Proceedings of the Fifth Meeting on Mathematics of Language (MOL-5),
Tilman Becker and Hans-Ulrich Krieger (eds), DFKI-D-97-02, 1997.

A Dependency-based Approach to Bounded & Unbounded Movement

Mark Hepple
Dept. of Computer Science
University of Sheffield
Regent Court, Portobello Street
Sheffield S1 4DP, UK
hepple@dcs.shef.ac.uk

1 Introduction

This paper addresses the treatment of movement phenomena within *multimodal* categorial, or type-logical, grammar systems. Multimodal approaches allow different modes of logical behaviour to be displayed within a single system. Intuitively, this characteristic corresponds to making available different modes of linguistic description within a single formalism. A key benefit of taking a multimodal approach is that it allows us to choose, for any linguistic phenomenon addressed, a level of description that encodes only the aspects of linguistic structure that are relevant to the treatment of that phenomenon. In practice, this means that we may lexically encode linguistic information which is relevant to one phenomenon but not another, but can discard such information where it is not needed. This characteristic means that the analysis of each phenomenon need focus only on relevant distinctions, allowing analyses to be simpler and more elegant.

In this paper, we are concerned particularly with how locality constraints on movement should be handled, both for bounded and unbounded movement cases. A central claim of the paper is that the treatment of such locality conditions requires representations that encode dependency (i.e. head-dependent distinctions).

2 Multimodal Categorial Grammar

The multimodal categorial approach used here makes available multiple modes of construction, realised in syntax via different product operators $\circ_\alpha$ (each with associated implicationals[1] $\stackrel{\alpha}{\rightarrow}$, $\stackrel{\alpha}{\leftarrow}$), whose behaviour reflects the axioms (e.g. associativity) governing the corresponding operator in the underlying interpretive semantics. Further axioms allow *interaction* between modes (e.g. $x \circ_i (y \circ_j z) = (x \circ_i y) \circ_j z$), and 'linkage' (e.g. $x \circ_i y \longrightarrow x \circ_j y$), i.e. movement from one mode to another. Axioms divide into three classes: (i) *mode internal axioms*, which involve only a single modality, e.g. the familiar associativity axiom $x \circ_i (y \circ_i z) = (x \circ_i y) \circ_i z$; (ii) *interaction axioms*, involving more than one modality, e.g. $x \circ_i (y \circ_j z) = (x \circ_i y) \circ_j z$; (iii) *linkage*

[1] These 'associated implications' correspond to the connectives that are typically notated as \ and / in familiar categorial systems such as the associative Lambek calculus, which is a *unimodal* system having a single associative product operator (typically notated as $\bullet$).

or *inclusion axioms*, allowing movement from one mode to another, e.g. $x \circ_i y \longrightarrow x \circ_j y$.[2] Intuitively, the move from one mode to another allowed by a linkage axiom is akin to movement from one description of a linguistic object to an alternative, less informative, description.

We adopt a *labelled* natural deduction formulation, employing inference rules (1–3) below.[3] Labelled formulae take the form: $m \vdash A : s$, with A a type, s a 'semantic' lambda term, and m a *marker* term, the latter being a structured object built up as deduction proceeds, that records information used in ensuring appropriate structure sensitivity in deduction. Hence this system is an instance of a *labelled deductive system* (Gabbay [2]). In linguistic derivations, lexical assumptions have lexically provided marker and semantic components (loosely, the word's 'string' or 'phonology' and its meaning). In all other assumptions (i.e. any additional assumptions, used in hypothetical reasoning, that are eventually discharged), these terms are a simple variable. The role of marker terms here, in recording the proof's significant structural information, closely parallels that of structured configurations of types in various sequent and natural deduction logical formulations, but differing perhaps in that they provide a somewhat more concise/readable representation of the proof's significant structural information. In a linguistic context, a marker may be viewed as providing a description of the linguistic structure of the object derived.

