DIMENSIONS OF VARIATION IN THE EXPRESSION OF FUNCTIONAL FEATURES: MODELLING DEFINITENESS IN LFG

Kersti Börjars and John Payne
The University of Manchester

Proceedings of the LFG13 Conference

Miriam Butt and Tracy Holloway King (Editors)

2013

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

The exponents of functional features show a range of variation that cannot be captured even by dichotomies such as the clitic-affix distinction, not even when extended to include categories such as phrasal affix. A multi-dimensional parallel correspondence theory such as Lexical-Functional Grammar is eminently suited to dealing with the full range of variation. However, some of the patterns we find in languages involve extensions to the standard constructional formalism of LFG. In this paper we look at some challenging patterns of distribution exemplified by the definiteness feature and compare analyses expressed in terms of standard LFG formalism with those captured in a specification language approach.

1 Introduction

Functional features can find exponence in a great many ways across languages. Traditionally, the focus has been directed at distinctions such as whether exponents take the form of words, clitics or affixes. This aspect of the variation in exponence is well explored, and has led to the postulation by linguists such as Zwicky and Pullum (1983), Lapointe (1990, 1992), Miller (1992), Miller and Halpern (1993) and a number of others to the postulation of further categories (phrasal affixes or edge-marking morphology) that combine properties of both clitics and affixes. The diversity in exponence may go beyond what can be captured even by this more subtle approach to morpho-syntactic categories (see for instance Börjars (2003), Spencer and Luís (2012) and Börjars et al. (2013)). A parallel correspondence theory like LFG, which represents dimensions of linguistic information separately and links them with mapping functions which permit many-to-one relations would seem eminently suited to capturing this kind of variation.

In this paper, we will consider the feature DEFINITENESS and explore the range of ways in which it can be manifested. Definiteness is a complex phenomenon, depending on properties such as identifiability and uniqueness. We will ignore these semantic subtleties here and simply assume that there are elements that can be described as (in)definite. We will capture this with a feature [DEF $\pm$]. Many languages do not mark definiteness specifically, but whether a noun phrase is interpreted as definite or not is left to pragmatics. Often definiteness arises as a by-product of some other element; for instance a possessor or a demonstrative may yield a definite interpretation or a particular word order give rise to a definite interpretation. However, in this paper, we will consider only elements which mark definiteness specifically. We will use the following abbreviations:

† We are grateful to the audiences of LFG2013 and the 2013 meeting of the Linguistics Associaction of Great Britain for their comments, especially Oleg Belyaev, Dag Haug and Ida Toivonen. For her comments at both conference, as well as her generosity in providing advice on LFG solutions, we are extremely grateful to Mary Dalrymple.

Dedicated definiteness marker (DDM)

a morphological or syntactic element which marks only definiteness [DEF $\pm$] (and possibly PERS/NUM/GEND features)

Functional definiteness marker (FDM)

a DDM whose presence is sufficient to make a noun phrase functionally definite: that is it will induce the presence of an f-structure feature [DEF $\pm$].

A noun phrase which has a value for the feature [DEF] in its f-structure will have the potential to function as a referential noun phrase, and in the cases we are considering here, the feature value will have originated from an FDM.

A cross-linguistic study of DDMs reveals different types of mismatches between form and function: in Dutch DDM finds exponence in the syntax — *de hond* 'DEF dog' — whereas in Icelandic it is part of a word — *hundinn* 'dog.DEF'. Danish has a prosodically independent DDM — *det* (*hus*) 'the house / it' — but in English it is prosodically dependent — *the house*.

In what follows, we will explore two cases which involve greater complexity than the mapping of definiteness from a syntactic DDM, as Dutch, Danish and English, or from a morphological DDM marked on a head noun, as in Icelandic. Both of the latter mappings allow the straightforward application of up and down designators in effect to pass the definiteness feature up the c-structure tree to the noun phrase node where it will map to an attribute in that noun phrase's f-structure. The two more complex cases are (i) definiteness marking on adjectives — as in Latvian, Hebrew and Swedish, and (ii) definiteness marking on a left edge — as in Ossetic, where the DDM is prosodic, and Bulgarian, where it is morphological. In interesting ways, both of these cases require extensions to the basic constructional notation of LFG. We therefore compare two different ways of formalising the mappings involved: standard LFG notation (with the aforementioned extensions), and a specification language approach. The former is familiar, but the latter may require some explanation. Blackburn and Gardent (1995, 39) argue that the elegant intuitions underlying the general LFG architecture are not captured in standard formalisations, in the sense that 'they make a detour via the LFG *construction algorithm*' (see also discussion in Kaplan (1995)). The specification language essentially describes the mapping between c-structure and f-structure directly: it is a propositional language enriched with operators ("modalities") which define the transitions between c-structure nodes, the transitions between f-structure attributes and their values, and the transitions between c-structure nodes and the corresponding f-structure attributes. The propositions which are needed for a particular language are then the axioms of that language

2 A specification language for LFG

Blackburn and Gardent (1995) demonstrate the construction of a basic specification language for LFG. The ontology of LFG is defined as a triple $<\mathcal{T}, \mathcal{F}, \mathcal{M}>$ where

$\mathcal{T}$ is the mathematical picture of c-structure, $\mathcal{F}$ is the mathematical picture of f-structure, and the function $\mathcal{M}$ defines the mapping between the two.[1] C-structures are modelled as standard trees in which each node is labelled with a category symbol and any applicable morphosyntactic features. Particularly relevant here will be the morphosyntactic feature *def*, which we assign to nouns, adjectives and other lexical categories which are morphologically marked as definite. F-structures are modelled as tree-like (properly multidominance) structures in which attributes are modalities associated with functions from one node to another node, and the symbol values of attributes are the labels of terminal nodes. Potts (2002) further specifies $\mathcal{M}$, the mapping between c-structure nodes and f-structure nodes, by pairing it with the binary modality <M> and assigning a unique name from the set {n1, n2, ... } to each node in a given c-structure, and a unique name from the set {n′1, n′2, ... } to each node in a given f-structure. We will henceforth adopt Potts' notation.

