

Chapter 20

Variation in the expression of information structure in eastern Bantu languages

Steve Nicolle

Canada Institute of Linguistics / SIL International

Despite significant typological similarities, eastern Bantu languages differ in how information structure is expressed, but much of this variation only becomes apparent when discourse considerations are taken into account. Using data from narrative texts in eleven eastern Bantu languages I highlight three parameters of variation. First, in some languages surveyed all topics must be left dislocated, while in others certain kinds of topics in specific discourse contexts may be right dislocated. Second, all languages surveyed express argument focus on non-subjects through cleft constructions and right dislocation. However, while argument focus can be expressed through right dislocation of the subject in most languages surveyed, it can be expressed only through the use of cleft constructions in three languages, one of which allows right dislocation of subjects in response to content questions. Third, whilst all of the languages havethetic (topicless) sentences with VS constituent order, most also allow SV constituent order in the orientation sections of narratives, and one language allows SVthetic sentences elsewhere as well.

1 Introduction

The eastern Bantu languages¹ are very similar as far as general morphosyntactic parameters are concerned: they all exhibit typical Bantu noun class systems with obligatory subject marking on verbs, optional object marking, and S>V>IO>DO>X constituent order in pragmatically neutral clauses. However, they also exhibit various kinds of morphosyntactic variation (Marten, Kula & Thwala 2007; van der Wal & Biberauer 2014), including variations in constituent order due to different instantiations of information structure (Zerbian 2006; Buell, Riedel & van der Wal 2011; Yoneda 2011; Downing & Hyman 2016). Most studies of information structural variation in Bantu languages to date

¹ The term *eastern* here indicates the geographical distribution of these languages within Africa rather than a genetic affiliation.



have relied predominantly on data collected through elicitation. Such data is well suited to isolating salient interpretational effects in individual sentences, but is less appropriate when investigating the effect of larger stretches of discourse on constituent order variation. This methodological bias is found in studies of African languages more generally, so that, for example, Güldemann, Zerbian & Zimmermann (2015: 157) in their review of research into information structure (IS) in African languages, stated, “we largely restrict the discussion to IS phenomena that hold within a sentence and exclude those in discourse units above this level.” In part, this situation reflects a lack of available text corpora. Mwamzandi’s (2014) study of variation in adnominal demonstratives and reciprocal constructions in Swahili is one of the few investigations into information structure in a Bantu language that is based on an extensive text corpus (the Helsinki Corpus of Swahili, consisting of 12.5 million words from written Swahili texts).

In this paper, I attempt to complement existing research into information structure in Bantu languages in two ways. First, this study is based on the expression of information structure in narrative texts rather than in elicited data.² There are a number of limitations but also benefits to be derived from using narrative texts as data sources. The most important limitations are that a) not every type of construction (contrastive topic, given topic, contrastive focus, corrective focus, verum focus, etc.) is attested in each text corpus, and texts representing other genres could well reveal additional constructions and parameters of variation; b) corpus data do not yield negative evidence – that is, textual data only reveals what is possible, not what is impossible (that is, ungrammatical or pragmatically implausible); and c) prosody is only available in cases where a narrative was transcribed from an oral source and that source is still available (punctuation only represents prosody indirectly and imperfectly, and there is no guarantee that prosody will be appropriately reproduced when written narratives are read aloud). The main benefits of a textual approach are that it reveals patterns beyond the sentence, and it is not dependent on speakers’ judgements, invented contexts, or translation. The second way in which this study is designed to complement existing research is that it describes variation in information structure in eleven Bantu languages, thereby complementing more detailed studies of individual languages.³ These languages and the associated data sources are listed in Table 1.

2 Information structural generalizations

Information structure concerns the way in which an utterance or text is structured to accommodate the (assumed) knowledge state of the addressees, thereby helping addressees to arrive at a coherent interpretation of the utterance or text. Information structure in eastern Bantu narrative texts is primarily expressed through variations in the relative order of subject, verb, object and oblique constituents in a sentence, although intonation,

² Unless otherwise stated, all examples are from texts that were either published in the sources listed in Table 1 or used in their preparation. All examples are in the orthographies used in the published sources.

³ The current study is part of a larger investigation into narrative discourse in eastern Bantu languages (Nicolle 2015a).

Table 1: Languages included in this study

Language	Clf. ^a	Location	Data	Source
Fuliiru	[flr] DJ63	South Kivu, Democratic Republic of Congo	153 texts studied; data drawn from 13 (1,000 clauses approximately)	Van Otterloo (2011; 2015)
Digo	[dig] E73	Kwale District, Kenya	5 lightly edited oral texts and 2 written texts (864 clauses)	Nicolle (2015b)
Jita	[jit] EJ25	Mara Rural District, Tanzania	10 lightly edited oral texts (1,096 clauses)	Pyle & Robin- son (2015)
Kwaya	[kya] EJ251	Mara Rural District, Tanzania	10 lightly edited oral texts (1,015 clauses)	Odom (2015)
Suba- Simbiti	[ssc] EJ403	Mara Rural District, Tanzania	8 lightly edited oral texts (513 clauses)	Masatu (2015)
Kabwa	[cwa] EJ405	Mara Rural District, Tanzania	9 lightly edited oral texts and 2 written texts (530 clauses)	Walker (2011)
Rangi (Langi)	[lag] F33	Kondoa District, Tanzania	66 texts (3,200 clauses)	Stegen (2011)
Mwani	[wmw] G403	Cabo Delgado Province and Quirimba archipelago, Mozambique	7 texts (number of clauses not known)	Floor (2005)
Bena	[bez] G63	Wanging'ombe District and Njombe District, Tanzania	3 written and 3 lightly edited oral texts (674 clauses)	Broomhall (2011); Eaton (2015a)
Malila	[mgq] M24	Mbeya Rural District, Tanzania	10 lightly edited oral texts and 4 written texts (755 clauses)	Eaton (2015b)
Makonde	[kde] P23	Makonde Plateau, Mozambique	8 lightly edited oral texts (585 clauses)	Leach (2015)

