PhD project

# A Quantitative Analysis of
# Information Structure in Natural Discourse

Viktoria Apel

Humboldt-Universität zu Berlin

---

# 0    PhD Thesis: Contents

# 1    Introduction

## *1.1  Fula and Pular*

- Classification: Niger-Congo > Atlantic-Congo > Atlantic
- Macro-language *Fula* is spoken by around 22 million speakers in 18 countries (Lewis et al. 2014)

---

[1]  The dissertation is realized in affiliation to Project B7 "Predicate-centered focus types: A sample-based typological study in African languages" of the SFB 632 'Information Structure' and is funded by the DFG and the Friedrich-Ebert-Stiftung.
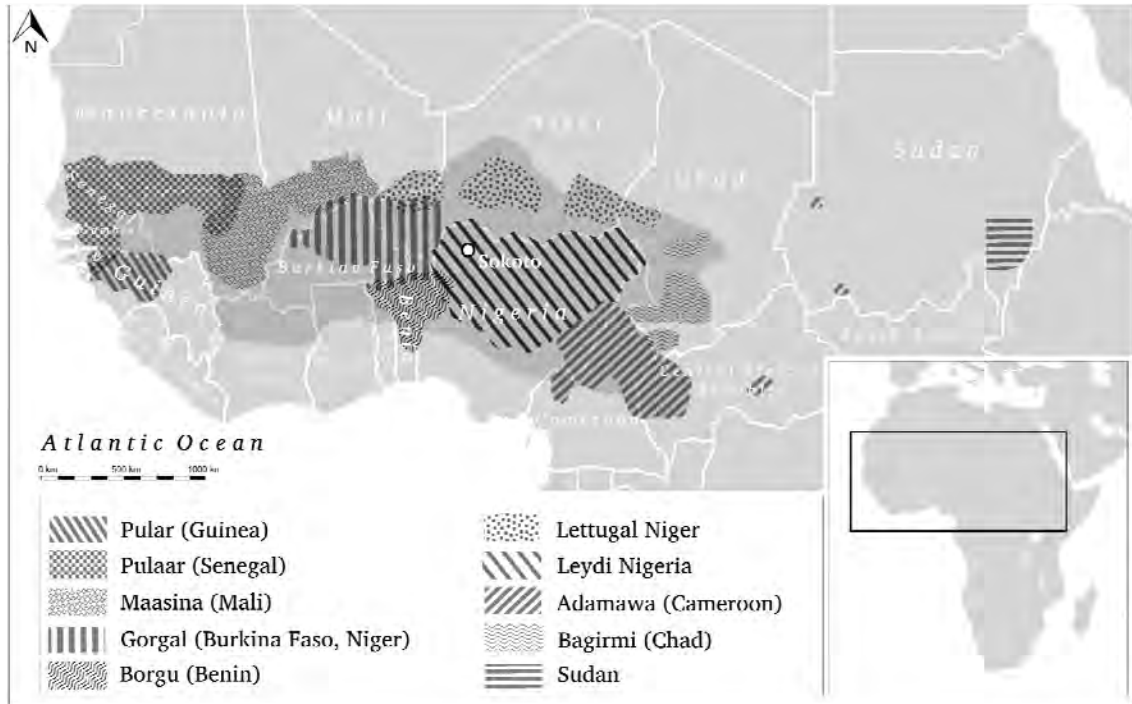
**Map 1:** Fula in Africa (map by Simon Argus, based on Harrison 2003)

- Guinean dialect (Fuuta Jaloo) *Pular* is spoken by around 3 million speakers in Guinea and neighboring countries
- iso 639-3: fuf
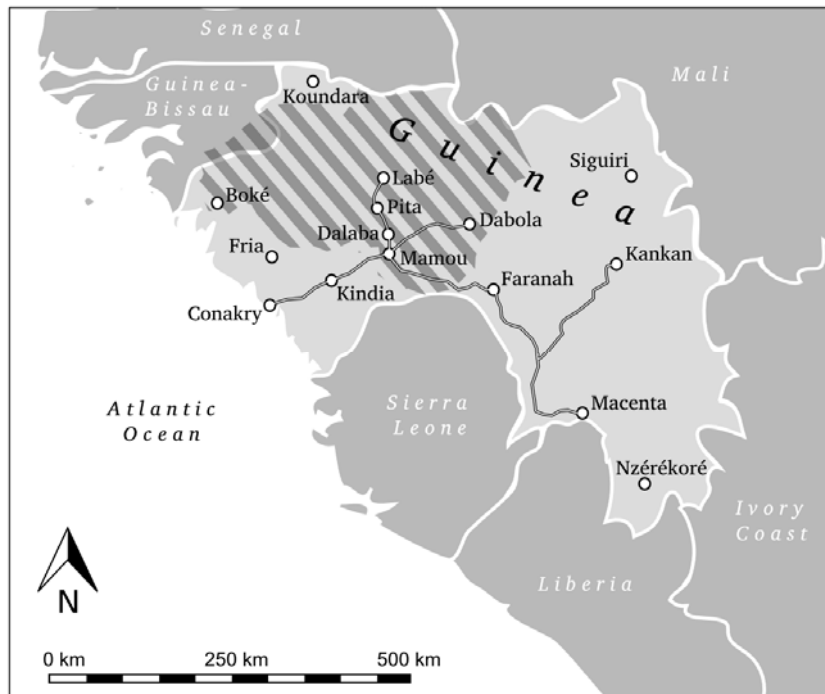


