PF–side" and "LF–side"—conceals a host of possible offenses, from prosodic inconsistency to referential circularity. Throw into the mix processing, pragmatics, and parsing, and the linguist's frustration multiplies.

1. Clitics Split Phrases. The present talk is a case in point. Initially, we set out to study the phenomenon of splitting of nominal expressions by clitics in the Croatian idiolect of one of the coauthors. The expectation of the other coauthor was that there would be nothing special to say about clitics inside nominal expressions. That is, the by now traditional wisdom is that splitting by clitics does *not* involve PF–side placement of the clitics internal to the nominal expression, but is rather

parasitic on more generally available processes that somehow "scramble" pieces of NPs. Thus, when clitics appear to interrupt, the clitics are really where they always are (wherever that is) and it is the pieces of the NPs that have done something unexpected. So, following this widely adopted claim—which, so far as we know, was first made in Franks and Progovac 1994 but has since been reiterated in various places and has been expanded upon in Bošković's work (especially Bošković 2001)—one should *not* treat clitic splitting data separately from splitting by other, fully lexical elements. Peti-Stantić (in press), for example, cites the Croatian examples in (1), where clitics split, and compares these to the corresponding examples in (2), where the intervening material is lexical:

(1)   a.   <u>Sestrina</u> **mi** <u>prijateljica</u> sutra dolazi u posjet.

'Sister's friend is coming to visit me tomorrow.'

b.   <u>Ivana</u> **sam** <u>Ivića</u> upoznala prije mjesec dana.

'I met Ivan Ivić a month ago.'

c.   O <u>izuzetno</u> **smo** <u>zanimljivoj knjizi</u> razgovarali.

'We were talking about an especially interesting book.'

d.   <u>Studenticu</u> **mi je** <u>teorijske fizike</u> predstavio.

'He introduced me to a student of theoretical physics.'

(2)   a.   <u>Sestrina</u> **sutra dolazi** <u>prijateljica</u>. (a ne neka druga)

'It is SISTER'S friend who is coming tomorrow.' (and not some other one)

b.   Odavno već <u>Ivana</u> **čitaju** <u>Ivića</u>. (a ne nekog drugog Ivića)

'I have been reading IVAN Ivić for a long time.' (and not some other Ivić)

c.   O <u>izuzetno</u> **razgovaramo** <u>zanimljivoj knjizi</u>. (a ne tek donekle zanimljivoj)

'We were talking about an ESPECIALLY interesting book.'

(and not just a moderately interesting one)

d.   <u>Studentica</u> **dolazi** <u>teorijske fizike</u>. (a ne studentica povijesti)

'A student OF THEORETICAL PHYSICS is coming.' (and not a student of history)

(For ease of reference, the split constituent is underlined and the splitting material is in boldface.) While one could imagine a so-called "Prosodic Inversion" account of (1a), whereby the clitic *mi* is left in initial position by the syntax, i.e., in front of *sestrina prijatelica*, and then undergoes a PF process so that it would be pronounced at the right edge of the first prosodic word, producing *sestrina mi prijatelica*, this is surely ludicrous for (2a). Here, not only is *sestrina prijatelica* separated by tonic material which would be perfectly happy wherever the syntax left it, but that material—*sutra dolazi*—is not even a constituent. Instead, *sutra* and *dolazi* are wherever they are and the pieces *sestra* and *prijateljica* for some reason end up being pronounced in a way that flanks *sutra* and *dolazi*. By

Occam's Razor, we eschew Prosodic Inversion and apply the same reasoning to (1a). (As an aside, one might think of this as the "adverb/negation" approach to clitics, in that, just as adverbs and negation in much of the standard literature are seen as demarcating the left edge of various verbal projections, the clitics are taken to occupy in some sense fixed positions, with other material revolving around them.)

2. Does Splitting Reflect Focus? It is noteworthy that, in offering the examples in (2), Peti-Stantić felt the need to add a parenthetical highlighting the contrastive focus nature of the construction. This makes sense, as scrambling in Slavic is by no means semantically vacuous, but rather reflects the exigencies of Functional Sentence Perspective. In (2a–c) it is the element that immediately precedes the splitting material that is focused (*sestrina*, *Ivana*, and *izuzetno*, respectively), but this does not seem to be a requirement, since in (2d) it is what follows, namely *teorijske fizike*, that bears contrastive focus. More generally, her judgments are that *either* side can (and *must*) be focused, so that in (2d) it could instead be *studentica* that is contrastive (e.g., *a ne profesorica*) and in (2a–c), respectively, it could instead be *prijateljica* (e.g., *a ne nastavnica*), *Ivića* (e.g., *a ne Gundulića*), or *zanimljivoj* (e.g., *a ne dosadnoj*).

However this splitting is ultimately implemented, the following question arises: Given that the splitting examples in (2) require contrastive focus to be felicitous, is the same true of (1)? According to the translations Progovac (2006) provides for the Serbian data in (3) and (4), this is indeed the case:

(3) <u>Vukina</u> **se** <u>ćerka</u> udala.   'It was VUKA's daughter who got married.'

(4) <u>Vukina</u> **odlazi** <u>ćerka</u>.   'It is VUKA's daughter who is leaving.'

These can presumably be opposed to examples with the subject NP *Vukina ćerka* intact, as in (5), which involve no special discourse presuppositions and are completely stylistically neutral:

(5) a. <u>Vukina ćerka</u> se udala.   'Vuka's daughter got married.'

  b. <u>Vukina ćerka</u> odlazi.   'Vuka's daughter is leaving.'

Peti-Stantić, however, points out a contrast between (3) and (4): whereas the former can be understood as perfectly neutral and does not necessary require focus, in (4) either *Vukina* or *ćerka* must be focused. These judgments are reflected in the additional translations provided in (3') and (4'):

(3') <u>Vukina</u> **se** <u>ćerka</u> udala. 'It was Vuka's DAUGHTER who got married.' AND ALSO ...

  'Vuka's daughter got married.'