^a Languages are classified using the ISO 639-3 language code as cited in *Ethnologue* (Lewis, Simons & Fennig 2015), indicated in square brackets in this table; and Maho's (2003) updated version of Guthrie's (1967-71) referential classification of the Bantu languages, indicated by capital letter(s) plus numeral.

pauses and – in certain languages – focus markers also play a role. Three parameters of variation will be discussed in this paper:

1. the position of the topic, specifically whether any topics can be right dislocated;
2. how argument focus on the subject is expressed;
3. howthetic sentences are expressed, specifically whether SV constituent order is possible.

Before looking in detail at how each of these parameters is instantiated in the data, I shall define the terms topic, focus andthetic sentence, and related concepts.

Although topic and focus are complementary notions, they belong to separate dichotomies (De Cat 2007: 66; Erteschik-Shir 2007: 42). The first dichotomy is topic-comment (corresponding to what Halliday (1967) and Dik (1981; 1989) call *theme-rheme*): the (conceptual) topic of a sentence is what a sentence provides information about, and the comment is whatever is predicated about the topic. Topic expressions⁴ may be syntactically integrated with the comment (internal topics) or syntactically non-integrated (external topics, also called *topic frames* or *themes*⁵). Both possibilities are illustrated in example (1) below. The context is that the animals have dug a well, but Hare did not help; Hare has repeatedly stolen water, so Tortoise plans to catch Hare and cut off his tail as a punishment. The topic of each clause is *oyo* ‘that’, referring to Hare. In the first clause this is a non-integrated (external) topic, and in the second clause it is an integrated (internal) topic, as it is coreferential with the object prefix *mu-*.

- (1) Jita (‘Well’ text, line 33)
- [Oyo]_{TOPIC} [munane era ripanga,]_{COMMENT}
that_one give(PL).me just machete
- [oyo]_{TOPIC} [enimugwata ara,]_{COMMENT}
that_one I.will.him.catch just
- ‘That one, just give me a machete, that one, I’ll just catch him.’

The second dichotomy is focus-presupposition: The presupposition of a sentence is the grammatically or lexically expressed information “which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered” (Lambrecht 1994: 52). The focus is defined in contrast to the presupposition as the grammatically or lexically expressed information “which cannot be taken for granted at the time of speech” (Lambrecht 1994: 207) and which, moreover, is considered by the speaker to be the most important information in the sentence (Dik 1989: 277). Within this broad definition, focus is often correlated with new information (Good 2010: 39), where information can be new either in relation to the discourse as a whole or in relation to the

⁴ Henceforth *topic* will be used to designate a topic expression (i.e. a linguistic form) and *conceptual topic* will be used where the referent rather than the referring expression is intended.

⁵ Dik (1981: 129–144) makes a distinction between *theme*, which is defined as an expression designating “a domain or universe of discourse with respect to which it is relevant to pronounce the following predication” (ibid. 130), and *topic*, which is part of that predication.

predicate in question (referred to in Gundel & Fretheim 2004 as referential givenness-newness and relational givenness-newness respectively), but also includes contrastive and identificational focus (see Gibson et al. To appear).⁶

Finally,thetic sentences are sentences in which neither the topic-comment dichotomy nor the focus-presupposition dichotomy is applicable. Thetic sentences are typically used to introduce new participants into a narrative, and to report events or situations in which neither the action nor the participants can be taken for granted, for example, in answer to a general question such as “What happened?” As such, they do not contain a topic. It has been argued that thetic sentences contain neither topic nor focus expressions (Yoneda 2011: 755), however following Lambrecht (1994; 2000) and Nicolle (2015a: 55), we will assume that, since the whole sentence is informationally prominent (that is, it is asserted), a thetic sentence exhibits sentence focus.

3 Obligatory and optional left dislocation of topics

The pre-verbal domain is restricted to lexical topics and non-focus subjects in all of the eastern Bantu languages surveyed; that is, focus elements may not occur pre-verbally. This may be true for most eastern Bantu languages (see Zerbian 2006; van der Wal 2009; Yoneda 2011), except for languages such as Kikuyu [E51], which has a pre-verbal focus marker that is derived historically from a copula construction (Schwarz 2003; 2007; van der Wal 2014). Pre-verbal topics are said to be *left dislocated*, meaning that they occur outside of the clause nucleus. However, there is variation concerning whether all topic expressions must be left dislocated or whether certain topics may be right dislocated in specific discourse contexts.

Left dislocation is used in a broad sense (as in Shaer et al. 2009) which subsumes both left dislocation proper and topicalization.⁷ Left dislocation proper involves a resumptive element, such as the class 9 object marker *i-* in example (2) below, which is coreferential with the topic *barabara* ‘road’. Topicalization involves a non-resumptive (‘gap’) construction, as in (3) where there is no object prefix corresponding to the topic *pesa* ‘money’.

- (2) Digo (Nicolle 2013: 237)
 [Barabara]_{TOPIC} ndiphō [a-ka-i-rich-a.]_{COMMENT}
 9.road then 3SG-SEQ-9.OM-leave-FV
 ‘The road, then, he left it.’

⁶ For a critical evaluation of the notion of focus as a universal linguistic category, see Matic & Wedgwood (2013).

⁷ In this paper, *topicalization* refers to a structural property and does not entail that the constituent exhibiting this property also functions as a topic in discourse. In the literature the term *fronting* is also used to refer to both topicalization and left dislocation proper (cf. Cohen 2009: 313).

