Relative clauses, left-periphery effects, and the dynamics of language processing

Ruth Kempson, Wilfried Meyer-Viol

King's College London

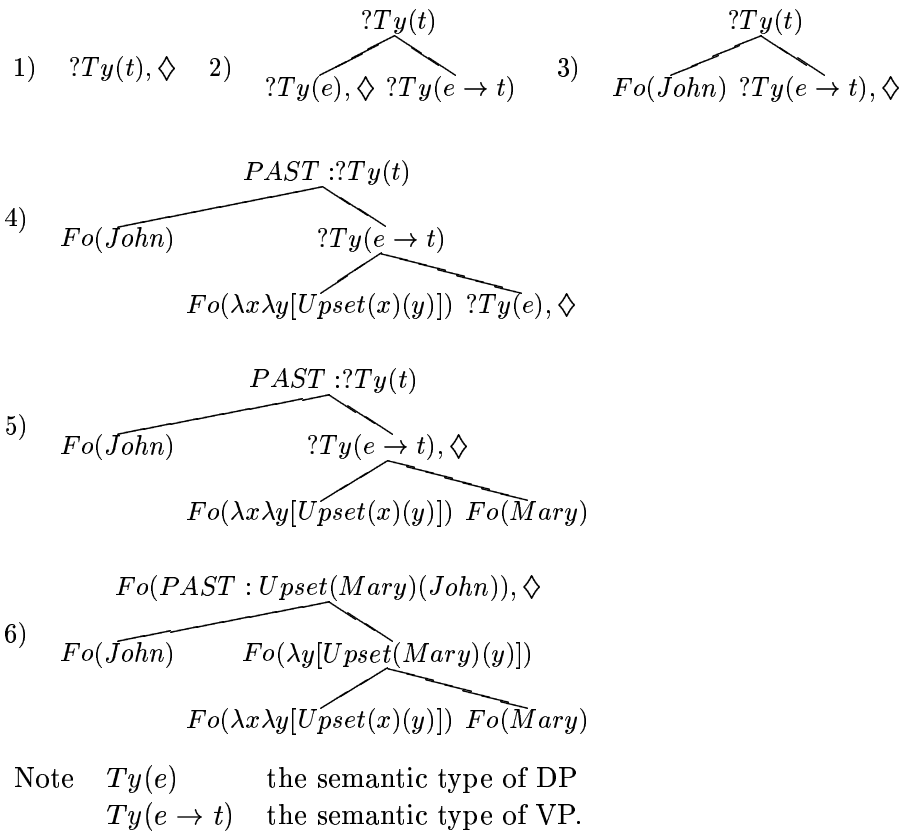
In this paper, we explore the extent to which Dynamic Syntax enables us to express different forms of left dislocation as a natural set of variants, in an account which builds on the analysis of relative clause construal. What we shall show is that characterising growth of a logical form, in which each logical form is represented as a tree structure, enables us to define the various left dislocation effects as a consequence of general principles of tree growth. The success of these analyses in delineating natural classifications of the full range of cross-linguistic data is, we shall argue, evidence that modelling the process of assigning interpretation to a string as a left-right process of tree growth provides a principled basis for natural-language syntax - which is the claim central to the Dynamic Syntax framework.

1 The Flow of Language Understanding

According to Dynamic Syntax, the process of natural language understanding is a monotonic process of building up a logical form following the left-right sequence of words, where the goal is to establish some propositional formula as interpretation ($= ?Ty(t)$) as the root of a tree.[1] The various nodes of this tree are decorated with sub-terms of this propositional formula. These are not words of the language under analysis; they are expressions of a typed lambda calculus – representations of content, each a value of the *Fo* predicate, (*Fo* for 'formula'). The process is one of incrementally building up interpretation in two stages: first, setting out the tree structure as directed by the overall goal and additional subgoals (so-called 'requirements'); second, annotating nonterminal nodes in the tree by steps of labelled type deduction defined over values of the *Ty* predicate (*Ty* for 'logical type') once there are appropriate decorations on the terminal nodes.

[1] This section of the paper is largely taken from Kempson and Meyer-Viol forthcoming where the analysis of resumptive pronouns is used as part of an analysis of crossover phenomena.

Intrinsic to this process are three forms of syntactic underspecification, each of which has to be resolved during the construction process. First, nodes are introduced with requirements to be filled later, as displayed in Figure.1, where, starting with a single node decorated a requirement indicating the overall goal to be achieved – $?Ty(t)$ – each node introduced has a requirement (the node marked with the pointer $\Diamond$ is the node currently being developed). Notice how in figure 1 each partial tree in the sequence is



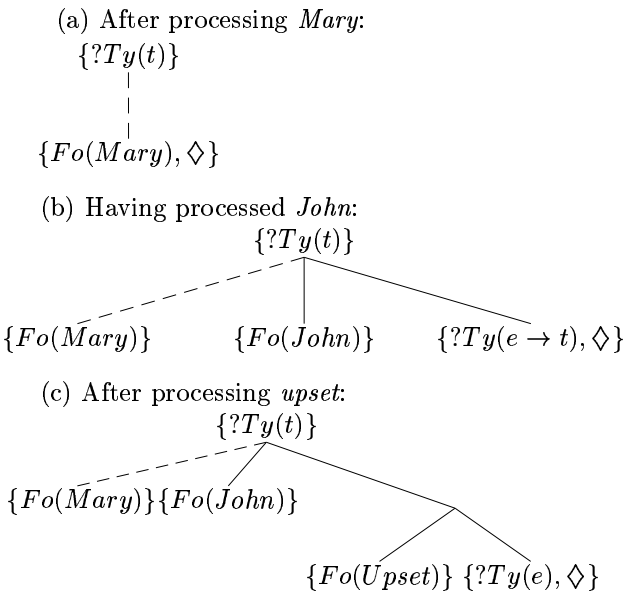
Note $Ty(e)$ the semantic type of DP
$Ty(e \rightarrow t)$ the semantic type of VP.

Figure 1: Parsing *John upset Mary*

a development of the previous one. It is characteristic of the requirements that drive this development process that they may be satisfied substantially later in the construction process than the point at which they are introduced. To pick out just the most extreme case, notice that the goal $?Ty(t)$ introduced at step (1) in figure 1 is not met until the final step.

