Extraction or splitting? How A’ movement phenomena in Lithuanian can shed light on the syntax-phonology interface and the NP-DP debate.∗

Janneke Van Hofwegen

March 25, 2013

0 Introduction

The phenomenon of Left Branch Extraction (LBE) is as of yet not fully understood, especially as it was at one time thought to be impossible. Ross (1967) proposed the Left Branch Condition, which blocks movement out of the leftmost constituent of the NP, after noting that such cases were impossible in English.

(1) a. *Whose did you buy [t1 necklace].
   b. *Which did you buy [t1 necklace].
   c. *That1 you bought [t1 necklace].
   d. *Beautiful1 you bought [t1 necklaces].
   e. *How many1 did you buy [t1 necklaces].

However, while this constraint is quite robust in English, it did not take long for Ross and others (Corver 1992) to note examples of LBE in other languages like Russian and Latin.

(2) a. Novuju Ivan kupil [t1 mašinu].
   New1 Ivan bought [t1 car]
   ‘Ivan bought a new car’
   Russian

   b. Cuiam1 amat Cicero [t1 puellam].
   whose1 loves Cicero [t1 girl]
   ‘Whose girl does Cicero love?’
   Latin

Even as more and more examples of grammatical LBE in other languages have emerged, the phenomenon still remains flagrantly unacceptable in English. The question remains: what are the relevant characteristics of these languages that enable them to allow LBE? To what extent might

∗I thank Rasa Matusiaviciūtė Masoudi for the many hours she spent sharing Lithuanian with me, and Simona Mačiulaitė-Berkner for providing me the data for the pilot portion of this study. I thank Vera Gribanova for all the weekly meetings and so much teaching and patient encouragement along the way. Thanks also to Tom Wasow for guidance and helpful feedback throughout, and to Penny Eckert for reading and comments. I’m also grateful to my student colleagues, especially James Collins and Dasha Popova, for many hours of discussion and lots of help on content and analysis. All errors are my responsibility.

1Example taken from Grebenyova (2005: p. 168)
2Example taken from Uriagereka (1988)
they differ in fundamental noun phrase structure to allow this sort of extraction? Alternatively, what might this say about movement/merge possibilities in LBE languages vis-à-vis non-LBE languages? Given LBE’s prevalence in the Slavic languages in particular, Slavic linguists (i.e., Bošković 2005, 2009, 2012; Rappaport 2001; Pereltsvaig 2007, 2008; Progovac 1998; Zlatic 1998; Trenkic 2004) have readily embarked on the task of analyzing LBE, with two distinctive perspectives emerging.

The first perspective, espoused by Bošković and others, concentrates first on the structure of the noun phrase and then on particular movement constraints for the elements involved. It does not address the motivation for movement of the extracted elements themselves. Fundamental to this approach is the putative observation that LBE can occur only in languages that lack definite articles, which Bošković concludes indicates no Determiner Phrase (DP) structure. All the Slavic languages save for Macedonian and Bulgarian (for which LBE is disallowed, see below) would fall in this category. Combined with other evidence, the lack of a DP suggests a noun phrase structure where AP is adjoined to the NP. Bošković shows how adjective-like elements (including demonstratives, quantifiers, wh-words, etc. as well as adjectives) are adjuncts. This is in contrast to a structure whereby the NP is the complement of A (see Abney 1987 and this paper, section 2, for more about these structural differences). As the Bošković theory goes, prenominal elements (as adjuncts) are able to extract out and over the noun they modify. This is because there is no D to block their movement out of the VP. The theory posits that extraction is impossible in DP languages because D is a phase and blocks such movement.

The other perspective, introduced by Pereltsvaig, argues first against the notion that Slavic languages do not have a DP (Pereltsvaig 2007), and second that whether they do or not has no relevance to the extractability of their prenominal elements (Pereltsvaig 2008). Rather, the phenomenon is motivated by focus and information structure, such that the noun phrase may “split” into two or more pieces, which can then be pronounced in various positions. Under this analysis, “splitting” is actually a copy/merge phenomenon whereby the entire phrase is copied, moved, and pieces of it are pronounced in PF according to discourse/focus motivations.

While the two perspectives are incompatible with each other, it must be said that Bošković and Pereltsvaig are moving toward two distinctly different goals, albeit working on the same phenomenon (i.e., extraction/splitting). For Bošković, the investigation of left-branch extraction in Slavic languages is part of a much larger research project that is typological in nature. The typology that Bošković is arguing for includes a general distinction between so-called ‘NP’ languages and ‘DP’ languages. His work on LBE contributes toward a list of eighteen generalizations (see Bošković 2012) which theoretically distinguishes between DP/NP languages. Section 2.1 goes into more detail about some of those generalizations, but the general thrust of Bošković’s proposal is that the presence of D in the noun phrase has important structural and semantic ramifications for any language (e.g., English, German, Hebrew, Italian) that uses D as a category.

For Pereltsvaig, the task is different. Namely, she is not arguing for a typology of NP/DP languages; rather, she is simply examining a specific phenomenon (‘splitting’) in colloquial Russian from a syntactic, phonological, and pragmatic point of view.3 While Bošković’s analysis of left-branch extraction hinges upon his theoretical typological distinction between NP and DP languages, Pereltsvaig’s analysis does not touch upon this issue, and in fact doesn’t need to. Instead, she demonstrates how the Bošković account of LBE doesn’t hold for her data for other reasons. Bošković, meanwhile, has not yet attempted to refute or debate the specific mechanics of Pereltsvaig’s analysis. Thus, while the two of them are analyzing the same phenomenon, they are not directly engaging in a debate...
over what extraction/splitting is or how it happens; moreover, they differ in the extent to which they intend to make typological generalizations.

Though Bošković has extended his typology beyond Slavic languages (his 2012 paper cites about forty languages for which he claims it holds true), the bulk of his work and Pereltsvaig’s is focused on the Slavic language family and, accordingly, so is also the basis of his analysis of LBE. In light of the relative significance of Slavic languages to these questions, it is helpful to examine another language - Lithuanian. As a related Indo-European language, Lithuanian (of the Baltic branch) shares important structural properties with the Slavic languages, including rich morphology, similar sentence structure, and robust case-marking of nouns and their modifiers. However, Lithuanian differs from the majority of the Slavic languages in that it utilizes an overt definite article - definiteness marking on prenominal elements - and thus (by Bošković’s parameters) has a DP structure. Therefore, whether LBE and other extraction phenomena occur in Lithuanian would address the question of whether D’s absence is crucial for allowing their movement.

At the same time, whether or not a DP has anything to do with it, the phenomenon of extraction (or ‘phrase-splitting’ as Pereltsvaig terms it) is poorly understood. What elements are moveable/extractable? What structural possibilities allow for this type of movement, and what prohibits it? What other interfaces (i.e., phonological, pragmatic) are integral to the process? An investigation of extraction possibilities in Lithuanian, as this paper argues, can address these questions more satisfactorily than the Slavic languages have thus far.

This paper analyzes noun phrase structure in Lithuanian and the extraction/splitting possibilities thereof. Specifically, it compares analytical approaches used for Slavic languages offered by Bošković (see 2005, 2009, 2012) against those of Pereltsvaig (2008) to Lithuanian, in order to address first the phenomenon of left-branch extraction (LBE) from the noun phrase, and then other types of extraction/movement phenomena. The paper will proceed as follows: first, Section 1 will describe the properties of the Lithuanian Noun Phrase, specifically those relevant to the extraction/movement discussion. Then, Section 2 uses the Bošković analysis to examine the extent to which LBE can occur in Lithuanian, what it looks like structurally, and whether it interacts with/is blocked by the appearance of a determiner. Crucially, the Lithuanian evidence shows that LBE readily occurs regardless of the relative definiteness of the noun phrase (conflicting directly with Bošković’s assertions) and can even occur in more diverse contexts and manners than has previously been shown. Finally, in Section 3, an alternative view is proposed - that of ‘splitting’ (Pereltsvaig 2008). In Lithuanian, phrase splitting (as opposed to extraction) can occur regardless of definiteness parameters and appears to be a matter of focus and information structure. This section applies the Pereltsvaig analysis to Lithuanian and identifies: 1) its comparative suitability to the Lithuanian data vis-à-vis the Bošković analysis; and, 2) two important shortcomings that suggest the Pereltsvaig analysis is too unidimensional.

1 The Lithuanian Noun Phrase

For many reasons, Lithuanian is a suitable language in which to investigate these various analyses. The following section, 1.1, gives a bit of background on this understudied language and its proximity to the Slavic family. Then, 1.2 shows the basic structural characteristics of the typical Lithuanian

---

3This is not to say that Pereltsvaig has not engaged with Bošković in the NP/DP debate (see the back-and-forth argument on the topic documented in Bošković 2005, Pereltsvaig 2007, and Bošković 2009).
noun phrase, its elements, and its unmarked word ordering. Finally, 1.3 explains the feature that makes Lithuanian distinct from the majority of its Slavic neighbors - the definiteness paradigm.

1.1 Background

Lithuanian is a time capsule of sorts in the context of modern Indo-European languages. As one of only two surviving members of the Baltic branch (the other being Latvian), it is widely touted as the most conservative of the Indo-European family. As Lithuanian grammarian Vytautas Ambrazas (2006: p. 5) explains:

"...[Lithuanian] has preserved many archaic features which can be directly observed and investigated in their present-day usage...including its rich inflection...and an extremely variable word order which reflects the complicated relations between the communicative and the syntactic levels of discourse."

This conservatism of form and structure makes Lithuanian a rich vehicle through which linguists can study natural forces for linguistic conservation, both internal and external. In fact, what structural change has been documented in the language has challenged the usual assumptions for how morphological change occurs: not as a process of reduction or decreasing morphological complexity, but rather in the direction of richer, more complex inflectional systems. Stolz (2010: p. 218) writes:

"it turns out that...Lithuanian has been subject to a variety of morphologically substantial (although sometimes only ephemeral) changes. Many of these processes do not belong to the usual garden variety of reductive changes like loss and merger. On the contrary, they attest an increase in morphological complexity both of paradigms and word forms...Thus, the already heavily inflecting Lithuanian has become even more heavily inflecting over time."

In this way, modern Lithuanian is a language that explicitly tells its story by marking its internal syntactic relations, verbal aspect, definiteness, and other phenomena, through inflectional systems not always retained in other Indo-European languages, and in sometimes quite innovative ways. In spite of these idiosyncrasies, Lithuanian remains largely understudied in the field of linguistics outside of its immediate regional context and in generative syntax specifically.

Despite its archaic qualities, Lithuanian’s genetics reflect close parallel and interactional development with other Indo-European languages, most notably the Germanic and Slavic branches. In fact, regarding the Slavic languages in particular, there is some debate as to whether Lithuanian and Slavic languages are related under a Proto-Balto-Slavic ancestor. Regardless of whether this genetic relationship can be proven, linguists nevertheless concede that “many of the similarities shared by Baltic and Slavic reflect not just a period of common prehistory, but the fact that they were neighbors from Proto-Indo-European times to the present and thus kept influencing each other for millennia, both in structure and vocabulary” (Hock & Joseph 2009: p. 53-4). Therefore, it would seem as though Slavic linguistics can and should be informed by studies in the much lesser-known field of Baltic linguistics, an endeavor to which this paper seeks to contribute.

1.2 Description of the Lithuanian Typical Noun Phrase (TNP)

As a robustly inflected language, Lithuanian encodes the syntactic relations among words and phrases through endings for case, gender, and number. Relatedly, word/phrase order within sentences is relatively free, although the unmarked pattern is S-V-O. The head noun of the noun phrase
in Lithuanian carries whatever case marking is appropriate, while all modifiers to the noun phrase (save for the genitive attributive nouns) share agreement features with the head noun. The noun itself is always inflected for gender (masculine or feminine), and is most commonly inflected for number (singular or plural; some dual). Additionally, all nouns in a given sentence, save for some foreign loan words, are marked for case. Modern standard Lithuanian exhibits six functional cases (nominative, accusative, genitive, dative, instrumental, and locative) as well as a vocative case.

The specific morphology of the case endings is determined not only by the function of the noun, but also by its gender and number.

