Bantu word order between discourse and syntactic relations

Allen Asiimwe (Makerere University), Patrick Kanampiu (University of Edinburgh), Elisabeth Kerr (LUCL), Zhen Li (LUCL), Amani Lusekelo (University of Dar es Salaam), Ernest Nzhemezimana (University of Burundi) & Jenneke van der Wal (LUCL)[1]

This Time for Africa, Leiden University, 1st April 2021

1. Introduction

Background of the research questions:

- Discourse function as a strong factor in conditioning Bantu word order (Downing & Hyman 2015, Downing & Marten 2019)
- Examples of discourse function in determining the word order of Bantu languages: locative inversion (1), and dedicated focus positions – immediately before the verb (IBV) in (2):

(1) (What has happened on the bridge?)

A-ha-ru-tindó	ha-a-rabá=hó	e-mótóka	ny-î̂ngi.
AUG-16-11-bridge	16SM-N.PST-pass=16	AUG-9.car	9-many

'On the bridge have passed many cars.' [Rukiga]

(2) (Who attacked the hunter?)

Mu-bhii	ki-mbuli	ki-siim-i.
1-hunter	7-lion	7SM-attack-PST

'[The lion]$_{FOC}$ attacked the hunter.' [Teke-Kukuya]

- Bantu word order has been argued to be best captured by reference to **discourse roles**, e.g. Topic-Verb-Nontopic (see e.g. Good 2010 for Naki, Yoneda 2011 for Matengo, and Morimoto 2000, 2006 for Bantu discourse configurationality in general).
- …But traditional **grammatical roles** of 'subject 'and 'object 'continue to shape descriptions and analyses; "The default order of sentence constituents across Bantu is S (Aux) VO (Adjuncts)" (Nurse & Philippson, 2003:9)

Debate in the literature:

- Larger debate about word order and types of languages:
  - Li & Thompson (1976): Topic-prominent vs. Subject-prominent languages;
  - Hale (1983): languages can be non-configurational;
  - Mithun (1987): pragmatically based word order, later termed 'discourse configurational' (see É.Kiss 1995);
- Over the years the debate has shifted from 'Which type does language L belong to?' to '**To what extent** is word order in language L determined by discourse roles and **to what extent** by grammatical roles?':
  - Öhl (2010: 251): discourse-configurationality vs. 'relation-configurationality';

[1] The main analysis in this paper was developed by Allen, Elisabeth, Patrick, Zhen, and Jenneke; all authors supplied the data on which the generalisations are built.

- Morimoto (2006): many Bantu languages are in a **transitional stage** between topic-based and subject-based agreement systems, being neither fully topic-based nor subject-based;

Purpose of the talk:

- We want to determine where the Bantu languages are on this **continuum** between discourse roles and grammatical roles in their effect on word order, taking seriously the significance of **information structure**;
- We switch to a discourse-configurational approach, asking **How far can we get in describing Bantu word order without reference to syntactic roles?** This question helps us pinpoint precisely where we *do* need syntactic relations;
- The results show Bantu-internal **microvariation**, with some languages heavily discourse-based and others more affected by grammatical roles, arguing against a one-size-fits-all account of word order.

2. Methodology

- Detailed data on information structure is often not available in existing descriptions.
- The BaSIS project therefore gathered **new fieldwork data** focusing on syntax and information structure, using our own project methodology[2] alongside the Questionnaire on Information Structure (Skopeteas et al., 2006) and the elicitation of natural speech.
- This talk presents results from our studies on 8 of the languages in the project: **Tunen** (A44, Cameroon), **Teke-Kukuya** (B77, Republic of Congo), **Kîîtharaka** (E54, Kenya), **Kirundi** (JD62, Burundi), **Rukiga** (JE14, Uganda), **Kinyakyusa** (M31, Tanzania), **Makhuwa** (P31, Mozambique), and **Copi** (S61, Mozambique).[3] Geographical distribution of these languages is shown in the map in Figure 1:

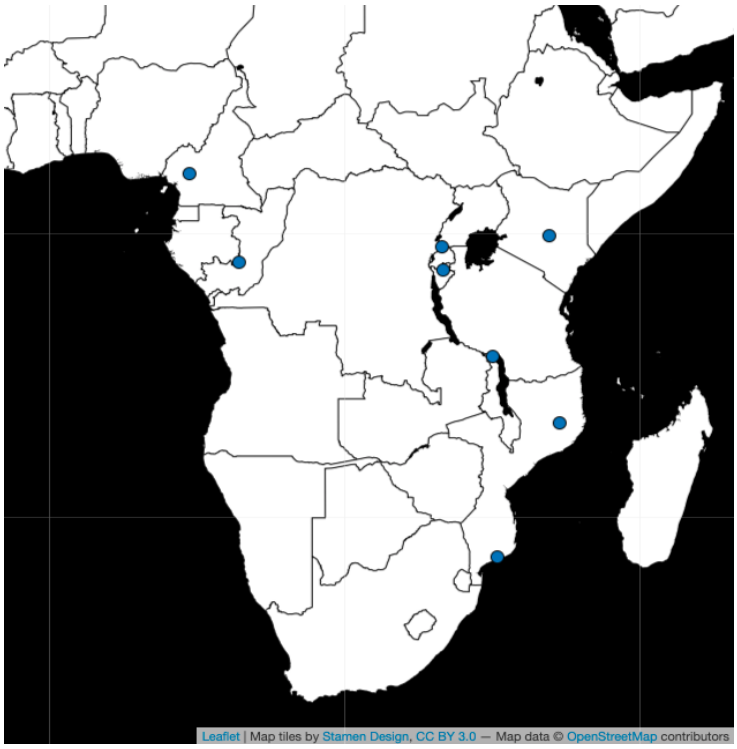

Figure 1. Map of languages in the BaSIS sample (co-ordinates from Glottolog).

[2] Available for download at https://bantusyntaxinformationstructure.com/methodology/.

[3] Guthrie classifications are given alongside language names and locations, from Maho (2009).

- We checked three factors of discourse-configurationality in word order, with subquestions that serve as checkpoints to set the parameter.
- For focus, we considered three diagnostics: wh-words (which are inherently focused), answers to wh-questions (simple/information focus), and modification by the focus-sensitive particle 'only' (exhaustive focus)

BaSIS word order parameters:

1. **Is there a dedicated focus position?**
   Checkpoints:
   a. Can the recipient and the theme be questioned in their canonical position?
   b. Can the recipient and the theme be an answer to a wh-question in their canonical positions?
   c. Can the recipient and the theme be modified by 'only' in their canonical positions?
   d. Can adverbs be questioned in their canonical position?

2. **Is the preverbal domain reserved for topics?**
   Checkpoints (for otherwise unmarked clauses, i.e. no cleft):
   a. Can preverbal arguments be questioned?
   b. Can preverbal arguments be the answer to a wh-question?
   c. Can preverbal arguments be modified by 'only'?
   d. Can thetic sentences be expressed as SV(O)? (where S is not an always-available referent)?
   e. Are topical arguments typically or preferrably expressed preverbally?
3. **Is there symmetry between grammatical roles relating to word order?**
   Checkpoints:
   a. Can the subject be questioned in the same position as the object?
   b. Can the subject as the answer to a subject wh-question occupy the same position as the object in the answer to an object wh-question?
   c. Can the subject be modified by 'only' in the same position(s) as the object?

Results:

- The colour coding indicates whether the answer provides evidence for the influence of grammatical roles (yellow) or discourse roles (blue) on word order;
- A column is added to show what the parameters would be for a language that entirely relies on grammatical roles (GR) or discourse roles (DR).

Parameter	Checkpoint	DR	GR		Tunen	Kukuya	Tharaka	Kirundi	Rukiga	Kinyakyusa	Makhuwa	Copi
1	a	N	Y		-	N	Y	N	N	Y	N	Y
	b	N	Y		-	N	Y	N	Y	Y	N	Y
	c	N	Y		Y/-	Y	Y	N	Y	Y	N	Y
	d	N	Y		Y	Y/N	Y	Y/N	Y/N	Y	N	Y
2	a	N	Y		Y	Y*	N	N	N	Y?	N	N
	b	N	Y		Y	Y*	N	N	N	Y?	N	N
	c	N	Y		Y	Y*	N	N	N	Y	N	Y/N?
	d	N	Y		Y	Y	Y	Y	Y	N	N	-
	e	Y	N		Y	Y	Y	Y	Y	Y	Y	Y
3	a	Y	N		N	Y	Y	Y	Y	Y	N	Y
	b	Y	N		N	Y	Y	Y	Y	Y	N	Y
	c	Y	N		N	Y	Y	Y	Y	Y	N	Y

Table 1. Parameter settings for word order in the BaSIS languages.

