Minimal phrase structure: a new formalized theory of phrase structure

John J. Lowe[1] *and Joseph Lovestrand*[2]
[1] University of Oxford
[2] SOAS, University of London

ABSTRACT

X' theory was a major milestone in the history of the development of generative grammar.[1] It enabled important insights to be made into the phrase structure of human language, but it had a number of weaknesses, and has been essentially replaced in Chomskyan generativism by Bare Phrase Structure (BPS), which assumes fewer theoretical primitives than X' theory, and also avoids several of the latter's weaknesses. However, Bare Phrase Structure has not been widely adopted outside the Minimalist Program (MP), rather, X' theory remains widespread. In this paper, we develop a new, fully formalized approach to phrase structure which incorporates insights and advances from BPS, but does not require the Minimalist-specific assumptions that come with BPS. We formulate our proposal within Lexical-Functional Grammar (LFG), providing an empirically and theoretically superior model for phrase structure compared with standard versions of X' theory current in LFG.

Keywords: phrase structure, X' theory, Bare Phrase Structure, Lexical-Functional Grammar

1 INTRODUCTION

X′ theory, first introduced in Chomsky (1970) and elaborated in Jackendoff (1977) among other works, was a major milestone in the history of the development of generative grammar. It provided, for the first time, a mechanism for capturing generalizations and constraints on possible phrase structures in language. X′ theory originated as a means of generalizing over sets of phrase structure rules (PSRs), but in the early 1980s, within the Principles & Parameters model, it led to the abandonment of PSRs as a part of the grammar of individual languages. X′ theory encapsulated important insights into the phrase structure of human language, but it had a number of weaknesses, and has been essentially replaced in Chomskyan generativism by Bare Phrase Structure (Chomsky 1995). Bare Phrase Structure (BPS) assumes fewer theoretical primitives than X′ theory, and is therefore preferable from a minimalist perspective; it also avoids several of the empirical and theoretical weaknesses of X′ theory. However, Bare Phrase Structure is unavoidably associated with a number of assumptions which are theory-specific to the Minimalist Program (MP) — most obviously perhaps, its derivational nature — and for this reason has not been widely adopted outside the MP.

Where Bare Phrase Structure is not adopted, X′ theory remains the most widespread approach to phrase structure, and it remains the standard means of approaching phrase structure in most introductory text books. The grammatical framework of Lexical-Functional Grammar (LFG: Kaplan and Bresnan 1982) retains X′ theory in largely its original form (i.e. as a set of cross-linguistic generalizations over PSRs in the grammars of individual languages), and thus retains both the benefits and weaknesses of this approach to phrase structure. We take the version of X′ theory currently utilized in LFG to be the most elaborate and precisely formalized version of X′ theory currently in use.

In this paper, we develop a new, fully formalized approach to phrase structure within LFG which avoids the major weaknesses of X′ theory and incorporates many of the advantages of BPS.[2] While formalized within LFG, our proposal is easily extensible to other theories.

---

[2] An early version of our proposal was made in Lovestrand and Lowe (2017). The present version differs in significant ways, most importantly in its use of distributive features (§3.3) to eliminate redundancy in labelling.

Our model has been tested within the computational implementation of LFG, the Xerox Linguistic Environment (XLE: Crouch *et al.* 2011).[3]

2   CONSTRAINING PHRASE STRUCTURE

Since the introduction of PSRs by Chomsky (1957) as a central component of the theory of formal syntax, there has been significant progress constraining this formal mechanism to approximate the actual types of phrase structures that are attested in languages, and to prevent the theory from being able to produce unattested phrase structures. The most significant milestone in the development of the theory of phrase structure was the development of X′ theory. However, X′ theory had a number of inadequacies which ultimately led to its replacement in the mainstream Chomskyan tradition. In this paper we focus on seven features lacking from X′ theory which should form a part of an adequate theory of phrase structure; most but not all of these are found in BPS. An adequate theory of phrase structure should (in contrast with existing formalized versions of X′ theory):

(1)   a.   Utilize only as much structure as required to model constituency, avoiding nonbranching dominance chains.

        b.   Avoid the assumption of massive/default optionality in PSRs.

        c.   Avoid redundancy in category labelling, ensuring that endocentric phrases necessarily share the category of their head without stipulation.

        d.   Lack a distinct notion of X′.

        e.   Incorporate a notion of X^{max} distinct from 'XP', and a notion of the highest projection distinct from X^{max}.

        f.   Incorporate a principled account of exocentricity.

        g.   Incorporate a principled account of nonprojecting categories.

---

[3] Being formulated within LFG, our model functions as a set of constraints on language-specific PSRs, but it is important to note that our proposal could without difficulty be reinterpreted within different frameworks purely as a set of constraints on phrase structure more generally, with no language-specific PSRs as such.

Most of the desiderata in (1) address specific issues that have arisen in the development of X′ theory. BPS has addressed many of these issues, though not all. The last two desiderata expand the coverage of the theory of phrase structure to include two types of non-X′-theoretic structures: nonprojecting words and exocentric structures are adopted in X′-theoretic approaches to phrase structure in LFG, but have not been formally incorporated in the theory. Our proposal below is the first fully formalized theory of phrase structure that satisfies all of the desiderata in (1).

In the following sections we discuss two contemporary approaches to phrase structure: the version of X′ theory current within LFG, which we take to be the most fully developed version of X′ theory currently in use; and BPS, the standard approach to phrase structure within the Chomskyan generative tradition.

2.1 *Current X′ theory in LFG*

X′ theory began as a means of stating generalizations over sets of PSRs.[4] Following Stowell (1981), X′ theory was reconceived within the Principles & Parameters framework as a set of universal constraints on phrase structure, and subsequently language-specific PSRs themselves were eliminated; language-specific characteristics of phrase structure were instead constrained by syntactic processes, such as the assignment of Case. This final step was not taken in LFG. In LFG, X′ theory remains a means of generalizing over and stating constraints on sets of PSRs. PSRs themselves cannot be eliminated, because they constitute the main body of non-lexical constraints in a grammar. A minimal Lexical-Functional Grammar consists of a set of lexical entries and a set of PSRs; grammatical structure is, and can only be built by the application of specific PSRs (which ultimately license insertion of lexical information).

The advantage of LFG's phrase-structure based approach to structure building is its computational efficiency: despite being a unification-based system, which therefore in principle has the power of an unrestricted rewriting system, the structure-building component of an LFG is a context-free phrase structure grammar; as shown by Maxwell and

---

[4] A detailed introduction to X′ theory and its development is provided by Carnie (2010, chapter 7). See also Carnie (2000) and Kornai and Pullum (1990).

Kaplan (1996), appropriate interleaving of context-free parsing and f-structure unification can be computed in cubic time.

Despite the obvious strengths which led to its great success, and which were largely adopted into BPS, X′ theory suffers from a number of weaknesses; see Kornai and Pullum (1990) for a detailed examination of the theoretical weaknesses of X′ theory. We focus here on X′ theory as it is currently conceived and used within LFG, which admits a number of extensions to and alterations of the strict principles of X′ theory in its original formulation.

We focus on four main weakness of X′ theory as utilised within LFG, all of which are evident in (2), a standard LFG constituent structure for the sentence *Spot runs*: nonbranching dominance chains, optionality of daughters (related to the existence of nonbranching dominance chains, of course, but including heads), redundancy in category labelling, and the need to assume intermediate (X′) nodes as an independent theoretical construct (1a–d). We discuss these issues in turn.

(2)

```
                IP
              /    \
           DP        I′
           |         |
           D′        VP
           |         |
           NP        V′
           |         |
           N′        V
           |         |
           N        runs
           |
          Spot
```
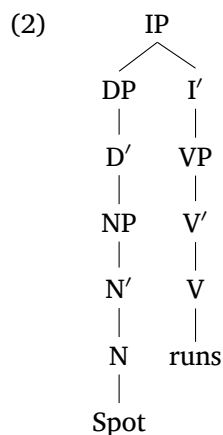
As discussed in §3.1, the LFG representation of phrase structure, c(onstituent)-structure, models only surface constituency relations, while functional syntactic relations are modelled at a separate level of structure, f(unctional)-structure. Thus phrases which consist of only one word, like the DP *Spot* and the VP *runs*, can only be modelled within LFG's approach to X′ theory by assuming nonbranching dominance chains, such as the DP chain in (2), where we have four nonbranching nodes dominating the N. There can be no silent specifier, head or complement positions hosting functional features to fill

out the tree, because such features are represented at f-structure and, as stated, the tree models only the surface constituency relations of the overt elements of the sentence.[5] Even within syntactic theories which admit empty nodes, adherence to X′ theory would still involve some nonbranching dominance chains (though perhaps not as long as in (2)).

Although nonbranching chains as in (2) do model relevant properties of the structure, such as the dual maximality (phrasality) and minimality of the individual words, the resulting structure, involving ten nonterminal nodes, seems inordinately complex as a representation of the surface constituency of a two word sentence. This constituency could be equally well captured by the tree in (3), which is considerably more in the spirit of BPS. Our proposal below licenses structures equivalent to (3).

(3)

```
          V
         / \
        N   V
        |   |
      Spot runs
```
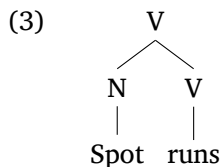
Related to this problem is the issue of optionality of phrase structure nodes (1b). Clearly, dominance chains like XP-X′-X require that specifier and complement positions be optional. But as can be seen in (2), heads can also be optional. This must be possible for functional categories like I and D, on the assumption, standard in LFG, that V and N are necessarily dominated by these categories (as in 2). But many analyses also require heads of lexical phrases to be optional. Most work in LFG, therefore, including the standard textbooks of Bresnan (2001) and Dalrymple (2001), assume that all phrase structure positions are in principle optional, heads and nonheads alike. However, there are

---

[5] There is some debate within LFG over the existence of traces, i.e. whether there may be some terminal nodes in a c-structure which do not correspond to any overt element. Arguments against traces were made by Kaplan and Zaenen (1989), and widely accepted within the LFG community; traces are accepted by Bresnan (1995, 1998, 2001) and Bresnan *et al.* (2016) only in order to account for weak crossover, but analyses of weak crossover which do not involve traces are offered by Dalrymple *et al.* (2001, 2007), Nadathur (2013) and Dalrymple and King (2013).

certain structures in some languages in which optionality must be suppressed; see Snijders (2012) and Dalrymple *et al.* (2015, 386-388) for detailed discussion of such cases.[6]

Optionality as the default situation, ruled out in certain circumstances, is widely assumed in existing LFG analyses, but has never been properly formalized: in LFG, the right-hand side of a PSR must be a regular expression; in regular expressions it is optionality (defined as disjunction with the empty set), not obligatoriness, which has to be specified. In contrast, it would be more intuitive, and PSRs would be considerably less ambiguous, if optionality were the exception, rather than the rule. The model we present below avoids the need for mass optionality, treating optionality as an occasional necessity, rather than a default.

A further weakness of X′ theory involves another type of redundancy in representation: each node is independently specified with a category label, but given the inherent constraints on X′-theoretic structures, each node in a projection chain necessarily has the same category label, meaning that it ought not to be necessary to specify this information more than once for each projection chain. That is, the notion that a phrasal node necessarily has the same category label as its head ought to fall out naturally, rather than by stipulation, which is essentially the way it has to be done in X′ theory. Our proposal makes use of the concept of *distributive features* to ensure that only a single instance of category labelling applies for each projection chain.

The fourth major weakness of X′ theory is that it entails the existence of the intermediate node type X′ as an independent theoretical construct (1d). However, a wealth of research has demonstrated that there is no clear evidence of syntactic processes which make reference to the X′ level, suggesting that it is not an independent concept in human language.[7]

2.2   *Further problems: augmenting X′ theory*

In attempting to provide a sufficiently flexible model of phrase structure to adequately capture the wide range of crosslinguistic variation

---

[6] See further Lovestrand and Lowe (2017, 289–290).

[7] Early arguments in Travis (1984), see also Carnie (2000, 2010).

in surface configurational syntactic structure, LFG has been forced to admit certain augmentations to the basic X′-theoretic structures it inherited. These augmentations are not problematic in themselves, but they have never been properly integrated into existing formal analyses of X′ theory.

In addition to endocentric phrase structures, LFG also admits *exocentric* structures, most commonly the exocentric clausal category S (Bresnan 1982; Kroeger 1993; Bresnan 2001). S is not subject to ordinary X′-theoretic constraints: it is a non-headed category that may contain a predicate along with any or all of its arguments. S is most commonly utilized in the analysis of non-configurational languages (Austin and Bresnan 1996; Nordlinger 1998), but it is also utilized in some analyses of languages with relatively fixed word order, such as Welsh (Sadler 1997) and Barayin (Lovestrand 2018).[8] While S, and sometimes other exocentric categories, are widely admitted in LFG, recent formalizations of X′ theory find no place for exocentricity, leaving it outside the formal system while nevertheless remaining crucial to actual grammars and analyses.