In a typical noun phrase (TNP), the head noun is usually the last element, and modifiers for the noun, including adjectives, participles, numerals, demonstratives, quantifiers precede the head noun as applicable. Additionally, genitive nouns also precede the head noun. When all are present, these modifiers typically follow a preferred order and in fact (particularly in the case of genitives) the order is more strict than is English in keeping the modifiers prenominal, resulting in considerable potential for constituent stacking. See example (3):

(3) a. \( \text{kat-\text{\text{-}}} \)
   \( \text{cat-F.SG.NOM} \)
   'cat’

b. \( \text{\text{\text{-}}} \text{sit-a} \text{\text{-}}} \)
   \( \text{\text{\text{-}}} \text{kat-\text{\text{-}}} \)
   \( \text{this-F.SG.NOM cat-F.SG.NOM} \)
   'this’

c. \( \text{\text{\text{-}}} \text{sit-a} \text{\text{-}}} \text{jaun-a} \text{\text{-}}} \)
   \( \text{\text{\text{-}}} \text{kat-\text{\text{-}}} \)
   \( \text{this-F.SG.NOM young-F.SG.NOM cat-F.SG.NOM} \)
   'this young cat’

d. \( \text{\text{\text{-}}} \text{sit-\text{\text{-}}} \text{keturi-os} \text{\text{-}}} \text{jaun-\text{\text{-}}} \text{\text{-}}} \text{kat-\text{\text{-}}} \)
   \( \text{\text{\text{-}}} \text{these-F.PL.NOM four-F.PL.NOM young-F.PL.NOM cats-F.PL.NOM} \)
   'these young cats’

e. \( \text{\text{\text{-}}} \text{vis-\text{\text{-}}} \text{sit-\text{\text{-}}} \text{keturi-os} \text{\text{-}}} \text{jaun-\text{\text{-}}} \text{\text{-}}} \text{kat-\text{\text{-}}} \)
   \( \text{\text{\text{-}}} \text{all-F.PL.NOM these-F.PL.NOM four-F.PL.NOM young-F.PL.NOM cats-F.PL.NOM} \)
   'all these young cats’

f. \( \text{\text{\text{-}}} \text{vis-\text{\text{-}}} \text{sit-\text{\text{-}}} \text{keturi-os} \text{\text{-}}} \text{jaun-\text{\text{-}}} \text{\text{-}}} \text{Jon-o} \)
   \( \text{\text{\text{-}}} \text{all-F.PL.NOM these-F.PL.NOM four-F.PL.NOM young-F.PL.NOM John’s-M.SG.GEN} \)
   \( \text{kat-\text{\text{-}}} \)
   \( \text{cats-F.PL.NOM} \)
   'all these young cats belonging to John’

Notably absent in the above Lithuanian noun phrases is the definite article. While the language has

---

4 Hock & Joseph (2009) note three additional cases: ‘illative,’ ‘allative,’ and ‘adessive,’ but these are considered obsolete, and are thus not addressed in modern Lithuanian grammars.

5 Glossing abbreviations throughout: 1 first person, 2 second person, 3 third person, ACC accusative case, DAT dative case, DEF definite, DIM diminutive, FUT future, GEN genitive case, INS instrumental, LOC locative case, NOM nominative case, PL plural, PRS present, PRF perfect, PST past, PTP participle, Q polar question, SG singular
a variety of means for marking definiteness on nouns it does not utilize definite/indefinite articles in the way that English and other languages do. Definiteness marking in Lithuanian will be discussed in more detail in Section 1.3.

While the orderings of elements in the NPs of example (3) are the most unmarked, they are not restricted to this order. In fact, the modifiers can be switched around in any number of ways, resulting in phrases that may be marked in some contexts, but are quite unmarked in others if emphasis and context are taken into account. For example, the possessive genitive noun does not have to immediately precede the noun, but can be intersected by an adjective, demonstrative, or a quantifier. In the following example, (4-a) represents the unmarked, typical ordering, but (4-b) is also perfectly possible.

(4) a. šit-as mam-os kambar-ys
    this-F.SG.NOM mother’s-F.SG.GEN room-M.SG.NOM
    ‘this room belonging to mother’

    b. mam-os šit-as kambar-ys
    mother’s-F.SG.GEN this-M.SG.NOM room-M.SG.NOM
    ‘this room belonging to mother’

The same re-ordering is possible for the possessive and a modifying adjective. (5-a) is an unmarked example, while (5-b) is marked but readily available.

(5) a. sen-as mam-os kambar-ys
    old-M.SG.NOM mother’s-F.SG.GEN room-M.SG.NOM
    ‘old room belonging to mother’

    b. mam-os sen-as kambar-ys
    mother’s-F.SG.GEN old-M.SG.NOM room-M.SG.NOM
    ‘old room belonging to mother’

Likewise, demonstratives, adjectives, and possessives can be rearranged. Examples in (6) are typical orderings, but their semantically identical counterparts in (7) are fine as well.

(6) a. vis-os šit-os nuotrauk-os
    all-F.PL.NOM these-F.PL.NOM pictures-F.PL.NOM
    ‘all these pictures’

    b. vis-os ses-ers nuotrauk-os
    all-F.PL.NOM sister’s-F.SG.GEN pictures-F.PL.NOM
    ‘all sister’s pictures’ or ‘all pictures belonging to sister’

    c. vis-os sen-os nuotrauk-os
    all-F.PL.NOM old-F.PL.NOM pictures-F.PL.NOM
    ‘all old pictures’

(7) a. šit-os vis-os nuotrauk-os
    these-F.PL.NOM all-F.PL.NOM pictures-F.PL.NOM
    ‘all these pictures’
b. *ses-ers vis-os nuotrauk-os*
   sist'er's-F.SG.GEN all-F.PL.NOM pictures-F.PL.NOM
   ‘all sister's pictures’ or ‘all pictures belonging to sister’

c. *sen-os vis-os nuotrauk-os*
   old-F.PL.NOM all-F.PL.NOM pictures-F.PL.NOM
   ‘all old pictures’

The more stacked these modifiers get (as in example (3)), the more unnatural they become, so it is admittedly hard to test all possible ordering permutations among them with a native speaker. However, the data clearly illustrate some scrambling of these elements in the prenominal context.

1.3 Definiteness in Lithuanian

As noted above, in noun phrase structures, modern standard Lithuanian does not regularly utilize definite or indefinite articles. In fact, learners of Lithuanian are regularly told the language is devoid of such features. Rather, the relative definiteness of any particular noun is most often inferred through conversational context. Gillon & Armoskaitė (2012) hypothesize that bare nouns in Lithuanian are by default (covertly) definite, and thus tend to use special constructions to confirm an indefinite reading. In the following examples, (8-a) is the default, unmarked S-V construction, and carries a definite reading, while (8-b), the inverted V-S structure, tends to read as indefinite.

(8) a. Šun-uk-as vaikščio-jo.
    dog-DIM-M.SG.NOM run-PST.3SG
    ‘The doggie ran.’

b. Vaikščio-jo šun-uk-as.
    run-PST.3SG dog-DIM-M.SG.NOM
    ‘A doggie ran.’

It is no accident that the indefinite reading corresponds to the V-S construction. As Trenkic (2000) notes, this word order is commonly used in Slavic languages and other ‘relatively free word-order’ (2000: p.13) languages as a means of differentiating definite and indefinite noun phrases. As Trenkic explains, a noun phrase in initial sentence position (as in (8-a)) is ‘given’ information, meaning it has already been introduced in the discourse, and the material proceeding after it is new. However, if the noun phrase is indefinite (i.e., not already provided in the discourse), it is often given final sentence position (as in (8-b)), indicating it is ‘new’ information, and therefore the preceding material is given. Though this pattern is common, it is not categorical; there are other ways of marking definiteness. Gillon & Armoskaitė (2012) find that while the V-S reads strongly as indefinite, the S-V construction only reads preferentially as definite. Among other things, this leads them to conclude that while the default Lithuanian noun is definite, bare nouns (i.e., indefinite with no overarching DP structure) are also possible in Lithuanian.

1.3.1 Definiteness Marking in Lithuanian

While the S-V/V-S means of providing definite readings of nouns is most prevalent in Lithuanian discourse, it is not the only means of doing so. In fact, Lithuanian utilizes overtly pronounced so-called *long* morphology that can be attached to adjectives, participles, demonstratives, ordinal
numerals (Ambrazas 2006), and some interrogatives and superlatives (Stolz 2010). When present on one of these prenominal modifiers, the long morphology unambiguously denotes definiteness on the head noun. Example (9) compares a non-definite-marked noun to (10)’s definite-marked noun (šunukas/’dog’) with respect to adjectives, participles, ordinal numerals, and interrogatives respectively.

(9) Non-definite marked

a. balt-as šun-uk-as
   white-M.SG.NOM dog-DIM-M.SG.NOM
   ‘a/the white doggie’

b. mieg-a nt-is šun-uk-as
   sleep-PRS.3SG-PTCP-M.SG.NOM dog-DIM-M.SG.NOM
   ‘a/the sleeping doggie’

c. pirm-as šun-uk-as
   first-M.SG.NOM dog-DIM-M.SG.NOM
   ‘a/the first doggie’

d. kelint-as šun-uk-as
   which-M.SG.NOM dog-DIM-M.SG.NOM
   ‘which doggie’

(10) Definite marked

a. balt-as-is šun-uk-as
   white-M.SG.NOM-DEF.M.SG.NOM dog-DIM-M.SG.NOM
   ‘the white doggie’

b. mieg-a nt-ys-is šun-uk-as
   sleep-PRS.3SG-PTCP-M.SG.NOM-DEF.M.SG.NOM dog-DIM-M.SG.NOM
   ‘the sleeping doggie’

c. pirm-as-is šun-uk-as
   first-M.SG.NOM-DEF.M.SG.NOM dog-DIM-M.SG.NOM
   ‘the first doggie’

d. kelint-as-is šun-uk-as
   which-M.SG.NOM-DEF.M.SG.NOM dog-DIM-M.SG.NOM
   ‘which particular doggie’

At the same time as Stolz (2010) says it is obligatory for definite contexts, other Lithuanian grammarians (e.g., Ambrazas 2006) observe that the long definite morphology is optional and is largely falling out of use save for classification/taxonomic purposes (e.g., saldie-ji pipirai means ‘Sweet pepper’ as a species of pepper, not that a particular pepper is sweet). In her dissertation, which

---

6The -ys morpheme here is a derivation of -is (the usual morpheme used for masculine, singular, nominative adjectives of the (i)a class (Ambrazas 2006). In the definite construction, this morpheme receives stress, and thus the vowel is tensed and raised, resulting in -ys.

---
surveyed *The Corpus of Spoken Lithuanian Language* and *The Corpus of Contemporary (Written) Lithuanian Language.⁷* Spraunienė (2011) notes that the classifying/taxonomic denotation is the most prevalent usage of definite morphology (58 percent of all occurrences) in the corpora. In fact, the taxonomic rendition is robustly productive in Lithuanian (cf. Rutkowski & Progovac 2006), while the simple definite denotation of the morpheme is falling out of use. Spraunienė observes that in the corpora she surveyed, only 23.2 percent of all of the adjectives modifying nouns with definite readings were marked with long morphology. Altogether, this means that roughly 13 percent of all of the adjectives modifying definite nouns in the corpora used the long morphology to denote definiteness.

Nevertheless, while the long-morphology convention appears to be falling out of use in Lithuanian, it is still alive in the grammar and still provides an unambiguously definite reading of a noun. Spraunienė (2011) and others observe that its use tends to be considered redundant if definiteness is readily interpreted through context. Thus a speaker would be likely to use examples like (9) by default if the *sunukas* (‘doggie’) has already been introduced in previous discourse. Not only pragmatically, but also structurally, the definiteness marker is very sensitive to redundancy. For example, it is constrained to occurring once in the noun phrase (on the leftmost element) if denoting definiteness to the head noun is necessary by context. If used once, it will never be used again for the same noun in the discourse, except in circumstances where contrastive emphasis would necessitate a renewed mention of the particularity of a given noun. In the following example, (11-a) illustrates a typical definite-marked adjective and noun. (11-b) shows an ungrammatical example, in which both the participle and the adjective are definite-marked. Finally, (11-c), likewise ungrammatical, depicts a case where a definite-marked element is not the leftmost element in the noun phrase.

(11) a. *balt-as-is*  *šun-uk-as*
    white-M.SG.NOM-DEF.M.SG.NOM  dog-DIM-M.SG.NOM
    ‘the white doggie’

b. *mieg-a-nt-ys-is*  *balt-as-is*
    sleep-PRS.3SG-PTCP-M.SG.NOM-DEF.M.SG.NOM  white-M.SG.NOM-DEF.M.SG.NOM  
    *šun-uk-as*
    dog-DIM-M.SG.NOM
    ‘the sleeping the white doggie’

c. *mieg-a-nt-is*  *balt-as-is*  *šun-uk-as*
    sleep-PRS.3SG-PTCP-M.SG.NOM  white-M.SG.NOM-DEF.M.SG.NOM  dog-DIM-M.SG.NOM
    ‘sleeping the white doggie’

If these are general descriptive characteristics of the definite-marking paradigm in Lithuanian, the question remains about how this paradigm relates to paradigms in related languages.

1.3.2 A little more about Lithuanian definiteness marking

So-called ‘long’ definite morphology on prenominal constituents is not unique to Lithuanian, but can also in fact be found in several Slavic languages. For example, Serbo-Croatian (cf. Progovac 1998), and Bulgarian (cf. Harizanov 2009) utilize ‘special’ morphology, which can arguably denote

---

⁷These corpora do not appear to be publicly accessible.
definiteness for the head noun of a noun phrase. Additionally, Slovenian (cf. Marušič & Žaucer to appear) uses a particle ta argued to be comparable to the long morphology of Serbo-Croatian. Of the three languages, Bulgarian’s marker is more widely recognized as a true definite article (Bošković 2012), while the article-like status of these morphemes in the other languages is up for debate (Marušič & Žaucer to appear). How might the Lithuanian definiteness marker compare to the markers in the aforementioned languages?