(4') <u>Vukina</u> **odlazi** <u>ćerka</u>. 'It is Vuka's DAUGHTER who is leaving.' BUT *NOT* ...

  *'Vuka's daughter is leaving.'

This difference, which has been corroborated by Serbian speakers as well, might be taken to suggest that, although both (3) and (4) both look like left-branch extraction (LBE), we are not in fact dealing with a monolithic phenomenon.

There are, on the other hand, many reasons to want to analyze splitting by clitics as a subcase of the more general case, with differences in preferred interpretations following from independent considerations. As revealed by even the most superficial of investigations (like this one), there is a high degree of variation in judgments about the acceptability of different kinds of splitting, both across speakers and across languages. (For Russian, a very nice overview of the issues has been recently compiled by Pereltsvaig 2006.) Nonetheless, if splitting by clitics is indeed parasitic on splitting in general, then we would predict that the felicity of splitting should be comparable, again both across speakers and across languages, regardless of whether the splitter is a clitic or a tonic element. This expectation is borne out by the Slovenian data in (6), which are the Slovenian versions of Serbian (3) through (5):

(6)   a.   *<u>Milojkina</u> **se je** <u>hči</u> poročila.   (cf. ✓<u>Milojkina hči</u> se je poročila.)

      b.   *<u>Milojkina</u> **odhaja** <u>hči</u>.   (cf. ✓<u>Milojkina hči</u> odhaja.)

The fact that neither (6a) nor (6b) is particularly felicitous confirms our expectation that these should be disallowed or favored to a similar extent, *modulo* of course independent considerations. For example, since in Bulgarian pronominal and verbal auxiliary clitics are adjacent to the verb, these can never split NPs, even to the extremely limited extent that extraneous material can ever interrupt the adjective noun sequence in those languages. (The interrogative clitic *li* works differently, as noted near the end of this talk.)

There exists a burgeoning literature on splitting (and scrambling) in languages like Russian and German, and even a cursory examination of this literature reveals a considerable amount of disagreement over data. Our informal queries about South Slavic suggest a range of latitude in acceptability of splitting here too. Some speakers of languages that basically lack this, such as Slovenian, Macedonian, and Bulgarian, will still accept splitting in "poetic" styles or in discourse contexts where the appeal to contrastive focus might be greatest . On the other hand, some speakers of languages like Serbian and Croatian, which basically have splitting, seem to accept it very broadly in these contexts. The Croatian coauthor of this paper is one such speaker: Anita admits splits that have elsewhere been deemed impossible. (Whether this reflects a change in progress, is a matter of dialectal or idiolectal variation, or is simply a difference in how judgments are assessed and reported should be the subject of a different, more sociolinguistically oriented study than ours.) Be that as it may, what

we have in general found, comparing Croatian sentences where clitics can split phrases with those which are as minimally different as we could construct but with something other than a clitic doing the splitting, is that, if there is a difference, the sentence where a non-clitic splits might be degraded with respect to the one where a clitic splits. Consider, for example, (7b) versus (7c):

(7) a. <u>Sestra i njen muž</u> **će mi ga** pokloniti. [no splitting]

      'Sister and her husband will give it to me.'

    b. ✓<u>Sestra</u> **će mi ga** <u>i njen muž</u> pokloniti.

      'It is SISTER and her husband who will give it to me.'

    c. ??<u>Sestra</u> **rado** <u>i njen muž</u> poklanjaju knjige školskoj knjižnici.

      'It is SISTER and her husband who gladly give books to the school library.'

    d. ✓<u>Sestra</u> **će** <u>i njen muž</u> doći u utorak.

      'It is SISTER and her husband who will come on Tuesday.'

Although (7b) is the sort of thing, with multiple clitics splitting a coordinated NP (a so-called "fortress"), that, ever since the seminal work of Browne (1975), has been cited as unacceptable (cf. e.g. Schütze 1994, Progovac 1996, Franks and King 2000), Anita and other Croats we have polled find it perfectly fine. (More to the point, they do not rate (7b) worse than (7d), with only one clitic, whereas Progovac (1996) reports a clear contrast in her Serbian, with (7d) marginal and (7b) outright ungrammatical.) Putting aside the (very interesting) questions this variation raises, our point is simply that, to reiterate, whenever there is a difference, it is the sentence where a non-clitic splits that will be worse than the one where just a clitic splits. It seems to us that the proper way to interpret these facts is as a matter of degree. This contrast, we believe, corresponds to the difference in how much splitting necessarily reflects focus, where, on the whole, a contrastive focus reading is far more expected when non-clitics split than when clitics do. This fact does not necessarily imply distinct mechanisms for splitting phrases. We rather favor a more parsimonous approach, in which such differences should tell us something about the design and operation of the mechanisms for deriving splitting themselves. So let us now turn to a survey of those mechanisms.

3. An Abundance of Mechanisms. Just as different languages, and speakers of those languages, vary in their tolerance for splitting (sometimes quite radically), so does there exist an abundance of technical ways one might obtain splitting. We cannot do justice to them all here, but refer you to Pereltsvaig (2006) for a more careful survey and to Bošković (2005) for discussion in particular of LBE.

3.1. PF–Side Placement. For clitic splitting, as mentioned, one possibility is literal PF–side placement, with at least the two conceptual variants in (8).

(8) a. **Move** prosodically unsupported enclitics to the edge of an adjacent P-word. OR ...

b. **Linearize** clitics at the prosodically appropriate edge of an adjacent P-word.

Choice between these variants depends on whether or not linear order is imposed in the syntax, before material is shipped off to PF. Either way, and independently of the additional problems (8) raises, it is simply not possible to generalize such Prosodic Inversion approaches to splitting by tonic material, as in (2). Nor is (8) likely to accommodate the sort of splitting in (9), based on (7) and also judged perfectly felicitous by Anita, since this appears not to target the first prosodic phrase:

(9) <u>Moja sestra</u> **su mi ga** <u>i njen muž</u> poklonili.

'It is my SISTER and her husband who gave it to me.'

