



## Morpho-syntactic variation in a Xhosa microvariation project: what can a discourse-centered approach offer?

Eva-Marie Bloom Ström, University of Gothenburg and Rhodes University

### 1. *Aim of paper and research questions*

- To argue for the use of spoken texts in analysing morphosyntax (complemented by careful elicitation)
- Give an example of a construction found in a spoken text but not in elicitation
- Show that there is Agreeing Inversion in Xhosa
  
- What is morpho-syntactic microvariation?
- What kinds of micro-variation studies are there? And in Bantu?
- What are the pros and cons of using elicitation vs. recorded speech?
- Reports results from a project on morpho-syntactic microvariation in Xhosa, a South African Bantu language of the Nguni subgroup

### 2. *What is morpho-syntactic microvariation?*

*A branch of linguistic research that studies subtle but systematic differences in grammar between very closely related linguistic varieties.*

#### 2.1 *Different from traditional dialect research*

- Intra-linguistic variation based on geography has been studied by linguists since the 19<sup>th</sup> century
- Focusses on lexicon and the production of dialect maps
  
- Describes how varieties differ and where they are spoken
- Compares iso-glosses to determine dialect borders and estimate how (much) dialects differ from each other

This is also the aim of certain microvariation projects like ASIS (Italian) and SAND (Dutch) (Cornips and Poletto 2005)

#### 2.2 *Different from sociolinguistic studies of variation*

- Variation also much dependent on social factors like class and gender
- Led to the development of sociolinguistic studies of variation (Labov 1994)
- Microvariation studies like ASIS and SAND aim at homogenising the sample and keep all other factors than geography as invariable as possible

### 3. *Kinds of microvariation and the Bantu language family*

- Large-scale dialect atlas projects as for example Northern Italian syntactic dialect atlas (ASIS) and the syntactic atlas of the Dutch dialects (SAND). Mainly focus on the linguistic landscape of Europe and America, and even more narrowly, on Germanic languages, simply because this is the best-documented genetic group (Kortmann 2004). Considerable work has also been done on the Romance languages.
- In certain micro-variation projects the point of departure is grammatical expressions – deviating from the standard – that linguists have taken note of and then analysed. For example a collection of contributions from different authors on American varieties of English, presented on a website (Yale\_University 2015). Regional, ethnic, or age-based variation in certain syntactic constructions is investigated. For example, in which areas constructions occur with *need* (or *want*, *like*) followed by a passive participle as in ‘The car needs repaired’, and by whom these constructions are used. In many cases, the constructions can be considered shibboleth of a certain area and they are possibly stigmatized, as for example negative inversion (*don’t nobody care about that*) in African American English (Horn 2014).

#### 3.1 ***What is there in Bantu?***

None of these kinds of microvariation studies – large-scale dialect syntax projects and the ones that focus on a certain construction and its distribution – are so far found in Bantu language studies.

- Also traditional dialect studies are rare in the Bantu language family and in such work, lexicon and phonology are compared rather than morpho-syntax, e.g. (Möhlig 1974).
- Morpho-syntactic work comparing certain constructions within the Bantu language family have a rather long history (e.g. Bresnan and Moshi 1990 ; Marlo 2015)

##### 3.1.1 *Parametric studies of variation*

- Certain studies based on parameters (Marten et al. 2007). Yes-No parameters like ‘can the object marker and the lexical object NP co-occur?’ are compared in order to calculate the distance between languages and to shed new light on these constructions.
- Later more and other parameters as well as data from other languages (Zeller and Ngoboka 2015 ; Marten and Kula 2012 ; van der Wal To appear ; Edelsten and Marten 2016). Often fairly distinct language varieties.
- Parametric studies also of more closely related varieties: of the Greater Ruvuma group in Tanzania including lexical information (Petzell and Hammarström 2013); Kongo (Dom and Bostoen 2015).

Some studies on Bantu variation might be considered macro-variation studies (interlinguistic) rather than micro-variation (intralinguistic)

Less basic research on Bantu languages than e.g. the Germanic group, therefore less basis to go even deeper into intralinguistic varieties.

But needed, especially for larger official languages like Xhosa where there is a social relevance.