Key: blue shading = evidence for discourse roles (DR) conditioning word order, yellow shading = evidence for grammatical role (GR) conditioning word order (a column for DR and GR is giving to make clear what the parameter settings would be for a purely DR-conditioned or a purely GR-conditioned language).

*Yes, but only in IBV position, and therefore still evidence for IS-conditioned word order.

3. Observations

3.1. Cross-linguistic variation

- There is a lot of variation between the languages in the sample → cross-Bantu variation.
- Better to stop making claims about word order in 'the Bantu languages' as if they were a monolithic entity; further research on more languages is needed in order to make reliable generalisations.

3.2. Universal tendency for preverbal topics

- Answers to parameter 2e were all Y - all languages show a preference for topics in the preverbal domain;

- supports Gundel (1988)'s generalization "*Every language has syntactic topic constructions in which an expression which refers to the topic of the sentence is adjoined to the left of a full sentence comment.*"

3.3. Variation in focus positions

- Three languages with a dedicated focus position, but all are different: **IBV** in Kukuya, **IAV** in Makhuwa, and **sentence-final** in Kirundi;
  - Kukuya's IBV focus position also means that parameter 2 is interpreted differently:
    - The IBV focus position is marked, based on that SVO languages are more likely to have postverbal focus positions (Kidwai, 1999; Horvath 1986);
    - The existence of IBV focus positions in the family may mean that parameter 2 needs to be tweaked to exclude focus positions.

3.4. Variation within parameters

- The answers to the checkpoints are the same in all languages for parameter 3, but not for parameters 1 and 2.
- Some interesting exceptions for parameters 1 and 2 can be understood in the following way:
  - 1a for Rukiga shows that a wh needs to be in the IAV position whereas other focus elements may appear in other positions, see (3):

(3) a. Waaha kí Jéini?
Wa-a-ha **ki** Jeini?
2SG.SBJ-N.PST-give-FV what 1.Jane
'What have you given Jane?'

b. Naaha Jéín' énkofiira.
Na-a-h-a Jeini **e-n-kofiira**
1SG.SBJ-N.PST-give-FV 1.Jane AUG-9-hat
'I have given a hat to Jane.' [Rukiga]

  - 1d for Kukuya, Kirundi, and Rukiga shows that not all adverbs behave the same way. For example, the adverb 'when' in Rukiga is **not** restricted to the IAV position (4) but 'how' is:

(4) a. Okaza Kampala ryari?
o-ka-za Kampala **ryari**
2SG.SM-F.PST-go 23.Kampala when
'When did you go to Kampala?'

b. Okaza ryarí Kampála?
O-ka-z-a **ryari** Kampala
2SG.SM-F.PST-go when 23.Kampala
'When did you go to Kampala?' [Rukiga]

(5) a. Obutúmwa bukahika búta purésidenti?
O-bu-tumwa bu-ka-hik-a **bu-ta** puresidenti?
AUG-4-message 14-F.PST-reach-FV 14-how 1.president
'How did the message reach the President?'

b. *Obutúmwa bukahika purésidenti **búta**?

c. Purésidenti óbutúmwa bukamuhika búta?
Puresidenti o-bu-tumwa bu-ka-mu-hik-a **bu-ta**
1.Puresidenti AUG-4-message 14-F.PST-1OM-reach-FV 14-how
'As for the president, how did the message reach him?' [Rukiga]

- 2d for Kîîtharaka, Kirundi and Rukiga shows that non-topical transitive subjects are allowed preverbally, while focused arguments are banned—the preverbal domain in these languages is not restricted to topics, but can be characterised as 'non-focal'.