This same pattern of possibilities is given in (10), where the split NP is contained within a PP:

(10) a. <u>U izuzetno veliku sobu</u> **sam** ušao. 'I entered an exceptionally large room.' [no splitting]

b. <u>U izuzetno</u> **sam** <u>veliku sobu</u> ušao.

'It was an EXCEPTIONALLY large room that I entered.' AND ALSO ...

'It was an exceptionally large room that I entered.'

c. <u>U izuzetno veliku</u> **sam** <u>sobu</u> ušao.

'It was an exceptionally LARGE room that I entered.' AND ALSO ...

'It was an exceptionally large room that I entered.'

(10c), like (9), cannot be handled by anything like the systems in (8). Of course, this does not mean we can be sure that something like (8) does not exist, but rather that, even if it does derive some instances of superficial splitting in some languages, we are still going to need other mechanisms that operate in the syntax proper, mechanisms that will, in fact, handle most of the cases. (Indeed, any element that splits that is not prosodically *enclitic* is not even a candidate, thus ruling (8) out for Slovenian, for example, and for everything but *li* in Bulgarian, let alone Macedonian.) Fortunately (or unfortunately, as the case may be) potentially applicable alternatives to (8) abound, each entailing its own set of assumptions. Here we have time to consider four of these.

3.2. Left-Branch Extraction. The first and most familiar way of deriving ostensible splitting is LBE. Under this account, the sentence in (4) can be schematized as in (11).

(11) a. [$_{AP}$ Vukina$_i$] odlazi [$_{NP}$ t_i ćerka].   b. [$_{AP}$ Vukina] odlazi [$_{NP}$ [$_{AP}$ ~~Vukina~~] ćerka]

This just involves old-fashioned movement, traditionally represented as in (11a) or expressed using the Copy-and-Delete/Remerge system in (11b); note that (11) assumes the NP-over-AP structure for nominal expressions. Bošković 2005 considers LBE in depth, reviewing a range of approaches from

the perspective of why some languages tolerate it and others do not. His paper puts forward two robust correlations: (i) LBE is disallowed in DP languages and (ii) LBE presupposes the possibility of scrambling. This is part of Bošković's general scheme (which takes off from Corver's implementation of Uriagereka's seminal insight that articles somehow block LBE), according to which nominal expressions project up to DP only in some languages, whereas in others the projection stops at NP. So, for example, Bulgarian (and Macedonian) differ from the other Slavic languages in having a DP, as indicated by the definite article *–ta* in Bulgarian (12a), and eschew LBE splitting, as shown in (12b):

(12) a. Petko prodade [$_{DP}$ [$_{AP}$ novata [$_{NP}$ kola]]].   [Bulgarian]

       'Petko sold the new car'

    b. *[$_{AP}$ Novata] prodade Petko [$_{DP}$ [$_{AP}$ ~~novata~~ [$_{NP}$ kola]]]?

       'It was the NEW car that Petko sold.' [Intended]

Similarly, word order in Bulgarian, although far from completely fixed, is not nearly as free as in the other languages, more on the lines of German. Bošković 2005 considers various ways to derive the impossibility of LBE in (12b) from the idea that Bulgarian nominal expressions are DPs, roughly as indicated in (12b), which assumes an Abney type DP-over-AP-over-NP analysis, so that there simply is no LB AP to extract. Thus contrasts with felicitous Serbian (11b), with the alternative nominal structure in which AP is contained within NP.

Bošković 2006 adduces several additional candidate correlations, buttressing his idea that the traditional noun phrase masks (at least) two types of phrases and that a host of differences between languages in some way or other might derive from this typology. While we tend to doubt that the nominal world could be as simple as NP vs. DP, since in particular it has been argued by Franks (and others) that some nomimal expressions in Russian are maximally QPs and that the Slavic pronominal clitics are heads of KPs, the strength of Bošković's correlations, it seems to us, cannot be denied. If correct, we might want to characterize the primary contrast in terms of whether or not definiteness and specificity features are introduced on a distinct functional head (i.e., a DP). Turning in this light to the difference noted earlier between Serbian or Croatian and Slovenian, one would be forced to conclude that modern, colloquial Slovenian, which arguably lacks LBE, must somehow have adopted a DP, presumably under German influence. We realize this is a somewhat contentious claim since, as an anonymous FDSL referee pointed out to us, "overtly the DPs in the two languages aren't that different." We would nonetheless argue that, if indeed Slovenian lacks LBE, Bošković's analysis leads to the conclusion that Slovenian projects to DP, *even if* D^0 *itself remains empty*. Bulgarian and Macedonian, after all, do not actually have anything overt in D^0, definiteness being marked inflectionally on the head below DP, but there is still a DP, with specificity features, and it is this that

presumably gives rise to the pattern of effects discussed by Bošković. Thus, one cannot simply rely on superficial evidence of an obligatory element in D^0 to decide whether or not there is a DP. (Possible evidence for DP within Slovenian, however, might be the invariant colloquial *ta* described by Marušič and Žaucer 2006 which, as they show, is also not itself a D element.) Colloquial Slovenian, we suspect, may be something like colloquial Finnish, which according to Ritva Laury (1997 and personal communication) has also recently taken on a new definite article, *se*, as in (13):

(13) a. <u>Punaisen</u>   **ostin**   <u>auton</u>   [literary Finnish, poetic style]

      red-acc   buy-pst-1sg   car-acc

    b. ?*<u>Punaisen</u>   **ostin**   (sen)   auton   [spoken Finnish]

      red-acc   buy-pst-1sg   the-acc   car-acc

Platzack (2006) has recently argued for such a reanalysis in the history of Icelandic. He notes that DP also arose in other Scandinavian languages, suggesting a concomitant switch from NP-over-AP to AP-over-NP.