- (3) Digo (Nicolle 2013: 228)
 [Pesa]_{TOPIC} sino [hu-na-Ø-hew-a.]_{COMMENT}
 10.money us 1PL-PRES-Ø-be.given-FV
 ‘Money, us we are given.’

The topics in the examples above correspond to the grammatical object of each clause. In Makonde, Fuliiru, Mwani, and possibly in other languages (but not in Digo), left dislocation of the object can also be used to give prominence to the final constituent of the sentence. In (4) below, moving the object *kiryā kijumba* ‘that box’ out of the comment emphasizes the fact that it was laid on the bed (as opposed to being opened, laid elsewhere, etc.). The object is left dislocated but follows the subject, which is also left dislocated, giving two topics and giving focal prominence to the comment.⁸

- (4) Fuliiru (Van Otterloo 2011: 350)
 [Ulya munyere]_{TOPIC1} [kiryā kijumba]_{TOPIC2} [a-na-ki-gwejez-a ku
 1.that 1.girl 7.that 7.box 3SG-SEQ-7-lay-FV on
 ngingo.]_{COMMENT}
 bed
 ‘That girl, that box, she laid it on the bed.’

Most topics, however, are subjects, and such topic-comment constructions conform to the canonical SVO constituent order. Nevertheless, subjects in topic position can be separated from the verb by a non-core element such as an exclamative or adverbial phrase, such as *bhuri rusiku* ‘every day’ in the following example.⁹

- (5) Jita (Pyle & Robinson 2015: 32)
 Eyo mw=ibhara [wamembe]_{TOPIC} [bhuri rusiku :aa-jaga¹⁰ mu=mugunda
 there in=forest hyena every day he.PST-went in=field
 gwaye.]_{COMMENT}
 his
 ‘There in the forest, Hyena, every day he went to his field.’

Topics may also be extracted from their host clause(s), which then intervene between the topic and the comment. In (6), the topic *rhibuyi eryo* ‘that rock’ is the object of a purpose clause, which in turn is the complement of a possessive clause, which is the complement of ‘see’ in a temporal clause. (The embedded clauses are indicated in square brackets. Extraction from multiple embedded clauses such as this may not be possible in all languages surveyed.)

⁸ For discussion of a similar construction in Matengo, see Yoneda (2011: 756–758).

⁹ In this and subsequent examples glossing will be by word rather than by morpheme as a morpheme-based gloss is not necessary for the analysis of constituent order. In long examples, for reasons of space, only a free translation will be provided.

¹⁰ In examples from Jita and Kwaya, the symbol <: > at the beginning of a verb indicates far past tense and the symbol < > at the beginning of a verb indicates narrative tense with 3SG subjects. The past anterior and the far past, and the 3SG form of the narrative and the 1SG form of the anterior are only distinguished through tone, which is not marked in the orthography. Thus, these symbols differentiate the forms.

- (6) Jita (Pyle & Robinson 2015: 32)

Ribhuyi eryo_i [ejire arora [atari na [ja kwasisya Ø_i]]] eeganirisya muno.
 rock that when he.saw he.is.not with of to.break Ø he.thought much
 ‘That rock_i, [when he saw [he has nothing [to break (it_i)]], he thought a lot.’

In all of the languages surveyed, topics are left dislocated at points of textual discontinuity. Textual discontinuity occurs at episode and paragraph breaks, when events are presented in a non-iconic order (including elaborations and parenthetical material), and when the topic is a *switch topic*. A switch topic (also called *shifted topic* and *link* in Vallduví 1992: 109–110) occurs when the conceptual topic of the sentence under consideration is different from the conceptual topic of the immediately preceding sentence. In the following example, the topic changes from *yuno* ‘that one’, referring to the speaker’s brother, to *go mafuha ga taa* ‘that lamp oil’.

- (7) Digo (‘Sababu ya Bahari Kuhenda Munyu’ text, line 11)

Pho [yuno]_{TOPIC} [kanipha vitu]_{COMMENT} ananipha mafuha ga taa
 so that.one he.does.not.give.me things he.gives.me oil of lamp
 bahi, [go mafuha ga taa]_{TOPIC} ndo [n’yarya?]_{COMMENT}
 only that oil of lamp EXCL I.go.eat
 ‘So that one doesn’t give me anything, he only gives me lamp oil, that lamp oil – can I go eat it?’

A switch topic is distinct from a continued topic, which is a topic that was also the topic of the previous sentence. Both switch topics and continued topics that occur at points of discontinuity are typically expressed lexically, and always in a pre-verbal position (with the exception of renewed topics in Mwani and temporary topics in Rangi, to be discussed below). Continued topics at points of discontinuity include conceptual topics that continue across paragraph and episode boundaries, and topics that are repeated when describing events narrated out of sequential order, such as elaborations. In the following example, the conceptual topic remains the man who is mentioned in the first sentence (indicated in bold). The sentence after the direct quotation starts a new paragraph, indicated by the use of the past tense *a-* in *warima* ‘he farmed’ as opposed to the consecutive tense *chi-* in *achiamba* ‘he said’ in the previous sentence. The sentence initiating the new paragraph starts with the continued topic *yuya bwana* ‘that gentleman’ in pre-verbal position:

- (8) Digo (Nicolle 2015b: 65)

Yuya mlume achiamba, “Mino rivyo nchirima tsula n’naphaha, phahi ndarima dza phapha na ko Mwamtsola, ili niphahe vitu vinji vyanjina niguze nigule ng’ombe.”

Yuya bwana warima munda uchifika dza Mazera, na hiku uchifika dza Malindi ela kaguwire hata tsere mwenga.