This ontology will obviously need to be expanded if we wish to model other LFG dimensions. In particular, we will need a quintuple $<\mathcal{T}, \mathcal{F}, \mathcal{P}, \mathcal{M}, \mathcal{S}>$ if we wish to include p-structure.[2] In this case, $\mathcal{P}$ will be the mathematical picture of p-structure and $\mathcal{S}$ the function which defines the mapping between c-structure and p-structure. We leave open the question as to whether p-structures are to be modelled as prosodic trees (Dalrymple and Mycock, 2011; Mycock and Lowe, 2013), or as AVMs (Butt and King, 1998): either route is compatible with the particular statement we will suggest for definiteness marking in Ossetic. If the AVM route proves to be optimal, then a similar mathematical construction can be employed as for $\mathcal{F}$. Extending Potts's treatment of $\mathcal{M}$, we will then pair $\mathcal{S}$, under either a tree or feature-structure model, with the binary modality <S> which links c-structure nodes with p-structure nodes.

The specification language L itself then consists of propositions which hold of each node in the structures defined by $\mathcal{T}$, $\mathcal{F}$ and $\mathcal{P}$. Examples of basic propositions are given in (1).

(1) a. *NP* — the category label NP is true of this c-structure node
b. <DEF> + — the modality <DEF> maps this f-structure node to a terminal f-structure node labelled +

The clumsy English glosses are deliberate, and reflect the propositional nature of L. As a propositional language, L is also naturally defined to contain the standard Boolean connectives, so that a conjunction of basic propositions can hold of a given node. We might then have propositions such as (2)

[1] Blackburn and Gardent (1995) refer to $\mathcal{M}$ as *zoomin*.
[2] The need for a quintuple, rather than a triple, is pointed out by Potts (2002, 31, fn 14).

(2)	*NP* ∧ <M><DEF> +	the category label NP is true of this c-structure node and this node is mapped by the modality <M> to an f-structure node which is mapped by the modality <DEF> to a terminal node labelled +

In addition to defining the notion of c-structure head, Potts (2002) also defines a useful set of c-structure modalities. The ones we will need are found in (3).[3]

(3)
a. <d_1> map this node to its leftmost daughter
b. <d> map this node to any of its daughters
c. <d_1*> map this node to the leftmost terminal node it dominates
d. <d*> map this node to any terminal node it dominates

If we wish to indicate the name of a node which is the value of such a modality, we will abbreviate the naming convention defined by Potts and simply attach a subscript to the modality itself. In other words, <d$*_k$>Φ will stand as an abbreviation for <d*>($n_k \wedge \Phi$), i.e. map this node to any terminal node it dominates which is named n_k and of which the proposition Φ holds. We use *italics* for propositions, categories and features associated with c-structure and SMALL CAPS for those associated with f-structure.

There is a considerable literature on the potential advantages of specification languages as a mode of description, as well as their potential expressive power in comparison to other formalisms (see in particular Pullum, 2013). One important result is that of Rogers (2003), who shows that the complexity of languages described by quite a rich specification language, that of weak monadic second-order logic (wMSO), depends essentially on the dimensions of the ontology. Thus tree-adjoining grammars can be defined in the wMSO theory of certain three-dimensional tree-like structures. It remains to be explored how this kind of approach to formal language complexity might extend to the multidimensional architecture of LFG, but we see no prima facie grounds to be pessimistic.

3 Definiteness marking on adjectives

It is not uncommon for definiteness to be marked on the adjective, three examples can be found in (4).[4]

(4) a. lielais koks (Latvian)
big.DEF tree

[3]Similar definitions can be provided for righnmost daughter and rightmost terminal node, but these are not needed here.

[4]The adjectival ending in Swedish is sometimes assigned a separate feature [WEAK/STRONG]. However, this distinction between the feature on adjectives and that on determiners and nouns is not warranted in modern Swedish. In earlier forms of the language, there was not a complete correlation between [WEAK/STRONG] and [DEF +/–] (Delsing, 1994), but in modern Swedish there is.

'the big tree'

b. ha-sefer ha-gadol (Hebrew)
DEF-book DEF-big
'the big book'

c. den stora boken (Swedish)
DEF big.DEF book.DEF
'the big book'

As the data illustrate, languages which mark definiteness on the adjective vary as to whether it is also marked on other elements of the noun phrase, Swedish and Hebrew show agreement across other categories, whereas Latvian does not. Though as the example in (5) shows, if there is more than one adjective in a noun phrase in Latvian, all adjectives need to be marked.