3.3. Base Generation + LF Composition. Other factors of course come into play, particularly morphological richness. For example, in Warlpiri, discontinuous constituents invoke multiple morphological marking, as shown by a comparison of (14a) and (14b), which Bošković cites from Hale (1981):

(14) a. Maliki ka- pala wajilipi- nyi   <u>kurdu wita-</u> *jarra- rlu.*   [Warlpiri]

      dog   pres dual chase   nonpast child small dual   erg

    b. <u>Kurdu</u>-*jarra- rlu*   ka- pala maliki wajilipi- nyi   <u>wita-</u> *jarra- rlu.*

      child dual   erg   pres dual dog   chase   nonpast small   dual   erg

      'The two small children are chasing a dog.'

There is a sense however in which each piece acts independently, as in (15), from Hale (1973):

(15) a. <u>Tjantu   wiri</u> *-ŋki* = tji   yalku -nu.   [Warlpiri]

      dog   big   erg   1sg.obj bite   past

    b. <u>Tjantu</u>   *-ŋku* = tju   yalku -nu   <u>wiri</u> *-ŋki*.

      dog   erg   1sg.obj bite   past   big   erg

    c. <u>Wiri</u>   *-ŋki* = tji   yalku   -nuu   <u>tjantu</u> *-ŋku*.

      big   erg   1sg.obj bite   past   dog   erg

      'The big dog bit me.'

Note that, curiously not unlike Croatian or Serbian, the Warlpiri object clitic goes after the first constituent, which is just the noun in (15b) and just the adjective in (15c). (Note that its form *tji* or *tju*, reflects the same word level vowel harmony as the ergative case marker *ŋki* or *ŋku*, even though only

the latter appears on each piece.) So perhaps, at least in a language with great freedom of word order like Warlpiri, the scrambled pieces are base generated separately and composed only in LF. This could be implemented along the LF-lowering lines of Bošković and Takahashi (1998), raising however non-trivial questions about the role of case features on the LF–side. If adapted to our Serbian sentence in (4), then something like (16) would be all that the syntax provides:

(16)   [AP Vukina] odlazi [NP ćerka].

Closer to home, one might wonder why German should allow splitting, as discussed in Fanselow and Ćavar (2002), when Dutch does not, if not for differences in nominal morphology. However, in Bošković's system neither language can have true LBE, since both are DP languages. That is, if the tradition stemming from Uriagereka and Corver is right, splitting as in German (17) *cannot* be analyzed as LBE:

(17) a. <u>Wieviel</u> hat er <u>Bücher</u> gelesen? 'How many books has he read?'

b. <u>Bücher</u> hat er <u>keine</u> gelesen.

'As for books, he has read none.'

(BUT cf. *<u>Keine</u> hat er <u>Bücher</u> gelesen.)

c. <u>Bücher</u> hat man damals <u>interessante</u> in den Osten <u>keine</u> mitnehmen dürfen.

'As for books, one could not take any interesting ones to the East then.'

It is worth noting that in German, like Warlpiri, the pieces inflect like independent DPs so that, in (18b), the strong form *deutsche* reemerges:

(18) a. Er hat <u>keine deutchen Bücher</u> gelesen. 'He has read no German books.'

b. <u>Deutche Bücher</u> hat er <u>keine</u> gelesen. 'As for German books, he hasn't read any.'

Facts such as these may be problematic for LBE and might suggest a base generation account. Finally, the Croatian splitting examples in (7), (9), and (10) present problems, since these do not *prima facie* look like instances of the simple extraction of a LB AP.

3.4. Remnant Movement. So, what are some alternatives? An at one time popular kind of derivation involves remnant movement, schematized for the same Serbian example in (19).

(19) a. [AP/DP/KP Vukina t_i] odlazi [NP ćerka$_i$].   b. [AP/DP/KP Vukina ~~ćerka~~] odlazi [NP ćerka].

The remnant movement analysis, put forward for Serbian in Franks and Progovac (1994) *inter alia*, is more or less the opposite of LBE: in (19a), or Copy-and-Delete version (19b), the NP *ćerka* first extracts from some larger nominal domain (which, to avoid commitment, we have labeled as "AP/DP/KP"), before that larger remnant nominal domain fronts, minus the extracted portion. Thus, whereas under the LBE account the lefthand portion of the split is expected to be a constituent, under the remnant movement account it is the righthand portion that should be a constituent. Thus, (9) or

(10c) could be handled by moving *i njen muž* (which is a Boolean Phrase BP headed by *i*) or the NP *sobu* out of their containing phrases, which then themselves front, roughly as in (20):

(20) a. [Moja sestra ~~i njen muž~~] su mi ga [BP i njen muž] poklonili.   [= (9)]

  b. [U izuzetno veliku ~~sobu~~] sam [NP sobu] ušao.   [= (10c)]

What about (10b), repeated as (21)?

(21) <u>U izuzetno</u> **sam** <u>veliku sobu</u> ušao.

Franks and Progovac (1994), in order to explain the contrast they report in Serbian (22), argue that NP but not AP can extract from the remnant:

(22) a. [U izuzetno veliku ~~sobu~~] je Jovan ušao [NP sobu].

  b. (*)[U izuzetno veliku ~~praznu sobu~~] je Jovan ušao [AP praznu sobu].

Note that this only works if, contra Bošković, one adopts the AP-over-NP structure. Consider also (23a), from Bošković 2005:

(23) a. (*)<u>Visoke</u> **je on** <u>lijepe djevojke</u> vidio.

     'It was TALL beautiful girls that he saw.' [intended, and OK for some]

  b. (*)<u>Visoke lijepe</u> **je on** <u>djevojke</u> vidio.

     'It was tall BEAUTIFUL girls that he saw.' [intended, and OK for some]

While (23a) might be bad if AP cannot move out before remnant movement and *lijepe djevojke* is an AP, one wonders why (23b), with just NP moving out, is then no better. Turning to (10b)/(21), this idea becomes even more problematic, since—regardless of whether one adopts the NP-over-AP or AP-over-NP structure—*izuzetno* should be a constituent with *veliku*, which it modifies. Franks and Progovac therefore suggested an LBE rather than remnant movement approach to examples like (10b)/(21), with the AdvP *izuzetno* fronting and the P *u* somehow attaching to the AdvP beforehand. So, even in that early paper, at least two mechanisms for splitting were called for! Bošković (2005) points out a number of additional technical problems for using remnant movement to handle splitting, which we will not reproduce here. Suffice it to say these problems are to some extent complementary to those arising under the LBE analysis, depending on constituency.