‘**That man** said [CONS], “Me when I farmed a termite mound I was getting (a lot of food), so I shall farm from here to Mwamtsola, so that I will get lots of things to sell so I can buy a cow.”

That gentleman farmed [PST] a field as large as from from Mazeras to Malindi, but he didn't get even one grain of maize.'

The majority of continued topics do not occur at points of discontinuity, and as most topics are subjects, continued topics are typically expressed using just a subject prefix on the verb. Since the subject prefix is obligatory in all of the languages surveyed, strictly speaking there is no overt *lexical* topic expression in such clauses, though there is an understood conceptual topic. Occasionally, however, a continued topic is expressed lexically even when there is no discontinuity in the text. When this occurs, eastern Bantu languages differ in how such topics are expressed. In most eastern Bantu languages, topics must occur pre-verbally, whereas in others, topics in certain discourse contexts may occur post-verbally.

In most languages surveyed, all topics are left dislocated, and any post-verbal element is interpreted as the focus. The following examples from Jita and Fuliiru texts illustrate left dislocated continued topics in which the topic (*wamembe* 'hyena' in Jita and *wandare* 'lion' in Fuliiru) is a subject (preserving the canonical SV constituent order); the following verbs contain subject agreement and are inflected for narrative tense (Jita) and sequential tense (Fuliiru):

- (9) Jita ('Hare and Hyena' text, lines 8-10, Allison Pyle & Holly Robinson, p.c.)
 Wamembe nayomba, wamembe nasurumbara, wamembe najira obhuramusi,
 hyena he.spoke hyena he.lamented hyena he.got decision
 bhwokuja-otema obhurembo, natura mumugundu gwaye.
 he.went-cut sap/birdlime he.put in.field his
 'Hyena spoke, hyena lamented, hyena made up his mind, he went and cut
 birdlime (tree sap), (and) he put (it) in his field.' (Free translation by Steve Nicolle)
- (10) Fuliiru (Van Otterloo 2011: 541)
 Wandare anayuwwa kwâkola mulirira umwana. Wandare anabwîra
 lion he.heard(SEQ) that it.is.crying.for child lion he.told(SEQ)
 uyo mushaaja...
 that old_man
 'Lion heard that it [the cow] is crying for (its) child. Lion told that old man..'

In Digo, Mwani and Rangî, topics may be right dislocated under certain conditions. It should be noted that right dislocated topics are distinct from post-verbal subjects which may occur inthetic sentences (see §5 below). Right dislocated topics refer to specific, identifiable participants, are often marked as such (for example, they are modified by demonstratives or are proper names), and are separated from the verb by a pause and sometimes by non-core elements. Like left dislocated constituents, right dislocated constituents are outside of the clause nucleus. In contrast, post-verbal subjects are grammatical subjects in their own right and are never separated from the verb by a pause or by non-core elements.

In Digo and Mwani, continued topics are right dislocated when there is textual continuity, that is, when events are presented in sequence within a single thematic unit (i.e.

a paragraph). In the following example, the continued topic *mutu yuyu* ‘this person’ is right dislocated. This is possible because there is no change of topic and no discontinuity; the use of the consecutive tense marker *chi-* in *achinyamala* ‘he stayed silent’ indicates a sequential action within a single paragraph:

- (11) Digo (Nicolle 2015b: 27–28)
 Achidziuza mwakpwe rohoni, “Pho munda, nkauhenda mkpwulu na sikaphaha hata tsere mwenga, kpwani nini?” Lakini [achinyamala]_{COMMENT} [mutu yuyu]_{TOPIC} wala kayagomba na mutu.
 ‘He asked himself [CONS] in his heart, “That field, I have made it big and I have not got even a single maize cob, but why?” But he stayed silent [CONS] **this man**, neither did he speak with anyone.’

In (12), from Mwani, the conceptual topic does not change but is referred to using a noun phrase and proper names in the last clause. There is no discontinuity in the text as the order of events is iconically represented: they were called and then they came, and so the topic is right dislocated.

- (12) Mwani (Floor 2005: 10)
 Wakati waifkire sumana yawasikizane, wakitiwa,
 time when.it.arrived week that.they.agree they.were.called
 [wakija]_{COMMENT} [wó-wawiri,]_{TOPIC} Anli na Ntendaji.
 they.came those-two NAME and NAME
 ‘When the week that they agreed upon arrived, they were called, they came those two, Anli and Ntendaji.’

In addition, in Mwani a topic may also be right dislocated if it is a renewed topic, that is, an element that has previously functioned as a topic but not in the immediately preceding clause (Floor 2005: 10–11). (It is not clear whether this is always the case or is an option.) Example (13) shows the difference between a right dislocated renewed topic, *muka ire* ‘that woman’, which was last mentioned two clauses earlier, and a left dislocated switch topic, *vinu vire* ‘those things’, which has not previously functioned as a topic.

- (13) Mwani (Floor 2005: 10–11)
 Sambu [akikála]_{COMMENT} [muka ire]_{TOPIC} na [vinu vire]_{TOPIC} [akipika
 now she.sat woman that and things those she.cooked
 akiwápa wanu.]_{COMMENT}
 she.gave.to.them people
 ‘Now she sat down that woman and those things she cooked (them) and gave to the people.’

In Rangi, a switch topic may be right dislocated if it is only temporary; that is, if it functions as the topic of a single clause but the previous conceptual topic is resumed

immediately after. This is optional, however, as not all temporary topics are right dislocated. In (14), the subject twice changes from the elder to the boys and immediately back to the elder. The first clause involving the boys has SV order but the second has VS order. The boys are a temporary topic whereas the elder is more permanent: he is introduced formally in the first clause and it is he who speaks at the end of this passage; these are typical features of major participants (Dooley & Levinsohn 2001: 119).

