Abstract

Chichewa, a Bantu language of East Central Africa, displays mixed properties of configurationality such as the existence of VP, on the one hand, and discontinuous constituents (DCs), on the other. In the present work we examine the discourse and syntactic properties of DCs, and show that DCs in Chichewa arise naturally from the discourse-configurational nature of the language. We argue that the fronted DCs in Chichewa are contrastive topics that appear in a left-dislocated external topic position, with the remnant part of the split NP in the right-dislocated topic position. Once the precise discourse functions of DCs are properly integrated into the syntactic analysis, all the facts and restrictions observed in Chichewa DCs can be explained in a straightforward fashion.

1 Mixed (Non-)Configurational Properties in Chichewa

Bantu languages display rich verbal agreement morphology comprising 16-18 noun classes that cross-reference the verb's core arguments (subject and primary object). As might be expected given such a elaborate system of noun classes, they exhibit properties of non-configurationality, but only partially. For example, the example in (1a) illustrates the unmarked word order (SVO) in a transitive sentence in Chichewa (Mchombo 2002). Here, the verb stem obligatorily inflects for the subject marker (SM) *zi-*, which agrees in the relevant ϕ-features with the subject NP *njuchi* 'bees'. The obligatory verbal agreement with the subject NP allows for relative freedom of word order, as illustrated in (1b).

(1) a. SVO: Njûchi zi-ná-lúm-a alenje. Chichewa
10bees 10-PAST-bite-INDIC 2hunters
'The bees bit the hunters.'

b. VOS: Zinálúma alenje njûchi.

Objects in Chichewa, on the other hand, are licensed configurationally inside VP. Thus, unlike the subject NP, whose ordering is relatively free, the object NP must be immediately postverbal. As shown in (2), any other patterns of word order are ungrammatical (Bresnan and Mchombo 1987: 744–745).

(2) a. OVS: *Alenje zi-ná-lúm-a njûchi.
2hunters 10-PAST2-bite-INDIC 10bees

b. VSO: *Zinálúma njûchi alenje.

c. SOV: *Njĉhi alenje zináluma.

d. OSV: *Alenje njûchi zináluma.

The non-VO orders are permitted only with the presence of an object marker (underlined), as shown in (3).

(3) a. SVO: Njûchi zi-ná-<u>wá</u>-lúm-a alenje.
bees SM-PAST-OM-bite-INDIC hunters
'The bees bit them, the hunters.'

b. VOS: Ziná<u>wá</u>lúma alenje njûchi.

c. OVS: Alenje ziná<u>wá</u>lúma njûchi.

d. VSO: Ziná<u>wá</u>lúma njûchi alenje.

e. SOV: Njĉhi alenje ziná<u>wá</u>luma.

f. OSV: Alenje njûchi ziná<u>wá</u>luma.

As argued thoroughly and conclusively by Bresnan and Mchombo (1987), these word order facts can be explained by the following assumptions: (i) the subject marker (SM) is functionally ambiguous between grammatical agreement and topic-anaphoric agreement; and (ii) the object marker (OM) is unambiguously used as a pronominal argument. Thus, as grammatical agreement, the verbal agreement with a clause-internal subject NP is obligatorily present, as in (1). The assumption in (i) that the SM can also function as a pronominal argument is verified by examples like that in (4), where the formally identical SM functions as a topic-anaphoric pronoun.

(4) Zi-na-wá-lúma.
10SM-PST-2OM-bite
'They (bees) bit them (hunters).'

The SM as a pronoun can also be the antecedent for the reflexive, as illustrated in (5).[1] In (5a) the reflexive *dzi* is locally bound to its antecedent *mikango*. In (5b) the reflexive still has the nominal *mikángo* as its antecedent. However, the antecedent is not within the local domain. It is the SM in the embedded clause that is in an anaphoric relation with the nominal *mikángo* 'lions'.

(5) a. Mikángo í-ma-dzi-kánd-a.
4lion 4-hab-refl-scratch-fv
'Lions scratch themselves.'

[1]In many Bantu languages the reflexive morpheme is invariant, and it appears in the morphological position of the OM.

b. Mikángo i-ku-úz-á anyáni kutí sí-í-ku-fún-á kutí
4lion 4-pres-tell-fv 2baboon that NEG-4-PRES-want that
njovu z-dzíw-é kutí í-ma-dzi-kánd-a.
10elephant 10-know-SUBJN that 4-HAB-REFL-scratch-fv
'The lions are telling the baboons that they don't want the elephants to know that they (lions) scratch themselves.'

Hale (1983) identifies three major properties associated with non-configurationality: (i) free word order, (ii) null anaphora, and (iii) the existence of (syntactically) discontinuous expressions. By 'null anaphora', Hale refers to the 'situation in which an argument (e.g. subject, object) is not expressed by an overt nominal expression in phrase structure' (Hale 1983:40). This is illustrated in examples like that in (4) for Chichewa. As already mentioned, the fact that the verbal agreement morphology also functions as a pronominal argument (in the case of the SM) allows for the freedom of word order (for the subject). There are also instances of discontinuous expressions, as we discuss in the remainder of this paper. In this sense, Chichewa (and Bantu in general) might be viewed as at least partially non-configurational. As it turns out, these non-configurational properties are always closely tied to the *discourse*-configurational nature of this language family, in which there are designated structural positions for discourse elements such as topic and focus. It is these discourse-related properties of referents that allow them to be freely displaced from their canonical syntactic positions. As discussed below, we see this not only at the sentential level, but also in the nominal domain, in which such discourse-driven restructuring results in discontinuous expressions.

The discussion in the rest of the paper proceeds as follows. In section 2 we present data on discontinuous constituents in Chichewa and provide a brief critical review of previous analyses of DCs. In section 3, we establish the discourse basis for our analysis of Chichewa DCs; and in section 4 we offer a structural analysis based on the crucial distinction between internal and external topic, the precise discourse properties of split NPs, and their structural correlates. The final section summarizes our discussion.

2 Discontinuous Constituents

The noun class concord in the verbal domain is also quite extensive in the nominal domain, as exemplified in (6). In (6a), the constituents of the complex NP meaning 'these foolish hunters' all agree with the noun class of the head (class 2). In (6b) the head noun *mikángó* 'lion' is class 4, and the modifiers also must agree.

(6) a. Njúchí izi zi-ná-lúm-á álenje awa ópúsa.
10.bees 10prox.dem 10-PST-bite-fv 2-hunter 2prox.dem 2-foolish

'These bees bit these foolish hunters.'

b. Mikángó i-tátu i-ná-gúmúl-á makólá ónse a-náyi.
4lions 4three 4PST-pull.down-fv 6corrals all 6four
'Three lions pulled down all the four corrals.'

Although parts of these complex NPs typically occur together with the head noun,[2] it is possible, to split these nominal constituents although this option is restricted. Example (7a) shows the canonical NP structure in Chichewa. As shown, it exhibits a strict head-initial structure with Head-Demonstrative-Adjective order. The examples in (7b)-(7f) show various patterns of discontinuity of that NP (boldfaced).

(7) a. Njúchií izi zi-ná-lúm-á **álenje awa**
10.bees 10.PROX.DEM 10-PST-bite-FV 2.hunter 2.PROX.DEM
ópúsa.
2-foolish
'These bees bit these foolish hunters.' ... [H D A]

b. **awa** njúchií izi zi-ná-wá-lúm-a **álenje ópúsa**. D ... [H A]

c. **álenje** njúchií izi zi-ná-wá-lúm-a **awa ópúsa**. H ... [D A]

d. **álenje awa** njúchií izi zi-ná-wá-lúm-a **ópúsa**. [H D] ... A

e. **awa ópúsa** njúchií izi zi-ná-wá-lúm-a **álenje**. [D A] ... H

f. **álenje ópúsa** njúchií izi zi-ná-wá-lúm-a **awa**. [H A] ... D

Note that all the instances of discontinuity of the object NP above are accompanied by the presence of the OM that is coreferential with the whole NP, regardless of which part of the object NP (head or modifier) is discontinuous. Without the OM the examples are ungrammatical:

[2]The integrity of the complex NPs in (6) can be shown by their occurrence in displaced positions such as passive, topicalization, and cleft (see Kathol and Rhodes 2000 for relevant observations).