3.5. Movement with Distributed Deletion. This brings us to our final mechanism, one in which the entire phrase moves but, in the spirit of Fanselow and Ćavar 2002, deletion of the phrase's subparts is distributed (or scattered) across its various occurrences. This powerful mechanism is actually a second PF–side approach, but one (unlike PF movement) that capitalizes on an independently well-motivated aspect of the syntax–PF interface: in mapping from syntax into PF, decisions about which copies to pronounce and which to delete *must* be made . Returning once again to our example in (4), it is the

nominal portion *ćerka* of the higher copy which, under distributed deletion, is left unpronounced. This is shown in (24).

(24) [$_{NP}$ Vukina ~~ćerka~~] odlazi [$_{NP}$ ~~Vukina~~ ćerka].

Of course, the problem with such as system is constraining it. Nonetheless, while any of the methods considered might turn out to be viable for something like *Vukina odlazi ćerka*, it seems to us that—of the plethora of mechanisms countenanced by recent linguistic theory—distributed deletion is likely to be the best way to handle the striking types of splits which Anita finds acceptable in her Croatian speech. So, looking back at (22b) and (23), it should be noted that the asterisks were placed in parentheses because, once again, for Anita these are perfectly felicitous with the material preceding the split focused. It thus may be that different speakers and/or different dialects vary in how liberally they implement distributed deletion and in how willing they are to tolerate its application, but beyond that, we have little to say at this point about the discrepancies in reported judgments.

4. An Abundance of Judgments. Let us now look briefly at some additional examples of unrelated) kinds of problems which the data present:

(25) a. Splitting by clitics is less constrained (both in occurrence and in need to impose a focus
           interpretation) than splitting by other material.

      b. The pieces of the split need not be syntactic constituents.

      c. Clitics can split even very low in the structure.

The first problem suggests either distinct mechanisms or, preferably, some independent reason to expect clitic splitting to be less marked. This probably follows from the fact that there are always going to be copies above and below the clitics, if these move to high head position whose Spec must be filled, but not so with other material, where movement more clearly reflects Functional Sentence Perspective considerations. The second problem suggests a PF–side approach to deletion. The third problem is unique to Anita's data.

First, as noted and contrary to judgments reported elsewhere, many Croatian speakers judge to be perfect examples such as (7b) and (9), repeated in (26):

(26) a. <u>Sestra</u> **će mi ga** <u>i njen muž</u> pokloniti.

          'It is SISTER and her husband who will give it to me.'

      b. <u>Moja sestra</u> **su mi ga** <u>i njen muž</u> poklonili.

          'It is my SISTER and her husband who gave it to me.'

Syntactic accounts invoke constituency for movement; here, the separated pieces are indeed likely constituents. More problematic might be example (27):

(27) Moja **su mi ga** <u>sestra i njen muž</u> poklonili.

'It is MY sister and her husband who gave it to me.'

Under no syntactic analysis can *sestra i njen muž* be a constituent, since *moja* only scopes over *sestra*, so the remnant movement account is unlikely for (27). *Moja* could of course be targeted to move as a focus, although movement of *moja* alone might be a Coordinate Structure Constraint violation. Distributed deletion however would work, provided that the material following the focus need not be a constituent. Note that even the preserved focused portion is not required to be a constituent, given (10b), which was repeated as (21). Our distributed deletion proposal for (21) would exploit the fact that there are copies above and below the clitics. If the high copy bears focus features, it might look something like (28a). Similarly for (23b), which will look like (28b).

(28) a. [PP <u>U</u> [NP [AP <u>izuzetno</u> ~~veliku~~] ~~sobu~~]] [**sam** [PP ~~u~~ [NP [AP ~~izuzetno~~ <u>veliku</u>] <u>sobu</u>]] [ušao ...
           [+Foc]

   b. [NP <u>Visoke lijepe</u> ~~djevojke~~] **je on** [NP ~~visoke lijepe~~ <u>djevojke</u>] vidio.
           [+Foc]

When an element in a phrase XP bears the [+Foc] focus feature, that phrase moves to SpecFP (or wherever [+Foc] is checked). The operative principle is now that there can be no material within the fronted phrase that is to the right of the element bearing the [+Foc] feature. All material following the [+Foc] element is thus marked "flat" intonation and is subsequently deleted in the mapping to PF. This results in pronunciation of the next highest copy of that material. This is shown schematically in (29).

(29)
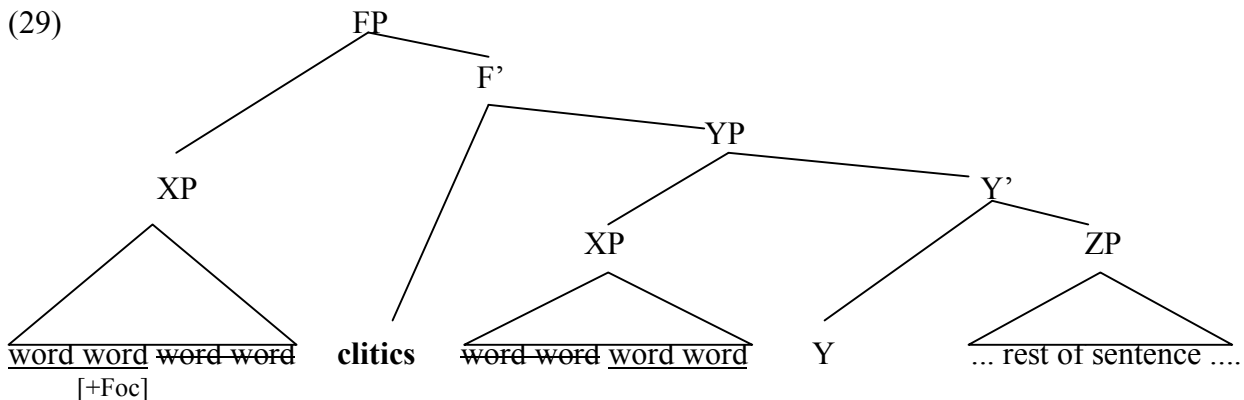
Crucially, as example (28a) shows most clearly, in this system material on neither side of the [+Foc] element is required to be a constituent: whatever precedes [+Foc] within a fact which is problematic for movement accounts. So, we need to say that the kind of deletion operative here is *not* constrained by constituency (unlike under traditional notions of ellipsis); instead, all that matters is linear adjacency. (Note that when non-clitics intervene there is more complex structure between the deletion and pronounced portions than indicated in (29), a factor which we believe makes focusing far more salient.)