**Slovenian and Serbo-Croatian** As Marušič & Žaucer (to appear) explain, the Slovenian ta is comparable to the Serbo-Croatian long adjectival morphology in that it cannot appear with bare nouns, but is restricted to marking adjectives, and only for noun phrases in the nominative case (additionally, Serbo-Croatian “definite” morphemes mark only adjectives modifying masculine nouns). It can appear multiple times in a given noun phrase, as in cases of stacked adjectives, the marker appears once on each adjective. The morpheme exhibits properties consistent with clitics rather than with affixes, as it cannot carry lexical stress (see Marušič & Žaucer to appear for relevant examples). Like the Lithuanian paradigm, the long morphology particle in Slovenian or Serbo-Croatian seems to be an optional part of adjectival morphology (it does not appear obligatorily for every semantically-definite noun phrase), but it differs in that it does not appear with other prenominal elements. Crucially, while it has been called a definiteness marker by Slavic linguists (cf. Progovac 1998), Marušič & Žaucer provide convincing evidence for its use (both for Slovenian and Serbo-Croatian) in explicitly indefinite noun phrases.

**Bulgarian** In contrast, the Bulgarian definite marker, Harizanov (2009) argues, is not tied to the adjectival structure alone, but instead will generally appear on the leftmost element of the noun phrase, whether that element be an adjective, a possessor, a quantifier, or even a bare noun. A notable exception is the demonstrative, which cannot carry the marker. While the marker has the distributional properties of a clitic, Harizanov argues that it is instead a definiteness agreement morpheme - an affix - which originates in the D head, but it is not itself a syntactic head. The def agreement morphology is inserted post-syntactically, by “adjunction of agreement nodes to the agreeing stems” (Harizanov 2009: p. 21), which originate in a phonologically null D. Similarly to Lithuanian, the adjective containing the def marker agrees with the head noun in φ features first, but then undergoes specific morphological metamorphosis indicating it is +def. This evidence, Harizanov stipulates, predicts why the def morpheme is always the last suffix to attach to the stem. He goes on to say that +def agreement is a separate process from φ agreement, and thus prenominal elements are not expected to have +def agreement simply because they show φ agreement. Thus, the definiteness marker in Bulgarian, while it might share surface similarity to the long morphology of Slovenian or Serbo-Croatian, actually plays a more fundamental role in the grammar, and is used exclusively in semantically-definite domains.

**Lithuanian** While an in-depth investigation of the definite article status of Lithuanian long morphology is outside the scope of this paper, at minimum we can see that the marker in Lithuanian patterns more closely with the Bulgarian marker (which Bošković has no qualms about labeling as an article) than it does with the long morphology of Slovenian or Serbo-Croatian. At first glance, it does share some similarities with them, however. Like Serbo-Croatian and Slovenian, the definite marker in Lithuanian originates as a form of “long” morphology customarily placed on adjectives. There are mixed reports of its relative obligatoriness (see section 1.3.1), but certainly it is not mandatory in every definite context (although it would never emerge in indefinite contexts). Additionally, the long morphology in both cases arises genetically from the same source; it is a remnant
of a pronoun (third-person in the case of Serbo-Croatian and Lithuanian, demonstrative in the case of Slovenian) (cf. Stolz 2010).

However, this is where the similarities with Serbo-Croatian and Slovenian end, and its similarities to the Bulgarian-like system begin. In fact, save for the fact that it is not obligatory in all definite contexts and cannot attach to the head noun, it shares very close surface similarities with the Bulgarian marker in almost every other way. Like the Bulgarian marker, it occurs once in a given noun phrase, on the leftmost element. Like Bulgarian, the marker can occur on an array of nominal elements (except for the bare noun, a key difference), including the demonstrative, most likely because those elements in Lithuanian utilize adjective morphology. Additionally, it occurs in noun phrases marked for any of the six functional cases used in Lithuanian (the morpheme is added to the stem before case-marking), and the modifiers it attaches to are themselves case-marked accordingly.

Finally, like Bulgarian, the Lithuanian definite marker meets the Zwicky & Pullum (1983) parameters for affixes, not clitics, as a structure that is “morphological and/or lexical in character, being concerned with the substructure of a finite set of words,” (1983: p. 503). This contrasts sharply with the long morphology of Serbo-Croatian and Slovenian. For example, the Lithuanian marker shows a high degree of selection with respect to its hosts. It will only attach to prenominal, adjective-like modifiers, and only at a certain, consistent point in the word form (the last affix). Another affix parameter is that it exhibits arbitrary gaps in its selection criteria; there are some adjectives in the lexicon that remain bare, regardless of definite status. Additionally, it illustrates morphophonological and semantic idiosyncrasies and is subject to syntactic rules, further parameters for affixes, and further similarities with Bulgarian. Harizanov (2009)’s Bulgarian analysis is thus very suitable for Lithuanian, suggesting that the marker is not itself a D, but is instead an agreement feature of a phonologically null D, an agreement feature that only emerges in certain contexts and on certain words. In this way, its presence would be an overt indicator of what Gillon & Armoskaite (2012) call a ‘covert’ D, and very likely marks Lithuanian as DP language (under the Bošković framework).

The question for Lithuanian remains not only whether or not this marker is indeed indicative of a definite structure (DP) in Lithuanian syntax but also whether it can act as such given its non-obligatory nature. From a semantic perspective, Gillon & Armoskaite (2012) argue that Lithuanian is a DP language even in the (majority of) cases where definite morphology is not used. However, if Lithuanian looks and behaves like one of Bošković’s NP languages in every way except for when it uses the definite marker, perhaps it can be one or the other depending on the structural situation. If Bošković’s generalizations about DP/NP languages hold merit, we may expect Lithuanian to exhibit characteristics consistent with only DP languages, or NP languages most of the time and DP languages only in the cases when the definite marker is used.

2 Bošković, Lithuanian, and the DP

In much of his work, Bošković (cf. 2005, 2009, 2012) has contended that languages which lack the definite article (like the English the) do not have a Determiner Phrase (DP) as part of their
noun phrase structure. While primarily focusing on Serbo-Croatian, Bošković extends his analysis to the other Slavic languages which lack articles (i.e., all except Bulgarian and Macedonian), and even to languages outside of the Indo-European family (cf. Bošković & Hsieh 2012; Bošković 2012; Bošković & Gajewski To appear). Bošković calls these “NP” languages, in contrast to languages like English which are unambiguously “DP” languages in his framework. He is not alone in suggesting that languages like these are NP, not DP languages. Indeed, his views are shared by Zlatic (1998) and Trenkic (2004), to name a few, and the theory itself was first outlined in detail by Corver (1992).

The no-DP movement in Slavic linguistics is a reaction to a wide-ranging assertion that, cross-linguistically, the noun is not actually the head of the noun phrase. Following Abney (1987)’s ground-breaking dissertation, many linguists have maintained that the functional head of the noun phrase is a so-called ‘non-lexical’ category D. As Abney describes it, the noun phrase is a sentence-like constituent, composed of ‘functional elements’ and ‘thematic elements.’ The D is a functional element (and is therefore not an argument) while the N is a thematic element. Just as C selects IP, and I selects VP, and neither of them is an argument, D likewise selects for NPs. Crucially, as the theory goes, even as it heads the phrase, the D is functionally projected from its noun complement. In the analysis, Abney showed how D selects for (among other things) AP and NP, and A selects for AP and NP, and NP is the complement of any of the above. Thus, he concludes the fundamental noun phrase structure in English (and most other languages) is one where D is the head, which governs AP in turn, which then governs NP. This approach is called the AP-over-NP analysis.

In contrast, for many Slavic linguists holding to the no-DP position, languages without overt articles have important distinguishing characteristics which point toward an NP-over-AP structure. First, using Zwicky’s (1985) headedness parameters as guidelines, Zlatic (1998) demonstrates how the noun in Slavic languages serves as the head of the noun phrase. This is contrary to the Abney (1987) proposal that determiners are the heads of the noun phrase. As Abney shows in unambiguously DP languages like English, nominal phrases can be argued to not be the maximal projection of the lexical head N, but rather of the functional head D. For these languages, it is D that is the head of the nominal phrase structure, while N serves as D’s complement. The head of D in unambiguously DP languages, as Trenkic (2004) explains, can only be an article, while other determiner-like elements (demonstratives, possessives, etc.) occupy the spec-DP position. See (12-a) for an example of this structure for the English noun phrase ‘a proud stingy rude man.’ This is in contrast to an analysis whereby APs are adjuncts to the NP, a construction called NP-over-AP (see (12-b)).

(12) a. AP-over-NP construction\(^9\)
However, what if a language has no articles, and other determiner-like elements are actually adjectives from a functional and syntactic perspective? Trenkic (2004) argues that there are no elements in the article-less Slavic languages (i.e., all the Slavic languages save for Bulgarian and Macedonian) capable of projecting a DP category. Rather, prenominal elements in Slavic noun phrase structures are effectually adjectives in both form and function. As both Bošković (2005) and Zlatić (1998) point out, all of the prenominal elements in Slavic languages retain adjective-like morphology, are easily stackable, and have adjunctive properties (e.g., multiplicity, optionality, interchangeability, etc.). Adjunctive status and consistent adjectival morphology allow for the scramble-ability of these elements such as found in (6). These characteristics are sufficient for positing these elements to be adjectives of a sort. And, as adjuncts, they do not have the power to either project an overarching DP or to select for a noun, as happens in the (Abney 1987) AP-over-NP paradigm.

There are other reasons for contending that these languages are not AP-over-NP but are instead NP-over-AP. Bošković suggests that the lack of D in Slavic languages triggers the NP-over-AP structure because A cannot serve as an argument. So, as an adjunct in specifier position, it is a full constituent that is easily extractable in cases of Left-Branch Extraction. In AP-over-NP structures, on the other hand, A cannot be extracted as the NP is a necessary component of the constituency of AP. As the English example (13-a) shows, the adjective ‘tall’ heads the adjective phrase ‘tall spies,’ and the NP ‘spies’ is the complement of the AP. In this structure, the adjective cannot be extracted out of the noun phrase, as it is not a full constituent. As the tree indicates, this structure is AP-over-NP. For the Russian example (13-b), the adjective вьсоких (‘tall’) is an adjunct to the NP, of which шпионов (‘spies’) is the head. As an adjunct, it is a full constituent that can be moved in its entirety out of the noun phrase. This structure is NP-over-AP.

(13) a. DP
   D  AP
   /   /
   the  A  NP
   /    /
   tall  spies

b. DP
   D  NP
   /   a
   N'
   AP  N'
   /    /
   A  AP  N
   /  /  /
   proud  A  man
   /  /
   stingy  rude

---


English

11
Another piece of evidence for the NP-over-AP type of structure in some languages can be found in cases where adjectives select constituents that are not noun phrases, which can happen in Lithuanian as well as in Slavic languages. If AP is a full constituent that does not select automatically for an NP (as in the AP-over-NP structure), then it is free to select for other constituents, such as the PP in (14), below. This construction is not allowed in English, where an adjective must select for a noun phrase.

(14) Toli mas nuo cent-ro nam-as buvo pig-us
far from center-M.SG.GEN house-M.SG.NOM was cheap-M.SG.NOM
‘the far from the center house was cheap’

The above example aside, however, the NP-over-AP structure is only part of the theory, because even as it provides a hypothetical structure for allegedly NP languages, it does not provide an explanation for why extractions such as LBE are possible in them (and are decidedly not possible in DP languages like English) and what the extraction operation would look like. To account for this, Bošković cites both Chomsky’s (1999) Phase-Impenetrability Condition (PIC) and the anti-locality restriction (i.e., movement has to cross over at least one maximal projection) and, following Svenonius (2004) concludes that D must be a phase. What does this mean, exactly? First, we assume that a fundamental principle of Universal Grammar is a principle of locality - i.e., that movement of elements in a sentence is typically not (in all cases) unbounded (i.e., it is subject to a constraint of locality). As Chomsky (1999) observed, elements subject to movement are usually constrained by certain boundaries (phases) in the sentence structure. That is, in order to move, they must first move to the edge of a phase (the ‘edge’ being a specifier or an adjunct to XP). This phenomenon is central to the PIC, in that a phase, then, is the only phrasal gateway through which movement can occur. Chomsky observed that CP and VP are phases (e.g., wh-movement in English occurs across the CP boundary, via spec-CP, or the edge of CP). At the same time as it is a gateway for some types of movement, it is also a barrier for other types of movement. For example, as wh-elements in English move across CP phases, other elements (e.g., prenominal modifiers) cannot. Thus, a phase both allows for and blocks movement. Svenonius (2004) provides evidence in Scandinavian languages for DP being a phase like CP and VP. Bošković and others take up this argument for the Slavic DP debate, asserting that it is the lack of a DP phase in Slavic languages that allows for prenominal modifiers to move out of a DP.

Thus, if adjectives and other nominal modifiers (as adjuncts, crucially) are to move out of the NP, they cannot land in a spec-DP position, as it violates anti-locality. What is more, if D is a phase, they cannot move out of a DP, as it would violate the PIC. Therefore, the analysis goes, languages with a DP structure, such as English, cannot abide such leftward movement, and LBE is impossible for them. However, for languages without a DP structure (as Bošković argues are most of the

---

11Both examples adapted from Pereltsvaig (2007: p. 60)
Slavic languages), the PIC is not an issue, and the NP modifier is free to move to a higher leftward specifier position. Example (15-a) below depicts how AP movement out of a VP is blocked by the DP phase. However, in example (15-b), we see that this extraction is theoretically possible if there is no DP in the structure.