#### 4. Challenges with written and oral elicitation in the Xhosa project (cf. European projects)

##### 4.1 Homogenising the sample virtually impossible

For projects like SAND subjects are as much as possible i.a.: native speakers of local dialect; born in the same community as parents and lived there until adulthood; not left the community for longer than 7 years; aged between 55-70. (Cornips and Poletto 2005)

Important to take the sociolinguistic situation into account. For the Eastern Cape:

- NORM (Non-mobile, Older, Rural Male) is very mobile, exposed to urban, mixed varieties
- Marriages across clans; women move
- Underlying migration and resettlement in 19<sup>th</sup> century caused by colonisation and wars
- Mass deportations and forced migrations of the apartheid period
- Certain 'dialectal' traits are actually hlonipa (but this is mainly a lexical issue)
- Also sociolinguistically interesting: very strong attachment to clans, kingdoms (ethnonyms)

##### 4.2 Elicitation for large amounts of data

A widely used method in traditional linguistic fieldwork is elicitation. It is also widely criticized (e.g. Bowerman 2008)

Advantages (at the same time drawbacks of analysing recordings):

- a) provides insight in a speaker's competence more easily
- b) one can examine sentence types that occur rarely, hours of recordings might not reveal a single example of a certain construction, especially when studying low-frequency phenomena. Difficult to get all the different contexts in which it may appear.
- c) interaction with other relevant syntactic variables predicted by theory, excluding other variables
- d) we can check grammaticality, scale of grammaticality
- e) negative data, in recorded speech you cannot check if this specific construction is acceptable or not for a certain speaker, just that the speaker did accidentally not use it
- f) Syntactic properties are difficult to elicit through means of stimuli or other methods apart from translation, especially when the aim is to compare similar structures across varieties

(Bucheli and Glaser 2002 ; Cornips 2002 ; Cornips and Poletto 2005).

##### 4.3 Problems with elicitation in micro-variation

Dialect syntax projects often use elicitation for the collection of data.

Also questionnaires that are filled in by speakers (Bucheli and Glaser 2002 ; Cornips 2002). Not used in current project:

- a) No agreed way to write the home variety. Influence of the standard writing system should be avoided.
- b) Low literacy and education levels, unfamiliarity with questionnaires
- c) Dialects often do not have prestige.
- d) On the other hand, the important ethnonyms can lead to hypercorrection
- e) No tone marking in orthography.

- f) The possibility that speakers do not understand the aim of the questionnaire or the sentences, that there is a difference in interpretation, and the linguist cannot be sure that the phrase is a translation of what s/he intended.

In oral elicitation the linguist checks the meaning of the phrase, tries to form correct structures, can adapt the questionnaire if necessary. There are still drawbacks...

- A. A speaker - aware of the researcher's interest in dialects - might discard a construction based on the wrong use of a lexeme, although the construction itself is correct.  
 B. Results might be influenced by the interview language  
 C. It happens that speakers reject a construction that they previously judged as grammatical.

Illustration of C: In an elicitation session for focus constructions, the translation given of the phrase 'Who was chasing the cat?' was *ngubá.ní ebéle qa ika:ti?* The answer produced by the speaker was then *ngábántwa:na* 'the children'. Following this I asked 'is it also possible to phrase this starting with *ku...*? (We were working on similar constructions earlier, subject focus constructions with a default subject marker *ku-*) and the speaker said yes, you can say:

*Kúlé:qé abantwana*  
 17SM-chase.PRF 2.children  
 'The CHILDREN were chasing (the cat)' [NH150505E]

As this was an unexpected result, I asked the same questions again at a later session, in which it was rejected as ungrammatical by the speaker:

1. *\*Kúlê-qé abantwa:na*  
 [NH150907E]

Did the speaker in question made a mistake during the first or the second session? Is the phrase is marginally acceptable in certain contexts and in that case which contexts?

Especially in a micro-variation study, in which we study the subtle differences between how people speak, this forms a challenge.

- D. Moreover, I have no idea whether he would ever actually use such a phrase. Another example is the instrument inversion phrase *imoto ihamba uSipho* 'Sipho is travelling by car', literally 'the car travels Sipho'. This is judged as correct by (at least some) speakers, but I have never heard anyone use that expression and it does not occur in any of the recordings carried out for the corpus collection part of my project.

#### **4.4 Why collect texts for analysis?**

It has frequently been argued for the use of elicitation techniques and oral tasks in combination with systematic recording of spontaneous speech (Cornips 2002 ; Horn 2014 ; Cornips and Poletto 2005).

- To avoid drawbacks of elicitation.

- Apart from knowing what is possible in language, we ideally want to know how language is used.
- Moreover, language is what people use for linguistic interaction and social exchange and should therefore be studied in its socio-cultural setting (König et al. 2015).

In the current project, spoken texts are collected, transcribed and included in a corpus. For each place recordings are made of different activities (Allwood and Hendrikse 2005)

## 5. Using recorded language in micro-variation

This section looks into one of the aspects studied in this micro-variation project, namely subject inversion, from a methodological perspective.