A further concept widely adopted within LFG is that of nonprojecting categories. Toivonen (2003) argues that alongside the traditional projecting lexical categories, there exist also *nonprojecting* categories, represented as $\hat{X}$, which adjoin to X^0 (projecting) heads. Non-projecting words do not head phrases, and so it is not possible for another phrase to stand in a specifier, complement or adjunct relation to such a word. Non-projecting words are often particles and/or clitics. Toivonen argues in detail that verb particles in Swedish are nonprojecting $\hat{P}$s, giving the example in (4), and proposing the augmentation to X′ theory shown in (5).[9]

(4) Eric har slagit ihjäl   ormen
     Eric has beaten to.death snake.DEF

---

[8] See ex. 38 below.

[9] The comma in the templatic PSR in 5 is the 'shuffle' operator, indicating variable order of the sequences on either side. For its use in LFG see Dalrymple *et al.* (2019, 204–205).
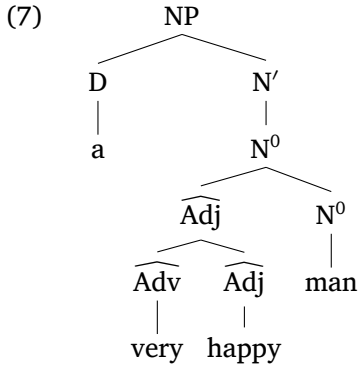
'Eric has beaten the snake to death.'
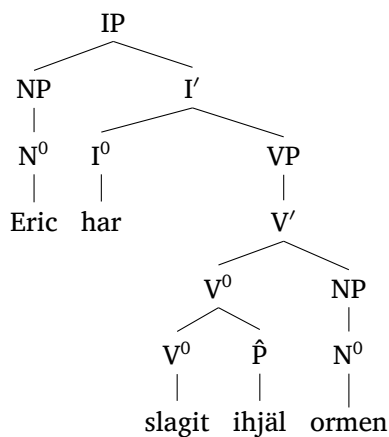
(5) $X^0 \rightarrow X^0 , \hat{Y}$

The possibilities for nonprojecting words have been further broadened by other authors, relaxing Toivonen's (2003) assumption that nonprojecting words adjoin only to X^0 heads. Spencer (2005) argues for adjunction of nonprojecting words to phrasal categories, as well as to X^0 heads, in order to capture the properties of case clitics in Hindi. Duncan (2007) and, more recently, Arnold and Sadler (2013), propose that nonprojecting categories may also adjoin to nonprojecting categories. Arnold and Sadler (2013) base their proposals on the relatively familiar features of prenominal modification in English. Building on work by Poser (1992) and Sadler and Arnold (1994), they argue that prenominal modification in English should be analysed in terms of nonprojecting categories; this accounts for the fact that prenominal adjectives cannot take postpositioned complements or modifers, unlike adjectives in other positions. But since prenominal modification is recursive, this requires that nonprojecting categories can be adjoined not only to X^0, but also to nonprojecting $\hat{X}$s. That is, we require a rule of the kind in (6); the analysis proposed by Arnold and Sadler (2013) for prenominal modification in English is shown in (7).

(6) $\hat{X} \rightarrow \hat{Y} \, \hat{X}$

(7)

```
                    NP
             ┌───────┴───────┐
             D               N′
             │               │
             a               N⁰
                      ┌───────┴───────┐
                     Adj              N⁰
                 ┌────┴────┐          │
                Adv       Adj        man
                 │         │
                very     happy
```

Here the nonprojecting category $\widehat{\text{Adj}}$ adjoins to N^0, while nonprojecting $\widehat{\text{Adv}}$ adjoins to $\widehat{\text{Adj}}$.

Once again, existing formalizations of X′ theory within LFG do not adequately account for nonprojecting categories. Our proposal does so, and we model our approach to nonprojecting categories with respect to English prenominal modification, adopting the proposals of Arnold and Sadler (2013) illustrated here. Our model also allows for adjunction of a non-projecting node (or any kind of node) to a phrasal category, XP, as proposed by Spencer (2005).[10]

2.3   *BPS*

The origins of BPS have been discussed in detail by a number of authors, including Carnie (2010, 135–167), and here we will focus only on the major innovations and insights which distinguish BPS from X′ theory.[11] In general, and in line with the Minimalist Program, BPS aims to incorporate the major insights of X′ theory not as stipulations but as the natural consequences of deeper principles. In doing this, certain problematic aspects of X′ theory have been discarded.

One early identification of a major weakness in X′ theory was by Fukui (1986), who shows that the amount of structure found with particular types of projection may vary crosslinguistically; in particular, in some languages functional categories lack specifiers. Fukui draws the conclusion that there is a difference between XP (understood as

---

[10] This possibility is not modeled below, but it could be achieved by modifying the adjunction rule in (36b) so that the template @LOM is replaced by @LPM.

[11] Formalizations of the principles of BPS are given by e.g. Stabler (1997), Gärtner (2002) and Collins and Stabler (2016). We discuss the latter work below.

X″) and X^{max}, a maximal projection: some maximal projections are equivalent to X′. Thus if there is cross- or even intra-language variation in the amount of structure admitted in different projections, X′ theory provides no coherent notion of a maximal projection. As noted by Lovestrand and Lowe (2017, 288–289) this weakness persists in X′ theory as utilized within LFG; for example, Bresnan *et al.* (2016, 130) permit phrases to lack specifiers "as a parametric choice", without addressing the formal problems this raises.

Similar problems with distinguishing X^{max} from the top projection, in cases of adjunction, are discussed by Hornstein and Nunes (2008): if the properties of mother and head daughter are identical in adjunction structures, then adjunction to X^{max} results in multiple X^{max} projections; only one X^{max} is the top projection, but this cannot be formally distinguished from the others.[12] Our proposal below can capture both the distinction between XP and X^{max}, and between X^{max} and the top projection.

The consequence of Fukui's separation of XP from X^{max} is a relativization of the notion of maximal category, and a concurrent weakening of the status of bar levels as absolute notions. A similarly relativized approach to projection levels was taken by Speas (1986). The underlying intuition is that the amount of structure in a phrase is only as much as needed to account for the constituency; maximal projections may correspond to X″, X′, or even X, depending on the phrase in question. Thus a node may be both maximal and minimal at the same time; it is primarily this intuition which motivates X′ theoretic structures like (2) to be simplified into structures more like (3). The relativized approach to X′ theoretic notions proposed by Speas (1986) provides a coherent definition of X^{max}, which is lacking in X′ theory.[13] But at the same time, this approach eliminates a coherent notion of X′. Speas (1986) shows that this is a valid elimination, since there are

---

[12] An alternative and more standard way of approaching adjunction within BPS involves the notion of 'pair-merge' (Chomsky 2001). We do not see how 'pair-merge' could be treated coherently within the framework adopted in this paper, and note that it has been criticized within the Chomskian tradition, e.g. by Hornstein and Nunes (2008).

[13] Speas' definition of maximal projection, as emended by Carnie (2010, 139), runs: "X = XP if $\exists$G, immediately dominating X, the head of G $\neq$ the head of X."

no syntactic phenomena which necessarily make reference to the X′ level (see also fn. 7).

The insights of Fukui (1986) and Speas (1986) fed into the theory of BPS as developed by Chomsky (1995). One of the fundamental features of BPS is the notion that all structure building can be attributed to a single basic syntactic operation, Merge. Merge takes two elements and forms them into a set, which is labelled with one of the two elements. The element which provides the label is the head.

The labelling mechanism is a further aspect of BPS relevant to the present discussion. For Chomsky (1995), the label of a merged structure is automatically derived from one of the merged elements. Thus labelling is a part of the definition of Merge, and as such the notion that a phrase necessarily has the same category label as its head falls out without further stipulation, given the definition of Merge. In contrast, as noted above, in X′ theory the fact that a head X necessarily heads a phrase XP (rather than YP) falls out only by stipulation: PSRs, or constraints on PSRs, are stated in such a way that this intuition is not violated, but in principle different rules or constraints might have been stated which did violate the intuition. Following Collins (2002), some approaches to BPS go further, attempting to eliminate labelling altogether. While this is not universally accepted, it reflects the deeper aims of the MP to eliminate as far as possible all redundant elements of analysis.

Another central element of BPS is the concern with accounting for linearization patterns, building on the work of Kayne (1994). In the PSR-based approach we use as the basis for our proposals in this paper, linear order is a given, stipulated in the PSRs wherever determinate, with variable ordering a marked possibility. We therefore do not consider this aspect of BPS further here.

2.4   *Conclusion*

In the foregoing discussion, we have identified seven main ways in which a theory of phrase structure should improve upon existing formalizations of X′ theory and/or should incorporate insights from BPS. A formal model of phrase structure should avoid non-branching chains, and the default optional nodes associated with them. It should not stipulate a mid-level X′ node, and should include a mechanism to distinguish a maximal node, in the sense of the mother of a struc-

ture including all specifiers and complements, from a higher node including adjunction structures. The theory should naturally produce endocentric structures in which heads and mothers share category information, while at the same time successfully modeling nonprojecting and exocentric structures.

3 A NEW MODEL: MINIMAL PHRASE STRUCTURE

3.1 *Underlying architecture*

As stated, our proposal is formalized within LFG. LFG is a constraint-based, non-derivational framework for grammatical analysis; handbooks include Dalrymple (2001), Falk (2001), Bresnan *et al.* (2016) and Dalrymple *et al.* (2019). A central aspect of the LFG framework is that it distinguishes different types of grammatical information and models them as distinct levels of grammatical representation. These levels are related to one another by means of *projection functions*.

One level of grammatical representation, central to the present topic, is the c(onstituent)-structure, which represents the phrasal structure of a clause. C-structure is represented as a phrase-structure tree, and constraints on possible c-structures are stated as PSRs. As discussed above, c-structure represents only the surface constituency of a clause or phrase, while more abstract functional syntactic properties and relations, such as grammatical functions, long-distance dependencies and agreement features, are dealt with at the level of f(unctional)-structure. F-structure is represented as an attribute-value matrix, and understood in set-theoretic terms as a set of attribute-value pairs (Dalrymple 2001, 30).

So, for the English sentence *Spot runs*, the c-structure can be represented as in (2), assuming for the moment standard X′ theoretic structures; the f-structure for the same sentence, representing the abstract grammatical structure of the clause, can be represented as in (8).[14]

---

[14] Following standard LFG conventions, we represent only those features of f-structure that are relevant for the discussion at hand, omitting features encoding information about person, number, gender, tense, aspect, and other grammatical information. More complex f-structures containing more features appear below, e.g. (23) and (24).

(8)
$$\begin{bmatrix} \text{PRED} & \text{'run}\langle\text{SUBJ}\rangle\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'Spot'} \end{bmatrix} \end{bmatrix}$$
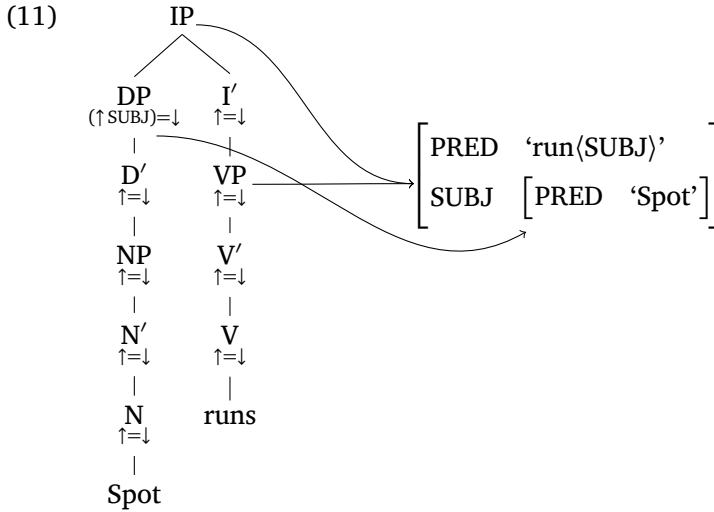
These two levels of grammatical representation are related via the projection function ϕ, which maps c-structure nodes to corresponding f-structures. Functional descriptions (f-descriptions) constrain the possible relations between c-structures and f-structures. The relations between c- and f-structure are stated by reference to c-structure nodes, their mothers, and the f-structures projected from those nodes and their mothers. So, any c-structure node can be referred to by the variable $*$, and its mother by the variable $\hat{*}$. The f-structure projected from any c-structure node is therefore obtained by the application of the function ϕ to the variable $*$, that is $\phi(*)$, and likewise the f-structure projected from a c-structure node's mother is obtained by the application of ϕ to $\hat{*}$, that is $\phi(\hat{*})$. These functions are abbreviated using the metavariables $\downarrow$ and $\uparrow$:

(9)   a. $\downarrow \equiv \phi(*)$

        b. $\uparrow \equiv \phi(\hat{*})$

Using these metavariables it is possible to concisely state constraints on the relation between c-structure and f-structure. For example, in English the specifier of IP is associated with the grammatical role of subject. The following PSR captures this constraint:

(10)   IP $\rightarrow$ DP   I$'$
                    $(\uparrow\text{SUBJ}) = \downarrow$ $\uparrow = \downarrow$

The annotation $(\uparrow\text{SUBJ}) = \downarrow$ on the specifier of IP states that the f-structure projected from the DP ($\downarrow$) is the value of the attribute SUBJ in the f-structure projected from the DP's mother ($\uparrow$). The annotation $\uparrow = \downarrow$ on the I$'$ states that the f-structure projected from the I$'$ ($\downarrow$) is the same f-structure as that projected from the IP ($\uparrow$). Ex. (11) repeats the c-structure in (2), but augmented with the functional descriptions specified for each node in the PSRs, and shows the projection function ϕ relating the c-structure to the f-structure (from 8) by means of arrows between the two structures.