A second form of underspecification is that nodes can be introduced into a tree as initially unfixed, characterised only as dominated by the top node, their position in the tree being fixed later in the construction process. This is the analysis proposed for the core cases of long-distance dependency.[2] Again, we display the phenomenon graphically:

(1) Mary, John upset



(d) Merging the unfixed node and the object node:

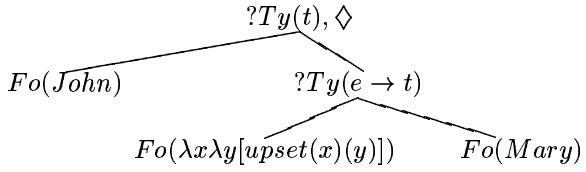


Figure 2: Parsing left-dislocation structures

At step (a), the node annotated by the word *Mary* is introduced as dominated by the top node without fuller specification of its relation to the topnode (hence at that juncture 'unfixed' within that tree). At step (d),

[2] This process bears close formal resemblance to the concept of 'functional uncertainty' of Kaplan and Zaenen 1989, articulated within LFG, but that framework lacks the dynamics of updating such uncertainty as part of the structural characterisation.

this node is merged with some further node whose relation to the topnode is determined, hence establishing its precise role in the logical structure. As figures (1) and (2) jointly display, different strings may lead to the same decorated tree, the difference between them arising from the different sequence of steps leading to that result.

Interacting with both these structural forms of underspecification is the third form of underspecification: the interpretation lexically provided by a pronoun. Reflecting the fact that the lexical content of a pronoun underspecifies its interpretation in context, a phenomenon which we model in representationalist spirit (cf Kempson et al 1999, Kempson et al 2001 ch.1 for arguments), a pronoun is defined to annotate a node with a meta-variable as *Fo* value to be substituted by some selected term. This process is taken to be a pragmatic one, restricted only in so far as locality considerations distinguishing individual anaphoric expressions preclude certain formulae as putative antecedent:

(2) Q Who upset Mary? Ans: John upset her.

Thus in processing the pronoun in (2), the object node is first decorated with a meta-variable **U** within $Fo(\mathbf{U})$, this being replaced by a copy of some other term, eg $Mary$, copied from the structure constituting the interpretation of the previous sentence. Notice that the substituend for the meta-variable is not the English word *Mary* but the term taken to represent the individual referred to by that word in the given context.

1.1 The Formal Framework

1.1.1 Decorated Partial Trees

As figures (1)-(2) have informally displayed, decorated partial trees are progressively constructed, each node of which is initially decorated with requirements and subsequently with annotations. Such trees are described by a modal logic of finite trees (LOFT - see Blackburn and Meyer-Viol 1994), with modal operators:

$$\langle\downarrow\rangle, \langle\downarrow_0\rangle, \langle\downarrow_1\rangle, \langle\uparrow\rangle, \langle\downarrow_*\rangle, \langle\uparrow_*\rangle, \langle L\rangle, \langle L^{-1}\rangle, \langle D\rangle, \langle U\rangle$$

In LOFT, modalities are interpreted on the nodes of the trees: e.g. the existential modality $\langle\downarrow\rangle$ is evaluated over the daughter relation, and $\langle\downarrow\rangle Ty(e \rightarrow t)$ 'holds' on a node n if there is a daughter where $Ty(e \rightarrow t)$ holds. More specifically, LOFT has $\langle\downarrow_0\rangle$ $\langle\downarrow_1\rangle$ interpreted over first and second daughters respectively, $\langle\downarrow_*\rangle$ over the reflexive transitive closure of the daughter

relation (dominance), $\langle\uparrow\rangle$ over the mother relation, $\langle\uparrow_*\rangle$ over the inverse of dominance, $\langle L\rangle$ over a relation of LINK between trees, $\langle L^{-1}\rangle$ over its inverse, and finally $\langle D\rangle$ interpreted over the reflexive transitive closure of the union of daughter and LINK relations. The decorations that may hold at a node include specification of a value for the formula predicate Fo, a type specification, expressed as an argument of the predicate Ty, a tree-node position, represented as an argument of the predicate Tn, and so on. Thus if $\langle D\rangle Fo(Run)$ holds at a node n, there is some node m that can be reached from n following daughter and link relations arbitrarily far, and $Fo(Run)$ holds at m. Included within possible specifications are meta-variables, being place-holders for some fixed value to be provided.

The specific and novel advantage of LOFT emerges from the use of the LOFT operators in combination with a generalisation of the concept of requirement $?X$ to any LOFT formula X. So the requirements that may be imposed are by no means restricted to nonmodal or simple modal requirements. To the contrary, any formula may be used to express a requirement. So while $\langle\downarrow_*\rangle Fo(\alpha)$ holding at a node n implies that n dominates a node where $Fo(\alpha)$ holds, $?\langle\downarrow_*\rangle Fo(\alpha)$ holding at n implies that $Fo(\alpha)$ is REQUIRED to hold at a node dominated by n. By this means requirements may constrain subsequent development of the tree; and this provides a mechanism for pairing noncontiguous expressions according as one imposes some requirement on a node which is fulfilled by an annotation on some discrete node supplied by the other. In particular, as we shall see, complementisers may impose complex modal requirements on the topnode of a newly introduced tree, to be met by some annotation in the subsequent construction of that tree.

1.1.2 The Dynamics of Tree Growth

LOFT is a language for describing (partial) trees. To to describe the tree growth process, we define transitions between partial trees. There are three types of action: computational actions, which are general (albeit possibly language-specific); lexical actions, which are associated with individual words; and pragmatic actions, which are substitution operations, using terms/structure antecedently available. To exemplify the pattern of computational actions defined, we list **-Adjunction* which licenses the introduction of an unfixed node:

***-Adjunction**

$$\frac{\{\{Tn(a), \ldots ?Ty(t), \Diamond\}\}}{\{\{Tn(a), \ldots , ?Ty(t)\} \quad \{\langle\uparrow_*\rangle Tn(a), \ldots , ?Ty(e), \Diamond\}\}}$$

A rule like this should be read as follows: this transition is defined as starting from a partial tree (described as a structured set of nodes) containing only one node (described by the set of formulae holding at that node), here some arbitrary tree node a with requirement $?Ty(t)$. The transition then adds to that one-member set the node described as being dominated by a, $\langle\uparrow_*\rangle Tn(a)$, requiring a type e decoration, with the pointer indicating that it becomes the node currently under development. (Notice how the rule conforms to the general pattern of defining information-preserving transitions from partial tree to partial tree.) Any node introduced by this rule has ultimately to be assigned a fixed tree position by a process, *Merge*, which unifies tree nodes. Characteristically *Merge* takes place (as displayed in figure 2), where co-present in a tree are an unfixed node annotated with a formula α of a certain type and a fixed node requiring that type.