(5) lielais skaistais koks (Latvian)
big.DEF beautiful.DEF tree
'the big beautiful tree'

There is a further distinction with respect to the definiteness marking on the adjective. Since only the adjective is marked for definiteness in (4-a), it is clear that in Latvian, the marking on the adjective is functional in nature: it makes the noun phrase as a whole functionally definite. No other element can be marked for definiteness, so that a noun on its own is ambiguous; *koks* can mean 'tree', 'a tree' or 'the tree'. In Hebrew and Swedish, other elements within the noun phrase are also marked for definiteness, but the adjective marking in the two languages turns out to behave quite differently. As (6-a) illustrates, the definiteness marking on the adjective in Hebrew is a functional feature in this sense, the *ha-* is an FDM, whereas (6-b) shows it is not in Swedish. A definite adjective in Swedish requires the presence of another definiteness marker to form a noun phrase, as in (6-c).[5]

(6) a. ha-gadol (Hebrew)
DEF-big
'the big one'

b. stora (Swedish)
big.DEF
*'the big one' (just means 'big')

c. den stora (Swedish)
DEF big.DEF
'the big one'

The conclusion for Swedish is that the feature DEF feeds into the f-structure of the noun phrase when it finds exponence on a determiner or a noun, but not when it occurs on an adjective. We seem to have variation in two dimensions: firstly, the

[5]For dialects of Swedish which have a different adjectival ending which does allow the definite adjective to function as a referential noun phrase, see Delsing (2003).

definiteness feature on the adjective can form part of an agreement pattern across the noun phrase, or it can be the sole exponent of definiteness; secondly it may or it may not feed into the f-structure of the phrase.

Discussions of agreement in the literature tend to see it as a relation between two elements within a phrase, but they differ as to the nature of the relation. Under one view, there is a directional relation, so that one of the elements is the SOURCE or the CONTROLLER, and the agreement is directed at a TARGET (see for instance Corbett, 2003). On the other view, agreement is non-directional, and the agreeing elements are assumed to CO-VARY (for example Pollard and Sag, 1994, 60–7). Agreement can also be viewed as a relation between an element and the phrase which contains it. For instance, Lehmann (1982, 204) says about definiteness, number and case that they are 'on the semantic level, categories of the nominal or NP and not of the noun'.

With the possible exception of gender, directional agreement does not provide the best way of accounting for noun phrase internal agreement. In the examples already considered, it is not clear what the Swedish adjective would agree with under a directional approach, or which of the two adjectives in (5) would be the source and which the target.

Furthermore, we think there are arguments in favour of taking a phrasal approach to agreement. In Hebrew construct state nominals, as exemplified by (7), the head noun is not marked for definiteness, so that even though the noun phrase is unambiguously definite, the head noun cannot have the definiteness marker *ha-*. Still, the elements within the phrase normally considered to be the complement of the noun must be marked for definiteness. It would then seem that the obligatory definiteness marker on *gadol* can only reasonably be attributed to agreement with the phrase as a whole.

(7) beyt Sophie ha-gadol (Hebrew)
house(M).CON Sophie DEF-big.M
'Sophie's big house'

Agreement in LFG is conceived of as an f-structure phenomenon.[6] Since an unbounded number of adjectives are in principle permitted within a noun phrase, the value of the grammatical function feature ADJ is a set. Following standard conventions within LFG, we would get the description in (8-a) for an ADJ. However, Dalrymple (2001) suggests the alternative convention in (8-b), where $\in$ is treated as the value of the attribute ADJ.

(8) a. $\downarrow\in (\uparrow \text{ADJ})$

b. $(\uparrow \text{ADJ} \in) = \downarrow$

[6] Falk (2006) argues that a separate dimension is required since aspects of agreement cannot be expressed satisfactorily in terms of f-structure. However, this proposal appears not to have been taken up by others.

These yield identical set-valued f-structures, but the latter formulation is a more convenient equivalence for 'writing constraints on set members, particularly in expressions involving *inside-out functional uncertainty*' (2001, 154). An account of the definiteness markers we have considered here would involve inside-out functional uncertainty: 'the f-structure within which the ADJ feature of which I am the value is embedded is definite'. Since each ADJ is embedded within a set, the approach used by Nordlinger (1998) cannot be straightforwardly derived from (8-a). However, from (8-b), we can derive the equation in (9), following standard definitions of inside-out functional uncertainty (e.g. Dalrymple, 2001, 145).

(9) $(\text{ADJ} \in \uparrow)$

Taking the Swedish adjectives first, since the definiteness marking on them does not make a functional contribution to the phrase, a natural LFG approach would involve the use of a constraining equation. The equation in (10) is constraining and hence does not build f-structure, but will ensure that when a definite determiner or definitely marked noun induces a functional definiteness feature in the f-structure of the moun phrase, then the adjectives must agree. Indefinite determiners and indefinite adjectives will not be able to occur in the same noun phrase as definite adjectives, since the presence of any element with the feature value [DEF –] would mean that the constraint is not satisfied. We will return to a discussion of constraining equations shortly. This constraining equation also predicts the ungrammaticality of (6-b) as a referential noun phrase.

(10) *stora* $(\uparrow\text{PRED}) = \text{`big'}$
$((\text{ADJ} \in \uparrow)\ \text{DEF}) =_c +$

The definiteness feature on the adjectives in Hebrew and Latvian, on the other hand, does make a functional contribution to the noun phrase and hence it needs to be expressed in terms of constructive morphology in the sense of Nordlinger (1998). We then get the equation in (11-a), which constructs the f-structure in (11-b).