- Shows that when using data ‘in the wild’ it can be very difficult to draw conclusions regarding the kinds of constructions used.
- In elicitation, we control the speakers by asking the questions that will for example lead to a certain focus construction, but in natural spoken language the speaker is not controlled in any way.
- I propose that we either need to base our analysis on much larger amounts of recorded data – very time consuming for languages for which there are no corpora – or we have to take shortcuts by checking the spoken data with the same and other speakers. Again, this gives the same drawbacks as for elicitation mentioned above.

### 5.1 Dislocated subjects or Agreeing Inversion?

Illustration from a narrative, folktale kind of story

The speaker frequently uses a construction in the past, whereby the subject follows the verb<sup>1</sup>.

2. *lâ:-ha:mb(a) izim*  
 5SM.PST-walk 5.giant  
 ‘The giant went’

#### 5.1.1 Analysed as a right-dislocated subject

The subject does not occur in the Immediately After Verb (IAV) position within the verb phrase

Here referred to as Agreeing Postverbal Subject (APS) for ease of reference, also sometimes referred to as afterthought constructions or anti-topic (Chafe 1976 ; Lambrecht 1994).

The dislocation is evidenced by the phonological phrasing (Jokweni 1995 ; Downing 2003 ; Zerbian 2004). The penultimate vowel of the verb *-hamba* ‘go’ is lengthened, indicating the end of the phonological phrase and that the subject *izim* ‘giant’ is phrased

---

<sup>1</sup> Parentheses indicate that the vowel is not heard, often written as *lahamb’ izim* in Xhosa texts. For glossing reasons, it is important to know what the missing vowel is. A colon indicates lengthening of the preceding vowel.

separately. As far as I know, the vowel liaison, although it only occurs phrase-medially, is not an indication of the close bond that occurs when an element is in IAV(??)

### 5.1.2 *Information Structure of these phrases.*

The dislocated subjects in APS are used for semi-active concepts, i.e. concepts that have been mentioned earlier in the text or that can be inferred from context. The giant in the story is, after introduction, picked up as topic through the APS construction, often after not having been mentioned for a couple of sentences:

3. *lâ:-fî:k(a) ízi:m lá-vál(a) e-mnyá:ngo*  
5SM.PST-arrive 5.giant 5SM.PST-open LOC-3.doorway  
'The giant arrived and opened the door'

### 5.1.3 *The phonological phrasing often 'fails'*

In the following example, the penultimate vowel of the verb is not long<sup>2</sup> and there is vowel liaison:

4. *wá-bâs(a) uNomaha:mle*  
1SM.PST-kindle uNomahamle  
'Nomahamle kindled the fire'

### 5.1.4 *Resembles Agreeing Inversion (AI).*

AI is a core inversion construction (Marten and van der Wal 2015)

Used for inactive, non-topical concepts (subject focus or presentational focus/thetic sentences).

The verb agrees with the post-verbal subject and is in a close bond with that subject: phonological phrasing, conjoint form in tenses where the conjoint/disjunct distinction is relevant.

AI has not been reported for Xhosa, and the language is listed as lacking relevant information for this construction in a recently published comparative study of subject inversion in Bantu languages (Marten and van der Wal 2015). The same study lists Zulu as not having AI.

### 5.1.5 *Default Agreeing Inversion*

AI resembles Default Agreeing Inversion (DAI), which is also used for subject focus or presentational focus, but in which the verb has a default subject marker. DAI is commonly used in Xhosa, with the default subject marker *ku-*, and also occurs in the story:

5. *kú-khála nê:-ntá:ka*  
17SM-sing and-10.birds (na+iintaka)  
'and the birds were singing..'

---

<sup>2</sup> The vowel length was measured manually in Praat.

### 5.1.6 *APS in past and future*

APS constructions are common in the story but mostly in the remote past (6), also in the future (7). Sometimes with penultimate lengthening, sometimes without.

6. *lá:-ba:lé:k(a) izim*

5SM.PST-run 5.giant

'The giant ran'

7. *bâ:-zo-phúm(a) ábantwana kw-ézi ngxo:wa*

2SM-FUT-come.out 2.children LOC-10.DEM.PRX 10.bag

'The children will come out through these holes'

How relevant is the lengthening of the penultimate vowel in the past?

For Zulu, it has been shown that the penultimate lengthening is relevant even in tenses in which there is no morphological marking of the forms as conjoint/disjoint (Zeller et al. To appear).

The study concerned object complements in IAV but the results for subjects in IAV should be similar.