- (14) Rangi (Stegen 2011: 532–533)
 Aho kali kwijaa na **moosi umwi**, afuma iundii, maa **vatavana** navo viyokuuja,
 maa akavasea, “Mpokeri isiri raani.” [Maa vakasiita]_{COMMENT} [vara
vatavana.]_{TOPIC} Maa [**ura moosi**]_{TOPIC} [akasea...]_{COMMENT}
 ‘In times of old there was **one elder**, he came from the field, and **the boys** and
 they, they are coming, and he told them, “Carry my hoe.” And they refused,
those boys. And **that elder** said...’

In other languages, temporary topics are left dislocated. In the Bena example in (15), *inyama* ‘meat’ is a left dislocated temporary topic; *u-Mbwa* ‘Dog’ occurs after the verb because it receives argument focus (see §4).

- (15) Bena (Eaton 2015a: 34)
 UDUuma aaheliye kwa mwipwawe kuhungila kivembo. [Inyama]_{TOPIC}
 Leopard he.went to uncle to.greet misfortune meat
 [aalekile,]_{COMMENT} [iloleela]_{PRESUPPOSITION} [uMbwa.]_{FOCUS}
 he.left he.looks.at Dog
 ‘Leopard went to console his uncle for his bereavement. He left the meat, Dog
 was looking after it.’

Given the available data, it seems that the majority of the languages surveyed pattern like Jita and Fuliiru in that all topics are obligatorily left dislocated. However, it is possible that evidence may emerge of specific discourse contexts that trigger right dislocation of topics in other languages.

4 The expression of subject argument focus

Argument (or *term*) focus arises when non-predictable information is expressed by a noun phrase. It is found in declarative sentences when a certain event or situation is presupposed, but the speaker assumes that the addressee does not know the identity of one of the participants in that event or situation. In (15) above, it is presupposed that Leopard will not leave his meat unattended and the post-verbal subject *u-Mbwa* ‘Dog’ identifies who has been left to guard it.

Argument focus in Bantu languages may be expressed in-situ, immediately after the verb (IAV), or in clause-final position (Yoneda 2011: 761–762; Gibson et al. To appear). In most eastern Bantu languages surveyed it appears that argument focus is associated with

clause-final constituents, an exception being Makonde which appears to use the IAV position, in line with most other Bantu languages exhibiting a conjoint/disjoint distinction (Gibson et al. To appear). There is evidence that elements in this position are right dislocated in at least some cases, as the occurrence of the adverbial *woori* ‘now’ between the presupposition and focus elements in (16) suggests. The context for this example is that the animals have dug a well, but Hare is coming at night and stealing the water. The animals plan to post a guard by the well, and this constitutes the presupposition; the focus identifies who was chosen.

- (16) Jita (Pyle & Robinson 2015: 18)
 [Mbamuta-ko]_{PRESUPPOSITION} woori [nyawatare.]_{FOCUS}
 they.put.him-there now lion
 ‘Now they put lion there.’

Makonde does not allow both an object and subject to occur after the verb, and so when there is a post-verbal subject the object is left dislocated (Leach 2015: 91). In Example (17) below, *shakulya* ‘staple food’ is always served with *imbogwa* ‘sauce’ and so it is assumed that someone will provide this; the post-verbal subject identifies this person as the speaker.

- (17) Makonde (Leach 2015: 92)
 Paukile ndawika kukaja kumwaulila ndyagwe do: “Ndyangu, taleka
 when he.went and.arrived he.told.her his.wife QUOT my.wife cook
 shakulya, [imbogwa namanya]_{PRESUPPOSITION} [nimwene.]_{FOCUS}”
 staple.food sauce I.will.know myself
 ‘When he got home he told his wife, “Get some food ready for me, wife – but as for the meat sauce, I’ll deal with that.”’

Argument focus can also be expressed through cleft constructions in all languages surveyed. In such constructions, the focused element and the presupposition – often in the form of a relative clause or verbless predicate – are connected using a copula or focus marker. The orders focus>presupposition and presupposition>focus are both found:

- (18) Kabwa (Walker 2011: 25)
 Kumbe [omukari wunu]_{FOCUS} ng’we [yankorera eng’ana yinu.]_{PRESUPPOSITION}
 EXCL woman this COP she.did.to.me thing this
 ‘Gosh, it was this woman who did this thing to me.’
- (19) Suba-Simbiti (Masatu 2015: 28)
 [Omoremo ghono yaamanyirë]_{PRESUPPOSITION} [no-bhötëghi ubhwa sinswe.]_{FOCUS}
 work REL he.knew FOC-trapping of fish
 ‘The work which he knew is fishing.’

All languages surveyed use both clause-final (or IAV) position and cleft constructions for non-subject argument focus, but there is cross-linguistic variation with subject focus

marking. The fact that subject focus behaves differently from other kinds of argument focus is not surprising. What Güldemann, Zerbian & Zimmermann (2015: 159) refer to as the “default topic-hood of subjects/agents” probably underlies the fact that focus marking on subjects sometimes differs from focus marking on other constituents, for example by requiring explicit focus marking even when a non-subject focus element is unmarked (ibid. 170; see also Fiedler et al. 2010 for a discussion of this phenomenon in Gur, Kwa and West Chadic languages).

In all the languages surveyed, argument focus on the subject can be indicated by cleft constructions. Argument focus on the subject can also be indicated by clause-final position in Jita, Kabwa, Kwaya, Suba-Simbiti, Bena, Malila and Rangì, and by IAV position in Makonde. The Jita example in (20) illustrates both strategies.¹¹ The post-verbal subject in the first clause identifies ‘women only’ as those who were living in the land (that the land was inhabited by someone is a presupposition); the cleft construction (with a copula clitic) in the second clause identifies these women as the ones who had stolen Mariro’s cows, an event of which the audience is already aware.