(11)

```
          IP
         /  \
       DP    I′
 (↑ SUBJ)=↓  ↑=↓
       |      |
       D′    VP
      ↑=↓    ↑=↓
       |      |
      NP     V′
     ↑=↓    ↑=↓
       |      |
      N′      V
     ↑=↓    ↑=↓
       |      |
       N    runs
      ↑=↓
       |
      Spot
```

$$\begin{bmatrix} \text{PRED} & \text{‘run}\langle\text{SUBJ}\rangle\text{’} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{‘Spot’} \end{bmatrix} \end{bmatrix}$$

Importantly, c-structure and f-structure are not the only two levels of grammatical representation, and ϕ is not the only projection function. For example, the function σ maps f-structures to s(emantic)-structures. Kaplan (1989) generalized the concept of projection functions between levels of grammatical representation, resulting in a 'projection architecture' of different levels of linguistic structure. Much recent work has debated the full inventory of projections and projection functions, including e.g. Bögel *et al.* (2009), Dalrymple and Mycock (2011), Dalrymple and Nikolaeva (2011), Giorgolo and Asudeh (2011), Asudeh (2012, 53), and Mycock and Lowe (2013).

For our purposes, the details of the projection architecture are not important. But one additional projection is vital to the present discussion. While c-structure representations standardly incorporate information on category labels and projection level in representing nodes as IP, N′, V etc., this is to be understood as a shorthand. Following Kaplan (1989), category information and projection level are not directly encoded in c-structure, but are projected from c-structure nodes via a projection λ. That is, the representation in (12) must be understood as a shorthand for something like (13). Note that the l-structures in (13) are for illustrative purposes only; the feature BAR is not an element of the analysis we propose below.[15]

---

[15] On BAR see §4.1 below.

(12)　　IP　　　　　(13)
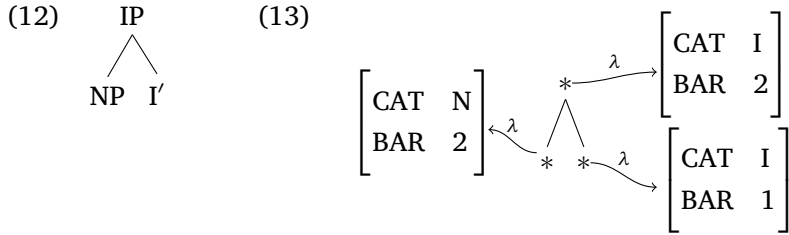
We refer to the structure projected by λ as the l-structure. Since projection level and category information are not actually a part of c-structure, but are projected from it just like f-structure features, it follows that projection level and category information must be constrained in PSRs by means of functional descriptions on nodes, rather than as inherent properties of nodes. For example, just as (12) is an abbreviation for (13), so the PSR in (14) can be understood as an abbreviation for something like (15); recall that $*$ represents a phrase structure node.

(14)　IP　$\rightarrow$　　　　NP　　　　　I$'$
　　　　　　　　　　($\uparrow$SUBJ)$=\downarrow$　　$\uparrow=\downarrow$

(15)　$*$　$\rightarrow$　　　　　　$*$　　　　　　　$*$
　　　　　　　　　　　　($\uparrow$SUBJ)$=\downarrow$　　　　$\uparrow=\downarrow$
　　　　　　　　　　　　$(\lambda(*)$CAT$)=$N　　$(\lambda(*)$CAT$)=$I
　　　　　　　　　　　　$(\lambda(*)$BAR$)=$2　　$(\lambda(*)$BAR$)=$1
　　　　　　　　　　　　　　　　　　　　$(\lambda(\hat{*})$CAT$)=$I
　　　　　　　　　　　　　　　　　　　　$(\lambda(\hat{*})$BAR$)=$2

3.2　　　　　　　　　　　　*Main features*

Clearly, the functional descriptions specifying category and projection level in (15) are highly inadequate, and fail to capture most or all of the desiderata for a formal model of phrase structure as set out above. In particular, the feature BAR with values 0, 1, 2, does no more than model the X$'$-theoretic distinction between X, X$'$ and XP, retaining all the problems with these notions discussed above.

　Our proposal goes beyond the basic assumptions in (13) in two major ways; the first of these will be discussed in this section, the second in §3.3. Firstly, we propose that a relatively minor alteration of the feature set seen in (13) is sufficient to license a model of phrase

structure which incorporates most of the desiderata set out above. We propose three l-structure features instead of two: CAT, which represents category labelling just as in (13); L, which intuitively represents the 'level' of any node, roughly corresponding in traditional terms to whether the node is a zero, one or two bar level node; and P, which intuitively represents the maximum projection level of the word/projection concerned.

(16)

$$\ast \xrightarrow{\quad \lambda \quad} \begin{bmatrix} \text{CAT} & \text{V} \\ \text{L} & 0/1/2 \\ \text{P} & 0/1/2 \end{bmatrix}$$

The values of L and P are integers, e.g. 0, 1, 2.[16] We assume that the value 2 is a sufficient maximum for English, but our formalization below does not enforce either a maximum or minimum value, meaning that if higher values are justified for some phrase types in some languages, or if some phrase types require only two values, 0 and 1 (for example because they lack specifiers), this will fall out unproblematically without further stipulation.

In order to make our proposal as clear as possible, we illustrate the l-structures we assume for the phrases *books*, *the book*, and *Bill's books*. However, the l-structure relations indicated here are not yet final, because we have not yet discussed our second innovation over (13); in order to simplify the presentation, we integrate that into our model separately, in §3.3.

The phrase *books* in the sentence *I read books* will have the following structure:

(17)

$$\ast \xrightarrow{\quad \lambda \quad} \begin{bmatrix} \text{CAT} & \text{N} \\ \text{L} & 0 \\ \text{P} & 0 \end{bmatrix}$$

books

As a phrase consisting of a single word, *books* is both maximal and minimal. In our system, the definition of a minimal projection is any node with the feature $\langle \text{L},0 \rangle$, while the definition of a maximal

---

[16] But see fn. 21.

projection is any node with the feature set $\{\langle L,n\rangle,\langle P,n\rangle\}$, that is any node whose L and P features have identical values. A node which is both maximal and minimal therefore has the feature set $\{\langle L,0\rangle,\langle P,0\rangle\}$

The phrase *the books* in the sentence *I read the books* will have the following (preliminary) structure:

(18)

$$
\begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 0 \\ \text{P} & 1 \end{bmatrix}
\qquad
\begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 1 \\ \text{P} & 1 \end{bmatrix}
\qquad
\begin{bmatrix} \text{CAT} & \text{N} \\ \text{L} & 0 \\ \text{P} & 0 \end{bmatrix}
$$

the   books
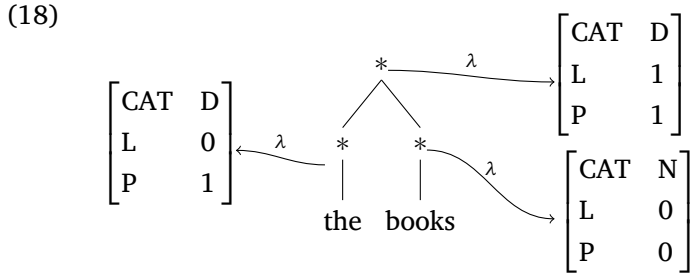
Once again, the noun *books* is both maximal and minimal as the noun phrase complement of D. The head D is a minimal projection, so has the feature $\langle L,0\rangle$, but it is not maximal. The maximal projection of the determiner phrase is the node that directly dominates the D head and the N complement. Since there are only two words in the phrase, we require only a single projection up from the preterminal nodes, just as in a BPS analysis. The maximal projection is one projection level up from the head; it therefore has the feature $\langle L,1\rangle$. As a maximal projection, its L and P values must be identical; it therefore also has the feature $\langle P,1\rangle$. The feature P represents the maximal projection level for the entire projection, and is shared by all nodes in the projection chain. Thus as the head of the determiner phrase, the head D must have the same P value as the maximal projection, meaning that it also has the feature $\langle P,1\rangle$.

Now consider the phrase *Bill's books*. Let us assume (purely for the sake of argument) that the possessive marker *'s* is a separate word which fills the head of the determiner phrase, and that *Bill* appears in the specifier of the determiner phrase.

(19)

$$
\begin{bmatrix} \text{CAT} & \text{N} \\ \text{L} & 0 \\ \text{P} & 0 \end{bmatrix}
\quad
\begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 0 \\ \text{P} & 2 \end{bmatrix}
\qquad
\begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 2 \\ \text{P} & 2 \end{bmatrix}
$$

$$
\begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 1 \\ \text{P} & 2 \end{bmatrix}
\qquad
\begin{bmatrix} \text{CAT} & \text{N} \\ \text{L} & 0 \\ \text{P} & 0 \end{bmatrix}
$$

Bill 's books
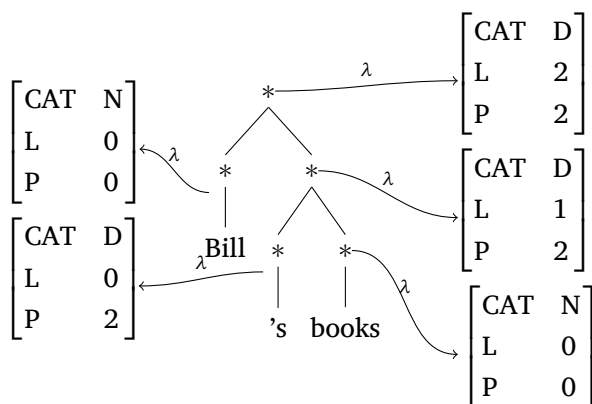
Once again the noun *books* is simultaneously maximal and minimal, and the same is true of the other noun in the phrase, *Bill*. But now the DP consists of three words, and thus necessarily has more structure. Since there is both a specifier and complement to D, the maximal projection is two projection levels higher than the head, and therefore has the feature set $\{\langle L,2\rangle,\langle P,2\rangle\}$. The head, as a minimal projection, has the feature $\langle L,0\rangle$, and since the maximal projection from the head has the feature $\langle P,2\rangle$, the head also has this feature. The intermediate node is one projection up from the head, and is part of a projection chain which extends two levels of projection above the head (i.e. which has the feature $\langle P,2\rangle$); the intermediate node therefore has the feature set $\{\langle L,1\rangle,\langle P,2\rangle\}$.

3.3   *Sets and distributive features*

Although the system illustrated in the previous section enables us to formalize an approach to phrase structure which eliminates non-branching dominance chains, and achieves several of the other desiderata set out above, it nevertheless incorporates a degree of redundancy, particularly as regards the CAT and P features. Essentially, in any projection chain the values for CAT and P for every node are identical, as e.g. with the three l-structures projected from the head, intermediate and maximal D projections in (19). It is possible to stipulate this identity, by means of constraints which require the head daughter of any node to have the same CAT and P values as its mother. But as discussed above, it would be preferable if the necessarily shared

properties of such nodes were shared as a natural consequence of the model (as in BPS), rather than by stipulation (as in X′ theory).

Happily, the LFG framework provides the mechanism we seek. L-structures are represented as attribute-value matrices, and just like f-structures, as discussed above, are understood in set-theoretic terms as sets of attribute-value pairs. It is also possible, and sometimes necessary, to assume sets of f-structures, that is sets of sets of attribute-value pairs. By extension, sets of l-structures are formally unproblematic.

Features (or attributes) interact with sets of f-structures in interesting ways, such that it becomes necessary to distinguish two types of features, *distributive* and *non-distributive* features. The need for this distinction has been most clearly demonstrated in relation to coordination and agreement; we therefore take a small detour to justify the difference between distributive and nondistributive features, before demonstrating their use for the present topic.

3.3.1 Agreement and (non)distributive features

Consider the following data, based on King and Dalrymple (2004):

(20) a. This boy and girl eat/*eats pizza.

   b. *These boy and girl eat/eats pizza.

   c. A boy and girl eat/*eats pizza.

   d. *This boy and girls eat/eats pizza.