- Kinyakyusa is exceptional for an eastern Bantu language in allowing preverbal focus, which appears to be an **areal innovation**, see Figure 2:

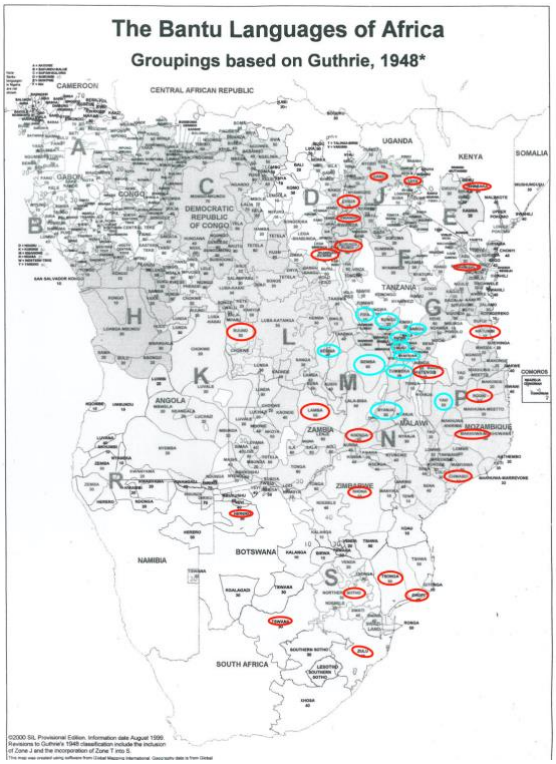

Figure 2 – Preverbal focus: blue = allowed, red = ban

3.5. Tunen as grammatical role-oriented

- Tunen is exceptional in showing **hardly any influence of information structure on word order**—only parameter 2e shows evidence for the influence of discourse role, all other parameter values showing evidence for grammatical role:[4]

[4] Parameter 1a/1b/1c are not filled in due to insufficient data; COVID-19 has delayed follow-up research here.

- The results for parameter 3 indicate that Tunen has word order asymmetries conditioned by grammatical role. For example, when using a cleft to express focus, the focus marker/copula *á* must **precede** focussed subjects, but it **follows** non-subjects.

(6) Context: Which politician died?
***(á)** Píɛ̀lə́ ánáwə̀
/á Piɛlə á-na-wə/
FOC Pierre FOC -die
'Pierre died.', 'It was Pierre who died.'

(7) Context: "What did Joseph cook?
bɛ́làmà ***(á)** Jósɛ̀pɛ̀ áná tálɛ̀àk
/bɛ-lama á Josɛpɛ a-na talɛa-aka/
8-vegetable FOC Joseph REL.SM.1-PAST2 cook-DUR
`Joseph cooked vegetables'

- This subject/non-subject asymmetry shows that **grammatical role has a larger effect on the grammar of Tunen** than discourse roles.
- This exceptional lack of influence of information structure could be explained due to the position of Tunen at the **Northwest of the Bantu family**; the **innovative nature** of Tunen's word order (Mous 1997) could explain why it patterns differently from the other languages.
- The **lack of inversion constructions** in Tunen may be an **areal pattern** for languages of the North-Western region (see Hamlaoui 2018, Hamloui & Makasso 2015).

4. Discussion

4.1. Working with parameters

- Working with parameters to compare languages has methodological pros and cons:
  - **Pros**: it aids comparison between languages in abstracting from the raw data and giving an overview of the different patterns that can be used to discover new generalisations (Baker 2010)[5]; the parameters can also guide future fieldworkers in investigating the influence of discourse roles and grammatical roles on word order;
  - **Cons**: parametric overviews can be misleading if the parameters are not well-defined and accurately coded; some language patterns may be too complicated to distil into a binary value of "Y"or "N" (Bickel, 2015; Haspelmath, 2018; Evans, 2020).

- To mitigate against these cons, we use checkpoints rather than single parameter settings, avoiding the risk of collapsing too much variation into a binary "Y" or "N" value; we allowed a non-binary "Y/N" value in cases which showed mixed behaviour.

4.2. Continuum of variation

- No language's word order was at the extreme of fully-conditioned by discourse roles, nor fully conditioned by grammatical roles; the word order patterns lie at different points of a **continuum**:

[5] This does not mean that the raw data should be ignored – see the appendix for the full list of data on which the parameter values were based. The parameters can serve as a guide to where to look in detail at the raw data.