It also showed, however, that there is more overlapping in vowel length between the morphologically unmarked forms. I.e. the medial and final forms differ less clearly in length in the remote past tense than in the present.

### 5.1.7 *APS in perfect*

8. *ú-th-é úVathíswá:*

1SM-say-PRF Vathiswa

'Vathiswa said:'

Occurs in the conjoint form

Only with -thi 'say': irregular verb

### 5.1.8 *Checking in elicitation*

Certain phrases from the story were also checked with another speaker.

In the context of the story, how would you say this? (Basically recording the phrase again, and separately)

Confirms the observation: also pronounces the phrases without penultimate lengthening in many cases. This is possible with or without vowel liaison:

9. *lá-súkum' izi:m*

*lá-súkúma i:zim*

5SM.PST-stand.up 5.giant

'the giant stood up' [NF160404E]

Changed to the present, penultimate lengthening is very clear:

10. *li-ya-súkû:ma i:zim*

5SM-DJT-stand.up 5.giant

'the giant stands up' [NF160404E]

### 5.1.9 Other evidence APS/AI?

If phonological phrasing does not give enough clues as to whether these constructions are APS or AI, what about other evidence for the close bond between a verb and a subject in IAV?

Other evidence that the subject is dislocated comes from the fact there can be an object marker on the verb. The penultimate vowel can be lengthened (example 11) or not (example 12)

11. *lá-m-fá:ka ízim*  
5sm.pst-1om-put 5.giant  
'The giant put her (inside)' [NF160404E]

12. *lá-m-fák(a) ízim*  
5SM.PST-1OM-put 5.giant  
'The giant put her (inside)' [NF160404E]

Clear evidence that this is a dislocated subject (APS)

Shows that phonological phrasing is not really evidence of the difference between APS and AI.

### 5.1.10 Thetic phrases

However, there are also a few examples in the story in which a new situation is introduced by a thetic sentence, and the agreeing subject follows the verb:

13. *kwâ:sa za-khál(a) ínkukhu*  
in.the.morning 10SM.PST-cry 10.rooster  
'In the morning the roosters crowed.'

14. *lâ:-tshona ilanga*  
5sm.pst-set 5.sun  
'The sun set'

Also, the speaker with which I checked the examples from the story accepts changing the thetic DAI in example 15 to a construction with AI in example 16:

15. *ku-khála né-ntá:ka*  
17SM-sing and-10.birds  
'and the birds were singing..' [NF160404E]

16. *za-khála né-nta:ka*  
10SM.PST-sing and-10.birds  
'and the birds sang' [NF160404E]

These examples show that AI occurs in Xhosa. However, it appears to only function as a thetic expression in Xhosa, and maybe also only as a certain type of thetic expression. AI does not occur at all in my data as subject focus. It also does not occur as answer to subject focus questions in elicitation. In elicitation, it is never given as a thetic sentence in answer to 'what happened?'



## 5.2 Conclusions

This shows the importance of considering spoken texts in order not to rely entirely on elicitation.

Also, the examples show that even when subjects are dislocated in APS constructions, phonological phrasing is not enough evidence for APS, at least not in a naturally spoken text.

Analysing spoken texts is necessary in order to find out what constructions are used in a language and how frequent they are. But can they be used to establish variation?

In this specific case, many more narratives spoken by different people need to be analysed in order to know if all speakers use APS to reactivate semi-active concepts, if phonological phrasing is different between speakers, and if AI is used by all speakers as a thetic construction. Here, also different kinds of thetic constructions can be considered. How feasible is such an endeavour when the researcher first has to collect the spoken language corpus before analysis can start? What are the options?