(15) a. VP
   \[ \begin{array}{c}
   V \\
   see \\
   DP \\
   D' \\
   D \\
   the \\
   AP \\
   beautiful \\
   NP \\
   houses \\
   \end{array} \]

   English

b. VP
   \[ \begin{array}{c}
   V \\
   video \\
   ‘see’ \\
   NP \\
   AP \\
   lijep e \\
   ‘b eautiful’ \\
   NP \\
   kuće \\
   ‘houses’ \\
   \end{array} \]

   Serbo-Croatian

In sum, Bošković’s view of the allowability of LBE in NP languages is based on two important propositions. First, prenominal elements in the the noun phrase are adjectival, and therefore adjunctive. If these elements are adjuncts, then they should in theory be moveable. However, if the AP is dominated by a DP, which is a phase (Svenonius 2004), its extraction is still constrained by: 1) the Phase Impenetrability Condition, which forces any movement to occur through the specifier position of a phase; and 2) the anti-locality constraint, in which movement must occur over at least one maximal projection (i.e., it must not be local). As example (15-a) illustrates, even if the AP ‘beautiful’ were an adjunct to the noun phrase (which we already know that Abney 1987 argues it cannot be for English), it would not be able to land at the spec-DP position because it is not more than one maximal projection away, and the anti-locality constraint would forbid this. In (15-b), however, there is no DP phase, so therefore there is no spec-DP projection and thus the AP lijep e (‘beautiful’) is free to move out of the noun phrase.

---

12 Note that this example assumes an NP-over-AP structure for English, which has been disputed since Abney (1987). However, this would be the only plausible way to depict such an extraction for English, as in the case of an AP-over-NP structure, extraction of A alone would be of a partial constituent, which is impossible.

13 The barrier indicated in example (15-b) is consistent with the discussion in Rappaport (2001). While arguably the boundary could be elsewhere, it doesn’t matter for the central point at hand: wherever the phase boundary is, it (under this theory) nevertheless blocks movement of an adjectival modifier out of the noun phrase.

14 Example adapted from Rappaport (2001: p. 6)

15 Example adapted from Bošković (2005: p. 2)
2.1 Bošković’s (2012) generalizations put to the test

While this sort of extraction is crucial to Bošković’s syntactic analysis of NP vis-à-vis DP languages and is a large part of the thrust of this paper, it is not the only diagnostic he uses to distinguish the two types of languages. Bošković (2012) provides a list of eighteen or so generalizations about NP/DP languages that he claims to be cross-linguistic in nature. While some of the diagnostics Bošković explores are not applicable to Lithuanian (e.g., Lithuanian is not a polysynthetic language, Bošković 2012: p. 9), others provide readily testable hypotheses. As Lithuanian is tested across these parameters, a mixed picture emerges, such that in some ways Lithuanian exhibits properties consistent with Bošković’s NP language criteria and in other ways consistent with DP languages. For the sake of space and focus, only a relevant subset of seven of Bošković’s generalizations will be reported upon here, sufficient to show the mixed behavior that Lithuanian exhibits. Table 1 provides a summary of Lithuanian’s characteristics with respect to the subset of seven parameters.

Table 1: Subset of seven of Bošković’s NP/DP generalizations applied to Lithuanian

<table>
<thead>
<tr>
<th>Generalization</th>
<th>NP</th>
<th>DP</th>
<th>Lithuanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left branch extraction</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Adjunct extraction from NP</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Scrambling</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Superiority and multiple wh-fronting</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Exhaustivity of possessors</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Adnominal genitive</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Superlative interpretation (majority reading of ‘most’)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

2.1.1 NP-like generalizations

For five generalizations out of the subset, Lithuanian behaves consistently with Bošković’s (2012) predictions for NP languages. These five generalizations are outlined in detail here.

Left-branch extraction  The first parameter is based on the phenomenon of left branch extraction (LBE). Bošković echoes Corver’s 1992 observation that only languages without articles allow left branch extraction. While Bošković focuses on adjectival LBE in his 2012 paper, he elsewhere extends his analysis to include demonstratives, quantifiers, possessives, and wh-modifiers (cf. Bošković 2009). As he contends, these elements are adjective-like in Slavic languages, so extraction properties are similar among all of them. In the examples below, we see how adjectives and demonstratives in Serbo-Croatian can be extracted, but in Macedonian (a DP language), they cannot.

(16)  a.  
\[\text{Skup\_a/Ta_1 je video [t_1 kola]}\]
expensive/that \_is seen \[t_1 \text{car}\]
‘He/she sees an expensive/that car’  LBE in Serbo-Croatian

b.  
\[\text{*Novata_1 ja prodade Petko [t_1 kola]}\]
new \_it sells \[Petko t_1 \text{car}\]

\[\text{Gillon & Armorskait\_e (to app ear) likewise test Lithuanian against Bošković's generalizations, with mixed, though not identical, results to those of this study.}^{16}\]

\[\text{The ‘NP’ and ‘DP’ columns in Table 1 depict Bošković’s predictions for NP and DP languages.}^{17}\]
Consistent with Serbo-Croatian and other Slavic languages, as the data below show, LBEs are allowable in Lithuanian, for a wide array of prenominal modifiers. As the examples below show, leftward extraction of possessives, demonstratives, wh-words, and adjectives out of a nominal phrase are possible in Lithuanian, data which contradict the Left Branch Condition proposed by Ross (1967). For example, the following sentences are grammatical in Lithuanian, but their English equivalents are disallowed.

(17)  
Kieno jis mat-é [t̩ mašin-q]?
whose he see-PST [t̩ car-F.SG.ACC]  
‘Whose car did he see?’

(18)  
Brangi-qi jis mat-é [t̩ mašin-q].
expensive-F.SG.ACC he see-PST [t̩ car-F.SG.ACC]  
‘He saw an expensive car.’

(19)  
Ši-qi jis mat-é [t̩ mašin-q].
this-F.SG.ACC he see-PST [t̩ car-F.SG.ACC]  
‘He saw this car.’

Additionally, multiple elements can be leftward extracted from the noun phrase. And, in the case of numerous prenominal elements, the leftmost ones can be extracted while leaving others behind, depending on pragmatic emphasis (see (20)).

(20)  
[Kieno mėlyn-qa] jis mat-é [t̩ brangi-qa mašin-qa]?
[whose blue-F.SG.ACC he see-PST [t̩ expensive-F.SG.ACC car-F.SG.ACC]  
‘Whose blue expensive car did he see?’

Adjunct extraction from NP  Another of Bošković’s parameters is that only languages without articles may allow adjunct extraction out of the noun phrase. Bošković argues that languages with articles categorically disallow this sort of extraction, and certainly English would be a case in point. Article-less Slavic languages, however, do allow it. As the examples below indicate, adjunct extraction is not acceptable in English, (21-a), and Bulgarian (a DP language), (21-b), while in Russian, (21-c), it is acceptable.

(21)  
a. Peter met girls from this city.
*[From which city] did Peter meet girls t̩?  
English

b. *[Ot koj grad] Petko srešta momićeta t̩?  
[from which city] Petko met girls t̩
‘Peter met girls from which city?’

Bulgarian

c. [Iz kakogo goroda] ty vstrechal devushek t_i?
[from which city] you met girls t_i
‘You met girls from which city?’

Russian

A direct translation of this sentence in Lithuanian, with the extracted PP, shows that it is acceptable (see example (22)).

(22) [Iš kurio miesto] Petras susipaužė su mergaitėmis
[from which city] Petras meet-3.PST with girls-f.pl.ins
‘Peter met girls from which city?’

Lithuanian

It is possible that the presence of the wh-word in the PP might be contributing to relative acceptability. Two other examples of adjunct extraction, sans wh-words, are also provided in (23) and (24).

(23) a. Marija myliu vyru su ilgais plaukais
Mary-f.sg.nom love-3.sg.prs man-m.pl.acc with long-m.pl.ins hair-m.pl.ins
‘Mary loves men with long hair.’ NP adjunct in situ

b. [su ilgais plaukais] Marija myliu vyru
[with long-m.pl.ins hair-m.pl.ins] Mary-f.sg.nom love-3.sg.prs man-m.pl.acc
‘Mary loves men with long hair.’ NP adjunct extracted

(24) a. aš geriu kava su cukru
1.sg.nom drink-1.sg.prs coffee-f.sg.acc with sugar-m.sg.ins
‘I drink coffee with sugar.’ NP adjunct in situ

b. [su cukru] aš geriu kava
[with sugar-m.sg.ins] 1.sg.nom drink-1.sg.prs coffee-f.sg.acc
‘I drink coffee with sugar.’ NP adjunct extracted

Sure enough, Lithuanian patterns right along with the examples that Bošković provides from Slavic languages, aligning with his NP generalizations. Adjunct extraction out of the noun phrase is acceptable, which Bošković assumes is due to the absence of a DP phase, which would theoretically block this kind of movement.21

20Examples are taken from Bošković (2012: p. 4). Arguably, (21-a) is acceptable in English, perhaps made more so by the presence of a wh-word in the extracted PP.

21Interestingly, Bošković suggests a correlative phase-impenetrability constraint on NPs. Thus, he posits in NP languages, while adjunct extraction is allowed, complement extraction is not (a phenomenon he calls ‘Deep-LBE.’ As the theory goes, while there is theoretically no DP to block adjunctive movement, an NP phase would still block movement of its complement. However, Pereltsvaig (2008) provides examples in colloquial Russian of such complement extraction, and Lithuanian can exhibit it too. This is just one piece of evidence that Bošković’s phase-based analysis of extraction does not generally hold.
**Scrambling**  In another generalization, Bošković notes that there is a correlation between a language’s lack of articles and the acceptability within it of scrambling. By ‘scrambling,’ he is working with a quite restrictive definition: “the kind of movement referred to as scrambling in Japanese...long-distance scrambling from finite clauses” Bošković (2012: p. 4). Bošković claims that for DP languages like English, this kind of scrambling is ungrammatical. However, he explains how it is acceptable in DP languages. In the case of Lithuanian, as example (25) illustrates, long-distance scrambling out of finite clauses is perfectly acceptable.

(25) a. student-qi Jon-as žin-o kad Marij-a mėgsta t
student-M.SG.ACC John-M.SG.NOM know-3.PRS that Mary-F.SG.NOM like-3.PRS t_i
‘John knows that Mary likes a student.’

b. Marij-a Jon-as žin-o kad t_i student-a mėgsta
Mary-F.SG.NOM John-M.SG.NOM know-3.PRS that t_i like-3.PRS student-ACC
‘John knows that Mary likes a student.’

**Multiple WH-fronting and superiority effects**  Bošković posits that only NP languages allow multiple wh-fronting with no superiority effects. NP languages, like Russian and Polish, he says, have multiple wh-fronting without any restrictions on ordering of wh-elements. In contrast, Bulgarian has superiority restrictions on ordering of wh-elements. See the Bulgarian example (26) below.22

(26) a. kogo vįžda?
who whom sees
‘Who sees whom?’

b. *kogo kogo vįžda?
whom whom sees
‘Who sees whom?’

Like Russian and other Slavic languages, Lithuanian is a multiple wh-fronting language. Typical Lithuanian examples are included in (27):

(27) a. Kas kq perk-a?
who what buy-3.PRS
‘Who buys what?’

b. Kas kur važiuoj-a?
who where drive-3.PRS
‘Who drives where?’

c. Kam kq Jonas dav-ę?
to-whom what John give-3.PST
‘John gave what to whom?’

While the above examples exhibit the preferred order for these wh-words, there are other possible orderings as well. Semantically equivalent sentences to the above are below (28):

(28) a. Kq kas perk-a?
what who buy-3.PRS

---

22Example taken from Bošković (2012: p.6).
‘What does who buy?’

b. Kur kas važiuoj-a?
   where who drive-3.PRS
   ‘Where does who drive?’

c. Kq kam Jonas dav-ė?
   what to-whom John give-3.PST
   ‘John gave what to whom?’

Regardless of their relative orders, these *wh*-elements must be fronted. Thus, the sentences (29) are unacceptable.

(29) a. *Kas perk-a kq?
   who buy-3.PRS what
   ‘Who buys what?’

b. *Kas važiuoj-a kur?
   who drive-3.PRS where
   ‘Who drives where?’

c. *Kam Jonas dav-ė kq?
   to-whom John give-3.PST what
   ‘To whom did John give what?’

Thus, once again, Lithuanian behaves in a way consistent Bošković’s parameters for an NP language - as a multiple *wh*-fronting language, it does not force *wh* elements into a restricted hierarchical order.

**Interpretation of possessors** If the above parameters have to do with syntactic structure and element extractability, there are other parameters in Bošković’s list that utilize clausal semantic parameters as well. In one of these Lithuanian behaves like an NP language. For example, the English translation of the Lithuanian noun phrases below yields only one reading of them - presupposing exhaustivity. As the following example shows, however, the default interpretation of the Lithuanian version of the noun phrase is not exhaustive. In fact, an exhaustive reading is only allowed with the insertion of the qualifier *vienintelai* (‘only’) into the phrase.

(30) a. Jon-o trys megtin-iai
    John-M.SG GEN three sweater-M.PL NOM
    ‘Three of John’s sweaters’
    *John’s three sweaters’

b. Jon-o vienintel-iai trys megtin-iai
    John-M.SG GEN only-M.PL NOM three sweater-M.PL NOM
    ‘John’s (only) three sweaters’

Beyond stating the generalization and providing some data, Bošković does not elaborate on how this generalization relates structurally to the NP/DP debate. He simply notes that it is a cross-linguistic observation.
2.1.2 DP-like generalizations

In the above five parameters, Lithuanian exhibits characteristics consistent with other NP generalizations of Bošković (2012). However, in the following two parameters, Lithuanian does not follow the NP-only generalizations. The following two sections describe how Lithuanian behaves with respect to the parameters of *Adnominal genitive* and *Superlative interpretation*.