(11) a. *lielais/ha-gadol* $(\uparrow\text{PRED}) = \text{`big'}$
$((\text{ADJ} \in \uparrow)\ \text{DEF}) = +$

b.

$$\begin{bmatrix} \text{DEF} & + \\ \text{ADJ} & \begin{bmatrix} \text{PRED} & \text{`big'} \end{bmatrix} \end{bmatrix}$$

The f-structure in (11-b) will be associated with the mother of the AP, presumably an N′, and through structure sharing with the phrase as a whole. On the assumption that a referential noun phrase is any nominal whose f-structure contains the feature value [DEF +], this equation will allow the adjective to form a referential noun phrase on its own and hence predict the grammaticality of (6-a) and its

Latvian equivalent.[7]. It will permit other elements in the noun phrase to be marked for [DEF +], but it will not require them to be so. It will of course rule out elements marked for [DEF –]. Agreement would then be enforced by the assumption that adjectives lacking the definiteness marking have the feature [DEF –] and that there are no adjectives unmarked for definiteness.

The difference between Latvian and Hebrew would lie in the feature properties of other elements within the noun phrase. In Latvian, as (4-a) showed, nouns are unmarked for [DEF] and can occur in both definite and indefinite noun phrases. As (4-b) illustrated, definite noun phrases in Hebrew require the noun also to be marked for definiteness. Again, we can assume that this is down to a potential feature clash because the noun without the definite *ha-* is marked as [DEF –]. The matter is however not as straightforward as with the Latvian adjectives since the bare noun also forms the stem with which *ha-* combines. Furthermore, the unmarked noun can occur in definite noun phrases such as (7); this is a fact to be dealt with as a property specific to the construct state constructions.

Within an LFG approach, adjectival agreement can thus be modelled as a feature with exponence on an adjective being co-specified as a feature of the noun phrase as a whole. This would seem to mean that the equations regulating adjectival agreement have the effect of non-directional phrasal agreement. To our minds, this is a desirable outcome.

We turn now to an account formulated within specification language. We consider the agreement in Scandinavian first. This is captured by (12).

(12) $(\mathit{NP} \wedge \text{<M><DEF>+} \wedge \text{<d*}_k\text{>}\ (\mathit{Adj} \wedge \text{<M><ADJUNCT>})) \rightarrow \text{<d*}_k\text{>}\ \mathit{def}$

If a node labelled NP maps to the f-structure attribute DEF with value + and it dominates a node k which is an adjective which maps to an f-structure attribute ADJUNCT, then this node k is labelled definite. Or in a very simplified form 'if a noun phrase is definite and it contains an adjective, this adjective must be definite'; the adjective agrees with its phrase. (12) ignores the fact that the value of ADJUNCT is set-valued, formally, $\text{<d*}_k\text{>}(\mathit{Adj} \wedge \text{<M><ADJUNCT>})$ should be $\text{<d*}_k\text{>}(\mathit{Adj} \wedge \text{<M>n}'_k \wedge \text{n}'_k \in \{\text{ADJUNCTS}\})$, so rather than referring to 'a node which maps to an f-structure attribute ADJUNCT' we need to refer to 'a node which maps onto an f-structure n'_k which is in the **ADJUNCT**s set'. This simplification is just for exposition.

The constructive definiteness in Hebrew and Latvian, on the other hand is captured by (13) (with a similar simplification to that in (12)).

(13) $(\mathit{NP} \wedge \text{<d*>}\ (\mathit{Adj} \wedge \mathit{def} \wedge \text{<M><ADJUNCT>})) \rightarrow \text{<M><DEF>+}$

If an NP node dominates a node which is an adjective, is definite and maps to an f-structure attribute ADJUNCT, then this NP node maps to an f-structure attribute DEF with value +. In a sense this is then the inverse of (12) 'if a noun phrase

[7] It will of course also be necessary to invoke some mechanism which creates a PRED value for the noun phrase as a whole.

contains a definite adjective, then the noun phrase is definite'; the feature works constructively.

The specification in (12) holds for a language in which there is agreement for definiteness. For a language with constructive definiteness on an adjective, (13) holds. In languages that have constructive definiteness on the adjectives and also agreement on the adjectives, both (12) and (13) hold.

Comparing the two approaches now, both can account for the data accurately. The fact that the value of ADJUNCT is set valued causes some inelegance in the specification language approach and requires an otherwise unwarranted notational convention. The LFG approach requires the use of constraining equations, a standard feature of LFG, but as Blackburn and Gardent (1995, 43–4) point out, constraining equations involve a departure from the declarative model generally central to LFG. A constraining equation provides a check on an attribute-value pair and hence requires the relevant f-structure already to have been built. The specification language cannot capture this notion, but Blackburn and Gardent suggest this may not be a drawback.

4 Definiteness on the edge

4.1 Prosodic marking

In Ossetic (Iron variety, Abaev (1959); Bagaev (1965)), the core noun phrase has phrasal stress which falls either on the first or second syllable:

- if the vowel of the first syllable is strong (/i, e, a, o, u/), then stress is on the first syllable
- if the vowel of the first syllable is weak (/æ, ə/), then stress is on the second syllable

There is no segmental marker of definiteness, but definiteness is indicated by a shift of stress to the noun-phrase initial syllable.[8]

Consider (14), in which the noun phrase consists just of the head noun *læppu* 'boy'.

(14) læp**pú** ∼ **lǽ**ppu (Ossetic)
boy boy.DEF
'a boy' 'the boy'

Noun phrases in Ossetic correspond to phonological phrases. The phonological phrase /*læppu*/ consists of two syllables, the first of which contains a weak vowel. Phrasal stress will therefore by default fall on the second syllable, giving /*læpppú*/.