This account extends to the analysis of splitting by focus *li* in Bulgarian in Franks 2006, based on ideas in Lambova 2003, even though Bulgarian as a DP language disallows true LBE. Thus (30a) and (30b) can be derived as in (31):

(30) a. <u>Novata</u> **li** <u>kniga na Ivan</u> vidja? 'Was it the NEW book by Ivan that you saw?'

    b. <u>V tozi</u> **li** <u>grad</u> si xodil? 'Have you been to THAT city?'

(31) a. [Novata [~~kniga na Ivan~~]] **li** [~~novata~~ [<u>kniga na Ivan</u>]] vidja?
       [+Foc]

    b. [V [<u>tozi</u> ~~grad~~]] **li** [~~v~~ [~~tozi~~ <u>grad</u>] si xodil?
        [+Foc]

It seems that material within a maximal projection that comes after the focus is deleted, causing it to be pronounced in a lower copy. The result is then we believe filtered out on the PF side, with considerable variation across speakers and languages in what can be tolerated.

    Peti-Stantić (in press), in examining whether spoken Croatian really adheres to second position for clitics, offers the examples in (32):

(32) a. Sva **su** moja djeca sasvim posebna.

      'All my children are completely special.'

    b. Sva moja **su** djeca sasvim posebna.

    c. Sva moja djeca **su** sasvim posebna.

    d. Sva moja djeca sasvim **su** posebna.

    e. Sva moja djeca sasvim posebna **su**.

She judges (32a, c, d) as "stylistically neutral" and (32b) as "slightly marked" and (32e) as more so. A perhaps more telling paradigm is given in (33):

(33) a. Svakog **ću** lijepog dana putovati vlakom svojoj kući.

      'Every nice day I will go to my house by train.'

    b. Svakog lijepog **ću** dana putovati vlakom svojoj kući.

    c. Svakog lijepog dana **ću** putovati vlakom svojoj kući.

    d. Svakog lijepog dana putovat **ću** vlakom svojoj kući.

    e. Svakog lijepog dana putovati vlakom **ću** svojoj kući.

    f. Svakog lijepog dana putovati vlakom svojoj **ću** kući.

Whereas Anita judges (33a, c, d) as "absolutely neutral", the second favors focus on *lijepog*, and the last two put some slight emphasis on the word immediately before (or after) the clitic.

    It is not easy to know how to derive such a range of possibilities. In a series of works, Franks and Bošković capitalize on clitics themselves moving and argue that lower clitic copy pronunciation is sometimes necessitated for PF reasons. These reasons are basically lack of prosodic support, either

because there is no host to the left of the enclitic or the host is not available because it is separated from the clitic by an intonational phrase boundary. So, at least for so-called "clitic third" (33c), it may be that lower pronunciation of *ću* is a PF phenomenon. Note, however, that since in none of these examples does Anita feel an intonation break, we may need to assume some degree of opacity in the derivation: the decision about what copy to pronounce is based on intonational information which may be subsequently erased (which makes some sense in that the presence of the clitic renders it recoverable). On the other hand, having the clitic *below* the verb is generally regarded as impossible; this is even true of English weak pronouns, as in (34).

(34) a. John threw out the books./John threw **'em** out.

b. John threw the books out./*John threw out **'em**.

It is however worth noting that Croatian (35) has exactly the same stylistic status as (33e):

(35) Svakog dana vlakom do svoje **putujem** kuće.

'Every day I go to my HOUSE by train.' OR 'Every day I go to MY house by train.'

The data in (33) and (35) thus reveal that, for Anita, clitics seem to be pronouncable much lower than expected, although when they do so, they are splitting just as tonic constituents do. So the first problem is getting the clitics below the verb: there should not even be a copy of *ću* that low in (33d) or (33e). But if we adopt the "adverb/negation" view of clitics, which (*modulo* PF considerations) tells is to treat them as the sun around which everything else revolves, then we are forced to conclude that (33d) and (33e) must derive from the syntactic structure in (36):

(36) [Svakog dana putovati vlakom svojoj kući] **ću** [svakog dana putovati vlakom svojoj kući]. Distributed deletion then applies, depending upon where the focus is.

PP-internal splitting provides another interesting illustration of the same general principles. Consider the range of splits in (37), again judged acceptable by Anita; readings provided are in decreasing order of preference:

(37) a. Od jučer **ga** prodaje <u>za velike novce</u>. [no splitting, neutral reading]

'Since yesterday (s)he's selling it for big bucks.'

b. <u>Za velike</u> **ga** <u>novce</u> prodaje od jučer.

'Since yesterday (s)he's selling it for big bucks.' [splitting, neutral reading preferred]

c. Od jučer prodaje <u>za velike</u> **ga** <u>novce</u>.

'Since yesterday (s)he's selling it for big BUCKS.' OR

'Since yesterday (s)he's selling it for BIG bucks.' OR JUST

'Since yesterday (s)he's selling it for big bucks.'

d. *Od jučer prodaje <u>za</u> **ga** <u>velike novce</u>.