**Adnominal genitive** Bošković claims that only languages with articles can allow for two genitive arguments for transitive nominals, where both the internal and external arguments to the noun can be marked with genitive case. In most cases, the external argument is realized as a genitive PP, as in the case for ‘of Rome’ in the English example (31) below.

(31) Hannibal’s conquest of Rome

Bošković provides similar examples from Polish and Czech to illustrate that this phenomenon is not possible for NP languages. See (32) below.

(32) *żničeńi Říma barbarū*
    destruction Rome-GEN barbarians-GEN
    ‘The barbarian’s destruction of Rome’

In contrast to Polish, Lithuanian allows for both arguments to the noun to carry conventional genitive morphology, and it is their relative prenominal ordering that gives indication as to their internal/external argumenthood. See (33) below.

(33) *Kanibal-o Rom-os sugriovim-as*
    Hannibal-M.SG.GEN Rome-F.SG.GEN destruction-M.SG.NOM
    ‘Hannibal’s destruction of Rome’

Thus, in terms of following Bošković’s parameters, here is another way in which Lithuanian does not conform to the generalizations for NP languages and instead aligns with DP languages.

**Superlative interpretation** Bošković has noticed that Serbo-Croatian in (34) has only one reading of the following sentence, while English and other DP languages allow for two possible readings.

(34) Največ *ljudi pije pivo.*
    most people drink beer
    ‘More people drink beer than drink any other beverage.’ (Plurality Reading)
    ‘More than half the people drink beer.’ (Majority Reading)

Lithuanian in (35) allows for both readings, which is consistent with English and other languages with articles.

(35) Daugiausia žmon-įg ger-ia ak-į.
    most person-F.PL.GEN drink-3.PRS beer-M.SG.ACC
    ‘More people drink beer than drink any other beverage.’ (Plurality Reading)
    ‘More than half the people drink beer.’ (Majority Reading)

2.2 The mixed case of Lithuanian

In summary, consistent with similar findings by Gillon & Armanskaitė (to appear), Lithuanian does not generally abide by Bošković’s generalizations for either an NP or a DP language. As 2.1 showed,
Lithuanian behaves like NP languages for the generalizations of Left branch extraction, Adjunct extraction from the NP, Scrambling, Multiple wh-fronting and superiority effects, and Interpretation of possessors. At the same time, it behaves like a DP language for the parameters of Adnominal genitive and Superlative interpretation. While Bošković’s theory does not explicitly allow for mixed-NP/DP languages, it is theoretically possible that Lithuanian, as a language with an optional definiteness marker, might show DP-like properties when the marker is used and NP-like properties when the marker is not used. If Bošković is correct and the determiner projects a phase-head D that blocks movement out of the DP, then elements within a definite-marked NP in Lithuanian (i.e., a DP) should not be extractable via Left-branch extraction, Adjunct extraction, or Scrambling. Thus, if we find that definite-marked elements are not extractable, but (as we’ve already seen) the non-definite marked elements are, then we can perhaps deem Lithuanian to be a mixed, or parameterized DP language of some kind, and Bošković’s theory can be upheld for Lithuanian. The following sections will test whether the above three operations are allowed in definite-marked Lithuanian NPs.

2.2.1 Extraction from definite-marked NPs (DPs) in Lithuanian

Below, we see that definiteness marking on elements of the Lithuanian noun phrase does not affect their extractability or scrambling abilities. Sentences with extracted definite-marked elements are just as acceptable as are those without the definiteness marker. Section 1.3 above presented the definiteness paradigm in Lithuanian, as well as the prenominal elements that can house the definite marker. It also provided several reasons for why Lithuanian’s long morphology definiteness paradigm can be considered to be more consistent with the Bulgarian (DP) system than with other Slavic (NP) long-morphology paradigms (as Section 1.3.2 attests). If this is the case, then, we can assume that the definiteness marker in Lithuanian is indeed a D. If, however, the comparison with the Bulgarian def marker is not convincing enough at this point, Gillon & Armoskaite (2012) have provided a very plausible semantic explanation for there being a covert D in Lithuanian. And, at the very least, we have already seen Lithuanian’s mixed performance regarding Bošković’s generalizations.

**Left branch extraction of definite-marked elements** According to Bošković’s theory, Lithuanian prenominal elements with a definite marker should not be extractable, and neither should other elements be able to be extracted over it. The following examples of LBE are acceptable in Lithuanian, even with the definiteness marker attached to them.

(36) *Kelint-a-ja*  
which-F.SG.ACC-DEF.F.SG.ACC:he see-3.PST [t1 car-F.SG.ACC]

‘Which (particular) car did he see?’  
Interrogative

(37) *Brangi-a-ja*  
expensive-F.SG.ACC-DEF.F.SG.ACC:he see-3.PST [t1 car-F.SG.ACC]

‘He saw the expensive car.’  
Adjectival

(38) *Ši-a-ja*  
this-F.SG.ACC-DEF.F.SG.ACC:he see-3.PST [t1 car-F.SG.ACC]

‘He saw this (particular) car.’  
Demonstrative

(39) *Pirm-a-ja*  
first-F.SG.ACC-DEF.F.SG.ACC:he see-3.PST [t1 car-F.SG.ACC]
‘He saw the first car.’

\[Vairu-o-ja-n\-iq-ja\] \(jis\) mat-\(é\) \([t_1\ mašin-a]\).

‘He saw the driven car.’

(40) \(\text{Extraction of definite-marked adjuncts}\) If, as Bošković’s theory goes, D is a phase, then adjuncts to definite-marked noun phrases should not be extractable out of the verb phrase. However, even with a definite-marked adjective modifying \(\text{mergaitēmis} (‘girls’)\), the adjunct could be extracted easily out of the DP.

\[(41) \ [Iš\ kuri-o\ miest-o_ji\ Petras\ susipāžin-o\ su\ from\ which-M.SG.GEN\ city-M.SG.GEN_ji\ Petras-M.SG.NOM\ meet-3.PST\ with\ gružio-sio-mis\ mergaitēmis\ t_i\ ?\ \text{beautiful-F.PL-DEF-INS}\ girl-F.PL.INS\ t_i\ \text{Peter met the beautiful girls from which city?}\]

\[\text{Extraction out of scrambled clauses}\) As Section 2.1.1 notes, Lithuanian allows for the sort of scrambling that Bošković believes is reserved for NP languages only (i.e., long-distance scrambling from finite clauses). As the theory goes, this sort of scrambling should be impossible for definite-marked NPs, as it is in English and other DP languages. However, in Lithuanian, scrambling of either the definite-marked subject or the definite-marked object is fine.

\[\text{(42) a. } [\text{proting-a-sis}\ \text{student-as}_j]_i\ Jon-as\ \text{žin-o}\ \text{Jon knows that the intelligent student likes Mary.}\]

Incidentally, left branch extraction of definite-marked elements can occur across clause boundaries.

\[(43) \text{a. } [\text{proting-a-sis}\ \text{student-as}_j]_i\ Jon-as\ \text{žin-o}\ \text{Jon knows that Mary likes the intelligent student.}\]

\textbf{2.3 Preliminary Summary}

In sum, data from Lithuanian provide a much more varied and nuanced picture of extraction phenomena than allowed for by the DP-NP theories of Bošković. Despite its lack of obligatory definite articles (as per languages like English, Italian, etc.) Lithuanian can arguably be described as a
DP language. Lithuanian definite markers (embodied as adjectival long morphology) behave in many ways similarly to definite articles in Bulgarian and Macedonian and not, by contrast, to long morphemes in Serbo-Croatian and Slovenian. Moreover, Gillon & Armokaitė (2012) have shown from a semantic perspective how bare nouns on Lithuanian carry definite (as opposed to indefinite) connotation. However, as a DP language, Lithuanian exhibits many characteristics that are supposedly reserved for NP languages only (at least as they are theorized by Bošković 2005, Bošković 2009, Bošković 2012, Bošković & Hsieh 2012, Bošković 2012; Bošković & Gajewski To appear, Zlatic 1998 and Trenkic 2004, among others). These characteristics include: 1) adjectival (adjunct-like) status and interchangeability of prenominal modifiers, characteristics which support an NP-over-AP structure; 2) left-branch and adjunct extraction; 3) scrambling; and 4) wh-fronting superiority effects (see sections 1.2, 2, and 2.1.1). Bošković contends that it is the presence of a DP that blocks extraction out of these environments, positing that D is a phase, and citing the phase impenetrability condition. However, each of these supposedly NP-language-only phenomena is also possible in Lithuanian, in indefinite, contextually-definite, and even explicitly-marked definite noun phrases (as shown in 2.2.1). So, clearly, the DP in Lithuanian is not acting as a movement-blocking phase, and another explanation for how and why this extraction occurs is needed.

3 Pereltsvaig, Lithuanian, and copy-theory movement

It turns out, as alluded to above, that extraction and word-ordering phenomena in Lithuanian are more fluid than are the prototypical extraction environments provided in the Bošković accounts. Moreover, the Bošković account does little to explain why extraction, if it occurs, is motivated. Hinging the extraction phenomena on the hypothetical lack of a DP provides only a structural account for how extraction can happen, and that account does not hold for Lithuanian. An alternate view is needed, and a plausible one is offered by Pereltsvaig (2008), which provides not only a structural account of extraction phenomena, but also a potential motivation for why it happens. Crucially, however, the Pereltsvaig account does not consider this phenomenon to be extraction in the same sense as does Bošković (what she calls ‘direct’ extraction). Rather, the extraction process noted in Pereltsvaig’s data is termed ‘phrase splitting.’ Pereltsvaig provides examples from corpora of colloquial Russian to show why this can be a better term to use. This theory proposes that what is happening is not movement characterized by extraction of one phrase into another location in the phrase structure (leaving a bound trace in its wake). Rather, phrases are copied and moved in their entirety, resulting in replicas at different points in the phrase structure. Phonological phenomena govern which portions of the copied phrases are pronounced at which points in the structure. The copy-movement process of phrases, Pereltsvaig suggests, is motivated by interpretable features (+contrast), and the splitting and variable pronunciation of those phrases is determined within the PF interface.

Section 3 describes the Pereltsvaig movement-and-partial-copy analysis of phrase splitting phenomena, and how it better suits Lithuanian data than do the direct extraction (e.g., Bošković) and non-movement accounts (e.g., Fanselow 1988; Bošković & Takahashi 1998).23 Following a Bailyn (cf. 2001) approach, Pereltsvaig assumes that phrasal movement occurs due to features of information structure. Below, a brief description of Lithuanian information structure will confirm that marked phrasal movement in Lithuanian is stimulated by contrastive features of the discourse. However, Lithuanian data suggest that motivation for splitting of the copied phrase into its various shapes and forms seems to be more than simply a factor of contrast. Additionally, the fact that
phrases can move out of strong islands, but cannot split across them suggests that PF alone as an explanation for when phrases can be split is inadequate. The Lithuanian data show that a splitting account that allows for both pragmatic and a structural constraints is needed.

3.1 Phrase splitting is not direct extraction

As mentioned in Section 2, Bošković’s structural account of left branch extraction hinges on the idea that prenominal elements have adjunctive, adjectival status, and that NP languages (lacking a DP) utilize an NP-over-AP structure instead of a AP-over-NP one. If this structural account is correct, then extraction of prenominal elements would not violate phrasal movement constraints (cf. Bresnan 1976) in that leftward adjectival extraction would be extraction of a full constituent. Movement of a non-constituent is generally prohibited.

However, Pereltsvaig notes in colloquial Russian many examples of fronted/moved elements of phrases that themselves are not constituents. For example, movement and splitting of prepositional phrases (PPs) is evidenced in both colloquial Russian and Lithuanian. See (44-b) and (44-c) below:

(44) a. А́я су́зн-ау [и́ш и́дом-и́г] ла́йкраш-и́гу|  

latvian phrasal order

1.SG NOM learn-1.SG.PST from interesting-M.PL GEN newspapers-M.PL GEN

‘I learned from interesting newspapers.’ Default word order

b. [и́ш и́дом-и́г]А а́я су́зн-ау [ла́йкраш-и́гу]Б

from interesting-M.PL GEN 1.SG_nom learn-1.SG.PST newspapers-M.PL GEN

‘I learned from interesting newspapers.’ Splitting

c. *[и́ш]А а́я су́зн-ау [идом-и́г] ла́йкраш-и́гу|Б

from 1.SG Nom learn-1.SG.PST interesting-M.PL GEN newspapers-M.PL GEN

‘I learned from interesting newspapers.’ Splitting

As (44) illustrates, the PP can be fronted in its entirety, and it can be split into two parts, the first (consisting of the preposition and adjective) fronted and the second (the object of the PP) remaining in situ. Note that (44-c) is unacceptable. One explanation for this could be the observation that the preposition и́ш (‘from’) is a light preposition that easily cliticizes to its adjacent host and is commonly lexicalized in verbal (e.g., и́сёiti ‘to leave’), adverbial (e.g., и́скёртинай ‘exceptionally’), adjectival (e.g., и́скальбис ‘eloquent’), and noun (e.g., и́се́йгинис ‘holiday’) morphology. A PP utilizing a heavy or lexical preposition has more versatile splitting potential; this is shown in (45) below.

