Siswati Clefts: The Meeting Ground of Context and Contrast

Nhlanhla Thwala, Ruth Kempson, Lutz Marten

Ruth Kempson
(primary author for correspondence):
Philosophy Department,
King's College London,
The Strand,
London, WC2R 2LS, UK.

email: ruth.kempson@kcl.ac.uk
FAX: (44)-0207-8482670

Nhlanhla Thwala and Lutz Marten:
Linguistics,
School of Oriental & African Studies,
Thornhaugh Streete,
Russell Square, London, WC1H OH8 H

August 15, 2006

1 Introduction: Siswati clefts

Cleft structures have proved to be a tantalising structure, not easily expressible in current formalisms. Traditionally, they have been taken as a canonical contrastive-focus device, with clearcut presuppositional properties indicating dependence on context of the non-focussed part (REFS), yet, as has emerged from corpus collections and other recent work (Prince 1988, Hedberg 2000, Buler 2005), there are many divergent uses of clefts, indicating a much more complex relationship between structure and interpretational effect. Despite the subtle variation in the extent and precise nature of the context-dependence implied, structurally, the syntactic properties of clefts robustly display a great deal of cross-linguistic uniformity despite minor language-particular idiosyncrasies. And in this uniformity hides another puzzle. Quite generally, they display extensive parallelism with relative clauses, even to the extent of using the same morphological indicators as a relative clause sequence; and yet at one and the same time they display properties of left-dislocation.

Siswati clefts display this straddling across relative-clause patterning and dislocation patterning particularly clearly. There are two main types of Siswati clefts; and they all display this split nature - with attributes in common with left-peripheral structures on the one hand, and with relative clause structures on the other. The simple type involve first marking the "clefted" element with a preceding *nge-* marker (for class 1 nouns), and secondly an associated relative clause sequence - relative clauses in Siswati are marked by a *la*-prefix as the first prefix in the verbal complex:

(1)

(Nge)-bafana	la-	ba-	to-	natsa	tjwala
It-boys	REL	1SM	FUT	drink	alcohol

It is the boys who will drink alcohol

The ordering of these two is free, with the relative-clause-like sequence occurring after or before the *nge-* marked NP, in the latter case with a D-linking kind of effect:

(2)

la-	ba-	to-	natsa	tjwala	Nge-bafana
REL	1SM	FUT	drink	alcohol	it-boys

They will drink alcohol, the boys

In both these cases, *nge-* marking is optional, yielding a further sub-type of cleft sentence, so that the only feature distinguishing these sequences from left- or right- dislocation structures may be the presence of the relative marker *-la*:

(3)

bafana	ba-	to-	natsa	tjwala
boys	1SM	FUT	drink	alcohol

The boys, they will drink alcohol

(4)

ba-	to-	natsa	tjwala	bafana
1SM	FUT	drink	alcohol	boys

They will drink alcohol, the boys

Indeed in non-human classes, there is no *nge* correlate at all, so there is the further problem that the only feature distinguishing these sequences from that of relative clauses is a distinctive tone marking on the "clefted" NP expression.

(5)

bu-twjala	la-	ba-	to-	bu	natsa	bafani
tjwala	REL	1SM	FUT	OM	drink	2-boys

The alcohol, the boys will drink it.

There is then, in addition, a complex form preceding such sequences involving what is diachronically a bleached locative pronoun. This bleached locative is followed by a Tense marker and then immediately by the *nge*-marked NP and the relative-marked sequence, the tense marking in this relative having to agree with the earlier tense marker:

(6)

ku-	to-	ba	nge-bafana	la-	ba-	to	natsa	tjwala
DSM	FUT	COP	it-boys	REL	1SM	FUT	drink	alcohol

It will be the boys who will drink alcohol

These more complex forms are not associated with any inversion of the clefted expression and the relative-marked sequence.

As these examples show, Siswati, like all Bantu languages has a rich verbal complex, in which a verb stem is prefixed with in order an obligatory subject marker, a tense marker, one (normally only one) object marker which is optional, and the verb. Independently of these cleft sequences, as long as both subject and object markers are selected, word order is very free. The subject, whose associated subject-marking prefix is obligatory, can occur after the verb, if so, right at the end of any VP sequence. The object NP can occur either before the subject or after the verbal complex, possibly after some postposed subject, if there is also an associated object-marking prefix; but if there is no such object-marking, which is optional, then the object NP occurs immediately after the verb, as in (3).

Relative clauses in Siswati, whose pattern these cleft sequences follow, are marked with a prefix on the verb which precedes all other prefixes of the verbal complex, and whenever the object position is relativised over, the object marker is obligatory:

(7)

bu-twjala	la-	ba-	to-	bu	natsa	bafani
tjwala	REL	1SM	FUT	OM	drink	2-boys

The alcohol which the boys will drink....

This is the regular pattern of languages with obligatory resumptive pronouns, and, following this pattern, locative NPs also require an obligatory resumptive element, but, in this case, there being no pre-verbal prefix, this takes the form of a demonstrative:

(8)
London lapho ngi- hlala khona
London where 1SM live DEMO
London REL-LOC I live there
London where I live,.....

(9)
*London lapho ngi- hlala..... London, where I live.....
London where 1SM live

The form *lapho* is a specific locative form of the relative pronoun (in fact a demonstrative); but equally licensed is the same simple form *la* as is used in the subject and object relatives:

(10)
London la ngi hlala khona
London REL 1SM live there
London, I live there

Some of the flexibility of independent clauses is retained in relative clauses: in particular there may be a subject preceding that complex, so that the relative clause marker is not necessarily the first element of the relative clause sequence:

(11)
bu-tjwala bafana la- ba- to- bu- natsa
14-alcohol boys REL 1SM FUT 14OM- drink
it-alcohol which the boys will drink

We have so far illustrated cleft structures with subject clefts. But the object and locative forms display equally marked parallelism with this relative pattern: object marking, which is otherwise optional, is obligatory; the subject may follow the verb; equally, the subject may precede the verbal complex:

(12)
(bu)-tjwala la- ba- to- bu- natsa bafana
14-alcohol REL 1SM FUT 14OM- drink boys
It-alcohol, REL they will drink it, the boys

(13)
la- ba- to- bu- natsa bafana (bu)-tjwala
REL 1SM FUT 14OM- drink boys alcohol
REL they will drink it, the boys, alcohol

(14)
bu-tjwala bafana la- ba- to- bu- natsa
14-alcohol boys REL 1SM FUT 14OM- drink
it-alcohol, boys REL they will drink it

(15)
?bafana la- ba- to- bu- natsa (bu)-tjwala
boys REL 1SM FUT 14OM- drink alcohol
REL the boys they will drink it, the alcohol

It's alcohol which the boys will drink

(16)
ku-se- London lapho ngi- hlala khona
DSM-LOC London where 1SM live DEMO
it-LOC London REL-LOC I live there
It's London where I live

With both object and locative clefts, as subject clefts, there is a further more complex form, the *nge* plus NP plus relative sequence being paired with an expletive *ku-* plus copula sequence:

(17)

ku	to-	ba	bu tjwala	la	ba	to	bu	natsa	bafana
DSM	FUT	COP	alcohol	REL	1SM	FUT	14OM	drink	boys

It will be alcohol which they will drink it the boys

Generally, this bleached locative form, *-ku* precedes the copula, but with past forms, the auxiliaries precedes the subject and in these, the expletive marker may be reduplicated as a form of agreement on the auxiliary:

(18)

kwa	ku-	ngu-	se-London	lapho	nga-
Expletive-Remote-Past	Expletive	NGE	LOC-London	where	1sg-REMOTE-PAST
ngi-	hlala	khona	na-make		
1SM	live	DEM	with mother		

It was in London where I lived with my mother

Overall then, cleft structures in Siswati show the marked parallelism with relative clauses that is characteristic of cleft structures. The question to be asked is: Given this striking parallelism, should these structures be analysed as relatives at some level of abstraction, and so be analysed as some kind of modifier of the focussed expression? Indeed, given the closeness of the complex and simple forms, should the simple clefts be analysed as expressing a covert copula, hence be analysed as like the complex form, with the relative sequence taken to be a relative clause modification of the focus head?

The issues raised by these cleft constructions are of quite general significance. Across all languages, so-called cleft structures display this mixture of hanging-topic-like structures, so in movement terms base-generated, and yet structures correlative with long-distance dependency, so, in such movement frameworks, requiring characterisation through some analogue of movement. This mixture is puzzling because these two analyses are incompatible. As a base-generated topic structure, the cleft expressed would be assigned as part of the highest TP projection, reflecting the clause-external status of the clefted expression: as an A' moved expression, it would be analysed as inhabiting some Spec CP position, hence clause-internal. No single structure can be analysed so that one and the same expression is analysed as simultaneously realised within both the highest possible projection and yet at the lower projection. These problems multiply with right-periphery phenomena, where it is much harder to reflect the parallelism with relative clause configurations; and extra VP-internal Focus and Topic projections are very generally assumed (see Mahajan 1995 for an early instance of this type of analysis).