(i) a. Álenje awa ópúsa a-ná-lúm-ídw-á ndí njúchí izi.
2hunter 2prox.dem 2foolish 2-PST-bite-PASS-fv by 10bees 10prox.dem
'These foolish hunters were bitten by these bees.'

b. Ndi makólá ónse anáyi améné mikángó itátu íná-gúmúl-á.
COP 6corrals all 6four 6replo 4lion 4three 4-PST-pull.down-fv
'It was all the four corrals that the three lions pulled down.'

(7′) b′. *__awa__ njúchií izi zi-ná-∅-lúm-a **álenje ópúsa**.

c′. *__álenje__ njúchií izi zi-ná-∅-lúm-a **awa ópúsa**.

d′. *__álenje awa__ njúchií izi zi-ná-∅-lúm-a **ópúsa**.

e′. *__awa ópúsa__ njúchií izi zi-ná-∅-lúm-a **álenje**.

f′. *__álenje ópúsa__ njúchií izi zi-ná-wá-lúm-a **awa**.

The presence of the OM is crucial in that those NPs that cannot be cross-referenced by the corresponding OM (or SM) cannot be discontinuous. For example, an instrumental phrase like *ndí makású awa óbúntha* ‘with these blunt hoes’ in (8) in a non-applicative construction cannot be discontinuous.

(8) a. Mikángó yókálamba i-ná-zí-gúmúl-a **ndí makású**
4lion 4aged 4SM-PST-10OM-demolish-fv with 6hoe
awa óbúntha nkhókwe.
6these 6blunt 10granary
'The aged lions pulled down the granaries with these blunt hoes.'

b. ***Awa óbúntha** mikángó yókálamba i-na-zí-gúmúl-a
6these 6blunt 4lion 4aged 4-PST-10OM-demolish-fv
ndí mákásu nkhókwe.
with 6hoe 10granary

Chicheŵa exhibits object asymmetry (see Alsina and Mchombo 1993; Bresnan and Moshi 1990; Ngonyani 1998). In an applicative construction, only the applied object has the properties associated with the primary object. For example, in (9), only the beneficiary object introduced by applicative *mikángó yókálamba* 'aged lions', and not the theme object *makású awa óbúntha* 'these blunt hoes', can be in anaphoric relation with the incorporated pronominal object.

(9) a. Anyání a-na-<u>í</u>-gúl-íl-á makású awa óbúntha
2baboon 2-PST-<u>4OM</u>-buy-APPL-fv 6hoe 6these 6blunt
mikángó yókálamba.
4lion 4aged
'The baboons bought (for) them these blunt hoes, (for) the aged lions.'

b. *Anyání a-na-<u>wa</u>-gúl-íl-á mikángó yókálamba makású
2baboon 2-PST-<u>6OM</u>-buy-APPL-fv 4lion 4aged 6hoe
awa óbúntha.
6these 6blunt
[Intended as:] 'The baboons bought them for the aged lions, these blunt hoes.'

The examples in (10) show that only the applied beneficiary, and not the theme object, can be discontinuous.

(10) a. **Yókálamba** anyání a-na-<u>í</u>-gúl-íl-á makású awa
4aged 2baboon 2-PST-<u>4OM</u>-buy-APPL-fv 6hoe 6these
óbúntha **mikángó**.
6blunt 4lion
'The baboons bought the aged lions these blunt hoes.'

b. *__Awa__ __óbúntha__ anyání a-na-<u>wa</u>-gúl-íl-á __makású__
6these 6blunt baboon 2-PST-<u>6OM</u>-buy-APPL-fv 6hoe
mikángó yókálamba.
4lion 4aged

Similarly, the oblique agent in a passive sentence cannot be cross-referenced by an OM and hence resists discontinuity, as shown in (11).

(11) a. Mikángó i-na-ph-édw-á ndí **alenje awa ó-dzí-kónd-a**.
4lion 4-PST-kill-PASS-fv by 2hunter 2these 2-REFL-love-fv
'The lions were killed by these selfish (self-loving) hunters.'

b. *__Ó-dzí-kónd-a__ mikángó i-na-ph-édw-á ndí **alenje awa**

c. *__Awa__ mikángó i-na-ph-édw-á ndí **alenje ó-dzí-kónd-a**

As expected from the obligatory presence of the topic-anaphoric OM with a discontinuous object NP, the DCs receive a topic interpretation. More precisely, our preliminary inquiry into discourse contexts of various instances of DCs suggests that the fronted element is often a contrastive topic equivalent to a left-dislocated topic, rather than simply given information, or continuing topic. Given the analysis of the Chicheŵa OM as a topic-anaphoric pronoun, the fact that the OM is required when part of the object NP is discontinuous shows that at least the fronted discontinuous part of the NP must be outside the minimal clausal domain.

In short, the discontinuous examples presented above share the following basic characteristics: (i) DCs in Chicheŵa occur clause-initially; and (ii) clause-initial DCs receive contrastive topic interpretation and require an anaphoric pronoun on the verb corresponding to the whole NP. The observation in (i) that DCs appear in the clause-peripheral position seems to be true for a majority of languages that allow such split NP construction (see Baker 1996 for polysynthetic languages; Dahlstrom 1987 for Algonquian languages in particular). Given that in many languages, clause-initial position is reserved for discourse-related elements such as topic and focus, the observation in (i) lends itself well to another aspect noted in (ii): that fronted DCs receive topic interpretation. In fact, we will show that 'topicalizability' is a precondition for any constituent to be discontinuous (at least in Bantu). As argued by Bresnan and Mchombo (1987), the Chicheŵa object marker is employed only as a pronominal argument anaphoric

to a floating topic outside the minimal clause nucleus (S/IP), never as grammatical agreement to a non-topical (clause-internal) NP. The observation in (ii) is therefore confirmed by the morphosyntax as well. In previous generative studies of DCs (e.g. Baker 1996, Jelinek 1984, Speas 1990), however, relatively little attention is given to the discourse function of DCs.

There is nonetheless some important work that recognizes the role of information structure in split constituents in general: Reinholtz (1999), for example, argues that clause-initial DCs in Swampy Cree have the discourse function of Focus, and that more generally, the Swampy Cree split NP construction has 'all of the hallmarks of *wh*-movement in so-called configurational languages' (p.202) in that '... both movement types show the ability to span several clauses, a limited application in relative clauses or embedded questions, and an inability to move any material out of adverbial constituents' (p.218). Reinholtz therefore argues that DCs arise as a result of *wh*-movement.

Fanselow (2001) examines split XP constructions in general, such as split VPs as in (12) and split DPs as in (13) in German.

(12) **Keine Bücher** hat er [__ gelesen].
no books has he read

(13) **Schrecklicher Morde an Studenten** ist er **vieler** beschuldigt worden.
horrible murders at students is he many accused been
'He has been accused of many horrible murders of students.'

Fanselow argues that such split XP constructions are generally associated with a particular pragmatic structure: 'in a split construction, the right part of XP must be focal, while the lefthand part may be a (link-)topic or a second focus' (p.85). Although the precise pragmatic nature of the fronted elements still deserves further discussion, these studies nonetheless suggest that the discourse-pragmatic functions of split constructions must be part of any analysis.

Two other observations are relevant for our analysis of the syntax of Chicheŵa DCs. First, regardless of the position, the ordering of *contiguous* elements is fixed – H(ead) > D(emonstrative) > A(djective) – as shown by the contrast between (7) and (14).