---

23 Pereltsvaig (2008) cites Fänselow (1988), who proposes a non-movement account of split phrases for German. He argues that they are derived through a merger of two independent noun phrases. Pereltsvaig shows how this analysis is not appropriate for colloquial Russian. In another account, Bošković & Takahashi (1998) argue that phrase scrambling in Russian is not movement per se, but rather that phrases are base-generated in their higher or lower places. Bailyn (2001) uses scope tests similar to those provided here in section 3.2 to show that the non-movement, base-generation analysis does not work for Russian data either.

24 In the examples provided in this section and following, a modification of Pereltsvaig (2008)’s notational style is used. The phrases in question are bracketed. Underlining and subscripts note the following: ‘____’ depicts a phrase in its unmarked location and any in situ elements of split phrases, ‘_____’ depicts a phrase and any split-off elements thereof in their copied/moved positions. In the cases of splits, their component elements are subscripted А or Б, (and sometimes С) depending on their conventional ordering in an intact phrase.
(45) 

a. aš gyven-u [arti didž-ios Nemun-o up-ės]
   1.sg.nom live-1.sg.prs near great-f.sg.gen Nemunas-m.sg.gen river-f.sg.gen
   'I live near the great Nemunas river.'

b. [arti didž-ios Nemun-o] aš gyven-u [up-ės]B
   Near great-f.sg.gen Nemunas-m.sg.gen 1.sg.nom live-1.prs river-f.sg.gen
   'I live near the great Nemunas river.'

Splitting

c. [arti didž-ios] aš gyven-u [Nemun-o up-ės]B
   Near great-f.sg.gen 1.sg.nom live-1.prs Nemunas-m.sg.gen river-f.sg.gen
   'I live near the greatest Nemunas river.'

Splitting

d. [arti] aš gyven-u [didž-ios Nemun-o up-ės]B
   Near 1.sg.nom live-1.prs great-f.sg.gen Nemunas-m.sg.gen river-f.sg.gen
   'I live near the great Nemunas river.'

Splitting

Here, the PP has quite a few splitting options, including fronted preposition stranding. Taken at face (i.e., split) value, examples like (44-b) and (45-b), (45-c), and (45-d) are problematic from a direct extraction perspective, because they seem to show movement/scrambling of non-constituents.

As Pereltsvaig (2008) points out, not only can non-constituents be moved out of a whole phrase, but the head of a phrase can also be moved without some of its modifying elements. For example, in (47), we see the AP has been fronted in the sentence, while its modifying adverb has remained in situ.

(46) 

a. Petr-as tur-i [vien-ą labai elegantišk-ą]
   Peter-m.sg.nom have-3.prs one-m.sg.acc very elegant-m.sg.acc 
   megztin-į
   sweater-m.sg.acc
   'Peter has one very elegant sweater.'

Default word order

b. [vien-ą] labai elegantišk-ą megztin-įC Petr-as tur-i
   one-m.sg.acc elegant-m.sg.acc sweater-m.sg.acc
   Peter-m.sg.nom have-3.prs
   [labai]B
   very
   'Peter has a one very elegant sweater.'

Splitting (adverb stranding)

Other times, Part B of the phrase is moved, while Part A of the phrase remains in situ. Pereltsvaig calls these ‘inverted’ splits. In still other cases, the phrase can undergo ‘triple’ splitting, even to the point where some of the elements therein are inverted. Examples (47-a), (47-b), and (47-c) illustrate each of these possible configurations.
b. \[
\begin{align*}
\text{\(\text{one-m.sg.acc Peter-m.sg.nom very elegant-m.sg.acc have-3.prs}\)}
\end{align*}
\]
\[
\text{sweater-m.sg.acc}
\]
`Peter has a one very elegant sweater.'

\text{Triple split}

c. \[
\begin{align*}
\text{\(\text{one-m.sg.acc Peter-m.sg.nom very elegant-m.sg.acc have-3.prs}\)}
\end{align*}
\]
\[
\text{sweater-m.sg.acc}
\]
`Peter has a one very elegant sweater.'

\text{Inverted triple split}

One other piece of evidence cited by Pereltsvaig (2008) for colloquial Russian is the fact that phrase splitting can occur across weak islands, including \textit{wh}-, negative, and factive islands. Data from Lithuanian confirms this phenomenon as well, and examples of each are found below in (48), (49), and (50).

(48) a. \[
\begin{align*}
\text{\(\text{tell-imp.form 1.sg.dat where big-m.sg.nom bridge-m.sg.nom}\)}
\end{align*}
\]
`Tell me please, where is the big bridge?'

\text{Default word order}

b. \[
\begin{align*}
\text{\(\text{tell-imp.form 1.sg.dat where bridge-m.sg.nom}\)}
\end{align*}
\]
`Tell me please, where is the big bridge?'

\text{Splitting out of a \textit{wh}- island}

(49) a. \[
\begin{align*}
\text{\(\text{2.sg no-go-fut 2.sg to more one-f.sg.acc store-f.sg.acc Vitalij-f.sg.nom}\)}
\end{align*}
\]
`You aren’t going into another store, Vitalija?'

\text{Default word order}

b. \[
\begin{align*}
\text{\(\text{to more one-f.sg.acc 2.sg no-go-fut 2.sg store-f.sg.acc Vitalij-f.sg.nom}\)}
\end{align*}
\]
`You aren’t going into another store, Vitalija?'

\text{Splitting out of a negative island}

(50) a. \[
\begin{align*}
\text{\(\text{Q 2.sg regret-prs.2.sg that eat-pst-2.sg rotten-f.pl.acc vegetable-f.pl.acc}\)}
\end{align*}
\]
`Do you regret that you ate rotten vegetables?'

\text{Default word order}

b. \[
\begin{align*}
\text{\(\text{Q 2.sg regret-prs.2.sg that eat-pst-2.sg vegetable-f.pl.acc}\)}
\end{align*}
\]
`Do you regret that you ate rotten vegetables?'

\text{Splitting out of a factive island}

Taken together, examples (44) through (50) are very problematic for a Bošković-style theory of extraction. They show that non-constituent elements of phrases can be extracted and moved, or can be left in situ while the phrase head moves out and over them. Moreover, weak islands do not serve as barriers to phrase movement and splitting. This, Pereltsvaig argues, is sufficient evidence against direct extraction theories, and an alternative theory of movement should be considered.
3.2 Is this movement or non-movement?

But, perhaps an obvious question should be asked first: How do we know that these examples demonstrate movement at all? What if these seemingly moved phrases and items are actually base-generated at different places in the sentence structure? The first simple answer is that the resulting split-phrased sentences remain semantically identical to their in situ counterparts. But, this should also be demonstrated empirically as well, and can be done through scope tests. For example, in (51) and (52) below, we see that the given sentences, in English, have two possible interpretations.

(51) John did not buy exactly three cars.
   a. **Interpretation 1:** It is not the case that John bought exactly three cars (i.e., he might have bought two or four).
   b. **Interpretation 2:** There were exactly three cars that John did not buy (i.e., he bought all the other cars).

(52) Peter must have eaten exactly two cakes.
   a. **Interpretation 1:** It is the case that Peter ate exactly two cakes (i.e., not two or four).
   b. **Interpretation 2:** There were exactly two cakes in particular that Peter must have eaten (i.e., he did not eat any of the others).

In English, both interpretations are possible for each sentence, even though the surface order might indicate otherwise. A surface linear scope interpretation for each produces a narrow interpretation of the quantifier ‘exactly,’ which produces the (a) readings above. However, in the context of quantifier raising (May 1985), ‘exactly’ can get a wider scope, resulting in the (b) interpretations. Assuming that scope can be represented at the syntactic level, the fact that both wide and narrow interpretations are possible given one surface form indicates that some sort of movement is occurring – whether it’s binding/trace movement or copy/deletion movement is not necessarily relevant. Notably, the Lithuanian versions of the above sentences likewise allow for two different interpretations in their surface form, see (53-a) and (54-a). However, Lithuanian word ordering freedom also allows for permutations that look like (53-b) and (54-b), which look like examples of left-branch extraction or phrasal movement, but could also theoretically be phrases that are base-generated in a higher position. A theory-neutral observation of these sentences raises the question of whether the quantifier phrase is in fact raise-able/moveable to this earlier position in the sentence (and would thus still maintain two possible interpretations), or whether it is base-generated and remains here (and therefore can only give a wide-scope interpretation). See (53) and (54) below.

(53) a. **Jon-as ne-pirk-o lygiai tri-ju mašiny**
   John-M.SG.NOM no-buy-3.PST exactly three-F.PL.GEN car-F.PL.GEN
   ‘John did not buy exactly three cars.’
   **Available interpretations:** 1 and 2

b. **lygiai tri-ju Jon-as ne-pirk-o lygiai tri-ju mašiny**
   exactly three-F.PL.GEN John-M.SG.NOM no-buy-3.PST car-F.PL.GEN
   ‘Exactly three John did not buy cars.’
   **Available interpretations:** 1 and 2
In (53-b) and (54-b) above, the quantifiers *lygiai triju* and *lygiai du*, respectively, are pronounced in a high position. If they were base-generated in this position, then only the wider interpretation (Interpretation 2) would be possible for both examples. However, this is not the case. Even with the quantifiers fronted in the sentence, there are still two permissible interpretations (wide and narrow) of the sentences. This suggests that the quantifier is actually base-generated in the lower position and moves higher via quantifier raising. Thus, what we have here is movement of a sort, and the question now is what kind of movement it is. The bulk of this section thus far has shown what kind of movement it is *not* (i.e., not direct extraction), so now we move to the task of showing what kind of movement it most likely is. Namely, the movement phenomena observed for both colloquial Russian (Pereltsvaig 2008) and Lithuanian is most consistent with a copy-deletion theory of movement (cf. Chomsky 1993).

### 3.3 Phrase splitting as copy-movement with partial PF interpretation

As Pereltsvaig (2008: p. 18) sums it up, “split phrases are derived through some sort of feature-driven movement of the whole phrase to be split...[which] creates multiple copies of the phrase to be split...[and] the actual splitting derives from partial interpretation of copies.” The partial interpretation of the copies at either the higher or lower points in the phrase’s instantiation is made through the PF interface.

So, what first has to be established is what kind of feature-driven movement Pereltsvaig is talking about. Following Bailyn (1995; 2001), Pereltsvaig calls this movement ‘scrambling,’ and uses it to refer to specifically A’ (as opposed to A) phrasal movement. Pereltsvaig does not provide specific evidence demonstrating that it is A’ movement necessarily, but by citing Bailyn (2001), who uses Russian examples, she is assuming this to be the case. Before extending her analysis to Lithuanian, however, we should also establish that the kind of movement in question is A’ and not A movement. Chomsky (1986) provides parameters for the two different kinds of movement, crucial being that A’ movement does not occur for reasons of theta-role or case assignment, and the landing sites for such movement are not places where theta/case assignment happens. A common example of A’ movement in English is *wh*-movement, in which *wh*-words move out and over (even multiple) clauses, creating long-distance dependencies. Long-distance scrambling as described by Bailyn (2001) is a similar sort of phenomenon. In fact, every one of the splitting examples above can be alternatively constructed as scrambling constructions. See (55) below.

\[(55) \quad \text{a. } [\text{Iš } \text{idom-\text{iy}} \quad \text{laikrašči\text{-\text{iy}}}] \quad \text{aš } \text{sužin-\text{au}} \quad \text{from interesting-M.PL.GEN newspapers-M.PL.GEN 1.SG.NOM learn-1.SG.PST} \]
'I learned from interesting newspapers.' see (44)

b. [artis didžios Nemun-o up-ės/ aš gyven-a] Near great-F.SG.GEN Nemunas-M.SG.GEN river-F.SG.GEN 1.SG.NOM live-1.SG.PRS 'I live near the great Nemunas river.' see (45)

c. [vien-a labai elegantisk-a megztin-i] Petr-as one-M.SG.ACC very elegant-M.SG.ACC sweater-M.SG.ACC Peter-M.SG.NOM tur-i have-3.PRS 'Peter has a one very elegant sweater.' see (47)

d. [didel-is til-as] pasaky-kite man kur. big-M.SG.NOM bridge-M.SG.NOM tell-IMP.FORM 1.SG.DAT where 'Tell me please, where is the big bridge?' see (48)

e. [ji dar vien-a parduotuv-ė] tu n-ei-si, Vitalij-a? to more one-F.SG.ACC store-F.SG.ACC Vitalij-F.SG.NOM 2.SG no-go-FUT.2.SG 'You aren’t going into another store, Vitalija?' see (49)

f. [supuus-ias daržov-es] ar tu gail-i, kad valg-e-i? rotten-F.PL.ACC vegetable-F.PL.ACC Q 2.SG.NOM. regret-PRS.2.SG that eat-PST-2.SG 'Do you regret that you ate rotten vegetables?' see (50)

And, we can go on and on. In fact, in Lithuanian, every example of phrase splitting can be represented as a process of A' phrase scrambling first. And, given the fact that split-off elements of phrases can behave in the same way as conventionally-accepted A' scrambling, it seems likely that an underlying (yet not fully pronounced) moved phrase is present. See (56) below for an example of phrase splitting that easily conforms to wh-like long-distance movement phenomena across finite clauses.