- (20) Jita (Pyle & Robinson 2015: 34)
 Echaaro echo [:bhaariga bheekaye-mo]_{PRESUPPOSITION} [abhagasi era,]_{FOCUS}
 land that they.were living-there women only
 [ni=bho]_{FOCUS} [:bhaariga bheebire jing’a ja Mariro.]_{PRESUPPOSITION}
 COP=3PL they.were they.had.stolen cows of Mariro
 ‘In that land were living women only, it was they who had stolen Mariro’s cows.’

Digo and Fuliiru do not allow post-verbal subjects to receive argument focus. In both languages, subject argument can only be expressed using cleft constructions. In Fuliiru, the cleft construction is formed with a ‘focus copula’ which is cliticized to the following verb:

- (21) Fuliiru (Van Otterloo 2015: 345)
 [Yábá bágéni]_{FOCUS} [bó=bàgírà yibi.]_{PRESUPPOSITION}
 these guests FOC=they.do these
 ‘These guests, they (are the ones who) did these things.’

In Digo, the cleft construction consists of the copula prefix *ndi* (*si* in the negative) plus a referential marker, and is typically, although not always, followed by a relative clause:

- (22) Digo (Nicolle 2015b: 55)
 Ndipho atu achimanya kukala [iye]_{FOCUS} [ndi=ye ariyehenda mambo
 then people they.knew that she COP=1.REF who.did things
 higo.]_{PRESUPPOSITION}
 those
 ‘Then people knew that it was her who did those things.’

¹¹ The precise differences in interpretation between different focus positions could not be determined on the basis of narrative corpus data alone.

Finally, in Mwani, argument focus is post-verbal in response to a content question, but when there is no prior question a cleft construction is used (Floor p.c. 8 April 2014; see also Floor 2005: 9):

(23) Mwani (Floor p.c.)

- a. Kitabu atwarire nani? [Katwala]_{PRESUPPOSITION} [Saidi.]_{FOCUS}
 book 3SG.REL.PST.take who(SUBJECT) 3SG.PST.take Saidi

‘Who took the book? Saidi took it.’

- a. [Atwarire]_{PRESUPPOSITION} ndi [Saidi.]_{FOCUS}
 3SG.REL.PST.take COP Saidi

‘It was Saidi who took it.’

5 VS and SVthetic sentences

As we have seen, the pre-verbal domain in eastern Bantu languages is restricted to topics and non-focus subjects, and so a sentence with canonical SV constituent order will normally be interpreted as expressing topic-comment sentence articulation. It is therefore not surprising thatthetic sentences, in which there is no topic, exhibit VS constituent order. These post-verbal subjects are not right dislocated; post-verbal subjects inthetic sentences are grammatical subjects in their own right, they are not topics (and so are generally non-specific or non-established), and they are never separated from the verb by a pause or by non-core elements. The following examples illustrate this.

(24) Kwaya (Odom 2015: 29)

Woori bhunu :aariga acheeganiirisha mmbe ^n-aa-j-a waarukerwe.
 now while he.was he.still.be_thinking so he.came frog

‘Now while he was still continuing to think, came a frog.’

(25) Digo (‘Mhegi wa Mihambo’ text, line 25b)

ratuluka fisi, rina chitswa dza cha mutu...
 it.emerged hyena it.has head of like person

‘...a hyena emerged, it had a head like a person’s...’

A common function ofthetic sentences is to introduce participants into a narrative. When participants are introduced using a verb of arrival (‘come’, ‘emerge’, ‘appear’, etc.), agreement is with the agent noun phrase in all languages (see the examples above) except Fuliiru. However, when participants are introduced using existential verbs, the verb agrees with the noun phrase in Makonde,¹² Bena, Malila, Jita, Kabwa, Kwaya, Suba-Simbiti and Ekoti, but with a locative noun class in Fuliiru, Digo and Rangi (see Nicolle 2015a: 17–20).

¹² van der Wal (2008) reports that languages which distinguish conjoint and disjoint verb forms – which includes Makonde – differ concerning which form is used inthetic sentences; for example, Sesotho [S32/33] uses a conjoint verb form, whereas Makhuwa [P31] uses a disjoint form. Makonde patterns like Makhuwa.

Inthetic sentences, only VS constituent order is found in the Digo, Jita and Kabwa text corpora. However, a few texts in the other languages surveyed begin with SV clauses where the subject is not a topic. Most of the subjects in these SV clauses are either well-known folk-tale or animal characters – as in (26) below from Bena – or they refer to non-specific participants, such as ‘children’ in the Kwaya example (27) and ‘one man’ in the Fuliiru example (28). This suggests that at the start of a narrative, where no established referents are available to be topics, SV constituent order may be used inthetic sentences when the subject is known to the audience or is (currently) non-specific.

- (26) Bena (Eaton 2015a: 54)
 Pa vutalilo [u-Mbwa nu Duuma]_S [vaali]_V nu wunyalumwinga.
 at start dog and leopard they.were with unity
 ‘In the beginning Dog and Leopard were together.’
- (27) Kwaya (Odom 2015: 60)
 Rusuku rumwi [abhaana]_S [mbaja]_V okureebha.
 day one children they.came to.herd
 ‘One day children went to herd.’
- (28) Fuliiru (Van Otterloo 2015: 104)
 [Mushoshi muguma]_S [akagira]_V lusiku likulu ha’mwage.
 man one he.held day great at.home
 ‘One man had a feast at his house.’