In English, a single determiner can occur with two conjoined singular nouns, and in this case the determiner must be singular. Yet the verb agreement with such a subject phrase must be plural. In LFG, coordinated phrases are analysed at f-structure as a set, whose members are the f-structures of the individual coordinated phrases. It is also possible for sets to have their own features, independent of the f-structures they contain; for example, a conjunction provides a feature such as ⟨CONJFORM,AND⟩, but this feature is a feature of the whole conjoined phrase, not of either (or both) of the embedded phrases. So for the sentence *this boy and girl eat pizza* the f-structure will look something like this:

(21) This boy and girl eat pizza.

$$
\begin{bmatrix}
\text{PRED} & \text{'eat}\langle\text{SUBJ,OBJ}\rangle\text{'} \\[2pt]
\text{SUBJ} & s:\begin{bmatrix}
\text{SPEC} & \text{THIS} \\
\text{CONJFORM} & \text{AND} \\
\left\{\begin{array}{l} b:\begin{bmatrix}\text{PRED} & \text{'boy'}\end{bmatrix} \\ g:\begin{bmatrix}\text{PRED} & \text{'girl'}\end{bmatrix}\end{array}\right\}
\end{bmatrix} \\[2pt]
\text{OBJ} & \begin{bmatrix}\text{PRED} & \text{'pizza'}\end{bmatrix}
\end{bmatrix}
$$
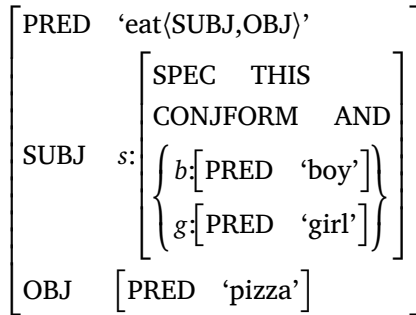
The structure labeled s is a *hybrid set*: it is a set containing both individual attribute-value pairs (features) and f-structures. The representation of s, with square brackets enclosing the features and braces enclosing the f-structures, is potentially misleading: it is not the case that the set of f-structures $\{b, g\}$ is contained within and distinct from s, but the square brackets and braces together identify the hybrid set s, which contains four elements, two features ($\langle\text{SPEC,THIS}\rangle$ and $\langle\text{CONJFORM,AND}\rangle$), and two f-structures (b and g).

In order to deal with the simultaneously singular and plural agreement of the conjoined noun phrase, King and Dalrymple (2004) adopt the proposal of Wechsler and Zlatić (2003) that there are actually two types of agreement feature for nouns: CONCORD and INDEX features. Informally, CONCORD is more morphological, and is generally relevant for agreement between nouns and their immediate specifiers and modifiers (e.g. determiners and adjectives). On the other hand, INDEX is more semantic, and is relevant for agreement outside the noun phrase, e.g. verb agreement.

Singular *this*, *boy* and *girl* specify both their CONCORD NUM and INDEX NUM as SG, while plural *these*, *boys* and *girls* specify their CONCORD NUM and INDEX NUM as PL. This is sufficient to account for the grammaticality/ungrammaticality of *this boy/these boys/*this boys/*these boy* etc. But to account for the grammaticality of *this boy and girl*, and the ungrammaticality of **these boy and girl*, we now require the distinction between distributive and nondistributive features. Distributive features are defined as follows (Dalrymple and Kaplan 2000):

(22) If a is a distributive feature and s is a set of f-structures, then

(s a = v) holds iff (f a) = v for all f-structures f which are members of s.

Informally, a nondistributive feature may hold of a set of f-structures (making the set a hybrid set) independently of whether it holds of each or any of the members of that set. In contrast distributive features cannot hold of a set independently, but must hold for every member of the set. If CONCORD agreement features are *distributive*, then any CONCORD feature specified of a set must hold of all f-structures within that set. So when *this* conjoins two nouns, and hence maps to a set of f-structures, its specification (↑ CONCORD NUM) = SG holds only if all f-structures within the set have the feature ⟨CONCORD NUM,SG⟩.

(23) This boy and girl eat pizza.

$$
\begin{bmatrix}
\text{PRED} & \text{`eat⟨SUBJ,OBJ⟩'} \\[2pt]
\text{SUBJ} \quad s: & \begin{bmatrix}
\text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`this'} \end{bmatrix} \\[4pt]
\text{CONJFORM} & \text{AND} \\[6pt]
\left\{
\begin{array}{l}
b: \begin{bmatrix} \text{PRED} & \text{`boy'} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix} \\[12pt]
g: \begin{bmatrix} \text{PRED} & \text{`girl'} \\ \text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}
\end{array}
\right\}
\end{bmatrix} \\[8pt]
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`pizza'} \end{bmatrix}
\end{bmatrix}
$$
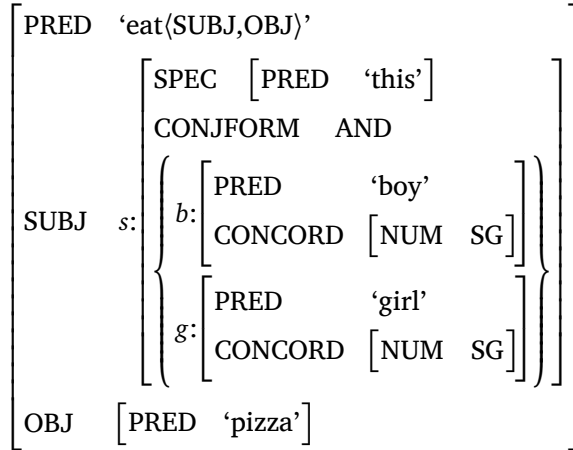
Correspondingly, *these boy and girl* is ruled out because *these* will require every member of its set to have the feature ⟨CONCORD NUM,PL⟩, which will not be compatible with the singular concord of the nouns. Singular or plural determiners with nouns of mismatched number, e.g. *this boy and girls*, are also ruled out, since the definition of distributivity requires *every* member of the set to have the same feature.

As for verb agreement, this depends on INDEX. INDEX is a *nondistributive* feature. Any non-3SG present tense verb specifies that the value of its SUBJ INDEX NUM is PL, or else that the value of its SUBJ

PERS is not 3; only the first disjunction is relevant here. If the subject is an ordinary, non-conjoined noun phrase, then the noun must be plural (since plural nouns specify their INDEX NUM as PL, while singular nouns specify it as SG, as discussed above). If the subject is a set, then the feature ⟨INDEX NUM,PL⟩ must hold of the set, but *need not* hold of any of the members of the set. Thus s has the feature ⟨INDEX NUM,PL⟩, which is different from the INDEX NUM feature of the members of s. This is exactly what we require to account for sentences like (20a):

(24) This boy and girl eat pizza.

$$
\begin{bmatrix}
\text{PRED} & \text{'eat}\langle\text{SUBJ,OBJ}\rangle\text{'} \\[4pt]
\text{SUBJ} \quad s: & \begin{bmatrix}
\text{SPEC} & \begin{bmatrix} \text{PRED} & \text{'this'} \end{bmatrix} \\[4pt]
\text{CONJFORM} & \text{AND} \\[4pt]
\text{INDEX} & \begin{bmatrix} \text{NUM} & \text{PL} \end{bmatrix} \\[4pt]
\left\{ \begin{array}{l}
b: \begin{bmatrix}
\text{PRED} & \text{'boy'} \\
\text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\
\text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix}
\end{bmatrix} \\[12pt]
g: \begin{bmatrix}
\text{PRED} & \text{'girl'} \\
\text{CONCORD} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix} \\
\text{INDEX} & \begin{bmatrix} \text{NUM} & \text{SG} \end{bmatrix}
\end{bmatrix}
\end{array} \right\}
\end{bmatrix} \\[4pt]
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'pizza'} \end{bmatrix}
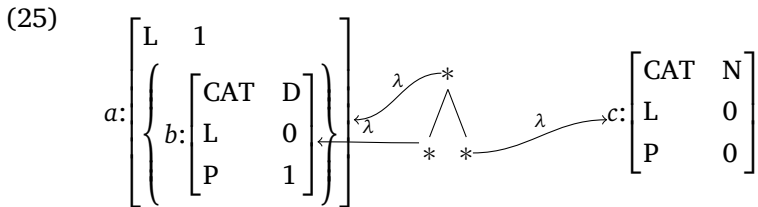\end{bmatrix}
$$

3.3.2 Back to phrase structure

How does the difference between distributive and nondistributive features help with modelling projection chains? Although, in coordination, sets of f-structures are necessarily sets of more than one f-structure, it is of course also possible to have singleton sets, i.e. sets containing a single member.[17] Now if a distributive feature applies to an f-structure, or l-structure, which is a singleton member of a set, that feature necessarily holds of the set as well. Likewise, if a distribu-

---

[17] This is a regular outcome in LFG analyses of adjunction.

tive feature is specified of a singleton set, it necessarily holds of the member of that set.[18]

Now let us revisit the projection structure for the phrase *the books*. In (18) we treated the three l-structures projected from the three nodes as structurally independent of each other. But now let us assume that in any projection chain the l-structure of the head daughter is contained within the l-structure of the mother, the mother's l-structure therefore being a hybrid set. The intuition we are trying to model is that CAT and P values are necessarily identical for any node in a projection chain.[19] If projection chains are modelled using set inclusion, then we can achieve the desired outcome simply by defining the relevant features as distributive. So instead of (18), we now propose:

(25)

$$
a: \begin{bmatrix} \text{L} & 1 \\ \left\{ b: \begin{bmatrix} \text{CAT} & \text{D} \\ \text{L} & 0 \\ \text{P} & 1 \end{bmatrix} \right\} \end{bmatrix} \quad c: \begin{bmatrix} \text{CAT} & \text{N} \\ \text{L} & 0 \\ \text{P} & 0 \end{bmatrix}
$$

That is, if CAT and P are distributive features, and if the l-structure for any head daughter is a member of the (hybrid, singleton) set that constitutes the mother l-structure, then CAT and P features are *necessarily* shared between any mother and head daughter. This means we require no stipulation to ensure that, say, a head of category D projects a phrase of category D: the distributive nature of the CAT feature and the nature of l-structure inclusion enforces this. The feature L, of course, must be defined as nondistributive, since mothers and daughters in a projection chain may have different values for this feature. Set inclusion can be recursive, so the principles illustrated in

---

[18] Recently, Andrews (2018) has explored the potential of singleton hybrid sets at f-structure for dealing with long-standing problems of scope in LFG, and our proposal is inspired by his work.

[19] We do not address coordination in this paper, but note that coordination of unlike categories is unproblematic, as we do not need to assume that set inclusion holds between coordinated nodes and their mother. To deal with unlike categories will require a more complex representation of categories, such as that proposed by Dalrymple (2017), which is entirely compatible with the model proposed here.

(25) will equally well account for a phrase which projects two levels (or more) above the head, as in *Bill's books*:

(26)
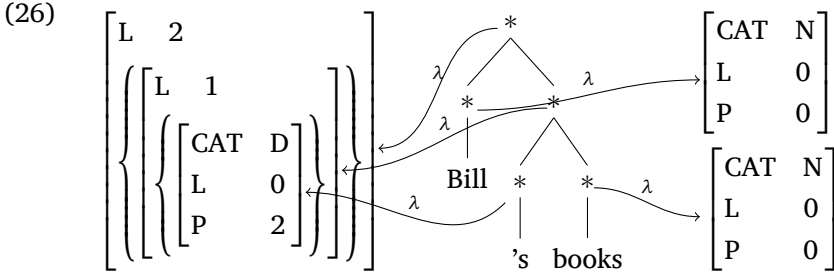
3.4   *Phrase structure rules and templates*

In the previous section we showed the desired outcome of our model. Now the question is how to state the relevant constraints which will realise that model. The constraints which derive l-structure values are realised as functional descriptions on PSRs and in lexical entries, i.e. the standard locus of constraints in LFG.

We require a fixed number of f-descriptions to model l-structure, which occur in different combinations in different contexts; in order to generalize over multiple instances of these f-descriptions, we define them as *templates* (Dalrymple *et al.* 2004; Asudeh *et al.* 2013); templates function like macros, allowing the same combinations of f-descriptions to be applied together wherever appropriate. For example, some projections require that the L and P values for a particular node are identical (i.e. a maximal projection), others require that the L value for a particular node is identical to the mother node's L value. We assume the following basic templates:[20]

---

[20] These templates use an alternative representation for projection functions from that introduced above: $*_\lambda$ is the same as $\lambda(*)$.

(27) Basic templates:

a. l-structure inclusion		LSTRIN $\equiv *_\lambda \in \hat{*}_\lambda$
b. Maximal phrase		LP $\equiv (*_\lambda$ L) $= (*_\lambda$ P)
c. Mother node is a maximal phrase		LPM $\equiv (\hat{*}_\lambda$ L) $= (\hat{*}_\lambda$ P)
d. L of node $=$ L of its mother		LUD $\equiv (\hat{*}_\lambda$ L) $= (*_\lambda$ L)
e. L of mother node $= 1$		LIM $\equiv (\hat{*}_\lambda$ L) $= 1$
f. L is one less than L of mother		LDOWN $\equiv (*_\lambda$ L) $= (\hat{*}_\lambda$ L) -1
g. L $= 0$		LO $\equiv (*_\lambda$ L) $= 0$
h. L of mother node $= 0$		L LOM $\equiv (\hat{*}_\lambda$ L) $= 0$
i. Mother node has a P value		PXM $\equiv (\hat{*}_\lambda$ P)
j. Node does not have a P value		PNX $\equiv \neg(*_\lambda$ P)
k. Mother does not have P value		PNXM $\equiv \neg(\hat{*}_\lambda$ P)

The first template here, LSTRIN, defines the l-structure inclusion relation: the l-structure of the current node is a member of the l-structure of the mother of the current node (the latter l-structure by consequence therefore being a set). Other templates refer directly to L and P values: they either specify that two features have the same value, or specify an absolute or relative value for a particular feature, or state existential constraints on the feature P.