Siswati relatives and clefts also pose a specific theoretical issue for the Kayne-style analysis of relative clauses: the specifics of that analysis appears to be inapplicable. Given that in object relatives, the subject may precede the whole verbal complex and so also the supposed relativising complementiser, it would seem that not only is the clefted expression external to the clausal sequence, but so too, in some sense, is the subject expression. But, given a Kayneian analysis in which the head starts life within the relative sequence and is progressively raised out of that structure to become the head, the presence of that supposedly structure external subject is awkward at best. The analysis would seem to require copying of the *la* relative marker at each phase of extraction in order to license the extraction of the

head not only out of the verbal projection, but across any such further structure external projection. Nevertheless what MUST be retained is not the record of that chain with appropriate feature-matching at the head of the chain by some copy of *la-*, but, rather, the copy which is the foot of any such chain, as part of the verbal complex. But this is contrary to other assumptions about movement, and so pure stipulation and highly problematic for Kayneian assumptions. This problem would then carry over to topic structures, and be replicated there.

In what follows, we develop a Dynamic Syntax perspective on Siswati, and then use that to set out an account of Siswati clefts in which none of these problems arise, while yet reflecting directly the dove-tailing of topic and long-distance dependency patterning. In particular we shall show that with the dynamic update perspective of Dynamic Syntax, the analogue of long-distance dependency which is the construction of an unfixed node can very simply be fed by a topic-structure building mechanism so that we can construct exactly the mixed effect, part topic-like, part focus-like, which cleft structures require. We shall close by arguing that the progressive growth of structure is all that is needed to capture the properties of Siswati clefts, leaving the system to license exactly the right degree of flexibility to yield the array of pragmatic effects relative to variation in context.

2 Dynamic Syntax: a characterisation of Bantu languages

The Dynamic Syntax model (DS) which we use as the framework for this analysis is radical in being a grammar formalism that reflects the step-wise way in which interpretation is built up during a parse sequence, with syntax defined as a sequence of parsing actions, inducing a process of tree growth across partial trees. A mapping is defined from words, as parsing actions, onto progressively enriched representations of content, until a fixed (in part, contextually established) interpretation is constructed. Interpretation is given as a semantically transparent tree structure, in which a logical formula decorates the top node, and the various sub-terms of that formula decorate the nodes it dominates. Individual nodes are decorated with *Formula* (Fo) and *Type* (Ty) values, reflecting semantic content in terms of expressions of some typed lambda calculus. The process of growth of semantic representation is the basis of all syntactic explanation: a sentence is defined to be well-formed just in case there is at least one possible route through that process.

2.1 The Parsing Mechanism: Siswati

Central to the process is the concept of requirement $?X$ for any decoration X. For example, decorations on nodes such as $?Ty(t)$, $?Ty(e)$, $?Ty(e \rightarrow t)$ etc. express requirements to construct formulae of the appropriate type on the nodes so decorated (propositions, terms and predicates respectively), and these drive the subsequent tree-construction process.[1] There are additional requirements too, on all aspects of underspecification, and all of these must be satisfied by the final step of completing a tree.

[1]The formal system underpinning the partial trees that are constructed is a logic of finite trees (LOFT: Blackburn & Meyer-Viol 1994). There are two basic modalities, $\langle\downarrow\rangle$ and $\langle\uparrow\rangle$, such that $\langle\downarrow\rangle\alpha$ holds at a node if α holds at its daughter, and its inverse, $\langle\uparrow\rangle\alpha$, holds at a node if α holds at its mother. Function and argument relations are distinguished by defining two types of daughter relation, $\langle\downarrow_0\rangle$ for argument daughters, $\langle\downarrow_1\rangle$ for functor daughters (with their inverses $\langle\uparrow_0\rangle, \langle\uparrow_1\rangle$). There is also an additional 'LINK' operator, $\langle L\rangle$, which relates paired trees, with a LINK relation from a node in one tree to the top node of another (see below). With these primitive relations, concepts of 'dominate' are definable in ways that are standard in formal tree-logic systems (see the concept of 'functional uncertainty' defined in LFG: Kaplan and Zaenen 1989). Thus $\langle\uparrow_*\rangle Tn(a)$ holds at a node if some node $Tn(a)$ is along some sequence of mother relations from this node.

Individual steps, as in all languages, take the parser from a tree with just a single root-node decorated with $?Ty(t)$, indicating the requirement (the assigned goal) of establishing a formula of type t, finally deriving a binary branching tree with all nodes decorated with formula values having used all the words in order (see the TWO trees in Figure 1). So in the parsing of (19), we have the initial tree and final tree as in Figure 1:

(19)

Bafana	la	to	natse	twjala
The boys	they	will	drink	alcohol

The boys will drink alcohol

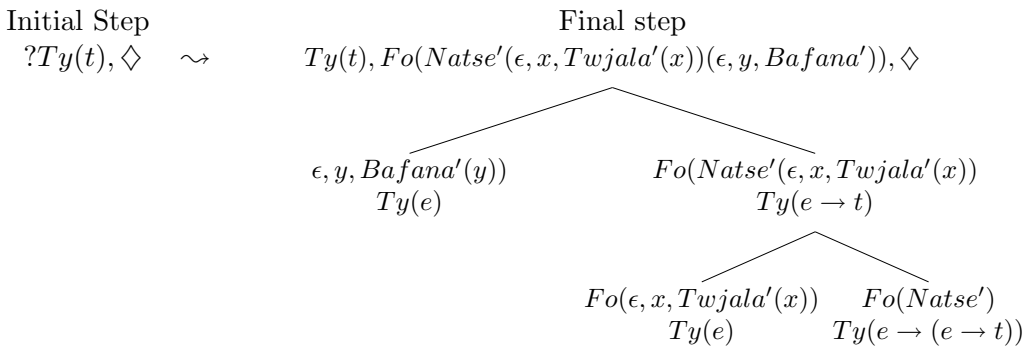


Figure 1: Parsing *Bafana la to natse twjala*

The pointer, $\Diamond$, indicates the node under development. So, at the initial step in any transition sequence, the pointer is at the initial (root) node in some emergent tree; and at the final step the pointer returns to that node in completing its decoration (see figure 1).

The intermediate steps in deriving such trees are determined either by general computational actions, such as anticipating a subject-predicate structure, or lexical actions triggered by parsing lexical items in the order in which they are presented in some string of words.[2] All such actions are procedures for making and decorating nodes in a tree, and moving around within a local subtree. Lexical specifications are defined as macros of such actions: they do not simply annotate nodes with information about semantic content, but may equally contribute to any aspects of tree building, both to decorating the nodes and to constructing them.

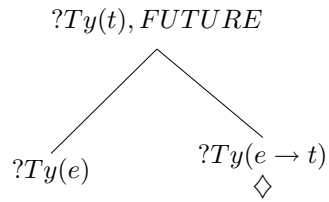
Languages vary in the balance between lexical and general computational actions. In English, for example, (see Cann et al. 2005: chapter 2), verbs project only structure internal to a predicate-requiring node. Pro-drop languages, on the other hand, project full propositional

[2]Quantification is expressed in terms of variable-binding term operators, so that quantifying NPs like all other NPs are of type e, with quantifiers analyzed in the manner of arbitrary names posited in predicate-logic proof steps: all scope effects are expressed within the evaluation of the restrictor of the term itself. The logic underpinning this is the epsilon calculus, of which the primary quantifying term is the epsilon term, the internal structure of such terms containing an epsilon binder,ϵ, a variable, and a restrictor: eg $\epsilon, x, Man'(x)$. Since in Latin, nouns project full specification of terms, the structure defined to be projected by *praemium* would be a subtree of which the quantifying term is the topnode, dominating a subtree decorated with a binder, a variable, and a restrictor specification. Furthermore, given the sensitivity to context in the way such bare nouns are understood, either as definite or as indefinite, this variation can be straightforwardly expressed simply by not requiring that they be assigned a fresh variable (unlike determiner-noun configurations in other languages: see Kempson et al 2001). We leave all details on one side.

structure, the binary structure given by subject node and predicate, as well as the predicate-internal structure, which the particular verb imposes. In Bantu languages, it is the tense marker which induces the subject node and predicate, these being diachronically bleached auxiliary verbs,[3] the subject and object markers which provide place-holding metavariables internal to the local propositional structure, and the verb, as in English, which induces the predicate-internal structure:

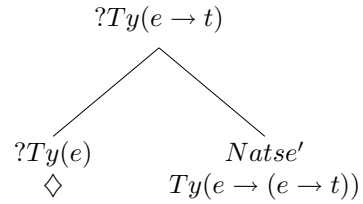
to_{AUX} 'future'