Lexical actions defining the contribution of individual words are, equally, procedures for updating partial tree descriptions. The lexical specifications are of the form $\langle\ IF\ \Sigma, THEN\ \alpha_1,\ ELSE\ \alpha_2\ \rangle$, with the 'IF' condition specifying decorations which must hold on the node at which the pointer resides if the actions given by 'THEN' is to be carried out. For example, the conditions for the actions induced by the English verb *upset* require the pointer to be at a node decorated with the requirement $?Ty(e \rightarrow t)$, from which it initiates the addition of a subtree – a daughter node annotated with $Fo(Upset)$ – and the addition to its mother of the requirement for a daughter:[3]

upset	
IF	$\{?Ty(e \rightarrow t)\}$
THEN	go($\langle\uparrow\rangle$), put(Tns(PAST)), go($\langle\downarrow_1\rangle$),
	make($\langle\downarrow_1\rangle$); go($\langle\downarrow_1\rangle$);
	put($Fo(Upset), Ty(e \rightarrow (e \rightarrow t)), [\downarrow]\bot$);
	go($\langle\uparrow_1\rangle$); put($?\langle\downarrow_0\rangle(Ty(e))$)
ELSE	ABORT

If the condition is not met, the current sequence of actions aborts.

[3] The detailed specification of condition and actions in the lexical specifications of verbs, including numbers of nodes to be constructed, varies from language to language, and indeed from verb to verb. This lexical description highlights the various basic actions DS uses to construct trees: put, go, make....

Pronouns illustrate how the structure projected by a word may underdetermine content. They supply a meta-variable which has to be replaced by some fixed value to yield a wellformed output.[4] Notice the specification of case, here nominative, as imposing a constraint on tree position that the mother node in the resulting tree be of type t:

he	
IF	$\{?Ty(e)\}$
THEN	put$(\{Fo(\mathbf{U}), Ty(e),$
	$?\langle\uparrow_0\rangle Ty(t), [\downarrow]\bot\})$
ELSE	**ABORT**

Both lexical specifications determine through the annotation '$[\downarrow]\bot$' that the annotated node in question is the terminal node of a tree, a general property of lexical items.

It is the interaction of computational, lexical and pragmatic processes which determines the assignment of interpretation to a string. A wellformed string is one for which at least one complete logical form can be constructed from the words in sequence within the context of a given class of computational and pragmatic actions WITH NO REQUIREMENTS OUTSTANDING. In consequence, as we shall see, the imposition of requirements and their subsequent satisfaction are central to explanations to be given.

1.2 Linked Structures and Relative Clauses

The Dynamic Syntax framework also licenses the construction of pairs of trees in tandem connected by a 'LINK' relation, described by the operator $\langle L\rangle$.[5] This adjunction introduces the top node of a new tree and copies information from one tree to the other. Taking nonrestrictive relatives as the most transparent case, consider the steps involved in projecting the construal of:

(3) John, who I like, chain-smokes.

Having processed the word *John* to yield a partial tree in which the formula $Fo(John)$ annotates a subject node (the 'head' node), a transition is licensed which builds a LINK relation from that node, introducing a new (LINKed)

[4] The pragmatic process of substitution is also used to model the incremental way in which some scope choices are established (see Kempson et al 2001).

[5] Analogous to the modal operators $\langle D\rangle$ (which is the union of $\langle\downarrow\rangle$ and $\langle L\rangle$ relations) $\langle U\rangle$ is the union of $\langle\uparrow\rangle$ and $\langle L^{-1}\rangle$ relations, hence ranging over any sequence of such relations in combination. It is these operators that allow a relation to be defined between any one node in a set of linked trees and any other node in such a set.

tree with topnode decorated with the requirement $?Ty(t)$ PLUS the requirement for a copy of the formula $Fo(John)$ without further specification as to where in that tree this formula might be located. This is expressed, for English, as the introduction of an unfixed node with a requirement of the form $?Fo(\alpha)$ for some $Fo(\alpha)$ annotating the head node:

Link Adjunction (English)[6]

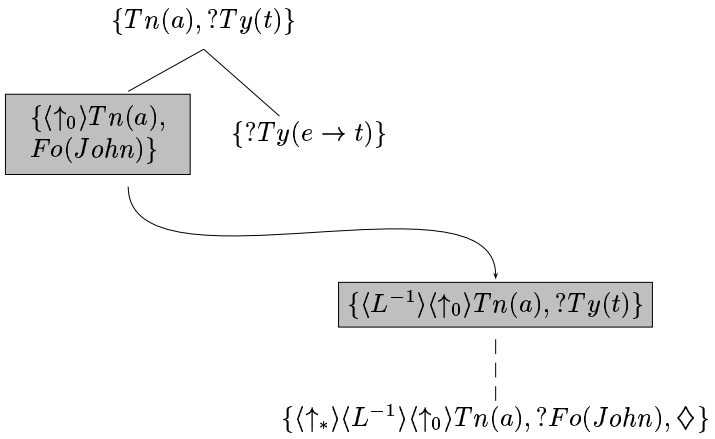


Figure 3: Building a *LINK* transition with *LINK Introduction*

$$\frac{\{..\overbrace{\{X, Fo(\alpha), Ty(e), \Diamond\}}^{\text{head}}\}}{\{..\underbrace{\{X, Fo(\alpha), Ty(e)\}}_{\text{head}}, \quad \underbrace{\{\langle L^{-1}\rangle X, ?Ty(t)\}}_{\text{linked node}} \underbrace{\{\langle\uparrow_*\rangle\langle L^{-1}\rangle X, ?Fo(\alpha), ?Ty(e), \Diamond\}}_{\text{unfixed node}}\}}$$

The relative pronoun itself then provides the required copy at the unfixed node. The subsequent construction of an interpretation for the relative clause follows the general pattern of left-dislocation structures; that is, it unifies the unfixed node with some node with an appropriate type require-

[6]The rule here does not extend to pied-piping cases, but this minor simplification is for purposes of exegesis. See Kempson et al 2001 for a fuller definition which applies to such more complex cases:
(i) A Givenchy shirt, the collar of which was faded, was in the sale.