(14) a. *Njúchií izi zi-ná-lúm-á **awa álenje ópúsa**. *... [D H A]

b. ***awa** njúchií izi zi-ná-wá-lúm-a **ópúsa álenje**. *D ... [A H]

c. ***álenje** njúchií izi zi-ná-wá-lúm-a **ópúsa awa**. *H ... [A D]

d. ***awa álenje** njúchií izi zi-ná-wá-lúm-a **ópúsa**. *[D H] ... A

e. ***ópúsa awa** njúchií izi zi-ná-wá-lúm-a **álenje**. *[A D] ... H

f. ***ópúsa álenje** njúchií izi zi-ná-wá-lúm-a **awa**. *[A H] ... D

The ordering restriction on the fronted elements suggests that they form a single constituent. This need not always be the case, however. For example, when the subject NP is left-dislocated, it can come between the two parts of the object DCs, as in (15). In such cases, these discontinuous parts of the object NPs may come in any order, each forming a separate constituent: as shown in (15), the canonical head-modifier ordering *mikángo (lion) ó-kálamb-a (aged)* is not maintained.

(15) **Yó-kálamb-a** anyaní **mikángo** a-na-í-gúl-íl-á makású
4aged 2baboons 4lion 2-PST-4-buy-APPL-fv 6hoes
awa ó-búnth-a.
6these 6-blunt-fv
'The aged lions$_j$, the baboons$_i$, they$_i$ bought them$_j$ these blunt hoes.'

The second additional observation concerns DCs involving complex possessive NPs. As shown by example (16), a possessive NP can be split in Chicheŵa.

(16) a. Anyaní á mísala a-ku-pwány-a **chipanda**
2-baboon 2ASSOC 4-madness 2-PRES-smash-fv 7-calabash
chá kazitápé.
7ASSOC 1-spy
'The mad baboons are smashing the calabash of the spy.'

b. **Chipanda** anyaní á mísala a-ku-chí-pwány-a **chá kazitápé**.
'The calabash, the mad baboons are smashing (it) of the spy'

c. **Chá kazitápé** anyaní á mísala a-ku-chí-pwány-a **chipanda**.
'Of the spy, the mad baboons are smashing (it) the calabash'

However, as soon as we add another layer of possessive NP, splitting gets more constrained. Consider the examples in (17). Example (17a) is a non-discontinuous example. The element in question, the object possessive NP, is in boldface. In (17b) we front the head noun of the possessive NP, and the result is ungrammatical.[3] In (17c) we front a possessor *a mfumu* 'of the chief'. Again

[3]Note that the example (17b) would be good if there were no OM. In this case, however, we only get the appositive interpretation of the fronted element. The absence of the corresponding OM thus suggests that nothing is out of the basic clause, and that the sentence-initial element is added on to the sentence as an appositive. We return to this contrast between (17b) and the appositive reading without an OM when we discuss the information structure of the non-fronted elements.

the example is rendered ungrammatical. Example (17d), on the other hand, shows that it is possible to front the entire possessor and leave the head noun postverbal.

(17) a. Anyaní a-na-mphwanya **chipanda chá alenje a mfumu**.
2baboons 2-PAST-smash 7calabash 7ASSOC 2hunter 2ASSOC 1chief
'The baboons smashed the calabash of the hunters of the chief.'

b. ***Chipanda**$_i$ anyaní a-na-chi-mphwanya $\_\_{}_i$ **chá alenje a mfumu**.
7calabash 2baboons 2-PAST-7-smash 7ASSOC 2hunter 2ASSOC 1chief
'The calabash, the baboons smashed of the hunters of the chief.'

c. ***A mfumu**$_i$ anyaní a-na-chi-mphwanya **chipanda chá alenje** $\_\_{}_i$.
2ASSOC 1chief 2baboons 2-PAST-7-smash 7calabash 7ASSOC 2hunter
'Of the chief, the baboons smashed the calabash of the hunters.'

d. **Chá alenje a mfumu**$_i$ anyaní a-na-chi-mphwanya **chipanda** $\_\_{}_i$.
7ASSOC 2hunter 2ASSOC 1chief 2baboons 2-PAST-7-smash 7calabash
'Of the hunters of the chief, the baboons smashed the calabash.'

At this point, we leave these facts simply as an additional observation about complex possessive NPs. In the analysis to follow, we suggest that the constraint that bans the examples in (17b, c) must be formulated in terms of the information structure and heaviness of the parts of the NP that remain postverbal rather than the syntax of complex possessive NPs. In the next section, we develop a base-generation account of our DC data, taking into consideration the information structure of both parts of the split NPs.

3 Chichewa DCs as an External Topic Construction

Based on the basic properties observed earlier that (i) DCs in Chichewa must occur clause-initially; and (ii) clause-initial DCs receive topic interpretations and require an anaphoric pronoun on the verb corresponding to the whole NP, we analyze the split NP constructions as instances of topicalization, in which the

clause-initial DCs are left-dislocated outside the minimal clause. The topicalization analysis of DCs is consistent with the fact that every instance of DCs requires the OM on the verb and with the analysis given by Bresnan and Mchombo (1987) that the OM in Chicheŵa is reserved only for topic-anaphoricity.

Additional data show that 'topicalizability' is in fact a pre-condition for a constituent to be discontinuous. For example, Chicheŵa has a number of verb-object idioms, in which the object is formally non-referential, as in example (18a). Non-referential NPs can never be topics, and, as such, they cannot be discontinuous, as demonstrated in (18b, c).

(18) a. Nd-a-gwil-a mwendo wáko.
1SG-PREF-grab-fv 3leg 3your
(lit.) 'I have grabbed (your) leg.' = 'I apologize.'

b. *Wáko nd-a-gwil-a mwendo.

c. *Mwendo nd-a-gwil-a mwendo.

Similarly *wh*-phrases, which are inherently focused, cannot be fronted:

(19) a. Mikango u-na-gumula **nyumba ya yani**? *wh*-phrases
lion sm-past-destroy house of who
'Whose house did the lions destroy?'
b. ***ya yani** mikango u-na-gumula **nyumba**?

More precisely, following Morimoto's (2000) proposal that Bantu languages exhibit two types of topic, EXTERNAL and INTERNAL, we take these clause-initial DCs to be EXTERNAL TOPICS. Before we present our analysis of split NPs, in the next section we briefly discuss the nature of external and internal topics and the motivation for this distinction in Bantu languages.

3.1 Two Types of Topic in Bantu

The two types of topic, internal and external, are distinct both structurally and pragmatically, and are motivated for various unrelated languages. Core characteristics of these types of topic are summarized in Table 1.

As for the structural position, external topics (E-TOPs) are always outside the minimal nuclear clause; in many languages, they occupy adjoined positions – CP-adjoined, as in Russian (King 1995) or IP-adjoined, as in Malay (Alsagoff 1992) – or E(xpression) nodes as in Mayan languages (Aissen 1992). As such, the E-TOP is generally allowed only in matrix clauses. The internal topic (I-TOP), on the other hand, appears inside the minimal nuclear clause, often conflates with the grammatical subject in so-called 'subject-oriented' languages like English, and hence, is typically not marked off intonationally from

	EXTERNAL TOPIC	INTERNAL TOPIC
Position	E(xpression) node (Mayan, Russian)	SpecCP (Mayan) IP-adjoined (Russian)
Bind argument?	no	yes
Resumptive pronoun?	yes	no
Island constraint?	no	yes
Discourse status	new topic (Mayan)	continuing topic (Mayan)
Definite?	yes	yes
Embedding?	no (Mayan, Russian)	yes (Mayan, Russian)

Table 1: Characteristics of external and internal topics

the rest of the clause. E-TOPs may be either arguments or adjuncts, while I-TOPs must be one of the core arguments (subject, primary object). Thus, there can naturally be multiple E-TOPs, while the I-TOP is restricted to only one per clause. Pragmatically, the E-TOP is used for contrastive/new topic while the I-TOP is old information, continuing topic.[4]

These two types of topic are also motivated independently of split constructions in Bantu (see Morimoto 2000). An external topic in left-dislocation is exemplified in (20) from Kinyarwanda. (21) exemplifies multiple left-dislocated topics in Kirundi (Sabimana 1986). Being characteristic of E-TOP, these topics are marked off intonationally in Bantu; they are in an anaphoric relation to the corresponding pronouns in the clause. For example, in (21), the class 1 subject marker *y* on the verb corresponds to Mudúga, the class 7 object marker *ki* corresponds to the secondary object *igitabo* 'book', and the class 2 object marker *bá* corresponds to the primary object *abâna* 'children'. They are pronominal arguments coreferential with the dislocated elements.