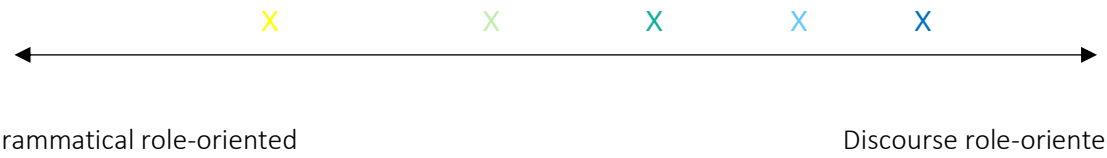

Figure 3. Diagram representing how the languages' word orders are at different positions on a continuum between grammatical roles and discourse roles being most influential. Note that no language is found at the extremes.

- The examples in Figure 2 are left abstract (represented by "x") rather than using language names, as the exact position on the continuum is influenced by the weighting of the parameters – it would only be an assumption to suggest our current parameters should be evenly weighted – and the current project only investigated word order, not considering other potential variables;

4.3. Innovation

- The variation within Bantu also elucidates language contact and language change, potentially under the influence of contact.
- The presence/absence of preverbal focus we discovered is an innovation, overlapping in its centre with two waves of innovation for object marking (changing from non-doubling to doubling, and from multiple to single object markers, Van der Wal to appear), see also Marten et al. (2007) on Bemba and Chichewa as the centre.
- Our study confirms the absence of inversion constructions in the north west (Hamlaoui 2018, Hamlaoui & Makasso 2015), which may be an areal feature too, possibly correlating with other features such as object marking (as Hamlaoui 2018 proposes).

5. Conclusion and further questions

- A very obvious conclusion is that our systematic overview confirms the interesting **microvariation** within the Bantu languages: no two of our languages show the same profile.
- It is not enough to describe the word order of Bantu languages in terms of information structural classifications such as "topic-nontopic", and the classical treatment of languages in terms of "SVO", "SOV" similarly fails to capture the full picture of word order. We therefore advocate a language-specific classification of the word order of each language.
- This means that **Theoretical models** need to capture the variation and the influence of both information structure and grammatical role in the grammar; there are two necessary questions that need to be answered on an individual basis:
    A. Which precise **notions** are active in a given language?—Makhuwa both 'topic' and 'focus'; Rukiga perhaps only 'focus'; or different notions are needed (contrast, unexpectedness, …);
    B. What type of **features** correspond to these notions? Semantic or syntactic features? A or A-bar features?
- Future work should involve at least:
    - The discourse-configurationality of agreement—for example whether the subject marker is perhaps better captured as a topic marker (Morimoto 2006);
    - Other morphological marking of information structure in determining the discourse-configurationality of a language (Gibson et al. 2017);

- What can explain the different focus positions as found in Kukuya (IBV), Makhuwa (IAV) and Kirundi (final)?

Abbreviations

Bare numerals (1, 2, 3, etc.) refer to Bantu noun class; when followed by SG/PL they indicate person.

ASSOC	associative marker
AUG	augment (pre-prefix)
FOC	focus marker/copula
N.PST	near-past tense
PST	past tense
SM	subject marker

Acknowledgements

This research was supported by a NWO Vidi grant 276-78-001 as part of the BaSIS 'Bantu Syntax and Information Structure 'project. We thank our speakers Bahati Laikon Mwakasege, Peter Mwasyika Mwaipyana, Yona Mwaipaja, Pamellah Geiga Birungi, Joel Tumusiime, Ronald Twesigomwe, Dennis Muriuki Katheru, Philip Murithi Nyamu, Onesmus Mvuambi Kamwara, Constancia Zaida Mussavele, Arlindo João Nhantumbo, Gomes David Chemane, Menrage Buananli, Joaquim Nazário, N'gamo Saida Aly, Pierre Molel, Daniel Mbel, Étienne Ondjem, Jeanne Ongmolaleba, Flore Ongmadack, and Patience Nambi for sharing their insights with us. We thank Helen Eaton, Hazel Gray, Heidrun Kröger, Julius Taji for checks on preverbal focus, Nelsa Nhantumbo for data collection on Cicopi, and student assistant Simon Msovela for help in data collection on Kinyakyusa.

References

Baker, Mark. (2010). Formal generative typology. In: Heine, Bernd; Narrog, Heiko. *The Oxford handbook of linguistic analysis*, 2nd edition. Oxford: Oxford University Press, 285-312.

Bickel, Balthasar 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In: Heine, Bernd; Narrog, Heiko. *The Oxford handbook of linguistic analysis*, 2nd edition. Oxford: Oxford University Press, pp. 901 - 923.