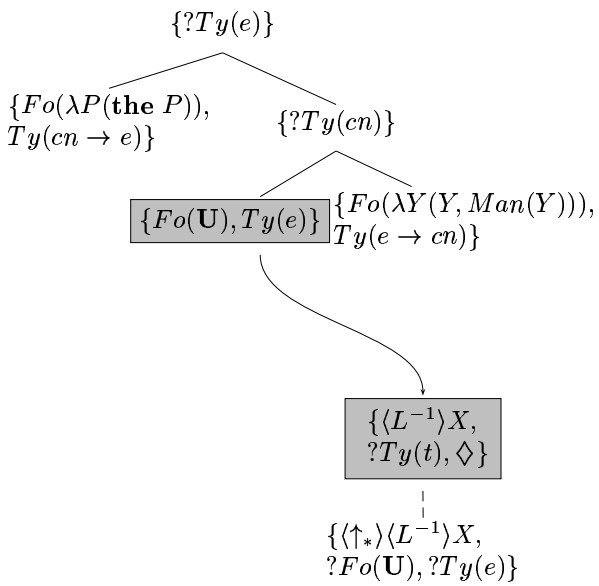


Figure 4: Structure resulting from *LINK-Adjunction* in (4)

ment.

This *LINK Adjunction* rule also applies to yield nonrestrictive relative clause construals. The internal structure resulting from processing an NP contains TWO nodes of type e: the node projected by the variable to be bound by the determiner, which is introduced in parsing the noun, and the node projected by the NP as a whole.[7] In both cases the *wh* relatives serves the anaphoric function of ensuring the presence of the copy in the LINKed structure. Restrictive relatives involve a copy of the variable: nonrestrictive relatives involve a copy of the formula decorating the containing node of type e. Figure 4 displays the process of *Link Adjunction* for the case of restrictive relative construal, imposing a requirement on an unfixed node in the newly introduced linked structure which is fulfilled by the relative pronoun *who* which copies the variable as the head.

(4) the man$_i$ who$_i$ Sue likes e$_i$

[7]Nonstandardly all NPs are taken to project expressions of type e with quantified expressions characterised as variable-binding term operators. That is, NP contents involve the building up of interpretation from a variable, a restrictor, and a variable-binding

The major alternative construction process for interpreting relative clauses involves the obligatory resumptive use of pronouns, as is displayed by Arabic. In (Egyptian) Arabic, a pronoun is essential in all non-subject positions for the strings to be wellformed:

(5) *il mudarris illi Magdi darab-u*
the teacher who Magdi hit him
'the teacher who Magdi hit' Egyptian Arabic

To reflect this distribution, we propose an analysis in which the complementiser induces the introduction of the required linked tree with its associated requirement for a copy. Unlike the English relative pronoun, it does not provide the required copy.[8] This requirement, which singularly lacks any restriction that the copy occur in some subtree, is expressed using the $\langle D \rangle$ operator, which is interpreted over an arbitrary sequence of daughter or LINK relations (see figure 5). This analysis of *illi* ensures that there is only one way of meeting the requirement which it imposes, and that is to use the regular copy process of the language - i.e. selecting as interpretation for some pronoun the value of the formula provided at the head node.[9] Such an interpretation is essential, since any other substituend will leave the LINKed structure with a requirement outstanding, hence not wellformed. In consequence, a pronominal MUST occur in the the subsequent string in a position from which an argument to the predicate can be directly constructed, and, moreover, MUST be interpreted as providing a copy of the formula annotating the head. This obligatory occurrence of a resumptively construed pronominal needs no separate stipulation, and the substitution process updating the pronominal remains a purely pragmatic process. It is merely its interaction with the modal form of requirement on the topnode of the LINKed tree which determines the result.

operator introduced by the determiner, which combines with some formula of *cn* type (an open formula constructed from variable plus restrictor) to yield a term of type *e*. See Kempson et al 2001 ch.4,7. We suppress here all details concerning the intrinsic content of the definite determiner.

[8] As a subject pro-drop language, we take verbs in Arabic to have a condition for lexical action of $?Ty(t)$. From this node, the subject node is constructed as well as the predicate-internal structure. In addition, that subject is annotated with a meta-variable needing substitution exactly as a lexical pronominal.

[9] In most variants of Arabic except classical Arabic, *wh*-questions are the only form of left-dislocated structure in which a resumptive pronoun is not required. This suggests that *Adjunction needs to be defined in Arabic to be sensitive to the presence of a +Q feature (see Kempson et al 2001 for analyses in detail, and all formal specifications of the framework).

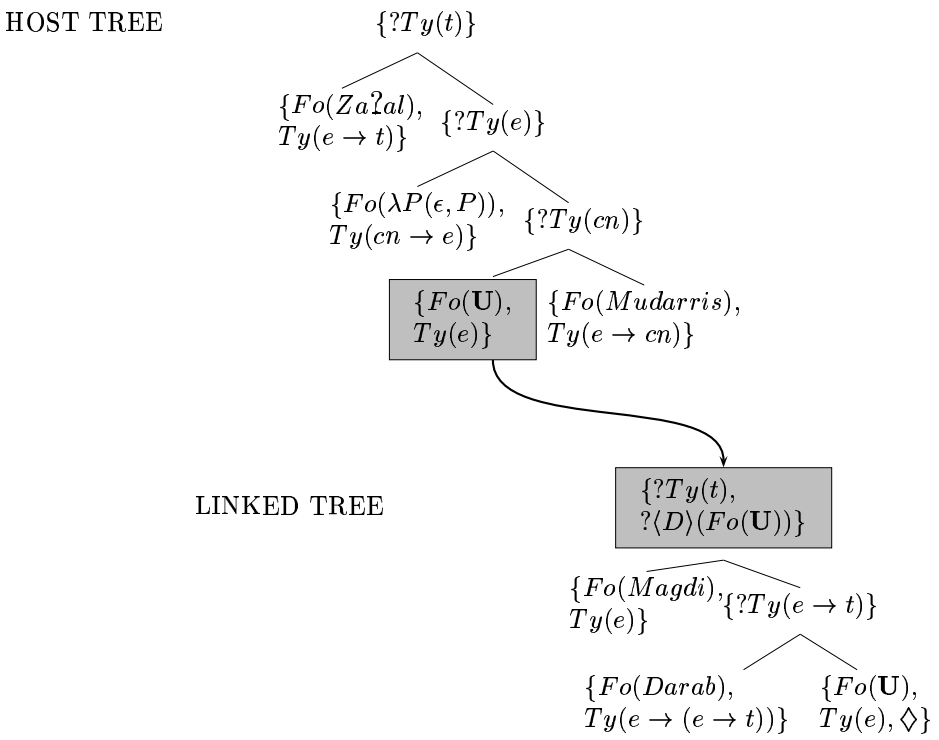


Figure 5: Projection of LINKed trees in Arabic

Just as individual languages display both anaphoric and weaker 'expletive' relativisers, so an individual language might contain both variations; and this is true of Hebrew, which, in having both a relative clause strategy with 'gaps' and one with free use of resumptive pronouns for the same subordinating particle, needs to be defined as having both an anaphoric-like form and an 'expletive' form of the complementiser. It is notable that the gap variant is subject to subjacency effects, whereas the resumptive form is not:[10]

(6) *ha-ʔ-iš* *še* *xašavt* *še* *Dani* *pagaš* *(ʔoto)*
the man that thought$_{2.Fem}$ that Dani met (him)
'the man that you thought Dani met'

[10]Unlike Arabic, Hebrew allows left-dislocation structures which lack a resumptive pronominal.