IF	$Ty(t)$		
THEN	IF	$\langle\downarrow\rangle\top$	the initial occurrence constraint
	THEN	Abort	
	ELSE	$\texttt{put}(FUTURE)$;	the tense specification
		$\texttt{make}(\langle\downarrow_0\rangle);\texttt{go}(\langle\downarrow_0\rangle);$	building the subject node
		$\texttt{put}(?Ty(e));\texttt{go}(\langle\uparrow_0\rangle)$	
		$\texttt{make}(\langle\downarrow_1\rangle);\texttt{go}(\langle\downarrow_1\rangle);$	building the predicate node
		$\texttt{put}(?Ty(e \rightarrow t))$	
ELSE	Abort		



natse 'drink'

IF	$?Ty(e \rightarrow t)$		
THEN	$\texttt{make}(\langle\downarrow_1\rangle);\texttt{go}(\langle\downarrow_1\rangle);$		
	$\texttt{put}(Fo(\text{Natse}'), Ty(e \rightarrow e \rightarrow t), [\downarrow]\bot)$		Main Functor
	$\texttt{go}(\langle\uparrow_1\rangle);$		
	$\texttt{make}(\langle\downarrow_0\rangle) : \texttt{go}(\langle\downarrow_0\rangle);$		
	$\texttt{put}(?Ty(e))$		Object Requirement
ELSE	Abort		



The effect of the verbal complex together is, as in other languages, a projection of full propositional structure:

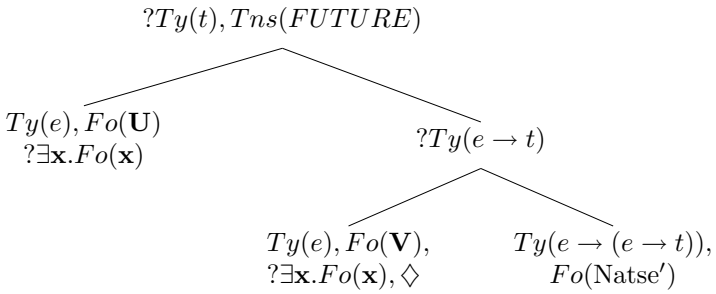


Figure 2: Result of running lexical actions of *ba-to-bu-natse*

Notice how this tree is the result of composing the two partial trees provided by the auxiliary and verb specifications (more strictly extending that provided by the auxiliary with that provided by the verb), with the two agreement devices, *ba* and *bu* ensuring that there are subject and object argument nodes that are decorated with place-holding metavariables, as anaphoric devices. Thus in parsing a transitive verb complex like *ba-to-bu-natse*, a tree structure is projected that expresses the fact that the predicate associated with *natse* takes two semantic arguments, and these are provided with concept-placeholders, meta-variables

[3]Note the agreement on the remote-past form in (18).

U, **V**, that stand for some value to be assigned from the context, exactly as though there were separable pronouns in the sequence.[4] This gives the first flavour of the DS commitment to articulating the projection of a semantic representation as involving articulation of concepts of underspecification and update, both of content and structure. Both pronouns and verb specifications project partial specifications of content through such metavariables, and these have to be replaced as part of the process of constructing an interpretation. This too is faithfully modelled in the system, since all such partial specifications have an associated requirement which ensures that they are replaced with a contentful value during the construction process. This substitution process is directly reflected in the system with a pragmatic process of substitution which enriches some lexically provided metavariable with a term that has already been established in context. But this has a further significance: the DS definition of well-formedness for a string involves the pairing of a string with a tree at output, with no outstanding requirements on any of its nodes, where that tree is derived from actions associated with the words taken in strict sequence. The effect is that the notion of well-formedness is itself context dependent (see Cann et al 2005 chapter 9).

2.2 The Left Periphery

2.2.1 Structural Underspecification and its update

This concept of underspecification plus update is extended well beyond the conventional recognition of anaphora as a content-based form of underspecification: central to Dynamic Syntax is the articulation of a structural form of underspecification plus update. In particular, discontinuity effects are expressed by licensing structural relations that are relatively weak, characterised as a dominance relation that only subsequently gets updated, with the point of update constituting the point at which the initial early partial specification becomes fully determined. For example, long-distance dependency effects are expressed by the construction of a node in some newly initiated logical structure to be developed downwards from a top type-*t*-requiring node:

(20)

babe,	emaphoyisa	a-	cabanga	kutsi	u	ta	tsenga	imoto.
Father,	6police	6SM	think	that	1SM	FUT	buy	9car

Father, the police think will buy the car.

So the node decorated by *babe* in (20) is specified only as dominated by the root node, its position within the unfolding tree being otherwise unfixed at this point in the construal process. Such nodes are annotated as $\langle\uparrow_*\rangle Tn(0)$, using the standard formal concept of 'dominate' (see footnote 5).[5] The point at which this relation is identified comes very considerably later in the interpretation process, in (20) when the verb *tsenga* is parsed. It is this move of building an unfixed node and defining a mechanism for passing it down through a tree under construction which enables the presentation of content being presented as a tree structure to be built up incrementally, for this unfixed relation needs, at some point in the construction process, to be fully specified in order to satisfy the requirement associated with all aspects

[4]The difference between the subject and object agreement markers, where the former is more agreement-like than the latter, can be expressed by the distinction between whether or not the decoration of the node in question has an associated terminal-node restriction in the manner of full lexical items. This is the decoration $[\downarrow]\bot$ included in the specification of *natse*. We ignore these details here: see Cann et al 2005.

[5]Tn is a predicate taking tree-node labels as value, e.g. $Tn(0)$ indicates the rootnode. The provided annotation then indicates that the rootnode dominates the current node.

of underspecification that they be updated before the interpretation process is completed.[6] There is locality in the sense that the unfixed node is constructed from some node which is then immediately developed, but the point of resolution updating that unfixed node may be indefinitely far ahead in the development of that tree.

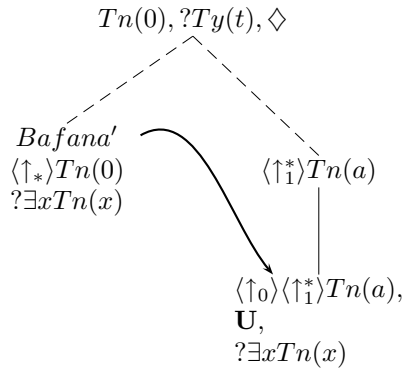
This constitutes a form of parallelism with anaphora since both are defined in terms of underspecification plus update; but whereas anaphoric expressions are underspecified in terms of content, the dislocated expression is expressed as an underspecification of tree relation, hence of structure. This parallelism is further pursued by distinguishing three different forms of structural underspecification in terms of the domain within which their update must be provided. There is the dominate relation that has to be updated within an individual structure but not necessarily locally (the general case already introduced). There is in addition a more locally restricted dominate relation which has to be updated within an essentially local minimal propositional structure, associated with nodes which are introduced as 'locally unfixed'. And, at the other extreme, a weakest form of dominate is defined, which allows update even across a sequence of trees. In free word order languages with very rich case morphology, such locally unfixed nodes may be introduced as a general tree-growth mechanism;[7] but in Bantu, where case is expressed only through these atrophied agreement markers, it is only the subject and object agreement markers which induce such locally unfixed nodes, where they play an important role in creating local pairings between full NP expressions and the argument node of the predicate whose thematic role they are construed as satisfying. For example in the parsing of (21), the first NP expression *bafana* is taken to decorate an unfixed (node diagrammatically displayed as an unfixed node:

$$Tn(0), ?Ty(t), \Diamond$$
$$\vdots$$
$$Bafana'$$
$$\langle\uparrow_*\rangle Tn(0)$$
$$?\exists x Tn(x)$$

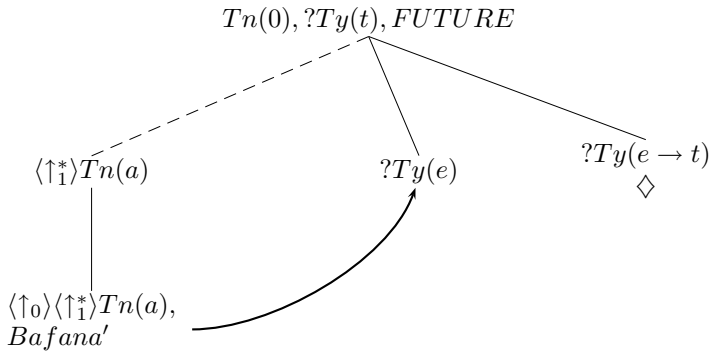
Ba then decorates a locally unfixed node from the local type-t-requiring node with which the first introduced node merges:

[6]There are obvious formal similarities between the DS system and all formal parsing systems which define such underspecified tree relations, (Marcus 1980 and many others since, including recently for Japanese Miyamoto 2002), but all these systems assume any such mechanism constitutes only a parsing mechanism to be defined with reference to some independent core grammar of the language, which has no such device.

[7]See Cann et al 2005 for application of this to the modelling of Japanese.



Then this, now locally unfixed, node is unified with structure provided by the immediately subsequent tense marker, which induces a binary-branching structure.