Although (37c) is quite unexpected under familiar models of clitic placement, if the entire clause is fronted past *ga*, it can be derived as in (38a). Alternatively, one might imagine a combination of deletion and scrambling. That is, the clitic could be placed as in (38b), within the initial PP *za velike novce*, and then scrambling reorders things, roughly as in (38b).

(38) a. [Od jučer prodaje za velike ~~novce~~] **ga** [~~od jučer prodaje za velike~~ novce]
            [+Foc]

    b. [Za velike ~~novce~~] **ga** [~~za velike~~ novce] [prodaje [od jučer]]. [= (37b)] ➔

       [Od jučer] prodaje [za velike **ga** novce].

However, since this latter kind of derivation involves PF–side scrambling, it would require significant motivation.

    The phenomenon of low splitting is not restricted to a single pronominal clitic interrupting near the end of the sentence, as shown by the examples in (39), all for Anita acceptable. (39a) involves the auxiliary clitic *je* rather than pronominal *ga*, and the results are identical. (39b) involves splitting by a non-clitic, *stan*. It is curious that Anita's preferred emphasis is different depending on whether the splitter is a clitic or not: emphasis on *novce* in (39b) would according to her require "special intonation" and is "very unlikely". Even more interesting, the neutral reading is not available here. Note in this regard that it is hard to imagine how a PF–movement kind of derivation could be invoked when *stan* splits, so perhaps something like (38b) might in fact turn out to be the source of the neutral reading when it occurs with clitics as splitters. (39c–e) present another sort of problem which also might be amenable to PF–side scrambling, as these involve sentence internal splitting (either by an argument, one clitic, or more than one, respectively). Alternatively, the piece after the split might be treated as not part of the fronted focus-containing phrase.

(39) a. Uvijek prodavao kuće za velike **je** novce.

      'He always sold houses for big BUCKS.'   OR

      'He always sold houses for BIG bucks.' OR JUST 'He always sold houses for big bucks.'

    b. Od jučer prodaje za velike **stan** novce.

      'Since yesterday (s)he's selling the apartment for BIG bucks.'

    c. Prodaje za velike **stan** novce od jučer.

      'Since yesterday (s)he's selling the apartment for BIG bucks.'

    d. Prodaje za velike **ga** novce od jučer.

      'Since yesterday (s)he's selling it for big BUCKS.' OR

      'Since yesterday (s)he's selling it for BIG bucks.' OR JUST

      'Since yesterday (s)he's selling it for big bucks.'

e. Jučer kupio <u>od tog</u> **ga je** <u>čovjeka</u> za velike novce.

'Yesterday he bought it from that MAN for big bucks.' OR

'Yesterday he bought it from THAT man for big bucks.' OR JUST

'Yesterday he bought it from that man for big bucks.'

The variation in focus, that is, either before or after the split, raises an additional concern, since so far all we have said is that [+Focus] causes material following it to be deleted. In order to implement this variation, we need to admit the possibility that the [+Focus] feature can actually be marked on a *lower* copy. Compare (40) with the derivation in (38a) of (37c).

(40) [Od jučer prodaje za velike ~~novce~~] **ga** [~~od jučer prodaje za velike~~ novce]
                                                                             [+Foc]

In (40), as opposed to (38a), it is the lower instance of *novce* which is focused, hence must be pronounced. This forces the higher copy of *novce* to remain silent. However, the material preceding it—*od jučer prodaje za velike*—is pronounced, following the general principle that higher copies are pronounced so long as they do not violate PF requirements. This causes the lower instance of *od jučer prodaje za velike* to be deleted. Why then is something like (40) not a possible scenario when the splitter is not a clitic, for example, in (39b)? Our speculation is that in order for a phrase to be fronted to the high position involved in splitting it *must* be because of information features such as [+Focus], but this is *not* true of movement to the left of (i.e., to the Spec of) phrases headed by clitics. Of course, (40) raises the far from insignificant question of how the two copies can be different and why in the higher one *novce* is not similarly the locus of a [+Focus] feature. One answer might be that [+Focus] in this case is not actually present in the syntax (since otherwise it should be copied), but exactly how to implement this insight eludes us.

Finally, what about (37d), in which the clitic cannot go immediately after the preposition? The syntactic source of the largely universal cohesion between preposition and complement is a widely researched topic, about which we will have nothing to say. However, if we are correct that whatever is going on here is a matter of PF–side deletion, one wonders why syntax should be relevant. And indeed, it seems as though on some level (37d) must be rejected out of PF considerations. The reason is that when such splittings involve a preposition that has an adverbial variant (i.e., need not be transitive), the splitting is not in fact filtered out by PF. Consider (41) and (42):

(41) a. <u>Ispred</u> **ga je** <u>ulaza</u> dočekala policija.

'The police were waiting for him in front of the exit.'

b. <u>Pored</u> **je** <u>tog čovjeka</u> sjela.

'She sat next to that man.'

(42) a. On je sjedio ispred/pored.

    b. Ispred/Pored je sjedio.

      'He was sitting in front/right by.'

Prepositions that do not admit intransitive usage block splitting, such as *prema* in (43):

(43) a. *<u>Prema</u> **je** <u>tom čovjeku</u> došao.   [cf. ✓<u>Prema tom čovjeku</u> **je** došao. OR

      'He was coming towards that person.'   ✓<u>Prema tom</u> **je** <u>čovjeku</u> došao.]

    b. *<u>Išao</u> **je** prema.   [Intended: 'He was going towards.']

If we are correct, the deletions in (44a) are valid but the ones in (44b) are not:

(44) a. ✓[Pored ~~tog čovjeka~~] **je** [~~pored~~ tog čovjeka] ...

    b. *[Prema ~~tom čovjeku~~] **je** [~~prema~~ tom čovjeku] ...

PF–side deletion can create both, but (44b) is filtered out, whereas (44a) is not. In general, our survey of prepositions shows that they fall into two classes along precisely these lines: splittability is enabled by potential (although not actual!) intransitivity. The correlation raises interesting questions we unfortunately do not have time to go into here.