[8]This is similar to Tongan, but Tongan also marks definiteness with an element which can be described as an article, see Poser (1985) and also Anderson (1992, 212–5) and (2005, 94–99).

If stress by default does not fall on the first syllable of the phrase, then stress on the first syllable indicates that the noun phrase is definite. If on the other hand a noun phrase consists just of a word such as *áxoræn* 'paint', whose first syllable contains a strong vowel, then stress will anyway fall on the first syllable and stress shift cannot apply. The noun phrase is in this case contextually interpreted as either definite or indefinite.

In (15), we see the operation of these principles at phrasal level.

(15) a. c'æx **á**xoræn (Ossetic)
blue paint
'blue paint'

b. **c'ǽx** axoræn
blue.DEF paint
'the blue paint'

In (15-a), the noun phrase consists of an adjective modifying a head noun. Since the adjective *c'ǽx* 'blue' consists of a single syllable with a weak vowel, phrasal stress will by default fall on the first syllable of he following noun. Shifting the stress to the adjective triggers a definite interpretation, as in (15-b). Note that phrasal stress does not always by default fall on the second syllable. Unassimilated Russian loan words, as in (16) can have the stress on some subsequent syllable even if the first one has a strong vowel and hence the stress shift can apply:

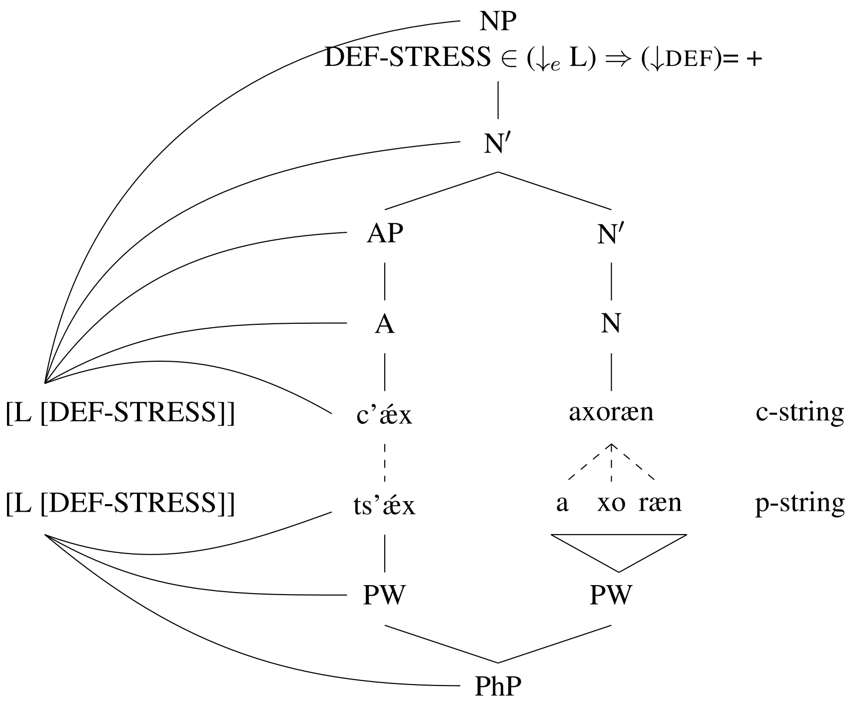
(16) special**íst** ∼ **spé**cialist (Ossetic)
'a specialist' 'the specialist'

The origin of the stress shift rule appears to be the former presence in Ossetic of an initial syntactic definite article consisting of a sugle vowel. This definite article is still attested in the more archaic Digor variety of Ossetic (Abaev (1959, 20), Isaev (1966, 33–4)).

In order to model definiteness marking in Ossetic, we require the f-structure feature DEF to be associated with a stress shift in the phonological phrase which corresponds to the noun phrase that maps to that attribute, i,e. reference needs to be made to both c-structure and p-structure.

A set of mechanisms for achieving this has been proposed by Dalrymple and Mycock (2011) and by Mycock and Lowe (2013). These proposals are in essence similar, and both require extensions to the standard LFG architecture. The basic idea is that a meaning-related feature projected from a c-structure phrase can be passed down the categories on the left or right edge of that phrase, landing ultimately on the leftmost or rightmost word. This word will in turn form part of a pairing between an s-string (syntactic string) and p-string (phonological string). The same meaning-related feature is also passed down the left or right edge of the corresponding prosodic structure from the level where it is operative to the leftmost

(17) Mapping for (15-b)



or rightmost prosodic word in the p-string. An interface harmony principle then ensures that the leftmost or rightmost elements in the c-string and p-string share the relevant feature. In the above-mentioned works, semantic features such as polar interrogation are passed down the right edge of c-structure and match the corresponding right-edge intonation pattern in p-structure. But the same mechanisms can equally apply to f-structure features.

In (17), we give a simplified representation of the various mappings we need for the definite interpretation illustrated by (15-b). This follows the proposals in Dalrymple and Mycock (2011), and uses a tree structure for the representation of prosody rather than an AVM.