This possibility makes available a further form of variation. The transfer of information from the head node to the new LINKed structure may, as in Arabic, take the form of a requirement for a copy of the nominal variable without any constraint on where in the subsequently constructed structure such a copy might occur, i.e. '$?\langle D\rangle Fo(\alpha)$' for some head formula α. But there are then two ways this requirement can be met. It may be met by construing a pronoun decorating a fixed tree node as providing the copy of α – in other words identifying both the pronoun's meta-variable as that of the head-formula and the tree position for that copy in one step (as in Arabic). On the other hand, the construction of a copy of the head-formula and the determination of where in the structure it should be located are independent, and, in Hebrew, may be separated. This separation will take place in case the meta-variable projected by some pronoun annotates an unfixed node. In that case, the meta-variable may be replaced by the head-formula:

(7) *ha-ʔ-iš* *še* *ʔani* *xošev* *še* *ʔalav* *ʔamarta* *še* *sara*
the man that I think that about him you said that sara
katva *šir*
wrote poem

(8) *ha-ʔ-iš* *še* *ʔalav* *ʔani* *xošev* *še* *ʔamarta* *še* *sara*
the man that about him I think that you said that sara
katva *šir*
wrote poem

So, as in (7)-(8), a pronoun construed as identical to the head may occur in initial position in any one of a sequence of clauses in the relative sequence.

As we have seen, the pattern displayed by Arabic arises in virtue of a relatively loose modal requirement imposed on the top node of the newly introduced LINKed structure. We can straightforwardly define a modal requirement with a more stringent locality condition. Such a restriction is motivated for Romanian relative clauses as introduced by the particle *care*, a structure in which, as in Arabic, resumptive pronouns are obligatory in all non-subject positions.[11] The only difference from Arabic is that such resumptive pronouns are required to occur locally to the complementiser *care*

[11]Romanian and Italian both have more than one relative pronoun, one of which demands a gap. These can be distinguished according to whether they project an annotation for an unfixed node, or merely a (modal) requirement for the required copy, lexically distinguishing two variants, unlike Hebrew in which a single morphological form has two divergent uses.

in the sense that they project a copy of the head formula in the same LINKed tree that the complementiser initiates, and cannot occur in a relative clause (in more common parlance across a strong island boundary):

(9) *băiatul* *pe* *care* *l-* *am* *văzut*
the boy pe which him have$_{1.SING}$ seen
'the boy that I saw' [Romanian]

(10) * *băiatul* *pe* *care* *am* *văzut*
the boy pe which have$_{1.SING}$ seen
'the boy that I saw'

(11) * *Omul* *pe* *care* *cunosc* *femeia* *care* *l-a*
the man$_j$ pe which$_j$ (I) know the woman$_i$ which$_i$ him$_j$ have
întîlnit
met e_j
'the man who I know the woman who met him'

Such a distribution is characterisable in the same terms as the Arabic restriction, except in the choice of modal operator with which the requirement is expressed - i.e. with a requirement on the top node of the LINKed structure of the form $?\langle\downarrow_*\rangle Fo(\alpha)$.

1.3 Topic structures as a pair of linked structures

The concept of building linked structures has so far been restricted to inducing a new tree (whose top node is duly decorated with a requirement $?Ty(t)$) from some node WITHIN a given partial structure. However, we can also straightforwardly define a process of LINK Adjunction between a tree with TOP node type e and some second structure of type t, duly imposing a requirement on that second structure that it contain an occurrence of the formula annotating the topnode of the first. Such a pair of trees can be used to model so-called topic structures in languages in which a left-peripheral NP, characteristically indicated to be separated from the following string by intonation, is associated with the presence in the following string of a coreferring pronoun:[12]

[12]Here, for simplicity, we assume that the compound preposition *as for* induces an annotation on a node of type e, in addition to the construction of the required LINK relation.

(12) As for Ali, I like him.

Notice what such an analysis of topic structures would lead us to expect. First, since there is no analogue to a relative pronoun in topic structures, their analysis as projecting linked trees would require the construal of the pronoun as identical to the interpretation assigned to the left-peripheral NP. This is because, given the modal form of requirement on the top node of the LINKed structure projected for interpreting the clause following that NP, together with the lack of any morphological expression analogous to an English relative pronoun, some pronoun MUST be interpreted as identical to the Fo value projected by that NP in order to yield a wellformed result. As in the case of Arabic relative clauses, this does not require any particular stipulation for the pronoun itself: it is a consequence of the interaction between requirements and the availability of placeholding devices subject to a pragmatic process of substitution.

Secondly, we anticipate an asymmetry between those languages which have a full relative pronoun and those which do not. Where the relativiser does not itself induce the required copy in the LINKed structure, as in languages such as Arabic and Romanian, there should be parallelism between topic structures and relative clauses, both requiring a suitably construed pronoun in some position within that string. In languages/structures in which a relative pronoun DOES secure the presence of the copy of the formula at an unfixed node within the introduced LINKed structure, there should be no such parallelism. This asymmetry is indeed reflected in the data. In Arabic and Romanian (in relative structures using *care*), the two structures display parallel effects. In Arabic for example, a suitably construed pronoun is obligatory in all nonsubject positions, as it is in Romanian:

(13) *l-bint illi ali ʔabilha.*
the girl that Ali met-her
'the girl who Ali met'

(14) *nadja, ali ʔabil-ha*
Nadia, Ali met her
'As for Nadia, Ali met her.'