Bickel, Balthasar. 2010. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.

Downing, Laura J., and Larry M. Hyman. (2015). Information structure in Bantu. In Féry, Caroline & Shinichiro Ishihara. *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Downing, Laura, and Lutz Marten. 2019. Clausal morphosyntax and information structure. In Van de Velde, Mark et al. (eds), *The Bantu Languages, second edition*, 270-307. London: Routledge.

É. Kiss, Katalin, ed. 1995. *Discourse Configurational Languages*. New York: Oxford University Press.

Evans, Nicholas. 2020. Introduction: Why the comparability problem is central in typology. Linguistic Typology 24(3). 417–425. DOI: https://doi.org/10.1515/lingty-2020-2055.

Gibson, Hannah, Andriana Koumbarou, Lutz Marten and Jenneke van der Wal. 2017. Locating the Bantu conjoint/disjointalternation in a typology of focus marking. In Van der Wal, Jenneke and

Larry Hyman (eds), *The conjoint/disjoint alternation in Bantu*, 61-99. Trends in Linguistics series. Berlin: Mouton de Gruyter.
Good, Jeff. 2010. Topic and focus fields in Naki. In Fiedler, Ines and Anne Schwarz (eds.), *The expression of information structure. A documentation of its diversity across Africa*, 35-68. Amsterdam: John Benjamins.
Hamlaoui, Fatima. 2018. Proto Bantu word order. Presentation at BGR2, Ghent.
Hamlaoui, Fatima, and Emmanuel-Moselly Makasso. 2015. Focus marking and the unavailability of inversion structures in the Bantu language Basaa (A43). *Lingua* 154. 35-64.
Hammarström, Harald. 2019. An inventory of Bantu languages. In Van de Velde, M. et al. (eds.), *The Bantu Languages, second edition*, 17-78. London: Routledge.
Haspelmath, Martin. (2018). How comparative concepts and descriptive linguistic categories are different. In Van Olmen D., Mortelmans T., & Brisard F. (Eds.), Aspects of Linguistic Variation (pp. 83-114). Berlin; Boston: De Gruyter. Online file: http://www.jstor.org/stable/j.ctvbkjwxf.6.
Li, Charles N., and Sandra Thompson. 1976. Subject and Topic: A New Typology of Language. In Charles N. Li (ed.), *Subject and Topic*, 457-90. New York: Academic Press.
Maho, Jouni F. 2009. NUGL Online: The online version of the New Updated Guthrie List, a referential classification of the Bantu languages. *Online file: https://brill.com/fileasset/downloads_products/35125_Bantu-New-updated-Guthrie-List.pdf* (version dated 04 June 2009).
Marten, Lutz, and Jenneke van der Wal. 2014. A typology of Bantu subject inversion. *Linguistic Variation* 14(2), 318-368.
Morimoto, Yukiko. 2000. Discourse Configurationality in Bantu Morphosyntax. Stanford University: Doctoral Dissertation.
Morimoto, Yukiko. 2006. Agreement properties and word order in comparative Bantu. *ZAS Papers in Linguistics* 43. 161-188.
Nurse, Derek and Gérard Philippson. 2003. Introduction. In Nurse, D. and G. Philippson (eds.), *The Bantu Languages*, 1-12. London: Routledge.
Öhl, Peter. 2010. Formal and Functional Constraints on Constituent Order and Their Universality. In Carsten Breul (ed.), *Comparative and Contrastive Studies of Information Structure*, 231-275. Amsterdam: John Benjamins.
Round, Erich R. and Corbett, Greville G. 2020. "Comparability and measurement in typological science: The bright future for linguistics" *Linguistic Typology* 24 (3), pp. 489-525, DOI: https://doi.org/10.1515/lingty-2020-2060
Skopeteas, Stavros, Fiedler, Ines, Hellmuth, Sam, Schwarz, Anne, Stoel, Ruben, Fanselow, Gisbert, Féry, Caroline and Manfred Krifka. 2006. *Questionnaire on Information Structure: Reference Manual.* ISIS Volume 3. Potsdam: Universitätsverlag Potsdam.
Van der Wal, Jenneke. To appear. A featural typology of Bantu agreement. OUP.
Yoneda, Nobuko. 2011. Word order in Matengo (N13): Topicality and informational roles. *Lingua* 121 (5). 754-771.