(1)
$$\frac{s \vdash \mathrm{A}\overset{\alpha}{\leftarrow}\mathrm{B} : a \qquad t \vdash \mathrm{B} : b}{(s \circ_\alpha t) \vdash \mathrm{A} : (a\ b)}\,\overset{\alpha}{\leftarrow}\mathrm{E} \qquad\qquad \frac{\begin{array}{c}[v \vdash \mathrm{B} : v] \\ (s \circ_\alpha v) \vdash \mathrm{A} : a\end{array}}{s \vdash \mathrm{A}\overset{\alpha}{\leftarrow}\mathrm{B} : \lambda v.a}\,\overset{\alpha}{\leftarrow}\mathrm{I}$$

(2)
$$\frac{t \vdash \mathrm{B} : b \qquad s \vdash \mathrm{B}\overset{\alpha}{\rightarrow}\mathrm{A} : a}{(t \circ_\alpha s) \vdash \mathrm{A} : (a\ b)}\,\overset{\alpha}{\rightarrow}\mathrm{E} \qquad\qquad \frac{\begin{array}{c}[v \vdash \mathrm{B} : v] \\ (v \circ_\alpha s) \vdash \mathrm{A} : a\end{array}}{s \vdash \mathrm{B}\overset{\alpha}{\rightarrow}\mathrm{A} : \lambda v.a}\,\overset{\alpha}{\rightarrow}\mathrm{I}$$

(3)
$$\frac{\begin{array}{l}[v \vdash \mathrm{B} : v],[w \vdash \mathrm{C} : w] \qquad t \vdash \mathrm{B}\circ_\alpha\mathrm{C} : b \\ s[(v \circ_\alpha w)] \vdash \mathrm{A} : a\end{array}}{s[t] \vdash \mathrm{A} : [b/(v \circ w)].a}\,\circ_\alpha\mathrm{E} \qquad\qquad \frac{s \vdash \mathrm{A} : a \qquad t \vdash \mathrm{B} : b}{(s \circ_\alpha t) \vdash \mathrm{A}\circ_\alpha\mathrm{B} : \langle a, b\rangle}\,\circ_\alpha\mathrm{I}$$

Additional *structural rules*, which directly reflect axioms of the underlying semantics, act to modify the form of the marker and thereby affect the derivability relation. For example, the associativity rule [a] in (4), which mirrors the associativity axiom $(x \circ_i (y \circ_i z)) = ((x \circ_i y) \circ_i z)$, is needed to enable derivation of the 'simple composition' theorem $\mathrm{X}\overset{i}{\leftarrow}\mathrm{Y},\ \mathrm{Y}\overset{i}{\leftarrow}\mathrm{Z} \Rightarrow \mathrm{X}\overset{i}{\leftarrow}\mathrm{Z}$, as illustrated in (7). Note that a system with only a single modality plus the rules (1–4) constitutes a formulation of the associative Lambek calculus. Further examples of structural rules (permutation and linkage) are shown in (5,6). Proof (8) illustrates how the linkage rule allows modality change within an implicational type.

[2]Hepple [4, 5] and Moortgat & Oehrle [9] introduce multimodal frameworks which, like the one to be described in this paper, allow juxtaposition of different levels of the substructural hierarchy of logics, with movement between levels allowed by linkage axioms. Interestingly, the two groups take precisely opposing views as to what constitutes an appropriate pattern of linkage between levels. Kurtonina [7] shows that both views are have well-founded interpretive semantics. There are other proposals that are also multimodal, in the sense of including multiple groups of operators within a single system, with patterns of derivability between different operators, e.g. Morrill [11], Morrill & Solias [12].

[3]A formula in square brackets here indicates an assumption that is discharged by a rule's use. For example, the [$\overset{\alpha}{\leftarrow}$I] rule indicates that given a proof of a formula of type A which rests on an assumption of type B, we can discharge that assumption to construct a proof of a formula with type $\mathrm{A}\overset{\alpha}{\leftarrow}\mathrm{B}$. Note that in (3), $s[(v \circ_\alpha w)]$ and $s[t]$ refer to marker terms that are identical except that where $(v \circ_\alpha w)$ appears as a subterm in the former, t appears instead in the latter.