The template LDOWN specifies a relative value for L: the value of L of the current node is one less than the value of L of the mother node. This crucial template is what drives the increase/decrease of L values up/down a projection chain. Note that technically natural numbers play no role in the LFG formalism; feature values like 0, 1, 2, are symbols, not natural numbers, so mathematical statements like L -1 are not strictly possible. It is, however, unproblematic to formalize addition/subtraction using the successor function, and we retain the mathematical statement as in (27f) for readability.[21]

The constraint in (27i) requires the feature P to exist in the l-structure of the mother node; PNX requires that P does not exist as a feature of the l-structure of the current node, and PNXM requires

---

[21] In Lovestrand (2018, 153) the @LDOWN template is defined as: @LDOWN $\equiv (\hat{*}_\lambda$ L PLUS) $= (*_\lambda$ L). In this approach, the value of L is either 0 or an attribute-value matrix with the attribute PLUS. In the l-structure, what is informally represented as the number 1 is formally represented as [L [PLUS 0]], the informal number 2 is formally [L [PLUS [PLUS 0]]], and so on.

the same of the mother's l-structure. These existential constraints are required to account for nonprojecting categories, as discussed in §3.6.

The constraints in (27) are the only constraints needed to model the phrase structure of natural language. Given these, and only these, constraints, certain features of the system fall out unproblematically. For example, in our system, intuitively, for any l-structure the value of L is never greater than the value of P: $\forall *_\lambda$, $P \geq L$. Given only the templates in (27), an l-structure that violates this intuitive general constraint cannot be generated, so the constraint need not be independently stated.

Common phrase structure positions require particular combinations of the constraints in (27). We therefore define further templates for convenience, which call combinations of the templates in (27).

(28) Complex templates:
a. Head of an endocentric projection: HEADX ≡ @LDOWN ∧ @LSTRIN
b. Head of an adjunction structure: HEADA ≡ @LUD ∧ @LSTRIN
c. Specifier or adjunct: EXT ≡ @LPM ∧ @LP
d. Complement: INT ≡ @LIM ∧ @LP
e. Non-projecting node: NONPRJ ≡ @LO ∧ @PNX
f. Non-projecting mother: NONPRJM ≡ @LOM ∧ @PNXM
g. Projecting mother: PRJM ≡ @LOM ∧ @PXM

HEADX applies to heads in specifier and complement structures, HEADA applies to heads in adjunction structures. EXT and INT apply to specifier/adjunct phrases and complement phrases respectively. We can now rewrite the standard schematic PSRs of X′ theory in our system:

(29) Schematic phrase structure rules:

a.	Specifier rule:	∗	→	∗	,	∗
				@EXT		@HEADX
b.	Complement rule:	∗	→	∗	,	∗
				@HEADX		@INT
c.	Adjunction rule:	∗	→	∗	,	∗
				@HEADA		@EXT

Notice the generality of these rules with respect to category sharing. There is no need for category label to be specified on the left-hand

side of a rule (or indeed on the right-hand side), because the category of the mother automatically follows from the category of the head daughter (by the constraint LSTRIN called by the templates HEADX and HEADA). In other words, once the head of an endocentric structure is identified by its template, there is no further need to stipulate what the category of the mother node is. However, this differs from exocentric structures, where the category of the mother node may need to be specified as an additional constraint on one of the daughters. Given this explicit formal restriction on the category of the mother node in our approach, the left-hand side of traditional PSRs, and the arrow, are redundant; we could equally well rewrite (29) as:[22]

(30)   Schematic phrase structure constraints:

     a.   Specifier structure: [*   ,   *   ]
                                  @EXT   @HEADX

     b.   Complement structure: [*   ,   *   ]
                                  @HEADX   @INT

     c.   Adjunction structure: [*   ,   *   ]
                                  @HEADA   @EXT

Such a representation accords more closely with the constraint-based conception of LFG, which interprets PSRs not as procedural rules, but as constraints on possible structures.

3.5   *Example*

As an illustration of our model, we give the necessary phrase structure constraints and lexical entries to derive the sentence *Bill read a book of poems*. In these constraints we specify category labels on the right-hand side in the traditional way, but this is to be understood as a shorthand for an f-description defining the CAT value of the relevant node's l-structure.

---

[22] The square brackets in (30) serve to indicate the left and right edges of the relevant constituents.

(31) Phrase structure constraints:

a. [{N|D}   {   I | V   }]
       (↑SUBJ) =↓   ↑ =↓   ↑ =↓
         @EXT   @HEADX   @INT
                              (λ(∗) CAT) = I

b. [V   {N|D}   ]   d. [N   P   ]
         ↑ =↓   (↑OBJ) =↓   ↑ =↓   (↑OBL) =↓
       @HEADX   @INT   @HEADX   @INT

c. [D   N   ]   e. [P   {N|D}   ]
         ↑ =↓   ↑ =↓   ↑ =↓   (↑OBJ) =↓
       @HEADX   @INT   @HEADX   @INT

The constraint in (31a) equates to a traditional specifier rule for IP; it is formulated so as to license optionality of the functional head I (notice that optionality is not a default, but has to be specifically licensed in this way). The constraint in (31b) equates to the complement rule for VP; that in (31c) equates to the complement rule for DP; (31d) is the complement rule for NP, and (31e) is the complement rule for PP. [23]
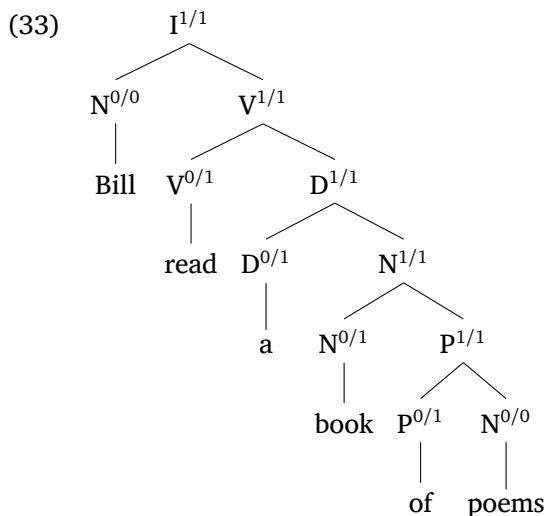
The PSRs in (31), together with the lexical entries in (32), produce the phrase structure in (33). Although we understand the features CAT, L and P as features within the 'l-structure' projected from

---

[23] Note that we adopt a simplified approach to category labels in this paper, treating N and D as fully distinct labels, but the rules provided here imply a more sophisticated approach, following e.g. Grimshaw (1991) and Bresnan (2001). We assume that in fact N and D share the same category label N, but are distinguished in terms of another feature ±F. The value of F may be specified in a given rule or not; so in (31a), {N|D} is really to be understood as N with underspecified value for F; but in (31c), which constrains the structure within a determiner phrase, the +F value of the head, and the −F value of the non-head, are crucial elements of the rule. The underspecification of certain nodes improves the resulting analyses by eliminating the need for certain nonbranching nodes. For example, with the subject position in (31a) underspecified, both *This* and *Bill* can serve as single word subject phrases requiring only a single c-structure node, $D^{0/0}$ in the former case, $N^{0/0}$ in the latter. For the present purposes, so as not to further complicate our presentation, we abstract away from the details of this, and present our analysis as though N and D are fully distinct categories, modelling the underspecification via optionality.

a node, for ease of representation in trees such as (33), we propose an abbreviatory notation whereby L and P values are shown as superscripts on category node labels. Each node, represented by its category label, appears with superscript numbers separated by a slash. The first number represents the L value, the second the P value for that node. So, a node which has the features $\langle\text{CAT,V}\rangle$, $\langle\text{L,0}\rangle$ and $\langle\text{P,1}\rangle$, will be represented as $V^{0/1}$.

(32) Lexical entries:
  a. Bill: N
              $(\uparrow\text{PRED})$ = 'Bill'
              @PRJM
  b. book: N
              $(\uparrow\text{PRED})$ = 'book$\langle\text{OBL}\rangle$'
              @PRJM
  c. poem: N
              $(\uparrow\text{PRED})$ = 'poem'
              @PRJM
  d. read: V
              $(\uparrow\text{PRED})$ = 'read$\langle\text{SUBJ,OBJ}\rangle$'
              @PRJM
  e. a: D
              $(\uparrow\text{SPEC})$ = 'a'
              @PRJM
  f. of: P
              $(\uparrow\text{PRED})$ = 'of$\langle\text{OBJ}\rangle$'
              @PRJM

(33)

$$I^{1/1}$$

Tree structure:

- $I^{1/1}$
  - $N^{0/0}$
    - Bill
  - $V^{1/1}$
    - $V^{0/1}$
      - read
    - $D^{1/1}$
      - $D^{0/1}$
        - a
      - $N^{1/1}$
        - $N^{0/1}$
          - book
        - $P^{1/1}$
          - $P^{0/1}$
            - of
          - $N^{0/0}$
            - poems

L and P values are determined 'bottom up'. So *poems* attaches to a node $N^{0/0}$, since there are no higher levels of projection in this phrase. In contrast, *book* attaches to a node $N^{0/1}$, since there is one level of projection above the head; the word *read* attaches to a node $V^{0/1}$, since there is one level of projection within the verb phrase. The L value is determined from the bottom up, with all words specifying L = 0 of their preterminal node. The head of an X′-theoretic projection is associated with the template LDOWN (via the template HEADX), meaning that every mother node in a headed projection chain has L value one greater than that of its head daughter.

The P value is determined by the number of projection levels in the phrase. All maximal projections are associated with the template LP, meaning that the P value for every maximal node is identical to the value of L for that node. So in a two level projection, where the preterminal head daughter has the feature $\langle L,0 \rangle$ and the mother therefore has the feature $\langle L,1 \rangle$ (by LDOWN), the value of P for the mother node will be the same as its L feature, i.e. 1. The inclusion relation specified for the l-structures of heads in a projection chain ensures that all nodes in any projection chain automatically and necessarily share the same value for CAT and P, as discussed above.

Regarding the top node, the constraint in (31a) licenses an I node with specifier and complement, but no head, daughters. This mod-

els the headless[24] IP structure standardly assumed in LFG for clauses without auxiliaries, but without requiring nonbranching projections. Only maximal projections (L = P) can have specifier daughters (as constrained by the template EXT); only nodes with the feature L = 1 can have complement daughters (as constrained by the template INT); the top node must therefore be $I^{1/1}$, satisfying both constraints simultaneously.[25]

3.6 *Dealing with nonprojecting categories*

As discussed in §2.2, no existing formalization of phrase structure adequately accounts for the existence of nonprojecting categories. Following Arnold and Sadler (2013), we model the difference between prenominal and nonprenominal adjectives in English in these terms: prenominal adjectives, which cannot take complements or other postmodifiers, and hence appear not to be able to head full phrases, are treated as nonprojecting adjectives, while adjectives in other positions (predicative or predicated) can head full phrases and so are projecting. Many adjectives in English can appear in both prenominal and other positions, e.g. *small*, and for such cases we assume that the grammar licenses both variants; some adjectives are restricted to one or the other position, however, and we analyse this by assuming that such adjectives have only nonprojecting (e.g. *former*), or only projecting (e.g. *asleep*), variants.

(34) a. The small dog eats biscuits.

b. The dog is small.

c. The former president eats biscuits.

d. *The president is former.

---

[24] Headless, but not exocentric, as the IP serves as the extended projection of the V.

[25] There is a partial parallel here between our approach and the exocentric treatment of CP by Jayaseelan (2008) and Putnam and Stroik (2009, 2010); our treatment of headless CP structures, which we do not have space to discuss here, would fully parallel the approach to headless IPs set out here, and would thus be very close to these exocentric treatments of CP. An alternative to the headless IP assumed here would be to adopt the older analysis of an exocentric clausal node S, as still assumed e.g. in HPSG (Pollard and Sag 1994) and in LFG by Bresnan *et al.* (2016).

e.   *The asleep dog eats biscuits.

f.   The dog is asleep.

Our model fully captures the grammaticality judgments in (34). The l-structure feature set of a nonprojecting category must be fully distinguishable from the possible feature sets available to projecting categories. For example, one might think that as a necessarily minimal and maximal projection, a nonprojecting category should necessarily have the features $\{\langle L,0\rangle, \langle P,0\rangle\}$ (as is assumed for clitics within BPS). However, this is a possible feature set for projecting words, whenever they happen to appear alone constituting a phrase. We must therefore allow adjectives in predicate position to have the features $\{\langle L,0\rangle, \langle P,0\rangle\}$, so this feature set cannot be attributed to nonprojecting adjectives, otherwise we would not be able to prevent e.g. *former* from appearing in predicate position.