(20) **bîno bitabo**$_i$, úmwáalimu a-ra-shaak-a ko
these books teacher SM-PRES-want-ASP that
tu-*(<u>bi</u>)-sóm-a $\_\_{}_i$.
SM-OM-read-ASP
(lit.) '**These books**, the teacher wants that we read them.' Kinyarwanda

(21) **Igitabo**, **Mudúga**, **abâna**, y-a-rá-ki-bá-ha:ye
7book 1Muduga 2children 1-PST-FOC-7-2-give
'The book, Muduga, (to) the children, he gave it to them.' Kirundi

The internal topic is observed most readily in inversion constructions. (In canonical word order subjects are often default I-TOPs.) For example Kirundi

[4]For a detailed discussion, see Aissen 1992, Alsagoff 1992, King 1995, and Morimoto 2000.

exhibits inversion of subject and object in a transitive sentence, where the canonically postverbal object appears in the preverbal subject position just in case it is a (continuing) topic, and the (focused) subject appears postverbally, as illustrated in (22).

(22) a. Uwo muhungu a-a-ra-gaburiye ubuyabu.
that boy 3s-PST-AF-feed.ASP cats
'That boy fed the cats.' Kirundi (Morimoto 2000)

b. Ubuyabu bu-a-gaburiye uwo muhungu.
cats 3pl-PAST-feed.ASP that boy
'(It's) That boy$_{FOC}$ (who) fed the cats$_{TOP}$.' subject-object reversal

In Chicheŵa a similar construction is observed involving locative, where the logical subject appears postverbally and the locative preverbally. The inverted locative is typically old information (Bresnan and Kanerva 1989).

(23) m-nkhalăngo mw-a-khal-á míkângo.
18-9forest 18-PERF-remain-FV 4-lion
'In the forest have remained lions.' Chicheŵa locative inversion

The distinct properties of the internal topic are manifested in various ways. Unlike the external topic, the internal topic is not marked off intonationally and is not accompanied by an object marker; instead (in the inversion constructions), the preverbal object agrees with the verb like the canonical subject (note the 3rd person plural agreement *bu* on the verb corresponding to *ubuyabu* 'cats' in (22b)). These facts suggest that the internal topic occupies a clause-internal position generally reserved for subjects. Pragmatically, the internal topic is a continuing (old) topic rather than a new or contrastive topic. The preverbal constituents in inversion constructions have exactly that pragmatic function. In many Bantu languages, subjects also have the restriction that they must be old information. In those languages, we may assume that subjects are always internal topics (see Morimoto 2000 for a discussion of Bantu subjects as default internal topics).

3.2 Discourse Functions and Syntactic Positioning of Split NPs

Having established the two types of topic for Bantu, we now turn to the discussion of the discourse functions and structure of the split construction in Chicheŵa. As mentioned earlier, the fronted part of a split NP is typically a contrastive topic, and this is in line with the assumption that functionally the fronted part is an external topic. According to C. Lee (1999a,b) while TOPIC is prototypically given, presupposed, and anchored in speech situation, CON-

TRASTIVE TOPIC has a focal part in contrast with the (aforementioned) discourse topic, and the speaker has the alternatives in contrast or contrast set in mind. While topic can be unaccented, contrastive topic shows a prominent intonation pattern cross-linguistically.

In Chichewa, the contrastive part of a topic constituent appears in the left-dislocated position, resulting in a split construction. For example, for the split example in (7d), repeated here in (24), the most likely context is where there are two sets of foolish people in prior discourse – these foolish hunters and those foolish fishermen. The expression *álenje awa* 'these hunters' is then contrasted with 'those fishermen' in the example. The 'foolish' part of the NP is old, non-contrastive information, and remains postverbal. Indeed, the fact that every instance of a discontinuous object NP requires the corresponding OM suggests that no part of the object NP remains inside the VP. This means that the remaining postverbal part of the object NP must be right-dislocated. This assumption is in line with the presumed discourse function of this part of the DC: it is the non-contrastive, given information.

(24) **álenje awa** njúchií izi zi-ná-wá-lúm-a
2.hunter 2.PROX.DEM 10.bees 10.PROX.DEM 10-PST-2-bite-FV
ópúsa.
2foolish
'These bees bit these foolish hunters.'

Preliminary investigation of the phonological phrasing of split examples in Chichewa (Féry, Mchombo, and Morimoto, in preparation) corroborates the observations regarding the discourse status and syntactic positions just noted. In a canonical SVO sentence in Chichewa, the subject forms its own phonological phrase separate from the VP, and the verb and object form one phonological phrase (see also Bresnan and Kanerva 1989), as schematically shown in (25).

(25) (subject)$_{pp}$ (SM-verb object)$_{pp}$ Canonical SVO sentence

The immediate postverbal object position is also a designated focus position in Chichewa (Bresnan and Kanerva 1989; Morimoto 2000). When the subject is focused and is in this postverbal focus position inside VP, as in the inversion construction we saw earlier in (23), the postverbal subject is phrased together with the verb like an object.

(26) (I-topic$_i$)$_{pp}$ (SM$_i$-verb focus)$_{pp}$ Logical subject in focus position

When the object is right-dislocated and the OM is present on the verb, the right-dislocated object forms its own phonological phrase separate from the verb:

(27) (subject)$_{pp}$ (SM-TNS-OM$_i$-verb)$_{pp}$ (object$_i$)$_{pp}$ Object right dislocation

Now in a split NP construction like that in (24), the fronted part forms its own phonological phrase, and then comes the subject, forming another phonological phrase as in the usual case. Then next comes the verb, again with its own phonological phrase, separate from the remaining part of the split object NP.

(28) (DC)$_{pp}$ (subject)$_{pp}$ (SM-TNS-OM-verb)$_{pp}$ (rest-of-DC)$_{pp}$
Split construction

The phonological phrasing of the split construction in (28) clearly shows that no part of the split NP is inside the minimal clause nucleus consisting of the subject and VP. In addition, we also have a preliminary result showing that the two contiguous parts of the fronted DC form one phonological phrase, suggesting that syntactically they also form one constituent.

To summarize, the available data suggest that DCs in Chicheŵa are best analyzed as an external topic construction, in which the dislocated elements are external to the minimal nuclear clause. Pragmatically they serve as a contrastive topic rather than continuing topic, as characteristic of external topics in other languages. The external topic analysis of DCs in Chicheŵa is not in line with Reinholtz's (1996) analysis that DCs have focus and arise by way of *wh*-movement. We suggest here that languages that permit split NP constructions make use of them for discourse purposes, but exactly what function DCs have may depend on the information structuring of an individual language. While focus (or discourse-prominent elements in general) may be expressed clause-initially in Algonquian languages (see Aissen 1992), in Bantu languages clause-initial position is strictly reserved for topic, and focus is expressed postverbally (see Morimoto 2000). Thus, given the patterns of information structuring in Bantu, clause-initial DCs would naturally receive a topic interpretation.