(4) $$\frac{s[(x\circ_i(y\circ_i z))]\vdash \mathrm{A}:a}{s[((x\circ_i y)\circ_i z)]\vdash \mathrm{A}:a}\,[\mathrm{a}]$$

(5) $$\frac{s[(x\circ_i y)]\vdash \mathrm{A}:a}{s[(x\circ_j y)]\vdash \mathrm{A}:a}\,[i/j]$$

(6) $$\frac{s[(x\circ_i y)]\vdash \mathrm{A}:a}{s[(y\circ_i x)]\vdash \mathrm{A}:a}\,[\mathrm{p}]$$

(7) $$\dfrac{\dfrac{\dfrac{x\vdash \mathrm{X}\overset{i}{\leftarrow}\mathrm{Y}:x \qquad \dfrac{y\vdash \mathrm{Y}\overset{i}{\leftarrow}\mathrm{Z}:y \qquad [z\vdash \mathrm{Z}:z]}{(y\circ_i z)\vdash \mathrm{Y}:(yz)}}{(x\circ_i(y\circ_i z))\vdash \mathrm{X}:(x(yz))}}{((x\circ_i y)\circ_i z)\vdash \mathrm{X}:(x(yz))}\,[\mathrm{a}]}{(x\circ_i y)\vdash \mathrm{X}\overset{i}{\leftarrow}\mathrm{Z}:\lambda z.(x(yz))}$$

(8) $$\dfrac{\dfrac{\dfrac{x\vdash \mathrm{X}\overset{i}{\leftarrow}\mathrm{Y}:x \qquad [y\vdash \mathrm{Y}:y]}{(x\circ_i y)\vdash \mathrm{X}:(xy)}}{(x\circ_j y)\vdash \mathrm{X}:(xy)}\,[i/j]}{x\vdash \mathrm{X}\overset{j}{\leftarrow}\mathrm{Y}:\lambda y.(xy)}$$

Regarding the linear order (i.e. word order) consequences of proofs, note that we cannot simply look to the order of assumptions as they are written on the page, since not all modalities carry simple ordering import (and hence likewise their associated connectives). However, we can 'read off' order information from a proof's marker term, *provided* it is constructed only using modalities that do have simple linear import (i.e. are not subject to any permutative axioms). Only proofs that have such markers can serve adequately as linguistic derivations.[4]

3 Categorial Analysis of Movement

The basic treatment of movement rests on being able to derive a type of the form $\mathrm{Y}\overset{\alpha}{\leftarrow}\mathrm{Z}$ (informally a phrase Y missing a subphrase Z) for the material forming the extraction domain.[5] Movement is allowed by assigning the displaced element an additional 'movement' type such as $\mathrm{X}\overset{i}{\leftarrow}(\mathrm{Y}\overset{j}{\leftarrow}\mathrm{Z})$ which can (for leftward movement) prefix to the extraction domain. For example, a relative pronoun might have a type $\mathrm{Rel}\overset{i}{\leftarrow}(\mathrm{s}\overset{j}{\leftarrow}\mathrm{np})$, so it can combine with a 'sentence missing np' to give a relative clause.

We shall illustrate this approach in relation to a multimodal system, having three modalities: n (non-associative, non-permutative), a (associative, non-permutative), and c (associative, permutative), for which we assume the structural rules [a] and [p] above to be appropriately conditioned. Further, we assume that the schematic linkage rule $[i/j]$ has permissible instances $[n/a]$, $[n/c]$ and $[a/c]$. Consider a relative pronoun type $\mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{c}{\leftarrow}\mathrm{np})$, and the derivation (9) it allows of the relative clause *who saw kim*.