    Consider finally the contrast in (45) reported by Fanselow and Ćavar 2002:

(45) a. <u>Na veliko</u> **se Ivan** <u>drvo</u> popeo. 'It was on a BIG tree that Ivan climbed.'

    b. *<u>Drvo</u> **se Ivan** <u>na veliko</u> popeo.

    c. (?*) <u>Na veliko</u> **se Ivan popeo** <u>drvo</u>.

(45a) is derivable in the same way as before, assuming an intermediate copy of *na veliko drvo* at the left edge of the VP, so that it precedes *popeo* but follows *se* and *Ivan*, as in structure (46a). (45b), with the hypothetical distributed deletion analysis in (46b), is bad because "inverted" splits (as are common in German) are not in fact derivable through distributed deletion, but we claim only through the base generation mechanism, each part displaying full morphology. (45c) on the other hand raises some problems. By our account, it should involve the structure in (46c).

(46) a. [Na veliko ~~drvo~~] **se Ivan** [[~~na veliko~~ drvo] [popeo [~~na veliko drvo~~]].

    b. *[~~na veliko~~ Drvo] **se Ivan** [[na veliko ~~drvo~~] [popeo [~~na veliko drvo~~]].

    c. (?*)[Na veliko ~~drvo~~] **se Ivan** [[~~na veliko drvo~~] [**popeo** [~~na veliko~~ drvo]].

While speakers who have written about this phenomenon find (45c) severely degraded, Anita does not. The difference between (45a)/(46a) and (45b)/(46b) is clearly whether it is the low copy of *drvo* or the next highest one which is pronounced. And while the latter seems required for many speakers, the fact that Anita allows something like (46c) is at least consistent with the other judgments she reports.

5. Conclusion. Recent linguistic theory countenances a superfluity of mechanisms for generating discontinuous constituents. These include PF mechanisms of linearization and deletion, syntactic mechanims of movement either of the left-branch or of a remnant constituent, and the LF mechanism of lowering to form fully interpretable phrases. What we are in need of is a theory that limits the search space of possible hypotheses. Unfortunately, in this talk we have surely raised more questions than we have answered. (Since we haven't really answered any!) The data are complex, different languages in all likelihood employ different mechanisms, and there is considerable disagreement among individual speakers. Colloquial Croatian, in particular, seems to allow much more word order latitude than has been previously recognized. Acknowledging this fact will, we hope, be the first step towards coming to grips with the problem

References

Bašić, Monica. 2004. *Nominal Subextractions and the Structure of English*. MA Thesis. University of Tromsø.

Bošković, Željko. 2001. *On the nature of the syntax–phonology interface*. Amsterdam: Elsevier.

Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59:1–45.

Bošković, Željko. 2006. What will you have, DP or NP? Talk at NELS 37, University of Illinois at Urbana-Champaign.

Bošković, Željko and Daiko Takahashi. 1998. Scrambling and last resort. *Linguistic Inquiry* 29: 347–366.

Browne, Wayles. 1975. Contrastive patterns in English and Serbo-Croatian conjoining. In Rudolf Filipović, ed. *Contrastive analysis of English and Serbo-Croatian*, vol. 1, Zagreb: Zavod za lingvistiku, 135–142.

Corver, Norbert. 1992. Left branch extraction. *Proceedings of NELS* 22, 67–84.

Fanselow, Gisbert and Damir Ćavar. 2002. Distributed deletion. In A. Alexiadou, ed. *Theoretical approaches to universals*. Amsterdam: John Benjamins, 65–107.

Franks, Steven. 2006. Another look at *li* placement in Bulgarian. *The linguistic review* 23:161–210.

Franks, Steven and Tracy Holloway King. 2000. *A handbook of Slavic clitics*. New York: Oxford University Press.

Franks, Steven and Ljiljana Progovac. 1994. On the placement of Serbo-Croatian clitics. In *Indiana Slavic Studies 7, Proceedings of the 9th biennial bonference on Balkan and South Slavic linguistics, literature and folklore*, 69-78.

Hale, Kenneth. 1973. Person Marking in Walbiri. In *A festschrift for Morris Halle*, ed. by Paul Kiparsky, pp. 308–344. New York: Holt, Rinehart & Winston.

Hale, Kenneth 1981. *On the position of Walpiri in the typology of the base*. Bloomington: Indiana University Linguistics Club.

Lambova, Mariana. 2003. *On information structure and clausal architecture: Evidence from Bulgarian*. Ph.D. dissertation. University of Connecticut.

Marušič, Franc and Rok Žaucer. 2006. The 'definite article' TA in colloquial Slovenian. In *Formal approaches to Slavic linguistics: The Princeton meetin*g, ed. by J. Lavine, S. Franks, M. Tasseva-Kurktchieva, and H. Filip, pp. 189–204.

Laury, Ritva. 1997. *Demonstratives in Interaction: The emergence of a definite article in Finnish*. Amsterdam and Philadelphia: John Benjamins.

Pereltsvaig, Asya. 2006. Split phrases in colloquial Russian. Unpublished ms. Cornell University.

Peti-Stantić, Anita. in press. Wackernagel – dobar domaćin ili naporan gost? In *Sintaktičke kategorie – promjene i stanje*, ed. by Branko Kuna.

Platzack, Christer. 2006. Left Branch Extraction in Old Icelandic. Ms, University of Lund.

Progovac, Ljiljana. 1996. Clitics in Serbian/Croatian: Comp as the Second Position. In A. Halpern and A. Zwicky, eds. *Approaching second: Second position clitics and related phenomena*, CSLI Publications: Stanford, pp. 411–428.

Progovac, Ljiljana. 2006. *A syntax of Serbian: Clausal architecture*. Bloomington, Ind: Slavica.

Schütze, Carson. 1994. Serbo-Croatian second position clitic placement and the phonology syntax interface. *MIT working papers in linguistics*, vol. 21, 373–473.

Uriagereka, Juan. 1988. *On government*. Ph.D. dissertation. University of Connecticut.