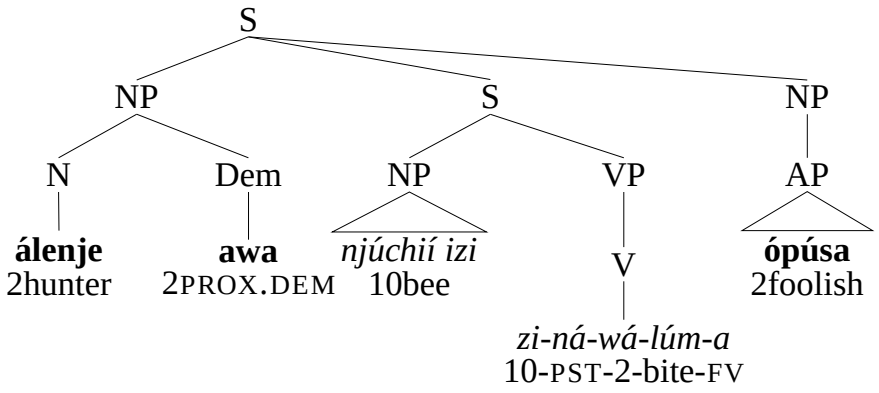
4 Discourse-Configurational Analysis

Taking the discourse functions and phonological phrasing as our basis, we now consider the syntactic structure of split NPs. The key analytical problems we wish to solve are the following: (i) functional identification of the DCs with the associated argument function; and (ii) configurational identification of the types of topic involved the split construction – namely the external, contrastive topic in the left-periphery and the afterthought topic in the right-periphery. We first lay out some theoretical assumptions in our analysis.

4.1 Parallel Structures

Based on the previous discussion we propose a syntactic structure for the split construction as shown in (29). For illustration, we use the example in (24) above.

(29)



The proposed structure assumes the basic architectural properties of Lexical-Functional Grammar, a representational theory of grammar in which parallel levels of representations are related not through derivations but via a set of correspondence principles (see Bresnan 1982, 2001; Bresnan and Kaplan 1995). Grammatical principles postulated in this framework are thus interpreted as constraints on surface forms. The two parallel structures fundamental in LFG are C(ONSTITUENT) STRUCTURE and F(UNCTIONAL) STRUCTURE.

The c-structure is represented as a familiar phrase structure tree and encodes precedence and dominance relations among syntactic words and phrases (NP, VP, CP, etc). Unlike the phrase structure assumed in derivational approaches, the c-structure directly models the surface forms of language. Thus, no empty nodes and traces are represented. Consequently, languages can vary in the c-structure representation of a particular utterance/expression. In addition, LFG posits two types of clausal organization in natural languages: the endocentric clausal organization with headed XPs, and the exocentric one with S. As in the structure in (29), we make use of the exocentric category S for languages that lack independent evidence for I. In Bantu languages, all verbs inflect uniformly like main verbs, and there is no particular class of inflectional verbs that behave otherwise. For this reason, it has been proposed that Bantu clauses consist of the exocentric category S rather than IP (e.g. Bresnan and Mchombo 1987; Morimoto 2000, 2001).

The f-structure is represented as attribute-value pairs. Unlike c-structure, f-structure is unordered, and it encodes predicate-argument relations and other morphosyntactic and semantic information in language-independent form. F-structure attributes may be grammatical functions (SUBJ, OBJ) or morphosyntactic feature categories (TENSE, MOOD, CASE, NUMBER, PERSON). F-structure values can be another f-structure, semantic content (e.g. 'boy'), or atomic symbols (PAST, ACC, SG). Given the type of information it encodes, the f-structure is largely invariant across languages. And it is at this level of representation that we see how parts of the discontinuous topic are related to the OM on the verb.

These parallel structures are related not through movement, but through correspondence principles. For illustration, let us take the minimal clause S in the

structure in (29), 'These bees bit them (the hunters)'. First we provide the semantic and morphosyntactic information carried by each lexical item and the functional schemata associated with it.

(30) a. Njúchií izi zi-ná-wá-lúm-á.
10.bees 10.PROX.DEM 10-PST-2-bite-FV
'These bees bit them.'

b.

njúchií	N	(↑ PRED) = 'bee'
		(↑ AGR) = 10
izi	D	(↑ PRED) = 'these'
		(↑ AGR) = 10
zi-ná-wá-lúm-á	V	(↑ PRED) = 'bite ¡SUBJ, OBJ¿'
		(↑ TENSE) = PAST
		(↑ SUBJ) = ↓
		(↓ AGR) = 10
		(↑ OBJ)
		(↓ PRED) = 'pro'
		(↓ AGR) = 2

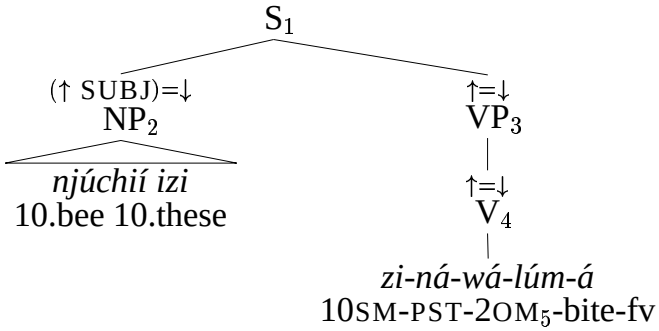
The ↑ arrows in the lexical entries in (30) are taken to refer to the f-structure of the preterminal node (e.g. N in the first lexical entry 'bee'). The ↓ arrows then refer to the f-structure of the terminal node. As shown, in addition to the verb (stem) which selects for subject and object in its lexical entry, the SM on the verb provides the equation '(↑ SUBJ) = ↓', stating that the verb's f-structure contains SUBJ, and the f-structure of the subject contains the AGR attribute whose value is class 2 ('(↓ AGR) = 2'). Similarly, the OM on the verb provides the equation '(↑ OBJ) = ↓', stating that the verb's f-structure contains OBJ. The OM contains a pronominal content and provides the PRED value 'pro', along with the AGR information.

The mapping from the c-structure to the corresponding f-structure is shown in (31)–(32). Each syntactic node is arbitrarily numbered to show which node maps to which f-structure. For example, the top-most node S_1 maps to the outermost f-structure also numbered 1; the subject NP numbered NP_2 maps to the f-structure numbered 2; and so forth. Formally this mapping from c- to f-structure is mediated by functional annotations on the syntactic nodes. The ↑ arrow points not to the mother node, but to the f-structure the mother node maps to. Likewise the ↓ arrow points to the f-structure the annotated node maps to, not the annotated node itself. Thus ↑ = ↓ on the VP means that the f-structure of the mother node S is the same as the f-structure of VP_3. In LFG, S_1 and VP_3 are said to be 'co-heads'.[5] In (31), S_1, VP_3 and V_4 are co-heads, and they all

[5]The notion of co-heads is similar to that of 'extended heads', as proposed by Grimshaw (1991, 1997): heads and intermediate projections within a projection line with the same cat-

map to the outermost f-structure. The annotation on NP_1 '(↑ SUBJ) = ↓' states that the f-structure of node 1 (S) has a SUBJ attribute whose value is identified with the f-structure of node 2 (subject NP).[6] The OM on the terminal node of V provides the functional descriptions shown in the lexical entry of the verb in (30) above, and these provide the well-formed f-structure of OBJ with the PRED value, numbered 5.

(31)



(32)

$$\begin{bmatrix} \text{PRED} & \text{'bite <SUBJ, OBJ>'} \\ \text{SUBJ}_{1,3,4} & _2\begin{bmatrix} \text{PRED} & \text{'bee'} \\ \text{AGR} & 10 \\ \text{PROX} & + \end{bmatrix} \\ \text{OBJ} & _5\begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{AGR} & 2 \end{bmatrix} \\ \text{TENSE} & \text{PAST} \end{bmatrix}$$

These standard functional application expressions are thus given the following interpretation:

(33) (*fa*) = *v* holds if and only if *f* is an f-structure, *a* is an attribute and the pair <*a v*> ∈ *f*.