The NP node will be associated with a conditional equation which says that if the epsilon structure (e-structure) corresponding to NP has a feature which we name DEF-STRESS in its set of L(eft) attirbutes, the f-structure corresponding to NP is valued + for the feature DEF. e-structure is a distinct dimension in which edge-features are gathered and shared with the categories along the edge of the phrase, ultimately including the adjective *c’æx* ‘blue’. A parallel chi-structure operates to pass the DEF-STRESS feature associated with the phonological phrase down the left-edge of the prosodic structure, where it will be manifested on the first syllable. Rules internal to the prosodic structure will determine that DEF-STRESS can only be assigned to a syllable which is not stressed by default. Consequently, the f-structure DEF feature is not, by these mechanisms, assigned to an NP in which the first word would carry stress by default. Finally, the interface harmony principle ensures that the leftmost elements in the paired c-string and p-string share the DEF-STRESS feature. Mycock and Lowe (2013) suggest an alternative to the location of edge features in separate e-structures and chi-structures, whose role is solely to pass the feature information down the relevant edges. Instead, they propose that the edge nodes in c-structure and p-structure be treated as AVMs, in which case the L and R attributes can be included alongside category information in the representation of these nodes. The basic mechanism of passing the interface features to the c-string and p-string is however essentially the same.

There are two striking aspects to these extensions to the basic LFG model, Firstly, as Mycock and Lowe make explicit, they involve the passing of feature information DOWN the edge of a tree. This is the reverse of the basic construction algorithm, which builds structure upwards. In fact, we note that it is probably useful to have features passing along edges in both directions. Whilst it makes sense to think of a feature such as polar interrogation originating on a clausal node and passing down to the right edge of the c-structure where it is eventually realised phonologically as a nuclear rise, it would make more sense to think of DEF-STRESS as originating at a p-structure edge and being passed upwards to the NP node in a manner analogous to the upward construction of the DEF feature by syntactic or morphological DDMs.

Secondly, the interface harmony principle which requires c-structure and p-structure features to harmonise has affinities with constraining equations. The only difference is that it applies to enforce agreement between nodes in different dimensions rather than between nodes in the same dimension. It gives rise therefore to the same conceptual concern: namely that structures are created which must then be ruled out by a principle which requires access to multiple representations in order to decide which are the acceptable ones. The specification language is in principle unable to state such principles, since it must state declaratively what the well-formed structures are.

As a specification language statement of definiteness marking in Ossetic, we propose (18). The new modality $<\sigma_1>$ will have the obvious interpretation: it will denote a transition from a phonological node to the node representing the first syllable dominated by that node.

(18) $$(\mathrm{NP} \wedge <\mathrm{S}><\sigma_1><\mathrm{STRESS}>+ \wedge <\mathrm{d}_1{}^*><\mathrm{S}><\sigma_1><\mathrm{STRESS}>-) \rightarrow <\mathrm{M}><\mathrm{DEF}>+$$

If an NP node which maps onto a phonological unit (phrase) whose first syllable is stressed and this NP dominates a leftmost node which maps onto a phonological unit (word) whose first syllable is unstressed, then this NP node maps to an f-structure attribute DEF with value +.

4.2 Segmental marking

The Bulgarian FDM appears to be a second position prosodically dependent element ("special clitic" according to Anderson (2005, 111)):

(19) a. knigi-te (Bulgarian)
books-DEF
'the books'

b. interesni-te knigi
interesting-DEF books
'the interesting books'

c. mnogo-to interesni knigi
many-DEF interesting books
'the many interesting books'

However, there are two major problems with this characterization. Firstly, the Bulgarian FDM shows morphophonological irregularities not predicted by this approach. The form of the FDM is dependent on partially arbitrary lexical, morphological and phonologic criteria and the FDM can trigger stem allomorphy (see Bermúdez-Otero and Payne (2011, 74–5) and Stojanov (1964)). A small subset of these irregularities is illustrated in (20).[9]

(20) a. grăk ∼ gărk-ăt vs. străk ∼ străk-ăt
Greek Greek-DEF stalk stalk-DEF (Bulgarian)

b. gnjav ∼ gnev-ằt vs. bljan ∼ bljan-ăt
anger anger-DEF dream dream-DEF

In (20-a) the addition of the FDM triggers metathesis in *grăk*, and in **??** iy triggers a syllable nucleus alternation in *gnjav*. The forms *străk* and *bljan* remain unaffected.

Secondly, as (21) illustrates, the positioning cannot be defined straightforwardly with respect to 'first word' (or even 'first phrase').

(21) a. naj-blizka-ta do pošta-ta kăšta (Bulgarian)
SUPERL-close-DEF to post office-DEF house
'the house closest to the post office'

b. tvărde interesna-ta kniga
very interesting-DEF book
'the very interesting book'

In (21), we see the placement of the FDM when we have a phrasal rather than lexical dependent of the noun as the leftmost category. In this example the leftmost category is an adjective phrase, but the same would apply to other phrasal dependents in this position, e.g. numeral phrases. In (21-a), the FDM must be located

[9] In contemporary Bulgarian there is some levelling in the case of (20-a), with some speakers preferring *grăk-ăt* (Bozhil Hristov, p.c.).

on the adjectival head, and not to the right of the phrase on the PP. In (21-b), it must likewise be marked on the adjectival head, and not on the preceding adverb modifier. An account of the FDM in Bulgarian needs therefore not only to account for the irregular morpho-phonological interaction with the host, but also to be able to make reference to the notion of 'head of leftmost daughter' for placement. This is not straightforward in any theory.