[4]There are some further aspects to a more complete presentation of the particular multimodal approach described here, which are described in detail in Hepple [6]. In particular, it is shown that explicit marker terms within proofs can be eliminated provided that proof term representations (i.e. the lambda terms encoding functional structure) are augmented with modality information, since marker terms can be directly computed from such enriched proof terms, and hence including explicit marker terms within proofs would then be redundant. The correctness tests on inference rule uses, performed above upon marker terms, can instead then be based upon the enriched proof terms. This development of the system allows for an approach were lexical items may be associated with string components that may be complex terms built using the operators of the proof term algebra (rather than just simple atoms), a move which amounts to allowing lexical encoding of partial proof structure.

[5]This general approach to extraction, depending on the 'flexible deduction' characteristic of many categorial systems, has been widely used within categorial work, and adapts ultimately from the proposals of Ades & Steedman [1].

(9)

$$\dfrac{who \vdash \mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{c}{\leftarrow}\mathrm{np}):\mathbf{who'} \qquad \dfrac{\dfrac{\dfrac{\dfrac{v \vdash \mathrm{np}:v \qquad \dfrac{saw \vdash (\mathrm{np}\overset{n}{\rightarrow}\mathrm{s})\overset{n}{\leftarrow}\mathrm{np}:\mathbf{saw'} \qquad kim \vdash \mathrm{np}:\mathbf{kim'}}{(saw \circ_n kim) \vdash \mathrm{np}\overset{n}{\rightarrow}\mathrm{s}:(\mathbf{saw'}\ \mathbf{kim'})}}{(v \circ_n (saw \circ_n kim)) \vdash \mathrm{s}:(\mathbf{saw'}\ \mathbf{kim'}\ v)}}{(v \circ_c (saw \circ_n kim)) \vdash \mathrm{s}:(\mathbf{saw'}\ \mathbf{kim'}\ v)}[n/c]}{((saw \circ_n kim) \circ_c v) \vdash \mathrm{s}:(\mathbf{saw'}\ \mathbf{kim'}\ v)}[\mathrm{p}]}{(saw \circ_n kim) \vdash \mathrm{s}\overset{c}{\leftarrow}\mathrm{np}:\lambda v.(\mathbf{saw'}\ \mathbf{kim'}\ v)}}{(who \circ_n (saw \circ_a kim)) \vdash \mathrm{Rel}:\mathbf{who'}(\lambda v.(\mathbf{saw'}\ \mathbf{kim'}\ v))}$$

Such extraction derivations involve an additional assumption (here of type np) appearing in the canonical place of the displaced phrase, which is subsequently discharged in an $[\overset{\alpha}{\leftarrow}\mathrm{I}]$ step, creating an implicational type $\mathrm{Y}\overset{\alpha}{\leftarrow}\mathrm{Z}$ — a 'Y missing Z'. The structure sensitivity of extraction depends crucially on this $[\overset{\alpha}{\leftarrow}\mathrm{I}]$ step, requiring the immediate subproof's marker to restructure to the form $(s \circ_\alpha v)$ (v being the discharged assumption's marker variable). Whether this restructuring is possible or not depends on the structural characteristics of the proof (or, if you prefer, the structure of the phrase), and on the specific modality α required, which is itself determined by the 'movement category' of the displaced phrase, e.g. modality c for the category $\mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{c}{\leftarrow}\mathrm{np})$ in the above proof. This modality allows linkage inferences that change the marker to a weaker description that allows use of the permutation rule, and hence this relative pronoun type can allow subject extraction, as it does in (9), but can equally well allow object extraction, as in the derivation (10) of *who kim saw*.