(21)

Bafana	ba	to	bu	natse
boys	2SM	FUTURE	14OM	drink

The boys will drink it.

bu then involves building a node from the predicate-requiring node, and this too is locally unfixed. This imposes the restriction of there being only one object-marker, even with ditransitives, a characteristic of many Bantu languages, and the precise argument relation is identifiable only upon parsing the verb.[8] So the sequence of actions involved in parsing a basic string involves building up local argument nodes and unifying them with the template provided by auxiliary and verb in combination. So in Siswati, as in other Bantu languages, the projection of a minimal propositional structure involves a complex interplay between the different prefixes and the verb, but the overall outcome is, as in all other languages, a fully decorated binary-branching propositional structure.

[8]The system itself imposes the restriction that there be only one unfixed node of a type at time. This constraint is a consequence of the tree logic underpinning the system (Blackburn & Meyer-Viol 1994). Formally, nothing prevents the construction of more than one unfixed node, but all nodes in a tree are identified by their relation to other nodes in a tree. This has the consequence that if more than one node is constructed from a dominating node characterized only as dominated by that node, these will collapse to a single node yielding inconsistent decorations, and the whole tree will be debarred.

There are number of huge issues to address here. First there is the issue of what it means to characterise the subject marker as introducing an unfixed node rather an inducing a fixed logical-subject relation. Indeed what the *ba-* marked subject gets identified as depends on the subsequent predicate, passive, locative etc imposing particular constraints on what constitutes the subject in the resulting structure (see Marten 2006 for a detailed account of Bantu Locative Inversion in these terms). From this it then follows that the framework provides a way of integrating what in movement frameworks are A and A' chains since while the latter involve an underspecified tree relation without any locality constraint, the latter equally involve an underspecified relation, but this is in addition subject to a locality constraint, and in virtue of being subject to such a constraint, the parse of some expression as subject-marked may serve to identify the initial construction of such a locality domain. In this paper however, we merely note these issues and take this sketch of Bantu syntax in the mean time as sufficient for addressing problems posed by clefts, leaving such issues for another time.

2.3 Building paired trees: topic structures as linked trees

In addition to such local parse strategies, paired trees can be built, so-called linked trees, which are subject to a restriction that they are anaphorically linked but otherwise independent. Relative clause construal is the canonical case for which this mechanism is introduced, an adjunct-like structure, sharing a term with the formula at the head from this structure is constructed, but it is by no means the only application of this tree growth mechanism, as we shall see.

Details of how such linked structures are built up may vary from language to language as long as they conform to the requirement of sharing a term in the resulting paired, "linked" semantic structures. In English the relative pronoun is an encoded anaphoric device which is associated directly with the provision of a copy of the formula from the relative-clause head in the ancillary linked structure: in English, resumptive use of pronouns is marginal at best (Kempson and Meyer-Viol 2002). In other languages the effect of the relative marker is weaker. Siswati is one such language. The relative marker, which occurs as the first element in the verbal complex, *la*, may induce presence of LINK transition – indeed it imposes a requirement for there to be an appropriate copy of the formula from the head from which the transition is defined – but it does not itself secure that copy (see Cann et al 2005 ch.4). Rather, it is the object and subject markers which play this role, and both subject and object agreement is essential in relative clauses in Siswati, as in many other Bantu languages:

(22)

bafana	la	be-	ba-	natsa	tjwala	ba-dzakiwe
2boys	REL	PAST	2SM	drink	alcohol	2SM - drunk

The boys who drank alcohol are drunk.

(23)

tjwala	la-	ba-	to-	bu	natsa	bafana....
the alcohol	REL-	2sm	FUT	OM	buy	boys....

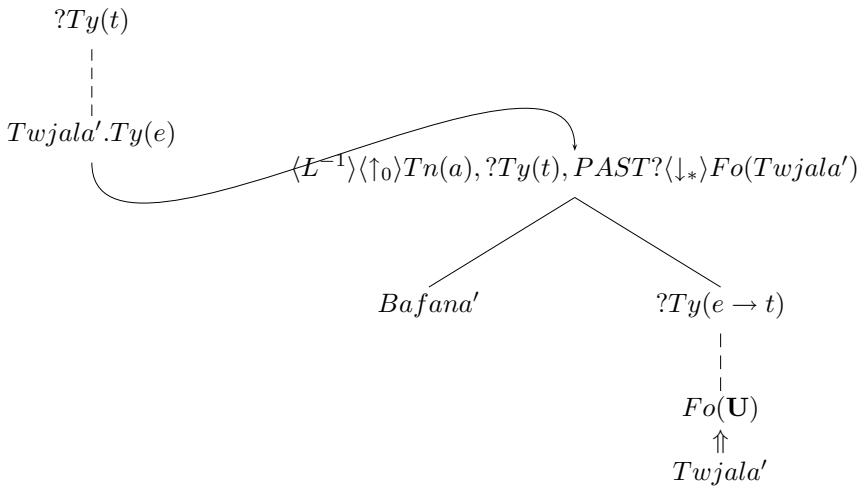
the alcohol which the boys will buy

All that *la* provides is information that the structure to which its processing contributes must be such a linked structure. This is a pattern familiar from languages in which resumptive pronouns are obligatory There is an immediate advantage to this analysis. Precisely because *la* doesn't itself provide either the transition onto an emergent linked tree or the required copy, its very weak specification can be compatible with an occurrence of the subject before it in the relative clause sequence:

(24)

tjwala	bafana	la-	ba-	to-	bu	natsa....
the alcohol	boys	REL-	2sm	FUT	OM	buy....

the alcohol which the boys will buy



All that *-la* provides is the information that the structure to which its processing contributes MUST be such a linked structure.[9]

In locative relatives, this pattern repeats itself. We analyse the first demonstrative as imposing a requirement for a copy, and the second, *khona*, as providing the requisite copy, hence its obligatory nature. Furthermore, the subject expression, may, as in object relatives, intervene between the head and the verbal complex which contains the relative marker:

(25)

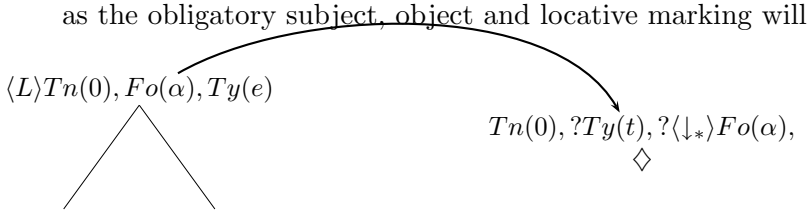
E-ndl-ini	la(pho)	bafana	ba-	lala	khona	ku-	khiyiwe
Loc-house-Loc	REL-LOC	2.boys	2SM-	sleep	DEM	SM-	locked

The house where the boys sleep there is locked

Notice that this account is compatible with the analysis of the agreement particles as inducing a locally unfixed node and updating it. In both (22) and (23) the subject and object agreement particles, despite the unfixed nature of the node they decorate, nevertheless serve to secure the occurrence of the formula in question in the emergent linked tree.

2.4 Left- and Right-Peripheral linked structures and unfixed-node building

As argued in detail in Cann et al 2005, this account immediately carries over to simple topic constructions if we assign initial topic constructions the same linked structural configuration, as the obligatory subject, object and locative marking will then be expected:[10]



(26)

bafana	ba-	to	natsa	tjwala.
2boys	2SM	FUT	drink	alcohol

The boys, they will drink alcohol.

[9]This account will straightforwardly extend to explain the emphla marking on Adjectives.

[10]We specify here the locality restricton on the new type-*t*-requiring node as imposing the requirement of a copy within an individual tree for simplicity.

(27)

tjwala	bafana	ba-	to-	bu	natsa....
the alcohol	boys	2sm	FUT	OM	buy....

the alcohol the boys will buy it

(28)

itolo	bafana	ba-nats-e	tjwala	khona
17yesterday	2boys	2SM-drink-PAST	alcohol	there17

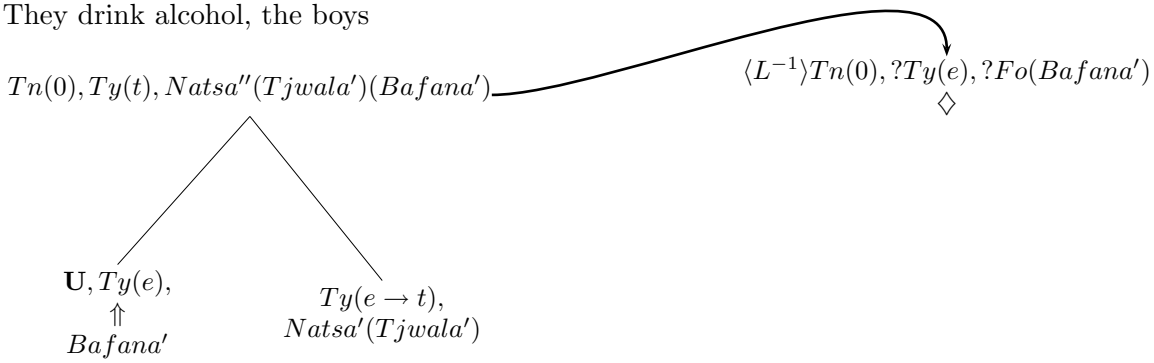
Yesterday the boys drank alcohol then

These strings are the direct analogue of Hanging Topic Left Dislocation postulated for the Romance languages, since there is no obligatory occurrence of *-ku*. Analogously we will expect to induce similar strategies in the closing stages of building up propositional structure, with the option of projecting a linked structure in the final stages of the interpretation of some propositional type, giving rise to a backgrounding effect, again without occurrence of *-ku*(see Cann et al for detailed arguments):