We propose that as necessarily minimal projections, nonprojecting adjectives have the feature $\langle L,0\rangle$, but that as categories which necessarily do not project, they have no value for the feature P. This is the purpose of the templates PNX and PNXM in (27j–k). The lexical specification for a nonprojecting word includes the template PNXM (called by NONPRJM), which ensures that the preterminal c-structure node dominating the word lacks the feature P. The template PNX appears in PSRs (called by NONPRJ) on nodes which are restricted to nonprojecting categories.

We thus assume the following lexical entries for *small*, *former*, and *asleep*:

(35)   Lexical entries:

a.   small   Adj
            (↑PRED) = 'small'
            { @PRJM | @NONPRJM }

b.   former   Adj
            (↑PRED) = 'former'
            @NONPRJM

c.   asleep   Adj
            (↑PRED) = 'asleep'
            @PRJM

The constraints in (36) license predicate adjectives and prenominal adjectives. (36a) defines a standard complement structure, and therefore the Adj complement has the specification LP (called by @INT), meaning that nonprojecting adjectives cannot stand in predicate position. (36b) requires that a prenominal adjective lack a feature P, thereby restricting the prenominal position to nonprojecting adjectives. A tree illustrating a noun phrase with nonprojecting adjective is given in (37).

(36) a. Predicate adjective: [I   Adj   ]
                                      @HEADX   @INT

     b. Non-proj. adjunction: [Adj   N   ]
                                        @LOM   @HEADA
                                        @NONPRJ

(37)

$$D^{1/1}$$

$D^{0/1}$ $N^{0/0}$

the $Adj^{0/}$ $N^{0/0}$

small   dog

3.7   *Exocentric categories*

As discussed in §2.2, exocentric projections are another widely accepted possibility in LFG which have nevertheless never been adequately formalized within a theory of phrase structure.

Our proposal enables a neat and insightful analysis of exocentricity. For the purposes of illustration, we adopt the analysis of Welsh proposed by Sadler (1997), which involves the following basic clause structure (stated in traditional, X′-theoretic, terms):

(38)   IP

     I   S

         NP   VP

The clause-initial finite verb, often an auxiliary, appears in I, and the complement of I is an exocentric phrase which includes both the

subject phrase (the NP in (38)) and the VP (often containing the lexical verb, and any object, etc.).

In our model, S will be licensed as a complement daughter of I; the functional constraints placed on S in the PSR will be fully parallel to those placed on any other complement, so the template INT will apply to the S node:

(39)   [I   S   ]
                ↑=↓   ↑=↓
            @HEADX   @INT

The template INT calls the templates LIM and LP. The first specifies the value of L for the mother node, while the second requires that the values of L and P for the S node be identical. Now consider the rule that introduces the daughters of S. Since S is exocentric, no daughter of S is the head, nor is any daughter a specifier, a complement, or even an adjunct; therefore none of the standard endocentric templates above apply to any of the daughter nodes. The daughters of S may themselves be specified as necessarily projecting, but no daughter need make any specification about the L/P values of S.

(40)   [N   V   ]
              @LP   @LP
                      $(*_\lambda$ CAT$) = $ S

When we try to construct a tree based on these rules parallel to (38), it is impossible to assign values to S for its L and P features. As a complement of I, S must satisfy the requirement L=P, and in the absence of specific values, this can only be satisfied if neither value exists. That is, we get the following:

(41) $I^{1/1}$

        $I^{0/1}$ S
          |
        aux $N^{0/0}$ $V^{0/0}$
                     | |
                 subject   verb

Since S is an exocentric category, its daughters lack the typical endocentric specifications introduced above. The result is that S lacks

L/P values. Since these features are used to define and constrain endocentric projections, we take this to be an intuitive definition of exocentricity: exocentric categories lack L/P features.[26]

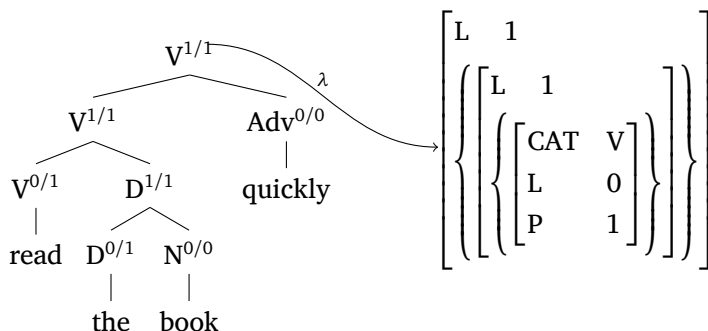3.8   *Comparison with traditional X′ theory*

In (1) we gave seven desiderata for a formal model of phrase structure. All seven are achieved by our model. The use of two features with numerical values, L and P, enable us to define PSRs in such a way that nonbranching chains are eliminated: a node can be both maximal and minimal at the same time, and more complex phrases have only the nodes required to model constituency. Our proposal also eliminates the need for default optionality in PSRs, as standardly assumed in LFG. Standard LFG takes optionality to be a default, because in any projection heads (particularly, but not only, functional heads), specifiers and complements may be absent. In our model, however, optionality is an exception rather than the rule: if a phrase lacks a complement and/or specifier, the PSRs introducing those positions are simply not utilized, and a simpler structure results.

      Our definition of a maximal phrase, $L = P$, avoids the problem raised by Fukui (1986) regarding the ambiguity of the label 'XP', since a maximal phrase may be e.g. $X^{0/0}$, $X^{1/1}$ or $X^{2/2}$. At the same time, our approach to projection chains, involving inclusion of l-structures, avoids the ambiguity between X^{max} and the top node of a projection chain noted by Hornstein and Nunes (2008). Consider the following VP:

---

[26] This conclusion is not without complication. According to the formulation of Kaplan and Bresnan (1982), a constraint such as $L = P$ is violated if neither L nor P are assigned a value, and thus a derivation based on (39) and (40) would fail. On the other hand, in the computational implementation of LFG, XLE (Crouch *et al.* 2011), the constraint $L = P$ is satisfied if $\neg L \wedge \neg P$. The rationale for Kaplan and Bresnan's approach is not clear to us; it has been suggested to us (Adam Przepiórkowski, p.c.) that the theory could be unproblematically emended to fall in line with XLE, and we adopt that emendation here.

(42)

$$
\begin{array}{c}
\text{V}^{1/1} \\
\diagdown \lambda \\
\text{V}^{1/1} \qquad \text{Adv}^{0/0} \\
\text{V}^{0/1} \quad \text{D}^{1/1} \qquad \text{quickly} \\
\text{read} \quad \text{D}^{0/1} \ \text{N}^{0/0} \\
\text{the} \quad \text{book}
\end{array}
\qquad
\begin{bmatrix}
\text{L} & 1 \\
\left\{ \begin{bmatrix} \text{L} & 1 \\ \left\{ \begin{bmatrix} \text{CAT} & \text{V} \\ \text{L} & 0 \\ \text{P} & 1 \end{bmatrix} \right\} \end{bmatrix} \right\}
\end{bmatrix}
$$

Both the top node of the VP and its head daughter are maximal nodes in the sense defined above (being $\text{V}^{1/1}$), but the top node is distinct (and therefore distinguishable), because its l-structure alone is not included within another l-structure. Thus the top node in any projection satisfies the equations $(*_\lambda \text{ L}) = (*_\lambda \text{ P})$ and the negative existential constraint $\neg(\in *_\lambda)$ (in words: there is no set of which my l-structure is a member), whereas other maximal nodes in a projection satisfy only the first.

Our proposal also lacks any distinct notion comparable to X′. Suppose we wanted to define adjunction to X′, i.e. adjunction of phrases closer to the head than any specifiers, but further from the head than any complements. A head which has a complement is, in our system, either 0/1 or 0/2 (depending on whether there is also a specifier). So nodes with the L/P values 0/1 and 0/2 must be excluded from the set of nodes to which X′ adjuncts could adjoin. But a head which has a specifier, but no complement, and to which we might therefore wish to permit X′ adjunction, will in our system be 0/1. Thus 0/1 nodes sometimes correspond to the size of an X′ and sometimes do not. Therefore a notion equivalent to X′ adjunction is unformalizable in our system, because there is no coherent set of L/P values which correspond to the traditional notion of X′.

Given our set-inclusion approach to projection chains, our model also reduces redundancy in category labelling and in specification of P values; that a phrase necessarily has the same values for CAT and P as its head is a necessary consequence of the model, requiring no additional stipulation. As shown in the previous sections, our model also affords principled accounts of nonprojecting categories and exocentric categories, which are lacking in existing formalizations of phrase

structure.

4   OTHER PROPOSALS

In this section we discuss three alternative formalizations of phrase structure, two within LFG and one within Minimalism. These approaches are simpler alternatives to the model presented above, in the sense that they use fewer formal features. However, their relative simplicity comes at the cost of failing to meet the theoretical desirata laid out in (1), and incomplete coverage of attested phrase structures types.

4.1   *Bresnan (2001)*

We take Bresnan (2001; unmodified in Bresnan *et al.* 2016) as representative of standard assumptions regarding the formal properties of phrase structure in LFG.[27] Bresnan (2001, 100) describes the formal properties of c-structure nodes thus: "Formally, X' categories can be analyzed as triples consisting of a categorical feature matrix, a level of structure, and a third, privative feature F, which flags a category as 'function' (F) or unspecified as to function (lexical)." The "level of structure" feature, which we call BAR following Andrews and Manning (1999), has three values: 0, 1, 2. These digits each correspond to a level of structure which is represented notationally using the traditional X-bar symbols: X^0 for [BAR 0], X' for [BAR 1], and XP for [BAR 2]. The use of integers in this context implies that, in an endocentric projection, a mother must have a BAR value higher than its daughter. The question of dominance is not discussed formally by Bresnan, but the familiar templatic description of X-bar principles (43) makes it clear that some undefined, and presumably stipulatory, mechanism is intended to enforce the dominance sequence.

(43)   a. **Specifier phrase structure rule**
           $XP \rightarrow X'$, YP

     b. **Complement phrase structure rule**
           $X' \rightarrow X^0$, ZP

---

[27] Bresnan's decomposition of syntactic categories builds on earlier work, e.g. King (1995).

This model is distinctly simpler than the model proposed in this paper, but it has a number of shortcomings. Most obviously, the use of a single numerically valued feature to model projection levels means that Bresnan's proposal is essentially a formalization of X′ theory, and thus it inherits all the failings of X′ theory. There is no principled distinction between XP and X^{max} (see Lovestrand and Lowe 2017, 288–289), nor betwen X^{max} and the top node of a projection; there is an independent notion corresponding to X′ (a node with ⟨BAR,1⟩); optionality is necessarily the default in PSRs, and by consequence nonbranching dominance chains are widespread.

Bresnan (2001, 91) realises that nonbranching dominance chains are unsatisfactory, and proposes a derivational process which Dalrymple *et al.* (2015) call "X′ Elision", to 'prune' unnecessary nodes from a well-formed c-structure so that it is as small as possible. As a derivational process this 'X′ elision' is not well integrated into the constraint-based assumptions of LFG, and although it does for the most part give the right results, it is preferable to avoid generating unnecessary nodes in the first place, as in our model, rather than generating them and then eliding them.

Bresnan's model also does not avoid redundancy in category labelling, and provides no formal account of exocentric or nonprojecting categories; the latter are not admitted in Bresnan (2001). They are admitted in Bresnan *et al.* (2016), but with no formal integration into the theory of phrase structure, which is unchanged from Bresnan (2001). There is no value of BAR which would both capture the minimality of nonprojecting categories and would not also render them indistinguishable from zero level categories. These issues are discussed further in Lovestrand and Lowe (2017).

4.2   *Marcotte (2014)*

Marcotte (2014) simplifies Bresnan's (2001) model by removing reference to bar levels, and thus requiring fewer theoretical primitives. Marcotte (2014, 417) explicitly likens his proposal to Chomsky's (1995) BPS; we therefore provide a detailed comparison of this proposal with our own.

While Marcotte's proposal can be said to reduce the number of formal devices needed to account for c-structure nodes, there are several syntactic structures that it cannot account for, and it does not

meet all the desiderata set out in (1).

4.2.1   Marcotte's proposal

Marcotte's proposal is to remove the BAR feature, and to instead define the relationships between nodes in terms of dominance relationships and shared category features. Marcotte proposes to label what we (following Kaplan 1989) have called l-structure as "x-structure", and assumes a function χ from nodes to x-structures, equivalent to our (and Kaplan's) λ. The function $\mathscr{M}$ is the function that relates the daughter node to its mother; as usual $*$ represents the node in question, and *n* represents any other node. There are three basic definitions of types of nodes.[28]

(44) **Marcotte's (2014) "Bare phrase structure for Lexical-Functional Grammar"**

    a. PROJECTING NODE: A node projects iff its x-structure is identical with its mother's x-structure.
       Proj$(*) \iff \chi(*) = \chi(\mathscr{M}(*))$

    b. MAXIMAL PROJECTION: A node is a maximal projection iff it is not a projecting node.
       Max$(*) \iff \neg$*Proj*$(*)$

    c. TERMINAL: A node is a terminal iff no node has it as a mother.
       Term$(*) \iff \neg\exists n.\mathscr{M}(n) = *$

In this system, there are four types of nodes, roughly equivalent to X^0, X', XP and $\hat{X}$. A projecting head (roughly equivalent to an X^0) is a node that meets the definitions of PROJECTING NODE (it has the same category as its mother) and the definition of TERMINAL (it is not the mother of any node).[29] A maximal projection (roughly equivalent to XP) is any node that meets the definition of MAXIMAL PROJECTION (it does not have a mother with identical features), and is not a TERMINAL. Intermediate nodes (roughly X') meet the definition of

---

[28] Marcotte's model is very similar to that of Speas (1986), cf. the definition of maximal projection in fn. 13.