Morphosyntactic edge phenomena, in contrast to prosodic edge phenomena, appear to have been somewhat neglected in LFG, and there appears to be no off-the-peg solution even for a straightforward case in which the first word of a phrase is targeted. We could in such straightforward cases hijack the L(eft) attribute proposed by Dalrymple and Mycock (2011) for semantic-prosodic interface phenomena, and allow the membership of L in principle to include morphosyntactic features such as *def* which would be passed down the edge of the noun phrase, either in e-structure or as part of the AVM extension to left-edge categories. But then it would be necessary to somehow distinguish such features from the interface features such as DEF-STRESS which are subject to the interface harmony principle. Assume this can be done, perhaps by having two types of L(eft) attribute, one of which contains morphosyntactic features and the other interface features. We could then attach a conditional equation to NP of the form: $def \in (\downarrow_e \text{L}) \Rightarrow (\downarrow\text{DEF}) = +$. Such an equation would ensure that if *def* were in the set of left-edge morphosyntactic features, the f-structure of the NP would contain the value + for the DEF feature. But there is still a problem, since this would not account for the head marking of adjective phrases in examples like (21-b). Possibly there are fixes which would work, e,g, by forcing L features in languages like Bulgarian to pass to the head rather than the edge whenever they encounter a phrasal node on the edge. This would entail an extension to the chi-structure or AVM model of L features.

The specification language alternative is given in (22).

(22) $(NP \wedge \langle \mathrm{d}_1\rangle def \vee \langle \mathrm{d}_1\rangle\langle \mathrm{d}^*\rangle(\mathrm{head} \wedge def)) \rightarrow \langle\mathrm{M}\rangle\langle\textsc{def}\rangle+$

If an NP node has either a first daughter labelled definite or a first daughter whose head is labelled definite, then this NP node maps to an f-structure attribute DEF with value +. The first clause within the parenthesis in the antecedent will allow definiteness marking on the head noun, if this happens to be the first element in the phrase, and the second clause will allow definiteness marking on the head of any initial phrasal dependent. The two clauses will never both be satisfied: since the exponents of definiteness marking in Bulgarian are words and not phrases, we cannot have a situation is which definiteness is marked both on an initial phrase and on its head.

Note however that (22) works constructively: only the left edge feature constructs definiteness. It does not however specifically exclude the possibility of multiple definiteness marking. Subsequent elements in the noun phrase, for example multiple adjectives, might optionally be marked with definiteness non-constructively. In order to block this possibility, we impose the further requirement (assuming binary branching) as in (23).

(23) $(NP \wedge \text{<M><}\textsc{def}\text{>+} \rightarrow \sim (< d_2 > def \vee < d_2 >< d^* > def)$

If an NP maps to a DEF attribute with value +, then neither its second daughter nor any terminal node that its second daughter dominates is marked definite. This is the negative counterpart of (12), blocking definiteness agreement rather than requiring it. We would only expect such clauses in edge-marking cases.

5 Conclusions

In this paper, we have used definiteness marking in noun phrases to illustrate two different approaches to the formalisation of constraints in LFG. In many cases, the basic construction algorithm with up and down designators straightforwardly accounts for the mapping between c-structure exponents of definiteness and the f-structure feature DEF. This is the case for example when the exponent of definiteness is a syntactic definiteness marker such as an article, or a morphological marker of definiteness on a head noun. In more complex cases, such as when definiteness is marked on adjectives, or is marked on an edge, the basic construction algorithm has to be augmented with a variety of further mechanisms. In particular, we need constraining equations to enforce the non-constructive definiteness agreement found in languages like Swedish. We need additional structures such as e-structure and chi-structure to provide a home for features which are passed along edges, or alternatively an alternative conception of c-structure nodes as AVMs which can contain such features. These features need, at least in some cases, to be passed down an edge, in the opposite direction to the standard construction algorithm. And finally, semantic or functional features which have an exponent in prosodic structure require an interface harmony principle which enforces the presence of this semantic or functional feature at the same position in the c-string and p-string.

These mechanisms may do the trick, but they look in some respects anomalous within the LFG architecture. The specification language which we have used to formulate the generalisations involved in these more complex cases is relatively perspicuous, but it also serves to highlight the anomalies which some of the augmentations involve. As Blackburn and Gardent (1995) note, constraining equations cannot be stated in the specification language format since they require access to multiple structures, some of which will be discarded. As we have noted, the interface harmony principle is a similar beast: it acts in effect as a constraining equation on c-strings and p-strings.

Our purpose in this article is therefore to suggest that the specification language approach to LFG might be an approach worth exploring in greater detail. It might have a particular advantage, for example, in cases where the analyst might be tempted to employ constraining mechanisms, or where it is necessary to pass features in ways which the standard construction algorithm does not allow, for example downwards rather than upwards, or from one dimension into another where there is no direct construction involved.

References

Abaev, V.I. 1959. *Grammatičeskij očerk osetinskogo jazyka [A Grammatical Sketch of Ossetic].*. Ordžonikidze: Severo-Osetinskoe Knižnoe Izdanie.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Bagaev, N.Ya. 1965. *Sovremennyj osetinskij yazyk, č.1 [Modern Ossetic language, pt. 1].*. Ordžonikidze: Severo-Osetinskoe Knižnoe Izdanie.

Bermúdez-Otero, Ricardo and Payne, John. 2011. There are no special clitics. In Alexandra Galani, Glyn Hicks and George Tsoulas (eds.), *Morphology and its interfaces*, pages 57–96, Amsterdam: John Benjamins.