(29)

ba-	to-	natsa	tjwala	bafana
2SM	FUT	drink	alcohol	boys

They drink alcohol, the boys



All that this requires is the assumption that the concept of structures with shared terms is not intrinsically directional, so we expect a transition between a proposialso to be available following the construction of a propositional structure, as well as before. However, there is an asymmetry in the construal of such end-placed NPs: the pronominal marking provided by the affix, given incrementality in the way interpretation is built up has to be identifiable as part of the completion of the first tree, i.e. from the context; and the building of the transition onto the linked structure then imposes a requirement on the construal of the full NP that IT be construed as sharing whatever interpretation is already asssigned to the metavariable projected from the pronominal affix. Hence the backgrounding effect (see Cann et al 2005, Cann et al 2004, for details). We also expect this type of structure to be constructable by carrying over a requirement of identity of any argument term in that first structure, hence licensing the doubling (and backgrounding) effects achieved with object pronominal agreement marking:

(30)

ba	to	bu	natsa	twjala
1.SUBJ	FUT	14.OBJ	drink	alcohol

They will drink it, the alcohol

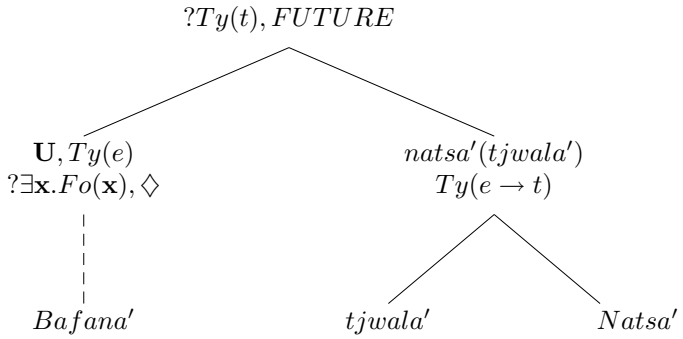
This is not the only device available for construing right-peripherally placed NPS. DS is a parsing-based system, and accordingly we epxect there to be more than one type of transition

for any string of words; and with all generally defined mechanisms being in principle available at any point in the construction process, late applicability of *Adjunction is also available as a means of building up interpretation post-verbally, even given that the parsing of the verb plus subject agreement has provided a provisional value for the subject node. There is asymmetry in the late application of *Adjunction has as there is in the late building of linked structures, but of a rather different sort. At any such late stage in the parsing of a clausal string in which a structure-building process is used to provide some add-on to what is already a full skeletal propositional structure, the only possible application will be the direct construction of an unfixed node of the same type as some fixed node already provided in the parse process, such as an argument (see Cann et al 2005 where this is defined as a discrete process: Late*Adjunction. This is in fact the configuration that results in processing a left periphery structure at the point at which unification of the two nodes is about to take place, so is a subvariant of *Adjunction. It is defined to apply to any node decorated with a metavariable, licensing the introduction of an unfixed node of the same type, which with the addition of a formula would then, without any passing down of any unfixed node through the tree, be able to unify directly with the previously formula-incomplete node to provide it with the required value (see the requirement $?\exists.xFo(x)$ at the fixed subject node). This is characteristic of Extraposition structures, for the development of expletive forms, but also postposed subject expressions:

(29)

ba-	to-	natsa	tjwala	bafana
2SM	FUT	drink	alcohol	boys

They drink alcohol, boys



Parsing *ba-to-natsa tjwala bafana*

In the sequence given by (29), the parsing of *ba* in conjunction with *to* yields a metavariable at the subject node,a nd once BOTH the verb and the ojbect are parsed, so completing the construction of a predicate formula, the pointer may return to the subject node, allowing that to be further developed, finally yielding a fixed subject value, and from that the value of the propositional formula. There are interpretational effects associated with delaying the parse of the subject in this way; and this order is often preferred to buttress a nonreferential interpretation of an indefinite, for reasons of scope choice, etc (see Kempson and Meyer-Viol 2004 for discussion of scope effects in connection with such delayed subject construal).

There is a consequence to this accumulation of strategies for building both unfixed nodes and introducing independent but anaphorically linked structures, which is that strings may, and characteristically will, be associated with more than one sequence of possible transitions. For example, subject expressions at both left and right peripheries will be able to be analysed as either decorating a linked structure, or an unfixed node introduced late on in the derivation. Given the design of DS as an architecture for a parsing system, this is however a bonus,

providing a basis for flexibility in parsing/production judgements enabling common ground between applications in real life processing sufficient for successful parsing without necessity of absolute success in judging the other party's intentions (see Ferreira, V 1995);[11] but this shift in methodology should not go unnoticed – there is NO commitment to one-to-one correspondence between string-interpretation pairs and sequence of actions used to build up that interpretation, a liberalisation which will have consequences in our analysis of the range of cleft structures in Siswati.

3 Siswati clefts

All of what has been set out so far is mere application of analyses independently argued for within DS; and we now have everything in place to provide the analysis of Siswati clefts. The intuition we wish to capture is that Siswati clefts display a range of strategies each of which is for the same end result, the building of a pair of linked structures, of which the "clefted" element is taken to decorate a structure of topnode type e, and the relative-clause sequence decorates a propositional structure linked to that node. This structure is built by a range of strategies of varying degrees of explicitness about the structure they induce. At one extreme, a linked structure can be projected with no specific morphological trigger to induce it: these are the regular so-called hanging topic structures. At the other extreme, there may be a complex sequence of expressions whose actions induce a full anticipation of the structure to be built prior to the subsequent fleshing out of that initially ghosted structure by the subsequent processing of the words. In between are a range of intermediate cases which provide some encoded indication of the process of building up such paired structure than is available from the minimal specification. The pattern preserved in ALL these available strategies, is the pairing of a structure that is decorated with a type e term and a propositional structure that is linked to it that, by definition, has somewhere within it that very same type e term.

The most simple of these isn't a so-called cleft structure, but the hanging topic left-dislocation structure that we have already sketched (with subject either before or after the verb):

(31)

Tjwala	bafana	ba-	to-	bu	natsa	
14alcohol	2boys	1SM	FUT	OM	drink	alcohol

the alcohol, the boys will drink it the boys

(32)

Tjwala	ba-	to-	bu	natsa	bafana
14alcohol	1SM	FUT	OM	drink	2boys

the alcohol, they will drink it the boys

The proposition-inducing sequence and the expression made to be peripheral are predicted to be able to occur in either order, across a pair of linked structures, with a right-placed term providing a backgrounding effect (see Cann et al 2005, chapter 5):

(33)

ba-	to-	bu	natsa	bafana	twjala
1SM	FUT	14-OM	drink	14-alcohol	2-boys

They will drink alcohol, the boys

[11] see Bouzouita forthcoming, who argues that it provides a natural basis for syntactic chanage.

The setting up of this pattern of interpretation can be made more explicit by the use of the *la* particle which explicitly encodes the presence of such a linked structure, again in either order (and again with the subject before or after the verb):[12]

(34)
Twjala	la	ba-	to-	bu	natsa	bafana
14alcohol	REL	1SM	FUT	OM	drink	2boys

It is the alcohol, which they will drink it the boys

(35)
la	ba	to	bu	natsa	bafana	twjala
REL	1SM	FUT	14OM	drink	2-boys	14-alcohol

It is the alcohol that they will drink it the boys

(36)
bafana	la	ba	to	bu	natsa	bafana	twjala
the boys	REL	1SM	FUT	14IN	drink	2-boys	14-alcohol

It is the alcohol that the boys they will drink it.