(56) a. aš žin-au kad jis man sak-ė kad jis girdė-j-o 1.SG.NOM know-1.SG.PST that he.NOM 1.SG.DAT say-3.PST that he.NOM hear-PST-3 kad ji puoš-ia-si /brang-iais papuošal-ais/. that she.NOM adorn-3.PRS-REFL expensive-M.PL.INST adornment-M.PL.INST 'I know he told me that he heard she wears expensive jewelry.' Default word order

b. [brang-iais papuošal-ais] aš žin-au kad jis expensive-M.PL.INST adornment-M.PL.INST 1.SG.NOM know-1.SG.PST that he.NOM man sak-ė kad jis girdė-j-o kad ji puoš-ia-si. 1.SG.DAT say-3.PST that he.NOM hear-PST-3 that she.NOM adorn-3.PRS-REFL 'I know he told me that he heard she wears expensive jewelry.' Scrambling

c. [brang-iais] aš žin-au kad jis man sak-ė expensive-M.PL.INST 1.SG.NOM know-1.SG.PST that he.NOM 1.SG.DAT say-3.PST

25In fact, observations of the Lithuanian data suggest that scrambling and phrase splitting are sequential steps in a pragmatically guided process, which will be discussed in more detail in 3.3.1
So, while there may be some pragmatic (and processing) difficulty in extending long-distance scrambling across too many clauses, the process is still syntactically grammatical even to the above extent.

At this juncture, however, it is important to note (as Pereltsvaig does) that intact, in situ placement of whole phrases is the unmarked and preferred variant of any of the constructions discussed in this paper. Metalinguistic observations from the Lithuanian informant reveal that the scrambled variants are admittedly quite marked, and the split variants are not only less frequent, but they also carry potentially stigmatizing ‘colloquial’ connotations (i.e., they are not considered an option in ‘standard’ prescriptive Lithuanian), and are thus even more marked. This raises a question: if these variants are indeed so marked, and are overwhelmingly outnumbered by semantically-identical phrases with conventional placement and word order, what features motivate scrambling in the first place, and furthermore, what about phrase splitting?

### 3.3.1 Information structure and scrambling

Bailyn (2001) notes that optionality is an undeniable feature of scrambling, and certainly Chomsky (1986) and others examining A’ movement note that it is for the most part non-obligatory for syntactic and semantic coherence (in contrast to A movement, which is predicated on the need for case or theta assignment). However, in order to make the case that scrambling and splitting are part and parcel of the same copy-movement process, we need to show how they occur and (if possible) why they occur. Accordingly, Bailyn offers what he calls an “alternate account of apparent optionality” (2001: p. 19). As he says, it is not for case or theta reasons; rather, he makes the claim that “A’ scrambling is related to the Topic-Focus or Information Structure component of the grammar” (2001: 21). This leads him to make two generalizations about scrambling phenomena: 1) scrambled word orders are *always* associated with certain discourse/informational interpretations; and 2) scrambling movement is motivated by discourse/informational considerations. For Bailyn, the specific discourse/informational factors he attributes to this process are those that indisputably disambiguate constructions that are semantically ambiguous in conventional word orderings. An example he gives for Russian (from 2001: p. 23) is provided in (57) The content of this example should be understood to be a CP complement to a verb in a higher clause.

(57) a. ....céto Ivan čitaet knigu
    ...that Ivan reads book
    ‘Ivan reads a book.’
    Russian

b. ....céto knigu Ivan čitaet
    ...that book Ivan reads
    ‘...that the book, Ivan reads.’
    Russian

Here, if we suppose the selecting verb is something comparable to ‘like,’ then the word order in (57-a) does not give a concrete clue as to what exactly is being liked. Does the subject like that Ivan (not Pjotr) is reading the book? That Ivan *reads* (rather than skims) a book? Or, that Ivan reads a *book* (not a newspaper)? When scrambled, this clause becomes disambiguated and only one clear meaning emerges (see (57-b). Disambiguation, Bailyn concludes, is the primary function of
scrambling.

While Bailyn’s contention makes sense in the above example, there are a few crucial aspects that need further elaboration. For one, why is the above clausal structure ambiguous, and what specific discourse function does scrambling serve? And second, how does the role of phonology (specifically, intonation) interact with scrambling in disambiguating the intact clause? Pereltsvaig (2008) builds on this work and that of Fanselow & Čavar (2002) and claims that the precise discourse feature in question is +/-contrast. Specifically, it is the +contrast feature that motivates scrambling. She then goes on to provide colloquial Russian examples of contrastive topic (CT) and contrastive focus (CF),26 illustrating how every case of scrambling or splitting found in her corpora exhibit these contrastive features.

Additionally, in every case of CT or CF, whether in situ, or scrambled, or split, there is always an accompanying intonational contour that identifies the phrase as +contrast. What is more, the same exact intonational tune is found on the contrastive element in all three cases. For Pereltsvaig, this consistent intonational contour is integral to confirming her copy-movement theory of scrambling, and her PF partial deletion theory of splitting. In simple terms, the theory goes as follows. See (58), reproduced from Pereltsvaig (2008: p. 18):

(58) a. 

možno kupit’ klubničnogo CF varen’ja
possible to.buy strawberry(ADJ) jam

‘It’s possible to buy strawberry jam.’

Default word order (merger position 1)

b. 

klubničnogo CF varen’ja možno kupit’ klubničnogo CF varen’ja
strawberry(ADJ) jam possible to.buy strawberry(ADJ) jam

‘It’s possible to buy strawberry jam.’

Step 1 (Feature-driven copy/movement)

c. 

klubničnogo CF varen’ja možno kupit’ klubničnogo varen’ja
strawberry(ADJ) jam possible to.buy strawberry(ADJ) jam

‘It’s possible to buy strawberry jam.’

Step 2 (PF interpretation)

(58-a) shows the unmarked, default word order, which Pereltsvaig labels ‘merger position 1.’ This sentence structure, however, differs pragmatically from its semantically identical, non-contrastive counterpart. Namely, there is a contrastively-focused element (in this case, klubničnogo), which receives pragmatic marking in the form of phonological (intonational) prominence. In the second step, the phrase containing the contrastive element is copied, and the copy is moved to an alternative (usually fronted) position in the sentence. Importantly, the sentence is never produced with both copies pronounced in whole. Instead, the PF interface has to interpret where and how each phrase will be pronounced. For scrambling, this PF stage means the phrase is pronounced in entirety in its moved location, all the while retaining the semantic and agreement features it had in its lower instantiation. In the case of splitting, however, only the contrastive elements of the moved phrase are pronounced, while the non-contrastive elements remain in situ. It is notable that a split phrase

26 Contrastive Topic (CT) and Contrastive Focus (CF) are used here consistently with the framework provided by Roberts (1996) and Lee (2003). That is, in discursive context, CT is preceded by a conjunctive question, whereas CF is preceded by an alternative disjunctive question. Basically, CT provides old information, but the possible alternatives are new. CF, on the other hand, provides new information, but the possible alternatives are old. Both CT and CF share the features of +contrast, in that they each provide information that is contrary to expectations in some way.

27 CF (contrastive focus) is noted here by the author, not by Pereltsvaig in the original version.
would never keep a contrastive element in situ, while moving a non-contrastive element. See (59) below.

(59) a. **brašk-✐ CF** **tu** **gal-i** **pirk-ti** **uogien-ės** **parduotuv-ėje**

strawberry-F.PL.NOM 2.SG.NOM can-2.SG.PRS buy-INF jam-F.SG.NOM store-F.SG.NOM

‘You can buy **strawberry** CF **jam** in the store.’ CF - **brašk** ̄iy ‘strawberry’ (fronted)

b. **uogien-ės** CF **tu** **gal-i** **pirk-ti** **brašk-✐** **parduotuv-ėje**

jam-F.SG.NOM 2.SG.NOM can-2.SG.PRS buy-INF strawberry-F.PL.NOM store-F.SG.NOM

‘You can buy **strawberry jam** CF in the store.’ CF - **uogienės** ‘jam’ (fronted)

c. *[uogien-ės** tu** **gal-i** **pirk-ti** **brašk-✐** CF** **parduotuv-ėje**

jam-F.SG.NOM 2.SG.NOM can-2.SG.PRS buy-INF strawberry-F.PL.NOM store-F.SG.NOM

‘You can buy **strawberry CF jam** in the store.’ CF - **brašk** ̄iy ‘strawberry’ (in situ)

In all three cases, then, (default word order, scrambling, and splitting) the discourse is such that the contrastively-marked element is given special intonational prominence, and it is that same intonational prominence that is maintained in any of the possible instantiations of the phrase. This identical intonational tune across multiple instantiations, Pereltsvaig argues, is sufficient evidence for the copy/movement/partial interpretation theory of phrase splitting in colloquial Russian.

**Information structure in Lithuanian** But, does this theory suffice for the Lithuanian data as well? In order to test it, we must first explore the basics of information structure in Lithuanian discourse. A fundamental observation is that Lithuanian typically places focused (or new) items at the end of an utterance. As shown above in section 1.3, example (8), this is one way in which Lithuanian speakers use word order for definite or indefinite reference, in lieu of an overt definite article. However, the preferred S-V-O clausal structure will often trump this general trend such that focused items are intonationally marked instead of being placed last. See (60) and (61) below.

(60) Focus (new information) marked on the last element of the sentence.

a. **Kam** **Judit-احة** **dav-է** **šokolad-ą?**

what.DAT Judita-F.SG.NOM give-3.PST chocolate-M.SG.ACC

‘To whom did Judita give chocolate?’

b. **Judit-احة** **dav-է** **šokolad-ą** **Jon-ui**

Judita-F.SG.NOM give-3.PST chocolate-M.SG.ACC Jonas-M.SG.DAT

‘Judita gave chocolate to Jonas.’

(61) Focus (new information) marked sentence-internally.

a. **Ką** **Judit-احة** **dav-է** **Jon-ui?**

what.DAT Judita-F.SG.NOM give-3.PST Jonas-M.SG.DAT

‘What did Judita give to Jonas?’

b. **Judit-احة** **dav-է** **šokolad-ą** **Jon-ui**

Judita-F.SG.NOM give-3.PST chocolate-M.SG.ACC Jonas-M.SG.DAT

‘Judita gave chocolate to Jonas.’
Note in the above examples that (b) is structurally and semantically identical for both, even as they are answering different questions. In each case, the specific element that answers the question posed (i.e., the focused or new element) receives intonational marking in the form of a L+H* pitch accent and elongated duration on the stressed syllable. Thus, Jonui in (60) and sokolad-q in (61) are marked in this way. Contrastive Topic likewise gets its own intonational tune, this time, with a more abrupt H* pitch accent. Contrastive Focus does not seem to be marked with a different intonational tune than Contrastive Topic, so both get an H* accent. In order to illustrate what these intonational contours look like, we must first look at what an unmarked declarative sentence looks like in Lithuanian. Figure 1 shows a sentence in a pragmatically unmarked (neutral, declarative) context.

![Pitch track illustrating contours unmarked for contrast or focus in Lithuanian. English gloss: ‘I live near the great Nemunas river.’](image)

In Figure 1, while we do see some fluctuations in pitch, there are no pitch accents comparable to the H* accents found in canonical CT/CF contexts. H* pitch accents can be elicited via examples such as (62), which shows a question-answer pair in which the answer provides contrastive topic in the form of a new alternative to old information. That is, mergaitės (‘girls’) is a new alternative even as its members are part of the group vaikai (‘children’), which as the question indicates is old information. On another hand, information that is purely new (i.e., it is not a new alternative within an already known set), is non-contrastive focus (F), which is also shown in (62). In this example, sijonus (‘skirts’) is new information that is not implied or expected as part of the pragmatic context of the question. A pitch track of (62-b) can be found in Figure 2.

(62) a. Ar vaik-ai dėv-ė-jo džins-us?
   Q child-M.PL.NOM wear-PST-3 jeans-M.SG.ACC
   ‘Are the children wearing jeans?’

   b. Mergait-ėsCT dėv-ė-jo sijon-usF
   girl-F.PL.NOM wear-PST-3 skirt-M.SG.ACC
   ‘The girlsCT are wearing skirtsF.’
Figure 2: Pitch track, with both Contrastive Topic (H*) and Focus (L+H*) contours in Lithuanian. English gloss: ‘The girls_{CT} are wearing skirts_{F}.’

The pitch track in Figure 2 illustrates the different canonical intonational patterns for CT/CF-marked and focus-marked elements. For the CT-marked word, *mergaitės*, the intonational contour makes a sharp H* peak on the stressed (penultimate) syllable of the word. The F-marked word, *sijonas*, reaches the H* peak on the stressed (penultimate) syllable after a low dip and a more gradual incline, resulting in a L+H* pitch accent. These distinctive pitch accents are enough to distinguish an element as +contrastive (either topic or focus) or generally (i.e., -contrastive) focused. Topic (or comment) discourse items receive no intonational prominence, characterized by a steady, relatively low, pitch contour.