(10)

$$\dfrac{who \vdash \mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{c}{\leftarrow}\mathrm{np}):\mathbf{who'} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{kim \vdash \mathrm{np}:\mathbf{kim'} \qquad \dfrac{saw \vdash (\mathrm{np}\overset{n}{\rightarrow}\mathrm{s})\overset{n}{\leftarrow}\mathrm{np}:\mathbf{saw'} \qquad v \vdash \mathrm{np}:v}{(saw \circ_n v) \vdash \mathrm{np}\overset{n}{\rightarrow}\mathrm{s}:(\mathbf{saw'}\ v)}}{(kim \circ_n (saw \circ_n v)) \vdash \mathrm{s}:(\mathbf{saw'}\ v\ \mathbf{kim'})}}{(kim \circ_a (saw \circ_a v)) \vdash \mathrm{s}:(\mathbf{saw'}\ v\ \mathbf{kim'})}[n/a]^*}{((kim \circ_a saw) \circ_a v) \vdash \mathrm{s}:(\mathbf{saw'}\ v\ \mathbf{kim'})}[\mathrm{a}]}{((kim \circ_a saw) \circ_c v) \vdash \mathrm{s}:(\mathbf{saw'}\ v\ \mathbf{kim'})}[a/c]}{(kim \circ_a saw) \vdash \mathrm{s}\overset{c}{\leftarrow}\mathrm{np}:\lambda v.(\mathbf{saw'}\ v\ \mathbf{kim'})}}{(who \circ_n (kim \circ_a saw)) \vdash \mathrm{Rel}:\mathbf{who'}(\lambda v.(\mathbf{saw'}\ v\ \mathbf{kim'}))}$$

Using instead a movement category $\mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{a}{\leftarrow}\mathrm{np})$, requiring modality a for the $[\leftarrow\mathrm{I}]$ step, we find that object extraction is still possible, as in (11), but that subject extraction is not since we cannot now move to a marker allowing permutation. The third alternative of using a movement category requiring modality n for the introduction step would allow neither subject nor object extraction (i.e. since neither [p] nor [a] could be used). Note then that the two connectives in a movement type such as $\mathrm{Rel}\overset{i}{\leftarrow}(\mathrm{s}\overset{j}{\leftarrow}\mathrm{np})$ play very different roles. The principle connective $(\overset{i}{\leftarrow})$ serves to *assign structure* in the usual way, whereas the second embedded connnective $(\overset{j}{\leftarrow})$ instead serves to *test structure*.

(11)

$$\dfrac{who \vdash \mathrm{Rel}\overset{n}{\leftarrow}(\mathrm{s}\overset{a}{\leftarrow}\mathrm{np}) : \mathbf{who'} \qquad \dfrac{\dfrac{\dfrac{\dfrac{kim \vdash \mathrm{np} : \mathbf{kim'} \qquad \dfrac{saw \vdash (\mathrm{np}\overset{n}{\rightarrow}\mathrm{s})\overset{n}{\leftarrow}\mathrm{np} : \mathbf{saw'} \qquad v \vdash \mathrm{np} : v}{(saw \circ_n v) \vdash \mathrm{np}\overset{n}{\rightarrow}\mathrm{s} : (\mathbf{saw'}\ v)}}{(kim \circ_n (saw \circ_n v)) \vdash \mathrm{s} : (\mathbf{saw'}\ v\ \mathbf{kim'})}}{(kim \circ_a (saw \circ_a v)) \vdash \mathrm{s} : (\mathbf{saw'}\ v\ \mathbf{kim'})}[n/a]^*}{((kim \circ_a saw) \circ_a v) \vdash \mathrm{s} : (\mathbf{saw'}\ v\ \mathbf{kim'})}[\mathrm{a}]}{(kim \circ_a saw) \vdash \mathrm{s}\overset{a}{\leftarrow}\mathrm{np} : \lambda v.(\mathbf{saw'}\ v\ \mathbf{kim'})}}{(who \circ_n (kim \circ_a saw)) \vdash \mathrm{Rel} : \mathbf{who'}(\lambda v.(\mathbf{saw'}\ v\ \mathbf{kim'}))}$$

Let us contrast this approach to handling movement constraints to what we perhaps might view as more 'standard' alternatives, which address a single description of linguistic structure and then apply some relatively complex constraint in deciding the permissibility of a given extraction in terms of the relation between the extracted phrase and its canonical position. For a tree-based approach to linguistic description, such a constraint could perhaps be formulated (or reformulated) as some non-trivial tree-traversing automaton. For the present approach, the absolute characterisation of a movement constraint is quite simple, i.e. does a given phrase have a description $(s \circ_\alpha v)$, for some α? The real complexity of the constraint can be seen to lie with the initial rich lexical encoding of linguistic structure and the system of alternative descriptions that the given multimodal approach allows, from which the individual description is selected.