Although VSthetic sentences are the norm in most of the languages surveyed, with SVthetic sentences occasionally at the start of narratives, this is not the case in the Suba-Simbiti text corpus. Only one Suba-Simbiti text starts with a VSthetic sentence; of the other texts, three start with SVOthetic sentences, one with a VOthetic sentence, and one with a subject in a copula construction followed by a past form of the verb *rë* ‘be’:

- (29) Suba-Simbiti (Masatu 2015: 16)
 [Musimbëtë na Mohaasha]_S m=bhaana abha enda ëmwë [bha-a-rë]_V
 Musimbiti and Mohaasha COP=2.children 2.Ass 9.stomach 9.one 3PL-PST-be
 ‘Msimbiti and Mohaasha were siblings, they were.’

Unusually for an eastern Bantu language, Suba-Simbiti also allows SVthetic sentences after the start of a narrative, as in (30). No buffalo has been mentioned previously in the text, and so a topic-comment reading is ruled out. (Two topic-comment sentences follow, in which the buffalo and then the youth function as topics.)

- (30) Suba-Simbiti (Masatu 2015: 44)
 Bhoono hano yaarëesyanga urusikö urwöndë, [eng’era]_S [ekaasha]_V
 now when he.was.herding day another buffalo it.came
 mu-rihisho irya waabho riyo.
 in-group of their.place that
 ‘Another day when he was herding, a buffalo came among their herd.’

6 Conclusions

Based on the available narrative texts, a number of generalizations can be made concerning information structure in the eastern Bantu languages surveyed.

- Obligatory and optional left dislocation of topics

All languages surveyed have left dislocated (pre-verbal) topics, and topic-comment sentence articulation is very common. When there is textual discontinuity (switch topics, episode or paragraph breaks, non-iconic order of events, etc.), topics in all languages are left dislocated. Moreover, all topics are left dislocated in Jita and Fuliiru, and probably in most other eastern Bantu languages. In Jita and Fuliiru, left dislocation of topics is never overridden by other discourse factors. However, certain types of topic are right dislocated in a few languages. Continued topics are right dislocated in Digo and Mwani when there is textual continuity (i.e. within a paragraph); renewed topics may be right dislocated in Mwani; and temporary topics are optionally right dislocated in Rangi. In Digo, Mwani and Rangi, therefore, the default left dislocation of topics can be overridden by discourse factors.

- The expression of subject argument focus

Clause-final or IAV constituents other than the subject may express argument focus in all languages. In Jita, Kabwa, Kwaya, Suba-Simbiti, Bena, Malila, Makonde and Rangi, argument focus on the subject can be expressed both through clause-final or IAV position and through cleft constructions. However, in Fuliiru and Digo, argument focus on the subject can only be expressed using a cleft construction; this appears to be a grammatical constraint and is not tied to specific discourse contexts. In Mwani, argument focus involves right dislocation in response to a content question but a cleft construction is used where there is no prior question.

- VS and SVthetic sentences

Post-verbal subjects are the norm inthetic sentences and are the only possibility in Digo, Jita and Kabwa. However, SVthetic sentences occur in Fuliiru, Kwaya, Suba-Simbiti, Rangi, Bena, Malila, and Makonde for the presentation of new participants at the start of narratives. The Suba-Simbiti corpus also includes SVthetic sentences elsewhere. This suggests that in most eastern Bantu languages surveyed, the default VS constituent inthetic sentences can be overridden by discourse factors.

Abbreviations

1PL	1 st person plural	OM	object marker
3PL	3 rd person plural	PRES	present tense
3SG	3 rd person singular	QUOT	quotative marker
COP	copula	REF	referential marker
EXCL	exclamation	REL	relative marker
FOC	focus marker	SEQ	sequential tense
FV	final vowel		

References

- Broomhall, Elizabeth. 2011. Bena text linguistics. Africa International University. Unpublished ms.
- Buell, Leston C., Kristina Riedel & Jenneke van der Wal. 2011. What the Bantu languages can tell us about word order and movement. *Lingua* 121. 689–701.
- Cohen, Ariel. 2009. Fronted quantificational adverbs. In Benjamin Shaer, Philippa Cook, Werner Frey & Claudia Maienborn (eds.), *Dislocated elements in discourse: Syntactic, semantic, and pragmatic perspectives*, 312–327. New York & London: Routledge.
- De Cat, Cécile. 2007. *French dislocation: Interpretation, syntax, acquisition*. Oxford: Oxford University Press.
- Dik, Simon C. 1981. *Functional grammar*. 3rd edn. Dordrecht: Foris.
- Dik, Simon C. 1989. *The theory of functional grammar, part 1: The structure of the clause*. Dordrecht: Foris.
- Dooley, Robert A. & Stephen H. Levinsohn. 2001. *Analyzing discourse: A manual of basic concepts*. Dallas, TX: SIL International.
- Downing, Laura J. & Larry M. Hyman. 2016. Information structure in Bantu. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 790–813. Oxford: Oxford University Press.
- Eaton, Helen. 2015a. *Bena narrative discourse* (SIL Language and Culture Documentation and Description 25). Dallas, TX: SIL International.
- Eaton, Helen. 2015b. *Malila narrative discourse* (SIL Language and Culture Documentation and Description 27). Dallas, TX: SIL International.
- Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.
- Fiedler, Ines, Katharina Hartmann, Brigitte Reineke, Anne Schwarz & Malte Zimmermann. 2010. Subject focus in West African languages. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 234–257. Oxford: Oxford University Press.
- Floor, Sebastian J. 2005. Marked word-order and focus in KiMwani. Unpublished ms.
- Gibson, Hannah, Andriana Koumbarou, Lutz Marten & Jenneke van der Wal. To appear. Locating the Bantu conjoint/disjoint alternation in a typology of focus marking. In Jen-