Blackburn, Patrick and Gardent, Claire. 1995. A specification language for Lexical Functional Grammars. In *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, pages 39–44, Assocation for Computational Linguistics.

Börjars, Kersti. 2003. Morphological status and (de)grammaticalisation: the Swedish possessive. *Nordic Journal of Linguistics* 26, 133—163.

Börjars, Kersti, Denison, David, Krajewski, Grzegorz and Scott, Alan. 2013. Expression of possession in English. The significance of the right edge. In Kersti Börjars, David Denison and Alan Scott (eds.), *Morphosyntactic categories and the expression of possession*, pages 123–148, Amsterdam: John Benjamins.

Butt, Miriam and King, Tracy Holloway. 1998. Interfacing phonology with LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, Stanford, CA: CSLI Publications.

Corbett, Greville G. 2003. The role of agreement in natural language. In William Earl Griffin (ed.), *TLS 5 Proceedings [Texas Linguistics Forum 53]*, pages 109–22, Texas Linguistics Society [http://uts.cc.utexas.edu/ tls/].

Dalrymple, Mary. 2001. *Lexical Functional Grammar*. New York: Academic Press.

Dalrymple, Mary and Mycock, Louise. 2011. The prosody-semantics interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*, pages 173–93, Stanford, CA: CSLI Publications.

Delsing, Lars-Olof. 1994. 'Hans sjukt ben' – om starka och svaga adjektiv i Fornsvenskan. In Nils Jörgensen, Christer Platzack and Jan Svensson (eds.), *Språkbruk, grammatik och språkförändring. En festskrift till Ulf Teleman*, pages 99–108, Lund: Lund University.

Delsing, Lars-Olof. 2003. Syntaktisk variation i nordiska nominalfraser. In Øystein Alexander Vangsnes, Anders Holmberg and Lars-Olof Delsing (eds.), *Dialektsyntaktiska studier av den nordiska nominalfrasen*, pages 65–84, Oslo: Novus.

Falk, Yehuda N. 2006. On the representation of case and agreement. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG06 Conference*, http://cslipublications.stanford.edu/LFG/11/lfg06.html: Stanford, CA: CSLI Publications.

Isaev, M I. 1966. *Digorskij dialekt osetinskogo jazyka*. Moskva: Nauka.

Kaplan, Ronald M. 1995. The formal architecture of Lexical-Functional Grammar. In Mary Dalrymple, Ronald M Kaplan, John T Maxwell III and Annie Zaenen (eds.), *Formal issues in LexicalFunctional Grammar*, pages 7–27, Stanford, CA: CSLI Publications.

Lapointe, Steven G. 1990. EDGE features in GPSG. In Michael Ziolkowski, Manuela Noske and Karen Deaton (eds.), *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, pages 221—35, Chicago, IL: Chicago Linguistic Society.

Lapointe, Steven G. 1992. Life on the edge: argument in favor of an Autolexical account of edge inflections. In Costas P Canakis, Grace P Chan and Jeanette Marshall Denton (eds.), *Papers from the 28th Regional Meeting of the Chicago Linguistic Society*, pages 318—33, Chicago, IL: Chicago Linguistic Society.

Lehmann, Christian. 1982. Universal and typological aspects of agreement. In Hansjakob Seiler and Franz J. Stachowiak (eds.), *Apprehension: das sprachliche Erfassen von Gegenständen II*, pages 201– 267, Tübingen: Narr.

Miller, Philip. 1992. Morphological marking misses the head. In Jonathan Mead (ed.), *11th West Coast Conference on Formal Linguistics*, Stanford, CA: CSLI Publications.

Miller, Philip and Halpern, Aaron. 1993. English possessives and the syntax of morphological features. In Laurel Smith Stvan (ed.), *FLSM III: Papers from the third Annual Meeting of the Formal Linguistics Society of MidAmerica*, pages 219—34, Bloomington, IN: Indiana University Linguistics Club.

Mycock, Louise and Lowe, John. 2013. The prosodic encoding of discourse functions. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, Stanford, CA: CSLI Publications.

Nordlinger, Rachel. 1998. *Constructive case: evidence from Australian languages*. Stanford, CA: CSLI Publications.

Pollard, Carl and Sag, Ivan. 1994. *Head-driven Phrase Structure Grammar*. Stanford, CA: CSLI Publications.

Poser, William J. 1985. Cliticization to NP in lexical phonology. In Jeffrey Goldberg, Susannah MacKaye and Michael T Wescoat (eds.), *Proceedings of WCCFL IV*, pages 262–272, Stanford Linguistics Association.

Potts, Christopher. 2002. Comparative Economy Conditions in Natural Language Syntax, paper presented at the North American Summer School in Logic, Language, and Information 1, Workshop on Model-Theoretic Syntax, Stanford University (June 28).

Pullum, Geoffrey K. 2013. The central question in comparative syntactic metatheory. *Mind & Language* 28(4), 492–521.

Rogers, James. 2003. wMSO theories as grammar formalisms. *Theoretical Computer Science* 293, 291–320.

Spencer, Andrew and Luís, Ana. 2012. Canonical clitics. In Dunstan Brown, Marina Chumakina and Grev G Corbett (eds.), *Canonical morphology and syntax*, pages 123–150, Oxford University Press.

Stojanov, St. 1964. *Gramatika na bălgarskija knižoven ezik*. Sofia: Nauka i Izkustvo.

Zwicky, Arnold M and Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English n't. *Language* 59, 502–513.