In what follows, our discussion of how to treat locality constraints within this general approach will not be illustrated by proofs for specific cases, but will instead focus directly on marker systems, and whether or not they allow restructuring of the above kind as is appropriate for characterising particular locality behaviour.

4 Dependency and Locality

An earlier account of locality constraints within a type-logical framework is offered by Hepple [3] (extending proposals due to Morrill [10]), which makes available multiple *unary* modalities, that can be marked on particular phrases to make them boundaries for movement. One criticism that can be levelled against this approach is that it is essentially stipulative, in that it allows boundaries to be specified in a way that is independent of other aspects of structure. In contrast, the aim here is to construct an account of locality that is rooted in a more broadly motivated account of linguistic structure, i.e. developing what appears to be an adequate description of linguistic structure for broader purposes, and then using it as the basis for formulating an account of locality constraints.[6]

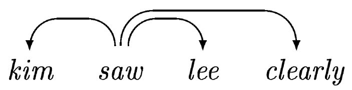
One syntactic distinction that might be encoded by modalities is dependency, i.e. the asymmetry between heads and dependents. This idea was introduced, within categorial work, by Moortgat & Morrill [8], who employ it in a type-logical encoding of metrical trees. A central claim of this paper is that, within a multimodal setting, structures encoding dependency provide an appropriate basis for addressing locality in movement. In particular, such structures allow for the use of a notion of 'domain' for locality which might be termed the 'domain of a head' — i.e. that domain consisting of a head plus its dependents — and allow us to distinguish the cases where the movement of some element does or does not stay within some given domain.

[6]The account of locality constraints to be presented here develops earlier proposals outlined in Hepple [4].

Let us notate left-headed modes using $\succ$, and right-headed ones using $\prec$ (a notation intended to be reminiscent of the 'arrow structures' of dependency grammar, with heads 'pointing' at their dependents), i.e. so that x is head in $(x \succ y)$. A derivation of (e.g.) *Kim saw Lee clearly* might yield a marker such as:

$$(\mathit{kim} \prec ((\mathit{saw} \succ \mathit{lee}) \succ \mathit{clearly})).$$

The use of binary operators here gives hierarchical structures in which some 'head' elements may be complex. This contrasts with the 'flatter' structures of dependency grammar, where all heads are lexical elements, as in e.g.:



To bridge this gap, it is useful to go beyond our purely binary structures to a recursive notion of R-*head* where a single atomic element has multiple dependents, e.g. in $((y \prec x) \succ z)$, atomic x is R-head, having R-dependents y, z (which are the 'immediate dependents' of the 'projections' of x). Additionally, let the 'maximal head' of an expression be the R-head within it that dominates all other R-heads (i.e. they are contained within its R-dependents).

Let us consider a multimodal approach whose modalities include ones that encode the head-dependent asymmetry, as indicated above. The modalities used in specifying lexical types will be ones that are structurally restrictive (and hence more 'informative'), encoding (we might expect) linear order and bracketting (i.e. being non-associative and non-commutative), as well as headedness (i.e. head-dependent asymmetry). This level can be linked to other structurally more-liberal levels, whose behaviour allows for different possibilities of dependency-sensitive movement. Let us imagine one such level (notated $\dot{\prec}$, $\dot{\succ}$), which maintains a head-dependent distinction, and consider how this level may be used in characterising constraints on movement. Note that we are only concerned with locality constraints, and not any putative order-related constraints, and so we shall firstly assume that the following axiom applies at this level, which freely reorders heads and dependents, whilst maintaining the distinction between them, and which hence serves to undermine any effects of order upon what is derivable:

$$\text{[ax1]:} \quad x \mathbin{\dot{\succ}} y \;=\; y \mathbin{\dot{\prec}} x$$