- neke van der Wal & Larry M. Hyman (eds.), *The conjoint/disjoint alternation in Bantu*. Berlin: De Gruyter.
- Good, Jeff. 2010. Topic and focus fields in Naki. In Ines Fiedler & Anne Schwarz (eds.), *The expression of information structure: A documentation of its diversity across Africa*, 35–67. Amsterdam/Philadelphia: John Benjamins.
- Güldemann, Tom, Sabine Zerbian & Malte Zimmermann. 2015. Variation in information structure with special reference to Africa. *Annual Review of Linguistics* 1. 155–178.
- Gundel, Jeanette K. & Thorstein Fretheim. 2004. Topic and focus. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 175–196. Oxford: Blackwell.
- Guthrie, Malcolm. 1967–71. *Comparative Bantu. An introduction to the comparative linguistics and prehistory of the Bantu languages*. (4 vols.) Farnborough: Gregg International.
- Halliday, M. A. K. 1967. Notes on transitivity and theme in English II. *Journal of Linguistics* 3. 199–244.
- Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2000. When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 24(3). 611–682.
- Leach, Rhoda Martyn. 2015. *Makonde narrative discourse* (SIL Language and Culture Documentation and Description 29). Dallas, TX: SIL International.
- Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2015. *Ethnologue: Languages of the world*. 18th edn. Dallas, TX: SIL International. <http://www.ethnologue.com>.
- Maho, Jouni. 2003. A classification of the Bantu languages: An update of Guthrie's referential system. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 639–651. London & New York: Routledge.
- Marten, Lutz, Nancy C. Kula & Nhlanhla Thwala. 2007. Parameters of morphosyntactic variation in Bantu. *Transactions of the Philological Society* 105. 253–338.
- Masatu, Amanda. 2015. *Suba-Simbiti narrative discourse* (SIL Language and Culture Documentation and Description 31). Dallas, TX: SIL International.
- Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49. 127–163.
- Mwamzandi, Mohamed Y. 2014. *Swahili word order choices: Insights from information structure*. Doctoral dissertation. Arlington: University of Texas at Arlington.
- Nicolle, Steve. 2013. *A grammar of Digo (a Bantu language of Kenya and Tanzania)*. Dallas, TX: SIL International.
- Nicolle, Steve. 2015a. *A comparative study of eastern Bantu narrative texts* (SIL Electronic Working Papers 2015–003). Dallas, TX: SIL International.
- Nicolle, Steve. 2015b. *Digo narrative discourse* (SIL Language and Culture Documentation and Description 26). Dallas, TX: SIL International.
- Odom, Shannon Ronit. 2015. *Kwaya narrative discourse* (SIL Language and Culture Documentation and Description 30). Dallas, TX: SIL International.

- Pyle, Allison & Holly Robinson. 2015. *Jita narrative discourse* (SIL Language and Culture Documentation and Description 28). Dallas, TX: SIL International.
- Schwarz, Florian. 2003. Focus marking in Kikuyu. *ZAS Papers in Linguistics* 30. 41–118.
- Schwarz, Florian. 2007. Ex-situ focus in Kikuyu. In Enoch O. Aboh, Katharina Hartmann & Malte Zimmermann (eds.), *Focus strategies in African languages: The interaction of focus and grammar in Niger-Congo and Afro-Asiatic*, 139–160. Berlin: Mouton de Gruyter.
- Shaer, Benjamin, Philippa Cook, Werner Frey & Claudia Maienborn. 2009. Introduction. In Benjamin Shaer, Philippa Cook, Werner Frey & Claudia Maienborn (eds.), *Dislocated elements in discourse: Syntactic, semantic, and pragmatic perspectives*, 1–27. New York & London: Routledge.
- Stegen, Oliver. 2011. *In quest of a vernacular writing style for the Rangi of Tanzania: Assumptions, processes, challenges*. Edinburgh: University of Edinburgh. (Doctoral dissertation.) <http://www.sil.org/resources/archives/42655>.
- Vallduví, Enric. 1992. *The informational component*. New York: Garland.
- van der Wal, Jenneke. 2008. Agreement inthetic VS sentences in Bantu and Romance. In Cécile De Cat & Katherine Demuth (eds.), *The Bantu-Romance connection. A comparative investigation of verbal agreement, DPs and information structure*, 323–350. Amsterdam/Philadelphia: John Benjamins.
- van der Wal, Jenneke. 2009. *Word order and information structure in Makhuwa-Enahara*. Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap) Publications.
- van der Wal, Jenneke. 2014. Morphological marking of term focus. (Paper presented at Information Structure in Head-Marking Languages workshop, Max Planck Institute Nijmegen, 28-29 March 2014.)
- van der Wal, Jenneke & Theresa Biberauer. 2014. From macroparameters to nanoparameters: A comparative Bantu case study. (Paper presented at ComSyn meetings, Leiden, 5 February 2014.)
- Van Otterloo, Roger. 2011. *The Kifuliiru language, vol. 2: A descriptive grammar*. Dallas, TX: SIL International.
- Van Otterloo, Roger. 2015. *Fuliiru narrative discourse* (SIL Language and Culture Documentation and Description 32). Dallas, TX: SIL International.
- Walker, John R. 2011. Discourse features of Kabwa and accompanying texts. Unpublished ms.
- Yoneda, Nobuko. 2011. Word order in Matengo (N13): Topicality and informational roles. *Lingua* 121. 754–771.
- Zerbian, Sabine. 2006. *Expression of information structure in Northern Sotho*. Doctoral dissertation. Berlin: Humboldt University.