In other words, a simple equation like (↑ SUBJ) = ↓ annotated on an NP is satisfied if and only if the f-structure corresponding to the mother node of that NP has an SUBJ attribute whose value is the f-structure corresponding to the annotated NP.

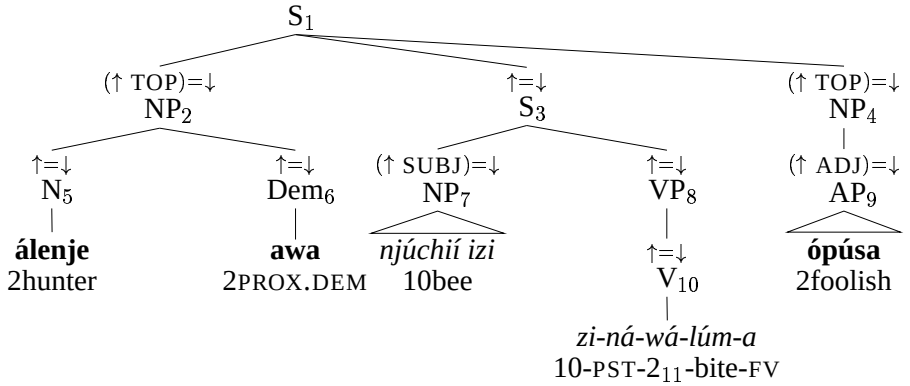
egorial features. For example, I is the head of IP, and V, V′, VP, and I′ are extended heads of IP.

[6]The f-structure of the SUBJ (numbered 2) is somewhat simplified.

4.2 Completeness, Coherence, and Inside-Out Function Application

Now we consider the whole structure with the split NP in (29), and see how the topic function of the DC can be formally identified in the current system. We annotate the earlier tree in (29) as in (34). The corresponding f-structure is shown in (35).

(34)



The annotation on the fronted DC (NP_2) and the remnant part of the DC (NP_4) '(↑ TOP) = ↓' states that the f-structure of the mother node (S_1) contains TOP, whose value is identified with the f-structure of the respective NP. The annotation on AP builds an inner f-structure of the ADJ(unct) function inside the f-structure of TOPIC. As we see, an f-structure can be constructed by information coming from multiple syntactic nodes.

(35)

$$\begin{bmatrix} & \text{TOP} & {}_{2,4,5,6}\begin{bmatrix} \text{PRED} & \text{'hunter'} \\ \text{AGR} & 2 \\ \text{PROX} & + \\ \text{ADJ} & {}_{9}\begin{bmatrix} \text{PRED} & \text{'foolish'} \\ \text{AGR} & 2 \end{bmatrix} \end{bmatrix} \\ {}_{1,3,8,10}\begin{bmatrix} \text{SUBJ} & {}_{7}\begin{bmatrix} \text{PRED} & \text{'bee'} \\ \text{AGR} & 10 \end{bmatrix} \\ \text{PRED} & \text{'bite'} \\ \text{TENSE} & \text{PAST} \\ \text{OBJ} & {}_{11}\begin{bmatrix} \text{PRED} & \text{'pro'} \\ \text{AGR} & 2 \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Completeness and Coherence: The functional identification of TOPIC with the argument function OBJ is ensured by the principles of COMPLETENESS and COHERENCE, or more precisely, EXTENDED COHERENCE. Completeness requires that every function designated by a predicate be present in the f-structure of that

predicate (Bresnan 2001:63). Thus, completeness rules out examples like that in (36), where all the arguments selected by the predicate *give* are not present.

(36) *John gave a book.

Note that completeness is a requirement that applies at the level of f-structure, and does not require that all the arguments be present at c-structure. Null argument languages like Japanese and Korean, for example, allow an utterance like that in (36), but at the level of f-structure, all the arguments selected by the predicate are represented and provide their morphosyntactic information and semantic content.

Now in examples like that in (34), part of the DC is the ADJUNCT function (AP) inside the object NP. Completeness is not sufficient to license such elements because it only requires that the selected arguments be properly represented in the f-structure. These adjuncts, not properly selected by the predicate, nonetheless must be properly integrated into the semantics of the predicate and its arguments. COHERENCE, or the EXTENDED COHERENCE CONDITION, on the other hand, ensures just this type of well-formedness. Coherence requires that every argument function in an f-structure be designated by a PRED. The principle rules out ill-formed examples like that in (37) (Bresnan 2001:63).

(37) *We talked *the man* about that problem for days.

$$\begin{bmatrix} \text{PRED} & \text{'talk ¡SUBJ, OBL¿'} \\ \text{SUBJ} & [\text{"we"}] \\ *\text{OBJ} & [\text{"the man"}] \\ \text{OBL} & [\text{"about that problem"}] \\ \text{ADJ} & [\text{"for days"}] \\ \text{TENSE} & \text{past} \end{bmatrix}$$

The intransitive verb *talk* takes an optional oblique argument, and PRED has the OBL designator in (37). It has no OBJ designator, however; having the extra argument violates the coherence condition and results in an ill-formed f-structure.

While the coherence condition applies only to argument functions (SUBJ, OBJ1, OBJ2, OBL), the extended coherence condition applies to all syntactic functions, requiring them to be appropriately integrated into an f-structure (Bresnan and Mchombo 1987; Fassi Fehri 1984; Zaenen 1985). As stated above, argument functions are integrated when they are designated by the PRED. Adjuncts are integrated if their immediate f-structure contains a PRED. The grammaticized discourse functions TOP and FOC are integrated if they are functionally identified, or anaphorically linked to, an integrated function.

Inside-Out Function Application: Returning to our example in (34), the

TOPIC function in the left- and right-periphery is properly integrated into the f-structure in (35) by the extended coherence condition, but completeness and extended coherence must be satisfied by one of the arguments identifying TOPIC as being associated with it. As we have seen, in a sentence with an object DC, the DC is cross-referenced by the obligatory presence of the object marker on the verb. In other words, the OM provides the information about a larger domain than that which contains the OM (VP) – namely that the object NP is dislocated outside the minimal clause containing the OM and has a topic-anaphoric function. This view of the OM is analogous to Nordlinger's (1998) constructive case, in which case markers are said to carry clause-level information. The central idea of constructive case is the use of inside-out function application, as stated in (38).

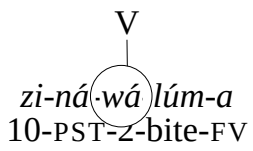
(38) **Inside-Out Function Application**: for any f-structure f' and attribute a, (af') designates the f-structure f such that $(fa) = f'$.

For an illustration, let a be OBJ. In the regular designator $(f\ \text{OBJ})$, f denotes the f-structure from which we can follow a path inwards through OBJ to another f-structure (f'), as in (39). This is the standard 'outside-in' function application defined earlier in (33). In the 'inside-out' function application, we have the designator $(\text{OBJ}\ f')$. f' denotes the f-structure from which we can follow a path outwards through OBJ to a higher f-structure (f).

(39) $f: \left[\text{OBJ} \quad f' \left[\ \right]\right]$

We can now add the appropriate annotations to the OM using inside-out function application, which will formally associate the TOP function with OBJ at the level of f-structure. The verb form in (34) is repeated below in (40).

(40) a.