The crucial division between where movement is restricted to be within local domains (in the 'domain of a head' sense indicated above) depends on our choice of axioms from the following two:

$$\text{[ax2]:} \quad x \mathbin{\dot{\prec}} (y \mathbin{\dot{\succ}} z) \;=\; (x \mathbin{\dot{\prec}} y) \mathbin{\dot{\succ}} z$$

$$\text{[ax3]:} \quad x \mathbin{\dot{\succ}} (y \mathbin{\dot{\succ}} z) \;=\; (x \mathbin{\dot{\succ}} y) \mathbin{\dot{\succ}} z$$

Although the axiom [ax2] does not preserve hierarchical head-dependent structure, it does preserve R-heads and R-dependents (i.e. the markers it equates will have the same R-heads, each of which will have the same set of R-dependents). In combination with [ax1], this axiom will permit restructuring that allows any dependent of the maximal head of an expression to 'move up' to topmost hierarchical position, so that a marker α may restructure to the form $\beta \mathbin{\dot{\succ}} x$ (for some β) *iff* x is a dependent of the maximal head of α. Consequently, the corresponding implication ($\overset{\dot{\succ}}{\leftarrow}$) could be used for a version of bounded movement of dependents, i.e. allowing an element to move to the periphery of the domain of its R-head but not beyond. However, this system appears too restrictive for most purposes. For example, it is well known that adverbial adjuncts cannot in general be extracted from embedded clauses (as illustrated in (12)). However, purely head-bounded movement is too restrictive for this phenomenon, as shown by (13b).

(12) a. John [$_{\text{vp}}$ [$_{\text{vp}}$ opened the box] [$_{\text{adv}}$ with a crowbar]]

b. How$_{\text{i}}$ did John [$_{\text{vp}}$ [$_{\text{vp}}$ open the box] $\_\_{i}$]

c. *How$_{\text{i}}$ do you remember that John [$_{\text{vp}}$ [$_{\text{vp}}$ opened the box] $\_\_{i}$]
(* under intended reading)

(13) a. John wants [$_{\text{vp}}$ [$_{\text{vp}}$ to leave] tomorrow]

b. When$_i$ does John want [$_{\text{vp}}$ [$_{\text{vp}}$ to leave] $\_\_{i}$]

In contrast to [ax2], the axiom [ax3] preserves neither heads nor dependents in general (nor, indeed, either R-heads or R-dependents). The move from $x \dot{\succ} (y \dot{\succ} z)$ to $(x \dot{\succ} y) \dot{\succ} z$ that it allows might be viewed as a 'non-local' restructuring whereby an 'embedded' dependent moves up a level. In conjunction with [ax1] and [ax2], this axiom will allow an embedded dependent to move up to topmost hierarchical position, and hence be extracted. More specifically, this system will allow a marker α to restructure to the form $\beta \dot{\succ} x$ (for some β) for *any* (atomic) x within α *except* its maximal head. Hence, such a system appears to be *too* liberal to be useful. A complementary observation, however, is that in a system with [ax1] and [ax3] (either with or without [ax2]), a marker α can restructure to the form $\beta \dot{\prec} x$ (for some β), with x atomic, *iff* x is the maximal head of α. Consequently, the corresponding implication ($\overset{\dot{\prec}}{\leftarrow}$) could be used in implementing a bounded form of head movement, allowing a head to move to the edge of its 'domain' (i.e. consisting of itself plus dependents) but not beyond. A possible use is in handling the bounded movement of the finite verb in the main clauses of V2 (Verb-Second) languages such as Dutch and German. The requisite movement types for finite verbs might be generated by a lexical rule such as:

$$\text{V} \Rightarrow \text{s}_m/(\text{s} \overset{\dot{\prec}}{\leftarrow} \text{V}) \qquad \text{(V a finite verb)}$$