This strategy may however be dovetailed with an instruction specifically to introduce some new term, by building at that initially introduced type *e* structure, an unfixed node to merge with it – this is the function of *nge*. *Nge* is defined, as we shall see, to co-occur with some linked structure, and to introduce into it some new term (via the construction of an unfixed node). Still with no imposition of specific ordering, the pairing of structures can be in either order:

(37)
nge	Bafana	la	ba-	to-	natsa	tjwala
NGE	2-boys	REL	1SM	FUT	drink	14-alcohol

It is the boys, who they will drink alcohol

(38)
la	ba-	to-	natsa	tjwala	nge	Bafana
REL	1SM	FUT	drink	14-alcohol	nge	2-boys

Note that here we are not following the English pattern of assuming that anything informally labelled as a "cleft" must have a copula at some overt level. To the contrary, on this analysis, the *nge--la* pairing is a codification of the linked structure, without assuming that the *nge*-marked term itself induces a propositional structure – in short, *nge* is not analysed as a copula.[13] Finally, the entire sequence of actions – the introducing of the independent node across a LINK transition, the introducing of the appropriate tense marking, and the introducing of a subject-predicate structure, may ALL be explicitly set up before any particular conceptual content is introduced, and so we get the complex cleft structures in which the frame to be built up is introduced first, and the particular propositional structure, only subsequently:

(49)
ku-	to-	ba	nge-bafana	la-	ba-	to	natsa	tjwala
DSM	FUT	COP	it-boys	REL	1SM	FUT	drink	alcohol

It will be the boys who will drink alcohol

So even here, we are not assuming in the end result that there is a bi-propositional structure: rather the sketch provided by the *ku-to-ba-nge X* sequence provides a single skeletal structure; and the subsequent string develops THAT structure.

[12] The occurrence of the subject before the verb is strongly dispreferred, as in (**??**), since this order fails to provide any means of disambiguation until very late on in the parse.

[13] *nge* also occurs in passive constructions. Apart from noting that this is compatible with our analysis under a construal of the passive subject as decorating an independent linked structure (as a form of topic), we leave this on one side.

In essence, the sketch just summarised gives the entire analysis: all types of clefts are seen as a topic plus predication structure, with the contrastiveness taken to be characteristic of clefts being brought out by the construction of a novel term within that structure. So the analysis is neither a reduction of the cleft structure to a relative clause "underlyingly"; nor a reduction to a simple long-distance dependency structure; nor, even, a reduction to a base-generation topic-style analysis. Indeed, these structures do not have to be seen as essentially discrete. The distinguishing property of the DS system which enables it to bring together these various strategies is its dynamic nature, a system for inducing and developing partial structural specifications. Its very design leads one to expect feeding relationships between one form of process and another. Thus, as we shall now see, the process of building a linked topic structure is able to feed into the mechanism for constructing some underspecified node without there being any inconsistency between these two forms of action.

3.1 The simple la-marked cleft

The analysis of the simple cleft involving just marking with the relative prefix *la*, lacking even the particle *nge* follows directly from the analyses already provided for left and right periphery effects. This cleft is no more than the use of the particle *la-* to make the transition required overt; and accordingly *la-* is defined as providing such a transition. Since , however, the relative sequence may precede the head in such cleft constructions, so the definition of actions for *la* needs to allow for the possibility of a relatively weak update, that is, simply to impose the requirement of such a LINK transition to a type-*t*-requiring node:

$$?\exists x(\langle L^{-1}\rangle Fo(x) \wedge \downarrow_* Fo(x))$$

Such a requirement will have the effect of ensuring that once the processing of *la* is carried out, at some possibly later stage, a LINK transition must be constructed such that the structure currently under construction is appropriately linked to it (with the two structures suitably sharing a term). This is what structures such as (39) provide evidence of:

(39)

la-	ba-	to-	natsa	tjwala	bafana
REL	1SM	FUT	drink	alcohol	boys

They will drink alcohol, the boys

In the use of such postposed topic constructions, despite the fact that this specification is extremely weak, it isn't non-trivial, since it ensures that there must be such a transition once *la* is parsed. But this analysis has the bonus of applying without more ado to cases where the requirement specific to the processing of *-la* is already met. These are the topic-initial cases where, with the building of a LINK transition being a free option, the transition may already have been implemented as a basis for parsing the subject. Indeed this transition must have already been constructed in cases where the subject expression precedes the verbal complex, since even the introduction of the unfixed node, which the subject will decorate, can only be made from an already introduced type-*t*-requiring node. So in this case, the update which *la* provides will be no more than confirmation of the type of structure already under development:

(14)

bu-tjwala	bafana	la-	ba-	to-	bu-	natsa
14-alcohol	boys	REL	1SM	FUT	14OM-	drink

it-alcohol, boys REL they will drink it

Overall, what is needed is a characterisation of *la* which expresses the parallelism between relative structure and topic structure, but which nevertheless allows *la* to play an extremely weak anticipatory role. This we express as:

la	IF	$Fo(\alpha), Ty(e)$		
	THEN	$\mathtt{make}(\langle L \rangle)$,		
		$\mathtt{go}(\langle L \rangle)$;		
		$\mathtt{put}(?Ty(t),)$		
		$\mathtt{put}(?\langle \downarrow_* \rangle Fo(\alpha))$		
	ELSE	IF	$?Ty(t)$,	
		THEN	$\mathtt{put}(?\exists x \langle \downarrow_* \rangle Fo(x) \wedge \langle L^{-1} \rangle Fo(x))$	
		ELSE	Abort	

The first subsequence of actions introduces the LINK relation as part of the actions associated with *-la*; the second subsequence merely ensures the presence of such a transition at some point in the derivation, equivalently for a relative clause and for a hanging-topic construction. The lexical entry could be simplified to express solely the very weak requirement it ensures, but we prefer to retain its constructive role for those cases where its presence ensures the transition itself.

3.2 The simple nge clefts

Now we can be more precise about what it is that *nge* provides, bearing in mind that it is optional. We wish to capture the way in which *nge* triggers the introduction of a novel item, novel in the sense of being constructed as part of the interpretation process, rather than essentially matching, or being picked out from, the context. What we have also to get right is the obligatoriness of some subsequent item after *nge*. *Nge* is not a demonstrative: it cannot occur on its own Moreover the item that follows it must be a full NP in its own right. We know this because of such as examples as:

(40)

nge	la	bafana	la	ba	to	natsa	tjwala
Nge	DET	2pl.boys	REL	2SM	FUT	drink	14.alcohol

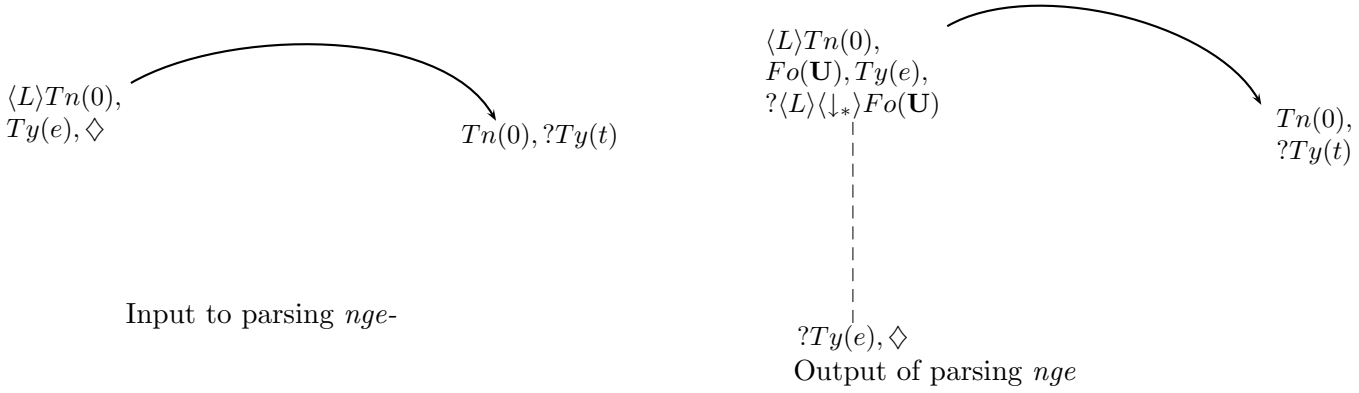
It is the boys who will drink alcohol

(41)

ngu	lobu	tjwala,	bafana	la	ba	to	bu	natsa
NGU	DEMO14	14alcohol	2boys	Rel	2SM	FUT	14OM	drink

It is this alcohol which the boys will drink

Assuming an introduction of an unfixed node in exact mimicry of Late*Adjunction but defined as an action triggered lexically by *nge-* will yield exactly the result we need. So we define *nge* as decorating a link structure already introduced and it itself triggering the process which will ensure that a new term is constructed as update:



nge

IF	$?Ty(e)$	
THEN	IF	$\langle L\rangle Tn(0)$
	THEN	$\texttt{put}(Fo(\mathbf{U}))$,
		$\texttt{put}(?\langle L\rangle\langle\downarrow_*\rangle Fo(\mathbf{U}))$
		$\texttt{make}(\langle\downarrow_*\rangle);\texttt{go}(\langle\downarrow_*\rangle)$
		$\texttt{put}(?Ty(e))$
	ELSE	Abort
ELSE	Abort	

Note the $?Ty(e)$ requirement on the unfixed node which the parsing of *-nge* induces: it is this that forces the next word to be parsed rather than getting some value from context. This analysis also correctly anticipates that *nge* is not itself part of the type e term to be provided: it merely imposes the requirement for actions given by the subsequent word(s) to satisfy it. The locality restriction on the required second copy correctly restricts this to within the tree from which the linked structure is constructed, as is made clear by the long-distance dependency in the construal of:

(42)

nge bantfu	sisini	le	si	tse	tinjua	ti	ba	coshile
NEG 2boys	7friend	REL	7SM	say	8dogs	8SM	2OM	chased

Its people who the friend said the dogs chased

We get immediate confirmation of this analysis of *nge* as introducing a decoration to a linked structure from the relative ordering at the right periphery of a *nge*-marked object NP and a postposed subject. In NON-cleft structures, if the subject is to be end-placed, it must be right at the end, and cannot be followed by some temporal adverbial:

(43)

Tjwala	la	ba	to	bu	natze	namuhla	bafana
alcohol	REL	2SM	FUT	14OM	drink	today	2boys

the alcohol, the boys will drink it today.