We can see more evidence of contrastive H* intonational patterns in another set of pitch tracks, each presenting semantically identical material, but with varying instances of contrastive focus marking and phrase movement/splitting. Figures 3 and 4 all take the example sentence from Figure 1 and modify it by adding the +contrastive pragmatic context found in the question-answer pair of example (63) below. In this contrastive context, the preposition *arti* (‘near’) is the contrastive-focused (CF) element and therefore receives H* intonational prominence on its stressed, or ultimate, syllable. In fact, crucially, as Figures 3 and 4 indicate, *arti* gets H* pitch accent wherever its position in the sentence. In surface linear order, Figures 1 and 3 are identical, but because of its contrastive context, Figure 3 differs intonationally through its use of an H* pitch accent. Additionally, when the phrase is split (as in Figure 4) the H* pitch accent is fronted right along with the CF-marked element. Interestingly, the dynamism of the contour is more striking in the split example as well – the H* peak is sharper and steeper.
Thus, what we see in Lithuanian is consistent in many ways with Pereltsvaig (2008)’s colloquial Russian data. The key points are these: 1) if and when we see phrases or items moved, it is for information-structure reasons; and 2) for information-structure reasons, the pragmatically-marked items receive special, consistent phonological treatment (in the form of intonational prominence) whether they remain in situ or moved. These data are consistent with Pereltsvaig’s argument in a nutshell: a copy theory of movement and a PF-driven force for pronunciation of elements in their respective places. The fact that the intonational contour remains consistent is key to this argument.

However, what has been notable in the process of collecting elicited examples from a native speaker is that the pragmatic intuitions accompanying each instantiation of a sentence such as the one in (63-b) are not the same, nor are they across any of the in situ, scrambled, or split representations of any of the sentences presented in this paper. As far as Lithuanian is concerned, Bailyn (2001) and Pereltsvaig (2008) are correct in that the only licensed cases of scrambling and splitting happen in cases of special (usually contrastive) discourse structure. However, that’s as far as their analysis goes, and for Lithuanian at least, there is more to the story. Especially for Lithuanian, where there are no necessary word-ordering modifications that go along with marking information structural factors (i.e., prosody in most cases is sufficient to do so), how do we account for scrambling and split data?

Figure 3: Pitch track illustrating in situ contrastive focus contours (H*) in Lithuanian. English gloss: ‘I live nearCF the great Nemunas river.’
3.3.2 Some problems with the copy-movement analysis.

Both Bailyn (2001) and Pereltsvaig (2008) acknowledge this to be an important question, and an as yet unanswerable one. Bailyn notes that scrambling enables the clarification of potentially ambiguous constructions (see (57)), but doesn’t address splitting. Pereltsvaig notes that only languages with rich case-marking morphology and NP-ellipsis (e.g., Russian and Lithuanian) allow splitting of scrambled phrases in the first place, but still cannot account for why a speaker would necessarily choose to keep a phrase in situ, scramble it, or split it. This is an important question that needs to be answered, but not one that falls within the scope of this paper. However, as far as the Lithuanian data are concerned, there is potentially something to be said for ambiguity as a motivator for scrambling and splitting. In fact, metapragmatic commentary from the Lithuanian consultant suggests that while in situ placement plus intonational marking is adequate for indicating contrastive meaning, the scrambling option is licensed in interactions rife with potential for misunderstanding. And, if ambiguity/misunderstanding still persists, then (and only then) is splitting allowed. The constraint hierarchy would hypothetically go something like in (64).

(64) Phrase placement as potential for pragmatic ambiguity increases $\rightarrow$:
    in situ $\rightarrow$ scrambled $\rightarrow$ split

More investigation is needed before the extent to which this prediction plays out pragmatically can be attested, but at this point, we can observe that splitting is more just than a PF phenomenon – it has to involve pragmatic conditioning as well.
In addition to the necessary pragmatic conditioning, the Lithuanian data offer evidence of the phenomenon being syntactically conditioned as well. One intriguing finding in the Lithuanian data is the fact that, while movement and splitting can occur out of weak islands (see examples (48), (49), and (50)), only movement can occur out of strong islands. In short, phrasal movement in Lithuanian can occur out of strong islands, but (crucially), these phrases cannot split. See examples for Subject islands in (65) and (66) below, followed by examples for Complex NP islands (67), and Adjunct islands (68).

(65) **Subject islands**

a. **sit-as trump-as pasakojim-as [apie didel-i**
   
   this-M.SG.NOM short-M.SG.NOM narrative-M.SG.NOM about big-M.SG.ACC
   
   *vilk-a] buupo pilnas junor-o
   
   wolf-M.SG.ACC was full-M.SG.NOM humor-M.SG.GEN
   ‘This short story about a big wolf was full of humor.’
   
   In situ

b. **[apie didel-i] vi*lk-a] sit-as trump-as**
   
   about big-M.SG.ACC wolf-M.SG.ACC this-M.SG.NOM short-M.SG.NOM
   
   pasakojim-as buvo pilnas junor-o
   
   narrative-M.SG.NOM was full-M.SG.NOM humor-M.SG.GEN
   ‘This short story about a big wolf was full of humor.’
   
   Scrambling

c. **[apie didel-i] vi*lk-a] sit-as trump-as pasakojim-as**
   
   about big-M.SG.ACC this-M.SG.NOM short-M.SG.NOM narrative-M.SG.NOM
   
   [vilk-a]B buvo pilnas junor-o
   
   wolf-M.SG.ACC was full-M.SG.NOM humor-M.SG.GEN
   ‘This short story about a big wolf was full of humor.’
   
   Splitting

(66) a. **Jono apsilankym-as [pas man-o jaun-cesn-j**
   
   John-M.SG.GEN visit-M.SG.NOM to 1.SG-GEN young-COMP-M.ACC
   
   broli] man-e labai nustebin-o
   
   brother-M.SG.ACC 1.SG-DAT very surprise-PST.3
   ‘John’s visit to my younger brother shocked me.’
   
   In situ

b. **[pas man-o jaun-cesn-j broli]** Jono
   
   to 1.SG-GEN young-COMP-M.ACC brother-M.SG.ACC John-M.SG.GEN
   
   apsilankym-as man-e labai nustebin-o
   
   visit-M.SG.NOM 1.SG-DAT very surprise-PST.3
   ‘John’s visit to my younger brother shocked me.’
   
   Scrambling

c. **[pas man-o]B Jono apsilankym-as [jaun-cesn-j**
   
   to 1.SG-GEN John-M.SG.GEN visit-M.SG.NOM young-COMP-M.ACC
   
   broli]B man-e labai nustebin-o
   
   brother-M.SG.ACC 1.SG-DAT very surprise-PST.3
   ‘John’s visit to my younger brother shocked me.’
   
   Splitting

(67) **Complex NP island**

38
a. \[ \text{aš gird-é-jau naujien-q kad Fred-as išsprend-é} \]
\[\text{I hear-PST.1SG news-M.SG.ACC that Fred-M.SG.NOM solve-PST.3SG}\]
second-M.ACC-DEF problem-M.SG.ACC
‘I heard the news that Fred solved the second problem.’ In situ

b. \[\text{[antr-a-ja] uždavim-jl aš gird-é-jau naujien-q kad}\]
second-M.ACC-DEF problem-M.SG.ACC I hear-PST.1SG news-M.SG.ACC that
\[Fred-as išsprend-é]\nFred-M.SG.NOM solve-PST.3SG
‘I heard the news that Fred solved the second problem.’ Scrambling

c. *\[\text{[antra-a-ja] aš gird-é-jau naujien-q kad Fred-as}\]
second-M.ACC-DEF problem-M.SG.ACC I hear-PST.1SG news-M.SG.ACC that Fred-M.SG.NOM
\[išsprend-é [uždavim-jl]\]
solve-PST.3SG problem-M.SG.ACC
‘I heard the news that Fred solved the second problem.’ Splitting

d. \[\text{[kur-i-j] aš gird-é-jau naujien-q kad Fred-as išsprend-é}\]
which-M.ACC problem-M.SG.ACC I hear-PST.1SG news-M.SG.ACC that Fred-M.SG.NOM solve-PST.3SG
\[[uždavim-jl]\]
‘Which did I hear the news that Fred solved the second problem.’ WH-movement
(more acceptable)

(68) **Adjunct island**

a. \[ tu iš-é-jai todel kad Marij-a kalb-é-jo \text{ apie}\]
2.SG from-leave-PST.2.SG because that Marij-F.SG.NOM speak-PST.3.SG about
\[\text{Greek-M.PL.GEN philosophy-F.SG.ACC}\]
‘You left because Marija was talking about Greek philosophy.’ In situ

b. \[\text{[api e} \text{ Grait-ü } \text{ filosofij-a]} tu iš-é-jai todel kad\]
about Greek-M.PL.GEN philosophy-F.SG.ACC 2.SG from-leave-PST.2.SG because that
\[Marij-a kalb-é-jo\]
Marij-F.SG.NOM speak-PST.3.SG
‘You left because Marija was talking about Greek philosophy.’ Scrambling

c. *\[\text{[api e} \text{ Grait-ü } \text{ filosofij-a]} tu iš-é-jai todel kad Marij-a\]
about Greek-M.PL.GEN 2.SG from-leave-PST.2.SG because that Marij-F.SG.NOM
\[kalb-é-jo [filosofij-a]\]
speak-PST.3.SG philosophy-F.SG.ACC
‘You left because Marija was talking about Greek philosophy.’ Splitting

\[28\] The examples for Complex NP islands presented here encompass noun complement cases; islands with relative clauses have not yet been tested.
If the Pereltsvaig account of splitting is correct, as fundamentally/solely interpreted through the PF interface in every instance where scrambling can occur, then two important questions remain. First, why is splitting not allowed in cases of A’ movement out of strong islands? Second, why is splitting allowed in A’ movement out of weak islands? These questions leave open the possibility that an important syntactic constraint is at work in the splitting process, such that it is sensitive to island effects that A’ movement alone is not. Again, an exploration of this question is beyond the scope of this paper, but it leaves an intriguing question for future investigations of splitting phenomena in Lithuanian.

4 Summary and Conclusion

This paper has set out to document for Lithuanian a phenomenon widely reported in (certain) Slavic languages - left branch extraction. As a language with close genetic, structural, and socio-historical ties to Slavic languages, Lithuanian is an appropriate case upon which to extend the study of extraction phenomena. Some Slavic linguists, most notably Bošković (see 2005, 2009, 2012), have connected extraction phenomena with syntactic structure, namely the presence or lack thereof of a determiner phrase (DP) in the noun phrase structure. Essentially, the theory goes that a DP structure prohibits extraction of elements out of a noun phrase, because the D acts as a phase, which blocks movement according to the Phase Impenetrability Condition (Chomsky 1999). Bošković and others have gone to great lengths to show how languages which lack articles (as do most of the Slavic languages save for Bulgarian and Macedonian) allow left branch extraction while languages with articles do not. Bošković has even created a list of generalizations describing so-called ‘NP’ languages vis-à-vis ‘DP’ languages.

Lithuanian proves an interesting case, because it possesses long morphology and a definiteness-marking paradigm similar to that of Bulgarian (a ‘DP’ language, according to the Bošković account). Yet, Lithuanian allows for left-branch extraction out of noun phrases that are marked with definite morphology, directly contradicting Bošković’s predictions. Additionally, Lithuanian shows mixed properties for the other Bošković generalizations for NP/DP languages – in some ways patterning with DP languages and in other ways patterning with NP languages. Thus, the DP-phase-based account is not an appropriate one through which to view extraction phenomena in Lithuanian.

Upon closer examination, we see that extraction phenomena in Lithuanian are actually more free and fluid than allowed for in prototypical examples of left-branch extraction. As noted by Pereltsvaig (2008) for Russian, elements of Lithuanian phrases can be moved even if they are non-constituents (e.g., PPs, PP complements, adverbials, etc.). Additionally, elements can even be moved out of weak islands. These are yet more pieces of evidence going against the Bošković account of extraction.

Thus, what is needed is an account that takes into consideration a more flexible understanding of extraction, and Pereltsvaig provides one, using what she calls a theory of “movement plus partial interpretation of copies” (2008: p. 17). Under this analysis, extraction is actually a process of copy-movement of phrases, at which point the PF interface decides which pieces of each copy to pronounce. Thus, the seemingly non-constituent nature of extracted elements are really underlyingly full phrasal constituents, which are partially pronounced at different points in the sentence.

Pereltsvaig supports this analysis by 1) showing how integrally related are scrambling/splitting
and contrastive information structure; and 2) how important PF is to the partial interpretation of moved copies. Splitting, in this account, is essentially a phonological interpretative feature, which can be illustrated through the retention of specialized intonational contours on the split elements of a +contrastive phrase.

The Pereltsvaig account suits the Lithuanian data nicely, as every instance of splitting in the data coincides within a pragmatic context of contrast. What is more, that contrast can be expressed through intonational contours that occur consistently across the board, whether in situ, in full phrasal scrambling, or on elements of split phrases. This analysis nicely characterizes a copy-movement and partial PF interpretation phenomenon. However, at least as far as Lithuanian is concerned, the analysis falls short in two crucial ways. First, the PF interface alone does not explain why most of the time phrases remain in situ, why sometimes they scramble, and and in even rarer occasions, why they split. An additional, possibly pragmatic, explanation is needed to account for this. And second, the PF interface alone would not explain why strong islands can be violated in Lithuanian through scrambling, but not through splitting. This is evidence for a syntactic explanation not yet explored in the Pereltsvaig (2008) analysis. These shortcomings provide fodder for further investigation of Lithuanian, and at the very minimum, they illustrate how the DP/NP (i.e., Bošković) analysis of extraction phenomena does not hold for this language.
References


Harizanov, Boris. 2009. The definite marker in Bulgarian.