[29] For Marcotte, the lexical information at the bottom of the tree is not considered a node, so his 'terminal' nodes are equivalent to what we (and Bresnan 2001) call preterminal nodes.

PROJECTING NODE, but do not meet the definition of TERMINAL (it is the mother of another node). A nonprojecting node (roughtly $\hat{X}$) is a node that is both a MAXIMAL PROJECTION and a TERMINAL.

Marcotte applies his approach to c-structure to the structure-function principles. He provides definitions for where it should be expected to find nodes that are functional co-heads with their sister (annotated as $\uparrow = \downarrow$), and definitions for where we should expect to find subjects, objects, obliques and possessors. Notably absent is a definition of adjuncts.[30]

(45) **Marcotte (2014) "Endocentric c- to f-structure mappings"**

    a. A projecting node shares the f-structure of its mother:
   $Proj(*) \Longrightarrow \uparrow = \downarrow$

    b. A SUBJ is a DP daughter of IP:
$$\begin{matrix} Max(*) & Max(\mathcal{M}(*)) \\ \chi = D & \chi(\mathcal{M}(*)) = I \end{matrix} \Longrightarrow (\uparrow SUBJ) = \downarrow$$

    c. An OBJ is a DP with a V(P) or P(P) mother:
$$\begin{matrix} Max(*) \\ \chi = D \end{matrix} \quad \{V \mid P\} \sqsubseteq \chi(\mathcal{M}(*)) \Longrightarrow (\uparrow OBJ) = \downarrow$$

    d. An OBL is a non-verbal/adjectival XP with a non-functional mother:
   $Max(*)\ \{V \mid A \mid I\} \not\sqsubseteq \chi(*)\ \{D \mid I\} \not\sqsubseteq \chi(\mathcal{M}(*)) \Longrightarrow (\uparrow OBL) = \downarrow$

    e. A POSS is a DP daughter of a DP:
$$\begin{matrix} Max(*) & Max(\mathcal{M}(*)) \\ \chi(*) = D & \chi(\mathcal{M}(*)) = D \end{matrix} \Longrightarrow (\uparrow POSS) = \downarrow$$

The first definition is the simplest. A PROJECTING NODE (any node that has the same category features as its mother) can be annotated as a functional co-head ($\uparrow = \downarrow$). A SUBJ annotation can be added to a MAXIMAL PROJECTION that has the category feature D. That node must also have a mother that is a MAXIMAL PROJECTION with the features of the category I. Likewise, the annotation for OBJ can be added to a MAXIMAL PROJECTION with the features of the category D. The mother of this node must have the feature V or P. An OBL annotation again requires a MAXIMAL PROJECTION. The node cannot

---

[30] Marcotte decomposes lexical and functional categories using his own system of privative features, *Pr, Tr* and *f* such that V: {*Pr, Tr*}, A: {*Pr*}, P: {*Tr*}, N: { }, I: {*Pr, Tr, f*}, and D: {*f*}. This part of his system has been simplified for readability.

be V, A or I, and its mother node must not be a functional node. Finally, for a POSS annotation, the MAXIMAL PROJECTION must have the features for the category D, and its mother must be a MAXIMAL PROJECTION with the feature for a D as well.

Marcotte ingeniously creates a set of structure-function association principles very similar to those proposed by Bresnan (2001), but without referring to any bar levels directly. He restricts himself to referring only to whether a node has identical category features to its mother (PROJECTING NODE) or not (MAXIMAL PROJECTION), and to what type of category features can be associated with which grammatical functions. However, there are problems with the proposal which would require significant modifications to the system in order to solve, modifications which would severely compromise the elegance of the system.

4.2.2   Issues

While it does have a notion of nonprojecting categories, Marcotte's (2014) model cannot deal with recursive adjunction of nonprojecting words to nonprojecting words, as we have assumed for nonprojecting adjectives in English, following Arnold and Sadler (2013). Ex. (46) shows a prenominal adjective structure in English translated into Marcotte's proposed system. The point of Arnold and Sadler's analysis is that there are a number of generalizations that can be made about prenominal modifiers in English which can be captured by analyzing those nodes as nonprojecting words. This generalization is lost in translation to Marcotte's system. The modifiers *very* and *happy* do not meet Marcotte's definition of nonprojecting words. According to Marcotte, a nonprojecting word is both a TERMINAL and a MAXIMAL PROJECTION. A MAXIMAL PROJECTION is defined as a node that does not have the same category features as its mother. In this particular structure, both modifiers have the same category as their mother, so, by definition, they are PROJECTING NODES, not MAXIMAL NODES.[31] Another difference between the two analyses is that in Arnold and Sadler's analysis, the node adjoined to the head noun

---

[31] This example would not be a problem if, following e.g. Payne *et al.* (2010), adjectives and adverbs were treated as part of separate c-structure categories. But the point remains valid, e.g. in the phrase *those really very happy people* the same problem would apply to the relationship between *really* and *very*.

is a nonprojecting node. This preserves their generalization that only nonprojecting modifiers can occur in this position. This generalization is lost in Marcotte's system. The mother of the two modifiers is, by definition, a MAXIMAL PROJECTION, but it is not a TERMINAL since it has a daughter node that shares the same features. In Marcotte's system, the mother of two nonprojecting nodes can never be a nonprojecting node itself.

(46)

$$N_{Max}$$

$D_{Max,Term}$ — those

N_{Proj}

A_{Max}

$N_{Proj,Term}$ — people

$A_{Proj,Term}$ — very

$A_{Proj,Term}$ — happy
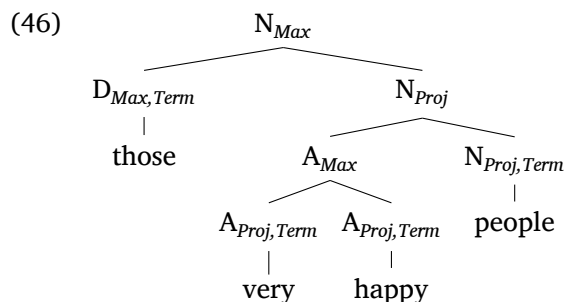
Marcotte (2014) also fails to make the correct distinction between X^{max} and the top node of a projection, as found in XP adjunction structures. In XP adjunction, a maximal phrase (XP) is the mother of another maximal phrase of the same category features. Such a structure is not possible in Marcotte's proposed system because, by definition, if a node's mother has the same category features, that node cannot be a MAXIMAL PROJECTION: it is a PROJECTING NODE. Ex. (47) shows the analysis of "topicalization" from Bresnan *et al.* (2016, 196) on the left, with a translation into Marcotte's proposed system on the right. The structure on the left has an IP in the topmost position dominating an identical IP node. On the right, in Marcotte's system, only the topmost I node is a MAXIMAL PROJECTION. The I node below the topmost node is, by definition, a PROJECTING NODE since it shares its category features with its mother. Thus although Marcotte (2014) does distinguish the top node of a projection from lower nodes, no node below the top node may be classified as X^{max}. This poses very real practical problems because, for example, in Marcotte's system the position annotated for SUBJ must be dominated by a MAXIMAL PROJECTION. In the tree on the left, the DP subject is dominated by IP, so it meets the structural requirement for a subject. On the right, the D node that dominates what should be the subject is *not* the daughter of

a MAXIMAL PROJECTION so it does not meet the structural require-ment for a subject position; this tree therefore cannot be generated in Marcotte's system.

(47)
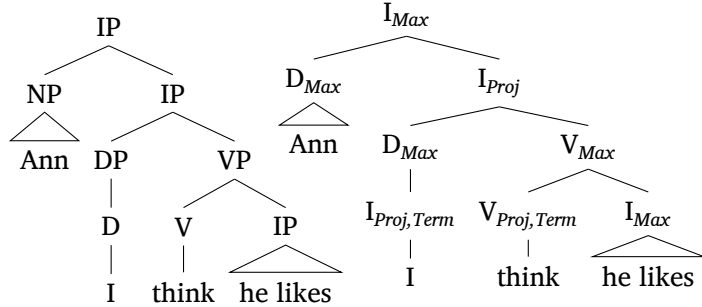
4.2.3 Summary

Marcotte's model is clearly an improvement on that of Bresnan (2001), capturing more of the desiderata we have been considering for a the-ory of phrase structure. His model eliminates nonbranching nodes, with trees that have only the structure required to model constituency. Marcotte also avoids the need for default optionality in phrase struc-ture rules, and lacks a distinct notion equivalent to X′. However, Mar-cotte's model does not correctly model the distinction between X^{max}, XP, and the top node of a projection, nor does it offer an adequate account of nonprojecting categories, nor any account of exocentric categories. It also does not avoid redundancy in category labelling.

Marcotte's proposal is in some respects very similar to that of Muysken (1982), who proposed to reformulate X′ theory in terms of two binary valued features, [±projection] and [±maximal]. The weak-nesses of Marcotte's model apply equally to the proposals of Muysken (1982).

4.3 Collins and Stabler (2016)

As part of their mathematically precise formalization of Minimalism, Collins and Stabler (2016, 62–66) define a labelling algorithm which they state allows natural definitions of all the X′ theory concepts. De-spite the fact that Collins and Stabler (2016) is one of the most com-plete and precise formalizations of mainstream Minimalism, their for-malization of phrase structure is limited in certain respects. Given the much less flexible approach to phrase structure adopted in minimal-

ism, some of the desiderata given in (1), notably the requirement for principled accounts of exocentricity and nonprojecting categories, are not relevant for Collins and Stabler (2016), and hence have no place in their system. As a formalization of BPS, it naturally captures most of the other desiderata. However, their system also lacks any account of adjunction, which would be crucial for a complete account of minimalist phrase structure, and does not provide any way to distinguish the highest projection from X^{max}.

Collins and Stabler (2016, 65) define a labelling function from syntactic objects to lexical item tokens, such that a. for all lexical item tokens LI, Label(LI) = LI, and b. for all complex syntactic objects, the label of the object is the label of its head. [32] As a labelling function this is similar to LFG's λ projection, but differs in a number of ways. The most important difference for the present purposes is that there are no distinct label such as N and C, but it is lexical item tokens themselves which function as labels.

Given this notion of labelling, Collins and Stabler (2016) define maximal, minimal and intermediate projections:

(48) a. For all C a syntactic object and LI a lexical item token both contained in a derivable workspace W, C is a *maximal projection* of LI iff Label(C) = LI and there is no D contained in W which immediately contains C such that Label(D) = Label(C).

   b. For all C, C is a *minimal projection* iff C is a lexical item token.

   c. For all syntactic objects C contained in workspace W, LI a lexical item token, C is an *intermediate projection* of LI iff Label(C) = LI, and C is neither a minimal nor a maximal projection in W.

They further define the complement as the first element merged with a head, and a specifier as any further element merged with a projection of the head.

The definition of maximal projection defines what we have called the highest projection, but does not allow any distinction between this

---

[32] In some sense b. is similar to our definition of CAT above as a distributive feature in order to propagate category information automatically from head daughter to mother.

and X^{max}. This distinction is relevant where adjunction to the maximal projection is admitted; since Collins and Stabler do not formalize adjunction, the failure to distinguish these notions is understandable. The definition of a minimal projection is unproblematic, and differs from the analysis presented above most significantly in that in BPS lexical item tokens are themselves the terminal nodes of the phrase structure, whereas in our model lexical item tokens are distinct from the terminal nodes of the c-structure.

The definition of complement as the first element merged with a head is not too dissimilar from our own definition, which effectively defines complement as the sister of a head with ⟨L,0⟩. In defining specifier as any further element merged with a projection of the head, Collins and Stabler license multiple specifiers, but leave little room for a notion of adjunction.

Overall, Collins and Stabler's formalization of BPS captures the most important notions of BPS discussed above and integrated into our model but, partly due to the less enriched notion of phrase structure which they are modelling, does not appear immediately extensible to cover adjunction and all the other phrase structure phenomena which we have attempted to model in this paper.

5   CONCLUSION

Hitherto, LFG has continued to utilize a model of phrase structure which is largely unchanged from the 1970s, and does not incorporate the insights and advances made within BPS and other theories. Our proposal offers a new model of phrase structure within LFG which captures the central insights of the last forty years of work on phrase structure in a fully formalized, and potentially theoretically broad, way.

REFERENCES

Avery D. ANDREWS (2018), Sets, heads and spreading in LFG, *Journal of Language Modelling*, 6(1):131–174.