(44)

*Tjwala	la	ba	to	bu	natze	bafana	namuhla
alcohol	REL	2SM	FUT	14OM	drink	2boys	today

the alcohol, the boys will drink it today.

However in these structures, with a *nge*-marked object (or locative), the *nge*-marked expression MUST be absolutely final, and cannot be followed by the subject:

(45)

*la	ba-	to-	bu	natsa	bu tjwala	bafana
REL	1SM	FUT	14OM	drink	alcohol	boys

*They will drink it the alcohol, the boys

This is all as the linked structure account would directly anticipate, as th *nge*-marked expression is taken to decorate a structurally independent tree. Indeed, as we would expect, sequences of arbitrary length and complexity can precede the *nge* marked expression: NT PLEASE CHECK]]

(46)

le-	ngi-	-m-	nike	kudla	nge	phinadze	nga-	m-	lelisa	ngu-	bafana
REL	I	OM	gave	food	I	again	I	OM	made-sleep	nge	boy

It's the child that I gave him food and again and made him sleep

Nevertheless, we also predict data of almost the reverse effect, that the *nge*-marked expression might occur apparently on its own, either in answer to a question or, even, in a discourse scenario where the predicate is recoverable directly from the situation:[14]

(47)

Ngu	bani	lu	to	natsa	emanti? Ngu	Jabalani
Ngu	Who	REL	FUT	drink	water? ngu	Jabalani

Who is drinking water? Jabulani is.

(48)

it Jabalani

eg'It's Jabalani who is here' (when someone knocks at your door)

At this point, we need to bring into the picture the relativity to context which the DS system allows: indeed Cann et al forthcoming argue that it should be explicitly defined as a context-relative mechanism (see Cann et al forthcoming), taking structure which the context provides as the INPUT to the structure-building process.[15] In question-answer exchanges, in particular, the structure is set up by the question, here by analysis a pair of linked structures, the *wh* form indicating a linked structure requiring provision of a substituend. All that this requires by way of an answer is the provision of a substituend, as this what is required to update that structure which the immediately previous parse has provided. Indeed, this is no more than direct use of context-provided structure, as the DS account would lead one to expect (see Purver et al 2006 for a detailed account of ellipsis). In the DS perspective, nothing requires that a requirement be met by structure provided solely by linguistic input or within that same sentence. In principle, even, if the structure can be provided by the dialogue as in (48), it may be so provided.[16]

3.3 Complex ku- clefts

It might seem that the complex clefts formed using a default subject marker (a bleached form of the locative demonstrative), a tense marker, a copula, all in addition to the *nge*-marked sequence would need a much more complex analysis, since, in this more complex form of

[14]The form *-lu* is due to coalescence of *la* and *-u*, the relativizer and class 1 subject marking.

[15]Cann et al forthcoming define a context-relative concept of wellformedness, so that fragments can be defined as well-formed as long as they are construed as providing in actions which, relative to some context, can yield a complete tree with no requirement outstanding.

[16]It is notable that this strategy of updating NP fragments from context without any linguistic antecedent from which to construct the appropriate interpretation is only available in highly routinised exchanges, such as request for directions, ordering food, buying tickets, etc. See Stainton 2004 for arguments that this type of case necessitate pragmatic reconstruction as though this was not a structure-building process. In the DS view, to the contrary, the concept of context is that of some non empty sequence of complete trees, so the recovery of any such open propositional value from the dialogue context is no less structural than the manipulation of a fragment relative to a previously parsed question (see Kempson et al 2006 for discussion).

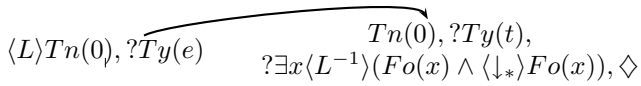
cleft, it would seem that the relative sequence, in following the *nge*-marked expression itself as complement to a copula, is serving as some kind of modifier of that expression, hence a true relative:

(49)

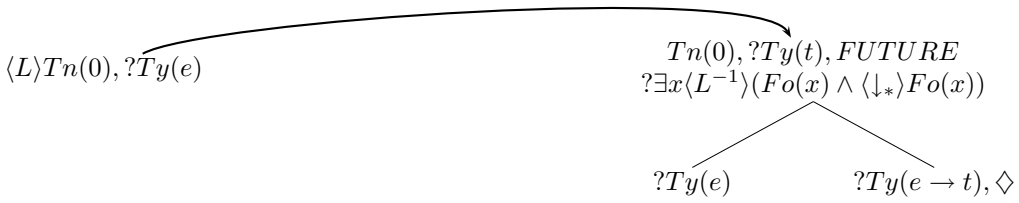
ku-	to-	ba	nge-bafana	la-	ba-	to	natsa	tjwala
EXPLETIVE	FUT	COP	it-boys	REL	1SM	FUT	drink	alcohol

It will be the boys who will drink alcohol

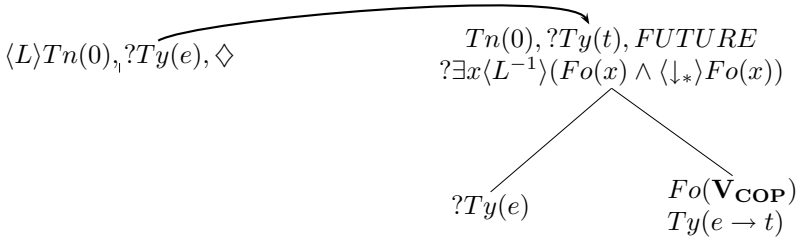
However what is striking is that all the parallelism indicative of the simple clefts carries over in full to the complex clefts. They display exactly the same form of patterns as the simpler *nge*-marked clefts, except for the one difference, that they do not seem to be used in reverse sequence, with the full complex form of opening actions *ku-to-ba nge-NP* occurring after the relative sequence that follows. This apart however, the patterns remain identical. So we have object clefts in which the sequences all occur with the preceding complex expletive formula, all requiring object argument marking, and allowing the subject either to precede or follow the verbal complex. This suggests an alternative analysis in terms of that opening sequence of parse actions setting up in outline the very sequence of actions which the simpler cleft form supplies without any such anticipatory opening gestures. And this indeed is our suggestion. We claim that *ku* is a form of type e expletive, introduced at a linked structure of topnode type e which does nothing more than require that this be a linked structure sharing a term with the propositional node to be subsequently developed, that *to* as a tense marker induces a subject-predicate structure without any formula decorations (see Marten and Kempson 2006), and that *ba* is a form of copula which allows the pointer to move back to the linked node originally introduced, enabling the parsing process to proceed exactly as though that skeletal structure were not there at all. We show the sequence of updates involved in parsing *ku*, *ku-to* and *ku to ba* in (49) and notice how this skeletal structure will be common to all sequences involving the future tense marker *-to*:



Parsing *ku*

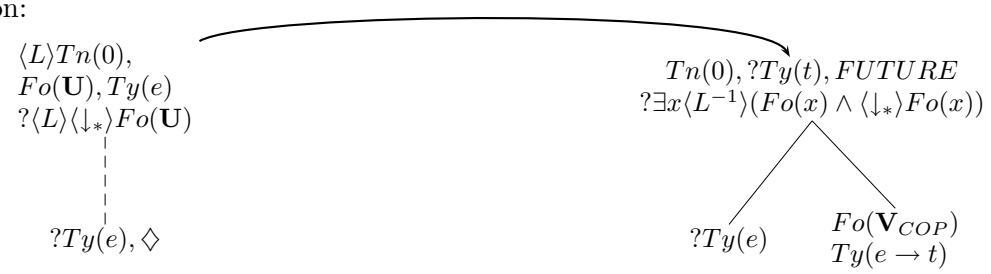


Parsing *ku-to*



Parsing *ku-to-ba*

From then on, the update actions will not merely PARALLEL that of the simpler cleft structure, they will be IDENTICAL. So, for example, the parsing of *ngu* will achieve the following transition, still with no formula content specific to the particular proposition under construction:



Parsing *ku-to-ba-ngu*

Then, exactly as before, the words will be parsed in sequence, achieving the fleshing out of this structure.