(15) *baiatul pe care l-am vazut* [Romanian]
the boy pe which him-have$_{1.SING}$ seen
'the boy that I saw'

(16) *Ion l-am intilnit anul trecut.* [Romanian]
John him-have$_{1.SING}$ met year last

John, him I met last year'

Furthermore, Hebrew, as the mixed case, should, like Arabic, display parallelism between relative clauses and topic structures (compare (7)-(8) with (17)-(18):

(17) *shalom,* *ʔani* *xošev* *še* *ʔalav* *ʔamarta* *še* *sara* *katva*
Shalom, I think that about him you said that sara wrote
šir
poem

(18) *shalom,* *ʔalav* *ʔani* *xošev* *še* *ʔamarta* *še* *sara* *katva*
Shalom, about him I think that you said that sara wrote
šir
poem

In eg English to the contrary, with its anaphoric complementiser, it is only in topic structures that a suitably construed pronoun is required (as in (12). In relative clauses it is not, and is merely an option associated with markedness effects:[13]

(19) The head of the department, who (even) he admits that he needs a holiday, is coming to the conference.

(20) That offensive professor, who I took great care to ensure that I didn't get HIM as a tutor, is complaining to the head that I don't go to his classes.

Setting relative clause sequences aside, the immediate consequence of an analysis of topic structures in terms of pairs of linked trees is that we have an additional strategy available for analysing left-dislocated NPs. Beside decorating an unfixed node within a single structure, we now also have them as annotating a head for a LINKed structure projected from the remainder of the sentence. And this gives us a natural basis for a left-dislocation typology – as we shall now see.

[13]Such resumptive pronouns are characteristically ignored in accounts of English relative clauses, see Sag 1997 which in other respects provides a comprehensive coverage of different kinds of relative clause. Nevertheless they occur commonly enough in all styles of speech. In this connection, we are grateful to Tami Kaplan of Blackwell's for passing me her collection made over a period of six years indicating that speakers of all ages use such structures. See Kempson et al 2001 for evidence that restrictions on their use in relative clauses in English is solely pragmatic.

2 Towards a Left-Dislocation Typology

The problem faced in analysing left-dislocation data is that there is more variation than orthodox assumptions about long-distance dependency effects would lead one to expect. There are, familiarly, languages with the left-dislocated expression paired with 'a gap' and displaying island restrictions:

(21) Mary, John thinks Tom had upset.

(22) *Mary, I dislike the man that married.

There are also languages/structures that display pairing of the left-dislocated expression with a pronoun with no subjacency effects:

(23) *il-kita:b da, 'inta tkallimt ma9a l-walad 'illi katab*
the-book this, you talked with the-boy who wrote
9aley-h [Egyptian Arabic]
on-it
'You talked with the boy who wrote on this book'

(24) As for Mary, I talked to the boy who had scribbled on her book.

However this is by no means all the types of variation allowed. There are left-dislocation structures which use a pronoun which nevertheless displays some of the properties of movement, eg strong island effects (as first explored for Italian in Cinque 1990):

(25) Ton Petro ton nostalgo poli [Greek]
The Peter$_{ACC}$, Cl$_{ACC}$ miss-1sg much
'I miss Peter a lot'

(26) *Tin Maria, gnorisa ton andra pu tin patreftike
'Mary, I met the man that her married'

Conversely, there are left-dislocation structures which lack subjacency restrictions without apparent reliance on a lexical pronoun, as in Japanese:

(27) *Ano hon-wa$_i$ Hanako-ga e_j t_i katta hit-o$_j$*
That book-TOPIC Hanako-NOM bought person-ACC
sagasite iru rasii [Japanese]
looking for seem
'It seems Hanako is looking for the person who bought that book'

There is also interaction with case effects. Left dislocated constituents may optionally display case-matching with some twinned pronominal; and, if so, the pairing displays subjacency restrictions (compare Greek (26) with (28)):

(28) *I Maria xtes gnorisa ton andra pu tin*
The Maria$_{NOM}$, yesterday I met the man who her
pantreftike [Greek]
married
'As for Maria, yesterday I met the man who married her.'

(29) *?*Tin Maria$_{ACC}$, xtex gnorisa ton andra pu tin*
The Maria$_{ACC}$, yesterda I met the man who her$_{ACC}$
pantreftike
married
'As for Maria, yesterday I met the man who married her.'

Then, yet further, there are mixed effects in which left-dislocated constituents may be paired with a pronoun which is itself dislocated:

(30) As for Shalom, he I think should be given the position.

(18) Shalom, ?ani xošev še ?alav ?amarta še sara katva šir
Shalom I think that about-him said-you that sara wrote poem[Hebrew]

And finally, clitic pronouns twinned with some left-dislocated element may depend on the particular form of the predicate as to whether or not they occur in some displaced position:

(31) Gianni, chidi-lo$_{ACC}$ domani
Gianni, ask him tomorrow [Italian]

(32) Gianni, lo vedro domani
Gianni, him$_{ACC}$ I will see tomorrow

This heterogenity appears to demand a number of different analyses with structure-specific stipulations and these data have been a focus of attention since the problems were first aired in detail for Italian in Cinque 1991 (see Anagnastopoulou 1996 for a representative range of analyses).

From a DS perspective, the range of possible effects can be seen as a consequence of interaction between two parameters for variation: on the one hand the distinction between annotation and requirement, and on the

other hand, various locality restrictions. First, we classify strings according to how the annotated formulae projected by the left-peripheral NP is introduced into the structure constituting the interpretation of the main clausal sequence. There are two possibilities. Either the left-peripheral NP projects an ANNOTATION on an unfixed node within a single tree. Or it is taken as annotating a fixed node of type e as head to which a tree interpreting the main clause is LINKed, hence imposing a REQUIREMENT on that second tree (columns (a) and (b) in figure 6). This distinction between unfixed annotation and modal requirement on a LINKed structure yields the distinction between strings for which no pronoun is required (the 'Move α' type of case and the various processes of scrambling as in Japanese, German etc (Saito 1985 and many others subsequently) and strings for which a pronoun IS required (covering both the topic structures corresponding to 'Hanging Topic Left Dislocation' (van Riemsdijk 1997) and Clitic Left Dislocation (Cinque 1990)).

Secondly, we classify strings according to where the underspecified aspects of the information associated with that left-peripheral NP are resolved – i.e. whether the update of such specification (either annotation or modal requirement) has to occur within the same tree, and in this sense locally restricted, or may be provided in an additional LINKed tree, hence possibly in a relative clause (rows (i) and (ii) in figure 6). This gives us a means of distinguishing languages in which topic structures are associated with a strong island restriction on the relation between left-dislocated expression and the twinned pronoun (identified in Romance as the Clitic Left Dislocation effect - Cinque 1990), and languages where the topic structure has no such restriction as in Arabic and the English *as for* construction (column (iib)). It also gives us the distinction between left-dislocation effects without an accompanying pronoun in a language such as English and topic structures in Japanese which, in the latter, may be resolved across a relative clause boundary, despite the lack of morphological pronoun.