As we have seen, the restructuring allowed by [ax2] alone is insufficient, but free involvement of [ax3] gives a system that is too liberal. However, an intermediate position between these two extremes is possible, which involves restricting the action of the 'non-local' axiom [ax3], and in particular linking its use to further distinctions encoded by modalities. Let us consider just one of many possible distinctions that might be invoked. Various linguistic approaches acknowledge a distinction between head-complement and head-adjunct relations. We might use different operators to encode these different dependencies, e.g. such as $\succ_c$, $\succ_a$ (c for complement, a for adjunct), so that our example *Kim saw Lee clearly* might yield a marker such as:

$$(\mathit{kim} \prec_c ((\mathit{saw} \succ_c \mathit{lee}) \succ_a \mathit{clearly})).$$

These modes might be linked to others ($\dot{\succ}_c$, $\dot{\succ}_a$) which preserve both headedness and the complement/adjunct distinction, but which are otherwise more liberal in being subject to variants of the axioms [ax1], [ax2] and [ax3]. A modified [ax3], in particular, might be restricted to apply in only certain cases, e.g. taking the form:

$$x \succ_i (y \succ_j z) \;=\; (x \succ_i y) \succ_j z)) \quad \text{where } \langle i, j\rangle \in \{\ldots\}$$

For cases of $\langle i, j \rangle$ pairs that are *not* allowed, the effect is that 'j-dependents' may not move up out of 'i-domains', so that 'i-domains' are islands to extraction of 'j-dependents'. For example, the island status of adjuncts, illustrated by (14), could be enforced by disallowing all pairs $\langle i, j \rangle$ in which i corresponds to a head-adjunct relation.

(14) a. Kim filed the articles without telling Lee.

b. *Who did Kim file the articles without telling?

The above example hopefully illustrates the point that this approach seeks to ground an analysis of locality constraints within a detailed representation of linguistic structure, exploiting the distinctions that this representation encodes rather than being merely a stipulative overlay.

References

[1] Ades, A.E. and Steedman, M.J. 1982. 'On the order of words.' *Linguistics and Philosophy*, **4**. 517–558.

[2] Gabbay, D.M. 1996. *Labelled deductive systems. Volume I.* Oxford University Press.

[3] Hepple, M. 1990. *The Grammar and Processing of Order and Dependency: A categorial approach.* PhD Thesis, University of Edinburgh.

[4] Hepple, M. 1993. 'A general framework for hybrid substructural categorial logics.' Ms, IRCS, UPenn. Available as IRCS Report 94-14.

[5] Hepple, M. 1995. 'Mixing Modes of Linguistic Description in Categorial Grammar.' *Proceedings EACL–7.*

[6] Hepple, M. 1996. 'Head Promotion and Obliqueness in a Multimodal Grammar', *Proofs and Linguistic Categories*, C. Casadio (ed), Proceedings of the Third Roma Workshop, Rome, April 1996.

[7] Kurtonina, N. 1994. *Frames and Labels: A Modal Analysis of Categorial Inference.* PhD thesis, Utrecht University.

[8] Moortgat, M. & Morrill, G. 1991. 'Heads and Phrases: Type Calculus for Dependency and Constituency.' To appear: *Journal of Language, Logic and Information.*

[9] Moortgat, M. & Oehrle, R. 1994. 'Adjacency, dependency and order'. *Proceedings of Ninth Amsterdam Colloquium.*

[10] Morrill, G. 1990. 'Intensionality and Boundedness,' *Linguistics and Philosophy*, **13**, 699–726.

[11] Morrill, G. 1995. 'Clausal Proofs and Discontinuity', *Bulletin of the Interest Group in Pure and Applied Logics*, Vol. 3, No. 2,3.

[12] Morrill, G. & Solias, M.T. 1993. 'Tuples, Discontinuity, and Gapping in Categorial Grammar.' *Proc. of EACL–6*, Utrecht.