((OBJ ↑) TOP) = ↑
(↑ PRED) = 'pro'
(↑ AGR) = 2

b. $\left[\text{OBJ} \quad \left[\begin{matrix}\text{PRED} & \text{“pro”}\\ \text{AGR} & 2\end{matrix}\right]\right]$

c. $\left[\begin{matrix}\text{TOP} & \left[\ldots\right]_i \\ \\ \text{OBJ} & \left[\begin{matrix}\text{PRED} & \text{“pro”}\\ \text{AGR} & 2\end{matrix}\right]_i\end{matrix}\right]$

The ↑ arrow in the designator (OBJ ↑) below the OM points to the f-structure of the OM. Starting from that f-structure, (OBJ ↑) states that the f-structure of

the OM is contained in the OBJ function, which is inside some larger f-structure. This will instantiate the f-structure shown in (40b). (OBJ ↑) TOP) states that this larger (outer) f-structure containing OBJ also contains the TOP attribute. The whole annotation ((OBJ ↑) TOP) = ↑ then means that the value of the TOP attribute, another f-structure, is identical to the f-structure of OM (= the OBJ function). The final f-structure instantiated by the annotation is shown in (40c). In other words, inside-out function application allows for the straightforward functional identification of TOPIC with OBJ at the level of f-structure and eliminates movement of various elements on c-structure, which may be difficult to motivate outside this particular construction.

4.3 From Information Structure to C-structure

In this section we consider the second analytic problem identified earlier for the configurational identification of the f-structure function TOPIC with particular types of discourse topics – namely the external, contrastive topic in the left-periphery and the afterthought topic in the right-periphery.[7]

Crosslinguistically, these types of discourse topic seem to be associated with the respective syntactic positions just noted. For example regarding the left-peripheral topic, in verb-initial languages, D. Payne (1990, 1992) identifies the preverbal position to be what she refers to as the 'pragmatically marked' (PM) position. The PM information is non-presupposed asserted new information, contrastive information (i.e. focus) as well as given, discourse-prominent information (topic). Payne shows that in strongly verb-initial languages, these pragmatically marked constituents, either focus or topic, appear sentence-initially. Cooreman (1992:244) essentially makes the same observation: the non-verb initial order in the canonically verb-initial language Chamorro is commonly found when 'the thematic unity of the [narrative] is disrupted', such as change of events, or when the paragraph theme is temporarily suspended. Cooreman's description of these sentence-initial elements in Chamorro is comparable to Aissen's (1992) description of the external topic – the new or contrastive topic. Subsequent work on verb-initial languages makes similar observations about the discourse function of the sentence-initial position (e.g. Harold 1995:50 for Biblical Hebrew; Tsimpli 1995 for Modern Greek).

In SVO languages, new or contrastive topics also appear at the left-periphery in a dislocated position. Birner and Ward (1998:256–257) show that among the various syntactic constructions that encode different types of discourse referents in English (e.g. inversion, *by*-phrase passive, topicalization, existential, left-dislocation, right-dislocation), new or contrastive topic (hearer-new or discourse-new in Birner and Ward's taxonomy) is expressed in the left-

[7]The discussion in this subsection is based on the fuller review of the cited literature given in Morimoto 2000, chapter 2.

dislocated position. As regards another SVO language, Tok Pisin, a creole language in Papua New Guinea, Sankoff (1993) provides an example showing that (what we would call) a new/contrastive topic appears in a left-dislocated position followed by an anaphoric pronoun.[8]

In SOV languages, where scrambling and case marking are common typological features, contrastive topics may not always appear in a left-dislocated position. They are nonetheless morphologically and prosodically clearly marked, according to C. Lee (1999a,b). In Korean, for example, even though topics with the topic marker *-(n)un* can scramble, the canonical position of these topics seems to be clause-initial (Choi 1999). In German, contrastive elements (topic or focus) occur in left-peripheral (SpecCP) position (see Berman 2000; Choi 1999 and earlier references cited in these works).

As for the right-dislocated topic, it is observed for a number of languages that the right-dislocated position is reserved for afterthought or discourse-old information – e.g. Takami 1995 for Japanese and English, Birner and Ward 1998 for English, Sells 1998 for Japanese, Kimenyi 1980 for Kinyarwanda; see also Morimoto 2000, chapters 4–5, which discusses the afterthought function of right-dislocated elements in Bantu languages.

These crosslinguistic studies of left- and right-topics collectively tell us that there is a robust tendency that these types of topics are structurally defined. As briefly discussed earlier, our preliminary findings on phonological phrasing of these left- and right-topics (Féry, Mchombo & Morimoto in preparation) indicate that they each form their own phonological phrase. These observations about the structural correlates at the syntactic and phonological level together suggest a grammatical architecture in which there is a flow of information, or mapping, (at least) between discourse or information structure ('i-structure') and c-structure, on the one hand, and i-structure and prosodic structure, on the other. Within the current model, we believe that it is the mapping between i-structure and c-structure that gives the f-structure notion of TOPIC particular discourse interpretations, where the left-peripheral external topic (c-structure notion) is interpreted as contrastive topic (discourse notion), and the right-peripheral topic as afterthought. Exactly how the mapping between these levels of structure should be represented will have to be left open for future research, but this line of research has been pursued by King (1995) and Choi (1999) for

[8] An example of a new/contrastive topic from Sankoff (1993) is given below (p.121). The dislocated topic is in small caps, and the anaphoric pronoun is underlined.

(i) kakaruk na pik wonem samting i-stap. Na OLGETA MAN IA ol i-poret long
chicken and pig what something stay and all people DET 3pl afraid of
guria na ol i-go pinis.
earthquake and 3pl go complete
'(Only) chickens and pigs and whatever were there. But ALL THE PEOPLE, they were afraid of the earthquake and they had all left.'

other types of discourse referents (e.g. given information, contrastive and completive foci).

4.4 Further Consequences of the Right-Dislocation Analysis of the 'Remnant'

Having presented the structural analysis of the parts of Chicheŵa split NPs, let us return to the restriction on the splitting of complex possessive NPs mentioned earlier in section 2, and see how the facts can be explained in our analysis. The relevant examples from (17) are repeated here in (41). The observation was that of the various splitting possibilities for a complex possessive NP, the only grammatical instance is where the head noun remains and the rest is fronted, as in (41d).

(41) a. Anyaní a-na-chi-mphwanya **chipanda chá alenje**
2baboons 2-PAST-7-smash 7calabash 7ASSOC 2hunter
a mfumu.
2ASSOC 1chief
'The baboons smashed the calabash of the hunters of the chief.'

b. ***Chipanda**$_i$ anyaní a-na-chi-mphwanya $\_\_{}_i$ **chá alenje**
7calabash 2baboons 2-PAST-7-smash 7ASSOC 2hunter
a mfumu.
2ASSOC 1chief
'The calabash, the baboons smashed of the hunters of the chief.'

c. ***A mfumu**$_i$ anyaní a-na-chi-mphwanya **chipanda**
2ASSOC 1chief 2baboons 2-PAST-7-smash 7calabash
chá alenje $\_\_{}_i$.
7ASSOC 2hunter
'Of the chief, the baboons smashed the calabash of the hunters.'

d. **Chá alenje a mfumu**$_i$ anyaní a-na-chi-mphwanya
7ASSOC 2hunter 2ASSOC 1chief 2baboons 2-PAST-7-smash
chipanda $\_\_{}_i$.
7calabash
'Of the hunters of the chief, the baboons smashed the calabash.'

Our speculation about this pattern is that it is not due to some syntactic constraint, but is constrained (at least partly) by phonological weight – namely that only one prosodic word is allowed in the right-dislocated position, where the constituent forms its own phonological phrase. A similar observation is made for non-discontinuous right-dislocation in other Bantu languages. For

example, in Kinyarwanda, Kimenyi (1980:203) observes that whereas multiple left-dislocated topics are possible (see example (21) from the closely related language Kirundi), right-dislocated topics are restricted to only one, as shown in (42). The right-dislocated topics are in boldface.

(42) *Umgabo y-a-ya-mu-haa-ye, **amafaraanga**, **umugóre**..
man 1SM-PAST-it-give-PERF money woman
'The man gave it to her, the money (to) the woman.'