Within these major divisions, interaction between pronoun construal and the decoration associated with the left-peripheral NP gives rise to further subdivisions. Taking first the class of LINKed structures (column (b)) with the *Formula* value constructed from the left-peripheral NP imposed as a requirement on their top node, there is the possibility of the two-step update from requirement to annotation at a fixed node VIA an annotation at an unfixed node, as is exemplified by Hebrew (17)-(18), and also by English (30).[14]

[14]In so far as these Hebrew and English strings involve restriction to having the copy of the head formula resolved within a single structure as in (i), this is due to the constraint

→ HOW introduced in structure for main clause ↓ WHERE resolved	(a) **ANNOTATION at unfixed node pronoun not required**		(b) **Requirement on root (from linked tree) pronoun essential no case match**	
(i) **LOCALLY RESTRICTED**	**without pronoun German English**(21) **Italian**	**with pronoun case-matched Greek**(25)	**in-situ clitic post-posed clitics**(31) **of Romance**	**unfixed relative to predicate pre-posed clitics**(32)
(ii) **NOT RESTRICTED.**	**Japanese Topic structure pronoun optional**(27)		**pronoun annotates fixed node Arabic** (23) **English Hebrew**	**pronoun annotates unfixed node Greek** (28) **English**(30) **Hebrew**(18)

Figure 6: A Left Dislocation Typology

Hence the subdivision in (iib). In languages in which the realisation of the LINK-imposed requirement of a copy of the head-formula is restricted to being WITHIN the same structure, as in the Romance Clitic Left-Dislocation effect, there is the further possibility that the annotation provided by the pronoun may be introduced into the tree description at a node which is unfixed relative to an individual predicate projected by the verb. In such a case the clitic pronouns will be processed prior to the verb, hence occur preceding it.[15] Hence the subdivision within the box (ib) of figure 6.

imposed by constructing an unfixed node which the pronoun is taken to annotate, not to a restriction imposed in the construction of the LINK relation:
(i) *As for John, him I met the woman admires.

[15]The distribution of clitic pronouns in Romance languages is sensitive to restrictions imposed by the particular form of the verb, with imperatives, for example, requiring

Taking next the set of structures in which the left-dislocated constituent is taken to annotate an unfixed node, there is also interaction between pronoun construal and the interpretation of the left-peripheral expression. Nothing precludes an unfixed node merging with a fixed node annotated with a meta-variable. So the tree-node position for an unfixed node may be established by unifying it with a node decorated by the meta-variable supplied by a pronoun. For successful application of *Merge* in these cases however, given that in general lexical items have an associated restriction that they decorate a terminal node (see section 1.1.2), one of two conditions must hold. Either the unfixed node in question must be a terminal node so this restriction is not violated (as in relative clauses where only a *Formula* value is copied over). Or the pronoun of the language in question must have a form which lacks the terminal-node restriction which words characteristically project. This second possibility provides a basis for including the case-matching effects of eg Greek left-dislocation structures within the overall picture – and the alternant non-matching of topicalised constituent and its paired pronominal. If a left-peripheral NP is marked with a case-requirement such as accusative indicating a relative position within a tree, then the NP MUST be analysed as decorating an unfixed node to be unified through *Merge*, for otherwise the requirement imposed by the case specification will not be met. To enable that node-unification process to take place however, the pronoun decorating the node to be merged must not impose a terminal-node restriction. On the other hand, if a left-peripheral NP is taken to induce the construction of an independent tree to which the remaining string provides a LINKed structure, then that NP cannot be assigned a case requirement indicating, for example, that the node must be immediately dominated by a predicate node (= accusative), as such a requirement would never be met. Such NPs are then either not marked morphologically for case at all, or they are marked for a case which is analysed in more abstract terms than that of some fixed tree-node relation.[16]

any clitic pronouns to occur after the verb. There are minor variations on this ordering between individual Romance languages, which have to be lexically defined. Note that though the clitic pronouns may have triggering conditions discrete from a discrete strong form of pronoun as to where in the tree they project a meta-variable as annotation, the *Fo* value lexically projected, hence the analysis of their lexically projected content, is the same as the full form.

[16]It is notable that analyses of many languages report that nominative case is morphologically null. The more abstract requirement of nominative case for such languages in which NOM is morphologically marked is that the node in question can be in a one-step relation to the root, whether that relation be L or $\downarrow$.

In this way, the framework can capture the asymmetry between languages such as Greek and English. In Greek, having lost the terminal-node restriction characteristic of lexical specifications, the clitic pronouns have shifted towards becoming an agreement device that merely imposes a requirement, with clitic dislocation being freely available for case-marked NPs as in (25).[17] In English to the contrary, the pronoun retains its terminal node restriction, and no such function is available for pronominal elements.[18]

Notice that this typology indicates possible analyses of strings and is not in itself a typology of individual languages. An individual string may indeed be subject to more than one analysis. For example, the clitic left dislocation data of Romance languages as in (31)-(32), in which there is no case marking on full NPs, classified in figure 6 as a pair of linked structures, may in addition be analysed with that left peripheral NP taken to decorate an unfixed node which is merged with the node decorated by the pronoun. In order to conform to the latter analysis, as in Greek, the pronoun must not imposed the constraint of decorating a terminal-node.[19] A second example of a string for which more than one analysis is available are Japanese topic constructions. Though classified as falling within (iia) in figure 6, i.e. as projecting an unfixed node, the *wa*- marked expression might also be analysed as projecting a tree to which the propositional formula subsequently projected is LINKed, hence leading to an interpretation in which the formula projected from the *wa*-marked expression is retained as an independent term. This availability of more than one analysis without any denotational distinction in the resulting interpretation is a bonus provided by a framework which

[17] Arguably, this is the formal reflex of a process of language change from a full anaphoric process towards one in which the element is defined as an agreement device (see footnote 20).

[18] See Kempson et al 2001 for an account of strong crossover in terms of the ability of construal of a pronoun to be provided by *Merge* in English. It is notable that in relative clauses, where a formula is copied across as an annotation of a node within the linked structure, no violation of the terminal-node restriction associated with the canonical definition of a pronoun is involved, unlike the case-matched left-dislocation structures of Greek.