To achieve this, the lexical specification of *ku* needs to be an action that introduces the requisite independent linked structure, and ensures that the two structures once finally completed have a shared structure. And we shall need to ensure that the pointer returns from parsing the predicate place-holder, in these examples *-ba*, directly to the decoration of the linked node. So, for *ku* we need some restriction, if stated at the linked type *e* node which would express:

$$?\exists x(Fo(x) \wedge \langle L\rangle\langle\downarrow_*\rangle Fo(x))$$

With this proposed sequence of actions, there are some immediate predictions, to which the account will commit us. We anticipate, for example, that there must be identical matching of tense specifications both preceding and following the *nge* specification. This is indeed correct - these invariably match:

(49)

ku-	to-	ba	nge-bafana	la-	ba-	to	natsa	tjwala
DSM	FUT	COP	it-boys	REL	1SM	FUT	drink	alcohol

It will be the boys who will drink alcohol

(50)

ku	to-	ba	bu tjwala	la	ba	to	bu	natsa	bafana
DSM	FUT	COP	alcohol	REL	1SM	FUT	14OM	drink	boys

It will be alcohol which they will drink it the boys

(51)

Kw	-a	ku-	ngu-se-London			
SDM-	RemotePast	SDM	it-LOC-London			
lapho	ng-	-a-	ngi-	hlala	khona	na-make
where	1sg-	PAST	1SM	live	DEM	with mother

It was in London where I lived with my mother

The question, then, is how to define the lexical specification of the expletive form *ku*. This bleached form of the locative marker isn't restricted to occurring strictly initially, as it is duplicated as an agreement marker on a form of auxiliary, the remote-past, which is restricted to occurring before the subject-:

(52)

kw	a	ku	nge	bafana	la	ba	natsa	twjala
DSM	Remote Past	DSM	Nge	2boys	REL	2SM	drink	alcohol

It was the boys that drank alcohol

Furthermore, the past tense form *be* occurs before any subject marking, so that with this form, *ku* occurs second in the string:[17]

(53)

be	ku	bu twjala	bafana	la	be	ba	bu	natsa	
PAST	DSM	Pro	14alcohol	2boys	REL	PAST	2SM	14OM	drink

It was alcohol that the boys were drinking

The definition we propose assumes that *ku* itself triggers the construction of an inverse LINK transition, in a duplication of this particle, this occurring trivially but harmlessly:

	IF	$?Ty(t)$
	THEN	$\texttt{make}(\langle L^{-1}\rangle); \texttt{go}(\langle L^{-1}\rangle);$
ku_{EXPLET}-		$\texttt{put}(?Ty(e)); \texttt{go}(\langle L\rangle);$
		$\texttt{put}(?\exists x \langle L^{-1}\rangle Fo(x) \wedge \langle \downarrow_* \rangle Fo(x))$
	ELSE	Abort

Indeed we expect that *ku* might be duplicated, which indeed it can:

(54)

ku-	to-	ba	ku	nge-bafana	la-	ba-	to	natsa	tjwala
DSM	FUT	COP	DSM	it-boys	REL	1SM	FUT	drink	alcohol

It will be the boys who will drink alcohol

Notice that this expresses the restriction of there being a shared term for the two structures being induced on the type-t-requiring node, rather than on the structure to which that node is linked. This is because the pointer has to reside at this node to parse the auxiliary and tense specification, leaving a totally open linked structure to which the pointer will have to return.

It might seem that there has to be stipulation to force the pointer back to the linked structure introduced in the parsing of *ku*. However, in fact this isn't so. Notice first, to fix the intuition that the formal constraint will match, that once having set up a skeletal structure for a shift in context, there must be development of that structure if this strategy isn't to be entirely empty. More formally, if we adopt the context-relative formalism being explored in Cann et al forthcoming in which all structures are developed with respect to some independent context, then those contexts by definition contain at least one complete tree. This restriction is precisely what is needed to immediately yield the right result for these structures: any derivation in which the pointer failed to return to any introduced LINK tree would fail, as violating this restriction.

4 Theoretical Signifcance

So the overall story is that all types of clefts are seen as a topic plus predication structure, with the contrastiveness taken to be characteristic of clefts being brought out by the construction of a novel term within that structure. The particular force of the analysis is got because the underlying syntax is not a mechanism that licenses structures, with each structure so licensed being in principle distinct from other structures, but rather a mechanism that licenses structure-building processes. So while the processes are defined as distinct from each other,

[17] To structurally ensure this distinction before *be* and *to*, all we need to do is for *be* to presume a trigger of $?Ty(t)$ which is required to have no nodes developed from it, introducing the binary subject-predicate split as with *ti* but moving the pointer back to the topnode, *to* contrarily to require that there be a locally unfixed node, introduced by the *subject* agreement marker, introducing the subject-predicate pair of nodes and leaving the pointer at prediate-requiring node.

there may nevertheless be feeding relations between them, giving rise to the overlap effect which is so characteristic of cleft structures, part-relative, part-long-distance-dependency, part-topic-like.

While we have restricted our attention here to Siswati clefts, there is of course the question of the extent to which this form of analysis will generalisable, as we would surely expect that it would.[18] However, there is a word of caution here: the structures which these processes induce are not structures that are inhabited by words. To the contrary, the DS system a system of constraints on possible tree growth processes to yield formulae in a conceptual representation system whose formulae include predicates and individual-denoting terms: a "language of thought", by definition a universal system of representations.[19] The a set of mechanisms for tree-growth, equally, are defined a priori, a fixed system of defining partial conceptual representations and possible forms of transition.[20] This in total constitutes the underlying form of syntax that all natural languages reflect. The sole structural source of variation that they allow is in the balance between what are taken as general forms of tree growth, and what are induced by individual lexical items, individually isolatable triggers for tree growth macros stored in the lexicon. Thus the only concept of syntax specific to natural languages themselves is in the transitions from one partial structure to another, and in this there may be variation according as one language encodes a given transition as an update that is lexically triggered, another leaves the lexicon relatively weakly specified, leaving the transition as a generally available option.[21]

This has consequences for language typologies. The DS system provides a basis for language typologies only via this indirect route of providing mechanisms for building paired structures as induced by words of the language. Furthermore, natural languages are by assumption systems that provide parsing-based strategies, in principle allowing several options for any single string output-structure pairing. As noted earlier, there is no one to one correspondence between string-construal of interpretation and licensed sequence of transitions. Clefts, relative clauses, etc, are simply packages of instructions for building LOT formulae: they are not themselves structures of LOT itself. We would not accordingly expect there to be a fixed single structure associated with a given syntactic pattern, eg clefts, relative clauses and so on. These just constitute different effects which the composite set of options

[18]This would necessitate an account of English clefts as involving compatible forms of enrichment for all cases where the auxiliary in the cleft sequence doesn't match that of the auxiliary in the relative sequence:
(i) It was John that should have been here.
In these past tense cases, given the analysis of auxiliaries as weak place-holding device, nothing would preclude an enrichment provided by the second auxiliary. The only problem concerns the use of the present tense, but there are grounds for analysing English present tense as essentially a-temporal, picking up on other time specifications as these are made available:
(ii) It is John that saw Mary.
Whether this can be sustained requires detailed DS analysis of auxiliary and modal perspectives, and we leave this for future work.

[19]The system presumes on a denotational semantics without this being defined as part of the natural-language system.

[20]Because the particular concepts themselves are acquired through the natural language to which the child is exposed as the means of codifying hence retaining the individuating concepts with which the child can then reason, these are culture-relative, a stock to which the human may add right through life, as exposed to new constructs needing to be individuated.

[21]How strictly this minimal set of assumptions can be sustained depends on the numbers of additional features that need to be posited as part of the DU language as triggers needed to appropriately filter the tree-growth actions. Gender specifications provide one problematic instance; and we note that with evey addition that has to be made to any such set of features, so the substance of this particularly strong universalist position will be weakened.

for building linked structures makes available. To ask for a single analysis of the set of strings that are said to constitute cleft structures in any language is thus to ask the wrong question. Languages have different strategies for yielding paired linked structures, and it is the very interweaving of these which may determine informational effects according as this interweaving yields one language ordering or another.

Nonetheless, we suggest that what are somewhat loosely called cleft structures involve just this composite combination of a context-creating backgrounding device (via the building of a linked structure) plus a device presenting essential novel information. On this expectation, we would expect that languages would vary depending on restrictions which the individual language might impose, with rigidly verb-final languages in particular not able to project the particular full skeletal-proposition-providing structure PRIOR to providing the contextual substituend element which non-verb-final languages so transparently can, lacking, as they do, devices for anticipatory construction of place-holding predicates, what we more familiarly call auxiliaries.

More generally, what the grammar formalism is providing directly is a universal base-mechanism for parsing structural representations of content. What is then inducible from it is a set of typological expectations of the type of interactions which the associated tree-growth mechanisms would lead one to expect, hence providing part of the stock of data relative to which the grammar can itself be tested. But the substance of the grammar itself does not lie in those abstractions: it lies solely in the claim that the grammar formalism constitutes the vehicle for language processing definitive of natural languages.