Avery D. ANDREWS and Christopher D. MANNING (1999), *Complex Predicates and Information Spreading in LFG*, CSLI Publications, Stanford, CA.

Doug ARNOLD and Louisa SADLER (2013), Displaced Dependent Constructions,

in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG13 Conference*, pp. 48–68, CSLI Publications, Stanford, CA.

Ash ASUDEH (2012), *The Logic of Pronominal Resumption*, Oxford University Press, Oxford.

Ash ASUDEH, Mary DALRYMPLE, and Ida TOIVONEN (2013), Constructions with Lexical Integrity, *Journal of Language Modelling*, 1(1):1–54.

Peter AUSTIN and Joan BRESNAN (1996), Non-Configurationality in Australian Aboriginal Languages, *Natural Language & Linguistic Theory*, 14(2):215–268.

Tina BÖGEL, Miriam BUTT, Ronald M. KAPLAN, Tracy Holloway KING, and III MAXWELL, John T. (2009), Prosodic Phonology in LFG: A New Proposal, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG09 Conference*, pp. 146–166, CSLI Publications, Stanford, CA.

Joan BRESNAN (1982), Control and Complementation, *Linguistic Inquiry*, 13(3):343–434, reprinted in Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, MIT Press, pp. 282–390.

Joan BRESNAN (1995), Linear order, syntactic rank, and empty categories: On weak crossover, in Mary DALRYMPLE, Ronald M. KAPLAN, John T. MAXWELL, and Annie ZAENEN, editors, *Formal issues in Lexical-Functional Grammar*, pp. 241–274, CSLI Publications, Stanford, CA.

Joan BRESNAN (1998), Morphology competes with syntax: Explaining typological variation in weak crossover effects, in Pilar BARBOSA, D. FOX, P. HAGSTROM, M. MCGINNIS, and D. PESETSKY, editors, *Is the Best Good Enough? Proceedings from the Workshop on Optimality in Syntax*, MIT Press, Cambridge, MA.

Joan BRESNAN (2001), *Lexical-Functional Syntax*, Blackwell, Oxford.

Joan BRESNAN, Ash ASUDEH, Ida TOIVONEN, and Stephen WECHSLER (2016), *Lexical-Functional Syntax*, Wiley-Blackwell, Oxford, second edition. First edition by Joan Bresnan, 2001, Blackwell.

Andrew CARNIE (2000), On the definition of X^0 and XP, *Syntax*, 3(2):59–106.

Andrew CARNIE (2010), *Constituent structure*, Oxford University Press, Oxford, second edition.

Noam CHOMSKY (1957), *Syntactic Structures*, Mouton, The Hague.

Noam CHOMSKY (1970), Remarks on Nominalization, in Roderick A. JACOBS and Peter S. ROSENBAUM, editors, *Readings in English Transformational Grammar*, pp. 184–221, Ginn, Waltham, MA, also in: Chomsky, Noam (1972), *Studies on Semantics in Generative Grammar*, The Hague: Mouton, pp. 11–61.

Noam CHOMSKY (1995), Bare Phrase Structure, in Héctor CAMPOS and Paula KEMPCHINSKY, editors, *Evolution and revolution in linguistic theory*, pp. 51–109, Georgetown University Press, Washington, D.C.

Noam CHOMSKY (2001), *Beyond explanatory adequacy* (MIT Occasional Papers in Linguistics 20), Department of Linguistics and Philosophy, MIT, Cambridge, MA.

Chris COLLINS (2002), Eliminating labels, in Samuel EPSTEIN and David SEELY, editors, *Derivation and explanation in the Minimalist Program*, pp. 42–64, Blackwell, Malden, MA.

Chris COLLINS and Edward STABLER (2016), A formalization of Minimalist syntax, *Syntax*, 19(1):43–78.

Dick CROUCH, Mary DALRYMPLE, Ronald M. KAPLAN, Tracy Holloway KING, John T. MAXWELL III, and Paula NEWMAN (2011), *XLE Documentation*, Palo Alto Research Center, Palo Alto, CA,
`http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html`.

Mary DALRYMPLE (2001), *Lexical Functional Grammar*, Academic Press, San Diego, CA.

Mary DALRYMPLE (2017), Unlike phrase structure category coordination, in Victoria ROSÉN and Koenraad DE SMEDT, editors, *The very model of a modern linguist: in honor of Helge Dyvik* (Bergen Language and Linguistics Studies (BeLLS), volume 8), pp. 33–55, University of Bergen, Bergen.

Mary DALRYMPLE and Ronald M. KAPLAN (2000), Feature Indeterminacy and Feature Resolution, *Language*, 76(4):759–798.

Mary DALRYMPLE, Ronald M. KAPLAN, and Tracy Holloway KING (2001), Weak crossover and the absence of traces, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG01 Conference*, CSLI Publications, Stanford, CA.

Mary DALRYMPLE, Ronald M. KAPLAN, and Tracy Holloway KING (2004), Linguistic Generalizations over Descriptions, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG04 Conference*, pp. 199–208, CSLI Publications, Stanford, CA.

Mary DALRYMPLE, Ronald M. KAPLAN, and Tracy Holloway KING (2007), The absence of traces: evidence From weak crossover, in Annie ZAENEN, Jane SIMPSON, Tracy Holloway KING, Jane GRIMSHAW, Joan MALING, and Christopher MANNING, editors, *Architectures, rules, and preferences: Variations on themes by Joan W. Bresnan*, CSLI Publications, Stanford, CA.

Mary DALRYMPLE, Ronald M. KAPLAN, and Tracy Holloway KING (2015), Economy of Expression as a principle of syntax, *Journal of Language Modelling*, 3(2):377–412.

Mary DALRYMPLE and Tracy Holloway KING (2013), Nested and Crossed Dependencies and the Existence of Traces, in Tracy Holloway KING and Valeria DE PAIVA, editors, *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, pp. 139–151, CSLI Publications, Stanford, CA.

Mary DALRYMPLE, John J. LOWE, and Louise MYCOCK (2019), *The Oxford reference guide to Lexical Functional Grammar*, Oxford University Press, Oxford.

Mary DALRYMPLE and Louise MYCOCK (2011), The Prosody-Semantics Interface, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG11 Conference*, pp. 173–193, CSLI Publications, Stanford, CA.

Mary DALRYMPLE and Irina NIKOLAEVA (2011), *Objects and Information Structure*, Cambridge University Press, Cambridge.

Lachlan DUNCAN (2007), Analytic Noun Incorporation in Chuj and K'ichee' Mayan, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG07 Conference*, pp. 163–183, CSLI Publications, Stanford, CA.

Yehuda N. FALK (2001), *Lexical-Functional Grammar: an Introduction to Parallel Constraint-Based Syntax*, CSLI Publications, Stanford, CA.

Naoki FUKUI (1986), *A theory of category projection and its applications*, Ph.D. thesis, MIT.

Hans-Martin GÄRTNER (2002), *Generalized transformations and beyond: Reflections on minimalist syntax*, de Gruyter, Berlin.

Gianluca GIORGOLO and Ash ASUDEH (2011), Multimodal Communication in LFG: Gestures and the Correspondence Architecture, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG11 Conference*, pp. 257–277, CSLI Publications, Stanford, CA.

Jane GRIMSHAW (1991), Extended Projections, mS, Brandeis University.

Norbert HORNSTEIN and Jairo NUNES (2008), Adjunction, labeling and Bare Phrase Structure, *Biolinguistics*, 2(1):57–86.

Ray JACKENDOFF (1977), *X̄ Syntax: A Study of Phrase Structure*, MIT Press, Cambridge, MA.

K. A. JAYASEELAN (2008), Bare phrase structure and specifier-less syntax, *Biolinguistics*, 2(1):87–106.

Ronald M. KAPLAN (1989), The Formal Architecture of Lexical-Functional Grammar, in Chu-Ren HUANG and Keh-Jiann CHEN, editors, *ROCLING II: Proceedings of the Computational Linguistics Conference*, pp. 3–18, The Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Tapei, also published in *Journal of Information Science and Engineering 5* (1989), pp. 305–322, and in *Formal Issues in Lexical-Functional Grammar*, ed. Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen, CSLI Publications, 1995, pp. 7–27.

Ronald M. KAPLAN and Joan BRESNAN (1982), Lexical-Functional Grammar: A Formal System for Grammatical Representation, in Joan BRESNAN, editor, *The Mental Representation of Grammatical Relations*, pp. 173–281, MIT Press, Cambridge, MA.

Ronald M. KAPLAN and Annie ZAENEN (1989), Long-Distance Dependencies, Constituent Structure, and Functional Uncertainty, in Mark R. BALTIN and Anthony S. KROCH, editors, *Alternative Conceptions of Phrase Structure*, pp. 17–42, University of Chicago Press, Chicago.

Richard S. KAYNE (1994), *The Antisymmetry of Syntax*, MIT Press, Cambridge, MA.

Tracy Holloway KING (1995), *Configuring Topic and Focus in Russian*, CSLI Publications, Stanford, CA.

Tracy Holloway KING and Mary DALRYMPLE (2004), Determiner agreement and noun conjunction, *Journal of Linguistics*, 40:69–104.

András KORNAI and Geoffrey K. PULLUM (1990), The X-bar theory of phrase structure, *Language*, 66:24–50.

Paul KROEGER (1993), *Phrase structure and grammatical relations in Tagalog*, CSLI Publications, Stanford, CA.

Joseph LOVESTRAND (2018), *Serial verb constructions in Barayin: Typology, description and Lexical-Functional Grammar*, Ph.D. thesis, University of Oxford.

Joseph LOVESTRAND and John J. LOWE (2017), Minimal c-structure: Rethinking projection in phrase structure, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG17 Conference*, pp. 285–305, CSLI Publications, Stanford, CA.

Jean-Philippe MARCOTTE (2014), Syntactic Categories in the Correspondence Architecture, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG14 Conference*, pp. 408–428, CSLI Publications, Stanford, CA.

Jr III MAXWELL, John T. and Ronald M. KAPLAN (1996), Unification-based parsers that automatically take advantage of context freeness, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG96 Conference*, CSLI Publications, Stanford, CA.

Pieter MUYSKEN (1982), Parametrizing the notion "head", *Journal of Linguistic Research*, 2(3):57–75.

Louise MYCOCK and John J. LOWE (2013), The Prosodic Marking of Discourse Functions, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG13 Conference*, pp. 440–460, CSLI Publications, Stanford, CA.

Prerna NADATHUR (2013), Weak crossover and the direct association hypothesis, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG13 Conference*, pp. 461–481, CSLI Publications, Stanford, CA.

Rachel NORDLINGER (1998), *A Grammar of Wambaya, Northern Australia*, Pacific Linguistics, Canberra.

John PAYNE, Rodney HUDDLESTON, and Geoffrey K. PULLUM (2010), The distribution and category status of adjectives and adverbs, *Word Structure*, 3(1):31–81.

Carl POLLARD and Ivan A. SAG (1994), *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago.

William J. POSER (1992), Blocking of phrasal constructions by lexical items, in Ivan A. SAG and Anna SZABOLCSI, editors, *Lexical Matters*, pp. 111–130, CSLI Publications, Stanford, CA.

Michael PUTNAM and Thomas STROIK (2009), Traveling without moving: the conceptual necessity of Survive-minimalism, in Michael PUTNAM, editor, *Towards a derivational syntax: Survive-minimalism*, pp. 3–20, Benjamins, Amsterdam.

Michael PUTNAM and Thomas STROIK (2010), Syntactic relations in Survive-minimalism, in Michael PUTNAM, editor, *Exploring crash-proof grammars*, pp. 143–166, Benjamins, Amsterdam.

Louisa SADLER (1997), Clitics and the Structure-Function Mapping, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG97 Conference*, CSLI Publications, Stanford, CA.

Louisa SADLER and Doug ARNOLD (1994), Prenominal Adjectives and the Phrasal/Lexical Distinction, *Journal of Linguistics*, 30:187–226.

Liselotte SNIJDERS (2012), Issues Concerning Constraints on Discontinuous NPs in Latin, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG12 Conference*, pp. 565–581, CSLI Publications, Stanford, CA.

M. SPEAS (1986), *Phrase structure in natural language*, Ph.D. thesis, MIT.

Andrew SPENCER (2005), Case in Hindi, in Miriam BUTT and Tracy Holloway KING, editors, *Proceedings of the LFG05 Conference*, pp. 429–446, CSLI Publications, Stanford, CA.

Edward STABLER (1997), Derivational minimalism, in C. RETORÉ, editor, *Logical aspects of computational linguistics*, pp. 68–95, Springer, New York.

Timothy Angus STOWELL (1981), *Origins of phrase structure*, Ph.D. thesis, Massachusetts Institute of Technology.

Ida TOIVONEN (2003), *Non-Projecting Words: A Case Study of Swedish Verbal Particles*, Kluwer, Dordrecht.

Lisa deMena TRAVIS (1984), *Parameters and effects of word order variation*, Ph.D. thesis, MIT.

Stephen WECHSLER and Larisa ZLATIĆ (2003), *The many faces of agreement*, CSLI Publications, Stanford, CA.