[19] That there is a gradient range of effects in the Romance languages is indicated by the Spanish dative clitic: while the other clitics in Spanish have lost this "bottom" restriction so that *Merge* provides a form of update, the dative clitic has shifted from projecting such a weak form of annotation on some node to an even weaker 'expletive' specification that projects only a requirement, hence obligatorily requiring further lexical input to provide the annotation on that node:
(i) Le rompi la pata a la mesa
it$_{3.sing.Dative}$ broke$_{1.sing.}$ the leg to the table
'I broke the table's leg.'

focusses on the process of building up interpretation and not just on the resulting interpretation.[20]

Stepping back now from the details, we can see why a framework in which natural language syntax is presented in terms of the sequential construction of decorated trees and a substitutional account of anaphora provides a natural classification of left-dislocation data. All classifications fall within limits determined by two general parameters – the distinction between annotation at an unfixed node and modal requirement at a topnode of a LINKed structure, and the distinction between whether some required form of update must take place within a given tree or may take place across a pair of trees: further phenomena merely lead to subdivisions within these limits.[21]

It should not go unnoticed that this characterisation of a left-dislocation typology has emerged without any need to identify a discrete variant of anaphora, or any special stipulations as in other analyses.[22] It might be countered that allowing a pronominally projected meta-variable to be assigned a value through *Merge* constitutes a different form of pronoun. However, though a discrete process of construal for pronouns could in principle be defined,[23] such added complexity would have no offsetting advantage.

[20] This flexibility provides a point of departure for an explanation of language change since discrete analyses may be available within any given system, hence allowing nonagreement over assigned analysis between speaker-hearer without either the risk of any breakdown in communication or the need to posit multiple grammars for an individual.

[21] The problematic double-clitic Italian dialects reported in Manzini and Savoia 1998 can be straightforwardly expressed, with the first such clitic being taken to be a requirement for the annotation provided (at an unfixed node) by the second clitic. It is notable that there are no dialects reported which license reiteration of more than two 'displaced' clitics, just as this account would lead us to expect. Though this is merely a promissory note for a fuller acoount, it is notably simpler than the multiple-functional-projections analysis of Manzini and Savoia.

It is notable that possible variation consequent upon the anaphoric/expletive properties of the relativising element is not available in left-dislocation structures, there being no analogue to a relative pronoun in left dislocation structures. However there may be particles such as the Romanian *pe* occuring in both left-dislocation and in relative clause structures, which arguably are a lexical encoding of the action of constructing a LINK transition simpliciter. This particle signally remains unanalysed in other frameworks: see Dobrovie-Sorin 1990.

[22] Examples of construction-specific stipulations abound in the literature – for example the widely held projection of one or more clitic-specific functional projections (following Sportiche 1992), or analyses in which a pronoun functions both as as an operator and as a variable within a single derivation (Demirdache 1991, Anagnastopoulou 1997).

[23] Such an account would involve defining a discrete resumptive form of pronoun that projects only some locality constraint and no annotation at all, in addition defining *Merge* more narrowly to preclude any unification of nodes unless the fixed node has a requirement of the form $?Ty(e)$.

To the contrary, on the present analysis, the distribution of resumptive uses of pronouns is a consequence of the analysis of pronouns as projecting a meta-variable. None of the major frameworks secures a comparable characterisation of resumptively construed pronouns as part of a unitary characterisation of anaphora – all orthodox frameworks stipulate some reflex of the analysis of *wh*-binding specific to these uses of pronouns. Furthermore no analysis within either Minimalist assumptions or HPSG provides any general basis for characterising the varying availability of the resumptive use of pronouns in relatives and left-dislocation structures either within or across languages. The only analysis in which such problems are extensively addressed is in the Optimality analysis within LFG of Bresnan 1999; but this approach involves multiple levels, each with its distinct vocabulary, and, with the additional optimality potential for expressing gradients of variation, is dangerously rich. In the DS framework to the contrary, a unitary account of anaphora is retained, the concepts of tree growth make available a number of clearcut means of differentiation, and the sometimes unclear variation in the data can be analysed as due to pragmatic constraints on how individual choices are made relative to the partial decorated structures projected.

Overall, modelling the process of interpreting strings as growth of semantic structure not only provides a restrictive mono-level concept of syntactic structure, but also a richly delineated typology of a core phenomenon of natural language syntax. Accordingly, we conclude that grammar formalisms for natural languages should directly reflect the dynamics of natural-language processing.

3 References

Anagnastopoulou, A. 1997. Clitic left-dislocation and contrastive left-dislocation. In eds. E. Anagnastopoulou, H. Riemsdijk and F. Zwarts, *Materials on Left Dislocation*, pp. 151-92. Amsterdam: John Benjamin.

Bresnan, J. 1999 .The emergence of the unmarked pronoun. ms. Stanford

Cinque, G. 1990. *Types of A' Dependencies.* Cambridge MA: MIT Press.

Cinque, G.1995. *Italian syntax and universal grammar.* Cambridge: Cambridge University Press.

Demirdache, H. 1992. Resumptive Chains in Restrictive Relatives, Appositives and Dislocation Structures. MIT Ph.D.

Dobrovie-Sorin, C. 1990. Clitic-doubling, WH movement and quantification in Romanian, *Linguistic Inquiry* 21, pp. 351-398.

Jespersen, O. 1931. *A Modern English Grammar on Historical Principles.* London: George Allen and Unwin, London.
Kaplan, R. and Zaenen, A. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In eds. M. Baltin and A. Kroch *Alternative Conceptions of Phrase Structure*, pp. 17-42, University of Chicago Press.
Kempson, R. Meyer-Viol, W, and Gabbay, D. 2001. *Dynamic Syntax: The Flow of Language Understanding.* Blackwell.
Kempson, R. and Meyer-Viol, W. forthcoming. The dynamics of syntax: anaphora, relative pronouns and crossover. In H. Simon and H. Wiese (eds.). Pronouns: Representation and Grammar. Benjamin.
Manzini, R. and Savoia, S. 1998. The syntax of object clitics in Italian dialects. ms. University of Florence.
Sag, I. 1997. English relative clauses. *Journal of Linguistics* 33, pp. 431-83.
Saito M. 1985. *Some Asymnmetries in Japanese and their Theoretical Implications.* Ph.D. MIT. Shlonsky, U. 1993. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23, pp. 443-368
Sperber, D. and Wilson, D. 1995. *Relevance: Cognition and Communication* (2nd edition).Blackwell.
Sportiche, D. 1992. Clitic Constructions. ms UCLA.
Suňer, 1998. Resumptive restrictive relatives: a cross-linguistic perspective. *Language* 74, 335-364.
van Riemsdijk, H. 1997. Left dislocation. In Anagnostopoulou, E. van Riemsdijk, H., Zwarts, F. (eds.) *Materials on Left Dislocation*, 1-12. Amsterdam: John Benjamin.