Furthermore, we noted earlier in note 3 that (41b) would be grammatical if the fronted head noun *Chipanda* 'calabash' had an appositive interpretation. Crucially, in that case the verb cannot have the OM. This suggests that what appears to be fronting with the appositive interpretation in fact involves neither fronting of any element nor right-dislocation of the 'remnant' element(s), and that the clause-initial appositive element is simply added on to a canonical SVO sentence. Therefore, assuming that our right-dislocation analysis of the remnant is correct, we conjecture that this right-dislocated position imposes a constraint on phonological weight, and DCs involving 'heavy' remnants are dispreferred.

5 Conclusion

In this paper, we have offered a discourse-configurational analysis of Chicheŵa split NPs, where the fronted element, the contrastive topic, occupies an external topic position, and the remnant part of the split NP, the afterthought, appears in right-dislocated position. The analysis is consistent with the fact that every instance of object DCs requires the corresponding object marker on the verb, whose function is topic-anaphoric (Bresnan and Mchombo 1987). The structural analysis is also supported by preliminary findings on phonological phrasing of DCs (Féry, Mchombo, and Morimoto, in preparation). Given the right-dislocated analysis of the remnant part of a split NP, we speculated that the constraint on splitting of complex possessive NP has to do with phonological weight – that heavy elements are dispreferred in right-dislocated position.

Examining DCs beyond the Bantu family would naturally require looking at various discourse functions that DCs serve in the languages in question and determining the structural correlates of such discourse elements. Nonetheless we hope that, in future research, our analysis of Chicheŵa split NP will be a step in the right direction towards taking into account multiple levels of representations (discourse, syntax, phonology) in order to provide a comprehensive analysis of split constructions.

References

Aissen, Judith. 1992. Topic and Focus in Mayan. *Language* 68(1): 43–80.

Alsagoff, Lubna. 1992. *Topic in Malay: The Other Subject*. Ph.D. diss., Stanford University, Stanford, CA.

Alsina, Alex, and Sam A. Mchombo. 1993. Object Asymmetries and the Chicheŵa Applicative Construction. In Sam A. Mchombo ed., *Theoretical Aspects of Bantu Grammar*, chapter 1, CSLI, Stanford, CA, pp. 17–45.

Baker, Mark C. 1996. *The Polysynthesis Parameter*. Oxford University Press, New York/Oxford.

Berman, Judith. 2000. *Topics in the Clausal Syntax of German*. Ph.D. diss., Universität Stuttgart, Stuttgart.

Birner, Betty J., and Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English*. John Benjamins, Amsterdam/Philadelphia.

Bresnan, Joan ed. 1982. *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, MA.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell Publishers, Oxford.

Bresnan, Joan, and Jonni M. Kanerva. 1989. Locative Inversion in Chicheŵa: A Case Study of Factorization of Grammar. *Linguistic Inquiry* 20(1): 1–50.

Bresnan, Joan, and Sam A. Mchombo. 1987. Topic, Pronoun, and Agreement in Chicheŵa. *Language* 63(4): 741–782.

Bresnan, Joan, and Lioba Moshi. 1990. Object Asymmetries in Comparative Bantu Syntax. *Linguistic Inquiry* 21(2): 147–185.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. CSLI, Stanford, CA.

Cooreman, Ann. 1992. The Pragmatics of Word Order Variation in Chamorro Narrative Text. In Payne (1992), 243–263.

Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell, and Annie Zaenen eds. 1995. *Formal Issues in Lexical-Functional Grammar*. CSLI, Stanford, CA.

Downing, Pamela, and Machael Noonan eds. 1995. *Word Order in Discourse*. Vol. 30 of *Typological Studies in Language*. John Benjamins, Amsterdam/Philadelphia.

É. Kiss, Katalin ed. 1995. *Discourse Configurational Language*. Oxford University Press, New York/Oxford.

Fassi Fehri, Abdelkader. 1984. Agreement in Arabic, Binding, and Coherence. In Michael Barlow and Charles A. Ferguson eds., *Agreement in Natural Language*, CSLI, Stanford, CA, pp. 107–158.

Féry, Caroline, Sam Mchombo, and Yukiko Morimoto. In preparation. Phonological and Syntactic Phrasing of Topic and Focus in Chicheŵa Split NPs. MS. Berlin-Potsdam/Berkeley.

Grimshaw, Jane. 1991. Extended Projection. MS. Department of Linguistics and Center for Cognitive Science, Rutgers University.

Grimshaw, Jane. 1997. Projection, Heads, and Optimality. *Linguistic Inquiry* 28(3): 373–422.

Hale, Ken. 1983. Warlpiri and the Grammar of Non-Configurational Languages. *Natural Language & Linguistic Theory* 1(1): 5–47.

Harold, Bruce B. 1995. Subject-Verb Word Order and the Function of Early Position. In Downing and Noonan (1995), pp. 137–161.

Jelinek, Eloise. 1984. Empty Categories, Case, and Configurationality. *Natural Language & Linguistic Theory* 2(1): 39–76.

Kathol, Andreas, and Richard Rhodes. 2000. Constituency and Linearization of Ojibwa Nominals. MS. University of California, Berkeley.

Kimenyi, Alexandre. 1980. *A Relational Grammar of Kinyarwanda*. University of California Press, Berkeley, CA.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. CSLI, Stanford, CA.

Lee, Chungmin. 1999a. Contrastive Topic: A Locus of the Interface. In K. Turner ed., *The Semantics/Pragmatics Interface from Different Points of View*. Elsevier Science.

Lee, Chungmin. 1999b. Topic, Contrastive Topic and Focus: What's on Our Minds. Plenary paper presented at the Second International Conference on Cognitive Science, Waseda, Tokyo.

Morimoto, Yukiko. 2000. *Discourse Configurationality in Bantu Morphosyntax*. Ph.D. diss., Stanford University, Stanford, CA.

Morimoto, Yukiko. 2001. Verb Raising and Phrase Structure Variation in OT. In Peter Sells ed., *Formal and Empirical Issues in Optimality-Theoretic Syntax*, CSLI, Stanford, CA, pp. 129–196.

Ngonyani, Deogratias. 1998. Properties of Applied Objects in Kiswahili and Kindendeule. *Studies in African Linguistics* 27: 67–95.

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. CSLI, Stanford, CA.

Payne, Doris L. 1990. *The Pragmatics of Word Order: Typological Dimensions of Verb Initial Languages*. Mouton de Gruyter, Berlin/New York.

Payne, Doris L. ed. 1992. *Pragmatics of Word Order Flexibility*. Vol. 22 of Typological Studies in Language. John Benjamins, Amsterdam/Philadelphia.

Sabimana, Firmard. 1986. *The Relational Structure of the Kirundi Verb*. Ph.D. diss., Indiana University, Indiana.

Sankoff, Gillian. 1993. Focus in Tok Pisin. In Francis Byrne and Donald Winford eds., *Focus and Grammatical Relations in Creole Languages*, Vol. 12 of *Creole Language Library*, John Benjamins, Amsterdam/Philadelphia, 117–140.

Sells, Peter. 1998. Japanese Postposing Involves No Movement. Paper presented at the meeting of the Linguistics Association of Great Britain, Lancaster, April 1998.

Speas, Margaret. 1990. *Phrase Structure in Natural Languages*. Vol. 21 of Studies in Natural Language and Linguistic Theory. Kluwer Academic Publishers, Dordrecht/Boston/London.

Takami, Ken-ichi ed. 1995. *Rightward Movement Constructions in English and Japanese*. Hituzi Syobo, Tokyo.

Tsimpli, Ianthi Maria. 1995. Focusing in Modern Greek. In É. Kiss (1995), pp. 176–206.

Zaenen, Annie. 1985. *Extraction Rules in Icelandic*. Garland, New York.