### References:

- Allwood, Jens and Rusandre Hendrikse. 2005. Guidelines for developing spoken language corpora. In: J. Allwood et al. (eds.) *Spoken African Language Corpora Series*, pp Pretoria, South Africa: UNISA, Dept of Linguistics.
- Barbiers, Sjef, Leonie Cornips and Susanne van der Kleij (eds.) 2002. *Syntactic Microvariation*. Amsterdam: SAND. <http://www.meertens.knaw.nl/books/synmic>.
- Bowern, Claire. 2008. *Linguistic Fieldwork: A Practical Guide*. Hampshire & New York: Palgrave Macmillan.
- Bresnan, Joan and Lioba P. Moshi. 1990. Object asymmetries in comparative Bantu syntax. *Linguistic inquiry* 21 (2), pp 147-185.
- Bucheli, Claudia and Elvira Glaser. 2002. The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems. In: S. Barbiers et al. (eds.) *Syntactic microvariation*, pp Amsterdam: SAND.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects and topics. In: C. N. Li (ed.) *Subject and topic*, pp 27-55. New York: Academic Press.
- Cornips, Leonie. 2002. Variation between the infinitival complementizers om/voor in spontaneous speech data compared to elicitation data. In: S. Barbiers et al. (eds.) *Syntactic microvariation* Amsterdam: SAND.
- Cornips, Leonie and Cecilia Poletto. 2005. On standardising syntactic elicitation techniques (part 1). *Lingua* 115, pp 939-957.
- Dom, Sebastian and Koen Bostoen. 2015. Examining variation in the expression of tense/aspect to classify the Kikongo Language Cluster. *Africana Linguistica* 21, pp 163-211.
- Downing, Laura J. 2003. Stress, tone and focus in Chichewa and Xhosa. In: R.-J. Anyanwu (ed.) *Stress and Tone – The African Experience*, pp Köln: Rüdiger Köppe Verlag.
- Edelsten, Peter and Lutz Marten. 2016. *Parametric approaches to morphosyntactic variation in Bantu* Paper presented at Bantu 6, Helsinki,
- Horn, Laurence. 2014. Afterword: Microvariation in Syntax and Beyond. In: R. Zanuttini and L. Horn (eds.) *Micro-Syntactic Variation in North American English*, pp Oxford: Oxford University Press.
- Jokweni, Mbulelo W. 1995. Aspects of IsiXhosa phrasal phonology, PhD thesis, University of Illinois, Urbana-Champaign
- Kortmann, Bernd. 2002. New prospects for the study of dialect syntax: Impetus from syntactic theory and language typology. In: S. Barbiers et al. (eds.) *Syntactic Microvariation*, pp 185-213. Amsterdam: SAND.
- . 2004. Introduction. In: B. Kortmann (ed.) *Dialectology meets Typology* Berlin: Mouton de Gruyter.

- König, Christa, Bernd Heine and Karsten Legère. 2015. *The Akie Language of Tanzania. A Sketch of Discourse Grammar*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo Univ. of Foreign Studies.
- Labov, William. 1994. *Principles of linguistic change*. Oxford: Blackwell.
- Lambrecht, Knud. 1994. *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Marlo, Michael. 2015. On the number of object markers in Bantu languages. *Journal of African Languages and Linguistics* 36 (1), pp 1-65.
- Marten, Lutz and Nancy C. Kula. 2012. Object marking and morphosyntactic variation in Bantu. *Southern African Linguistics and Applied Language Studies* 30 (2), pp 237-253.
- Marten, Lutz, Nancy C. Kula and Nhlanhla Thwala. 2007. Parameters of morphosyntactic variation in Bantu. *Transactions of the Philological Society* 105, pp 1-86.
- Marten, Lutz and Jenneke van der Wal. 2015. A typology of Bantu subject inversion. *Linguistic variation* 14 (2), pp 318–368.
- Möhlig, Wilhelm J. G. 1974. *Die Stellung der Bergdialekte im Osten des Mt. Kenya. Ein Beitrag zur Sprachgliederung im Bantu*. Berlin: Dietrich Reimer.
- Petzell, Malin and Harald Hammarström. 2013. Grammatical and lexical comparison of the Greater Ruvu Bantu languages. *Nordic Journal of African Studies* 22 (3), pp 129-157.
- van der Wal, Jenneke. To appear. What is the conjoint/disjoint alternation? Parameters of crosslinguistic variation. In: J. van der Wal and L. M. Hyman (eds.) *The conjoint/disjoint alternation in Bantu*, pp Berlin: Mouton de Gruyter.
- Yale\_University. 2015. *Yale grammatical diversity project: English in North America*. Retrieved from <http://microsyntax.sites.yale.edu/>.
- Zanuttini, Raffaella and Laurence Horn (eds.) 2014. *Micro-Syntactic Variation in North American English*. Oxford: Oxford University Press.
- Zeller, Jochen and Paul Ngoboka. 2015. On parametric variation in Bantu, with particular reference to Kinyarwanda. *Transactions of the Philological Society* 113 (2), pp 206-231.
- Zeller, Jochen, Sabine Zerbian and Toni Cook. To appear. Prosodic evidence for syntactic phrasing in Zulu. In: J. van der Wal and L. M. Hyman (eds.) *The conjoint/disjoint alternation in Bantu* Berlin: Mouton de Gruyter.
- Zerbian, Sabine. 2004. Phonological phrases in Xhosa (Southern Bantu). *ZAS Papers in Linguistics* 37, pp 71-99.