**Map 2:** Pular in Guinea (map by Simon Argus, based on Harrison 2003)

- **Phonology:**

  Pular uses the following consonants:

  | | | | | | |
  |---|---|---|---|---|---|
  | b [b] | ɗ [ɗ] | h [h] | m [m] | q [q] | x [ɣ] |
  | ɓ [ɓ] | nd [n͡d] | j [d͡ʒ] | n [n] | r [ʁ] | y [j] |
  | mb [m͡b] | f [f] | nj [n͡d͡ʒ] | ɲ [ɲ] | s [s] | ƴ [ʄ] |
  | c [t͡ʃ] | g [g] | k [k] | ŋ [ŋ] | t [t] | z [z] |
  | d [d] | ng [n͡g] | l [l] | p [p] | ' [ʔ] | |

  **Table 1:** Orthography and phonetic realization of consonants in Pular

  Vowels may be either short (a[a], e [e,ɛ], i [i], o [o,ɔ], u [u]), long (aa, ee, ii, oo, uu) or nasalized (an, en, in, on, un).

  Short vowels and consonants (V; C) are distinctive from long ones (VV; CC):

  (1) se**l**ugol     vs.     se**ll**ugol

      'to turn'              'to be healthy'

- **Nominal morphology:** 24 agreement classes ("noun classes")

  Except for proper names, loan words and generic nouns, the class is marked by a suffix on the noun stem:

  (2)[2] laaw-**ol**     vs.     laaw-**i**

      road-14              road-4

      'road'              'roads'

- **Verbal morphology:**
  - 3 Voices: Active, middle and passive
  - 3 Asserted perfective paradigms vs. 4 asserted imperfective paradigms

| Voice | PFV1 | PFV2 | PFV3 | PFV.NEG |
|---|---|---|---|---|
| Active | -u/-∅ | -i/-u | -ii | -aali |
| Middle | -i | -ii | -ike | -aaki |
| Passive | -a | -aa | -ama | -aaka |

**Table 2:** The perfective paradigms

| Voice | IPFV1 | IPFV2 | IPFV3 | IPFV4 | IPFV.NEG |
|---|---|---|---|---|---|
| Active | -u | -a | -ay | -ata | -ataa |
| Middle | -o | -oo | -oto | -otoo | -otaako |
| Passive | -e | -ee | -ete | -etee | -etaake |

**Table 3:** The imperfective paradigms

---

2    Class 14 designates mainly objects which are long. Class 4 is one of the two main plural classes for non-humans.

Derivational suffixes, TAM morphemes and pronouns are suffixed to the verb root in that order:

(3)  mi   ɲaam-it-ii=ndi
     1S   eat-REP-PFV3=6
     'I ate it (e.g. rice) again.'

   o  3 Locative periphrases (incl. a locative copula): Progressive, stative and habitual

(4)  Arkia  **hino** (<hi+no)  ɲaam-ude    vs.  **himo** (<hi+o)  ɲaam-ude
     PN     LOC.COP            eat-PROG          3S.LOC            eat-PROG
     'Arkia is eating'                          's/he is eating'

- **Syntax:** Basic word order is **SVO** (prepositional phrases follow the predicate, adverbs are placed either sentence-initially, after the predicate or sentence-finally)

(5)  S       V          O        (PP)
     Khaïry  sood-ay    liyy̆-i   (ka   luumo)
     PN      buy-IPFV3  fish-4   LOC  market.1
     'Khaïry will buy fish (at the market)'

## 1.2   Theory

- **Thetic statements:** *One* information unit (without internal information-structural division) (Sasse 1987, Lambrecht 1994)
- **Categorical statements:** *Two* information units, either topic-comment structure or focus-background structure
  - o  **(Sentence) Topic:** What the sentence is about (often, but not always the subject) (Dijk 1977)
  - o  **Comment:** What is added as information about the topic
  - o  **Focus:** The most salient information in the sentence (Dik 1997)
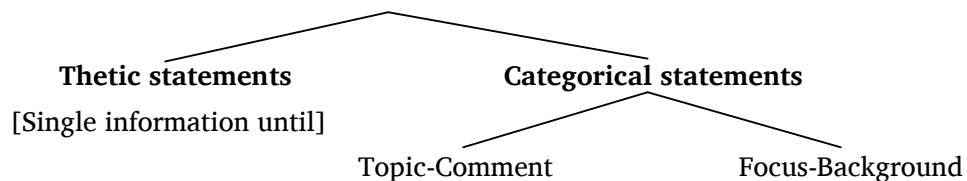


Thetic statement:      [——————————]
                       [There was a cow.]
Topic-Comment:         [——]<sub>Topic</sub>  [——————————]<sub>Comment</sub>
                       [She]      [was created first.]
Focus-Background:      [——————]<sub>Focus</sub>  [——————————————]<sub>Background</sub>
                       [It is the cow]      [which is superior to all beings.]

- Focus may be on a **term** (subject, object, adverbial), or on the **predicate** (state-of-affairs, truth value, TAM) (Dik 1997, Güldemann 2009)

## 2    The Corpus for This Study

- I collected data during two field trips (Dec 2012-Feb 2013 and Dec 2013-Mar 2014) in Dalaba, Guinea
- Besides the translation of questionnaires (on the verb system and on information structure) and experiments from QUIS (Skopeteas et al. 2006), I also collected natural discourse: folktales, cooking recipes, an event report, dialogues, controlled narratives after having seen a picture series or a short movie, songs
- Transcribed in the field, annotated (incl. pauses and ruptures) and translated in *Toolbox*
- Anomalies and queries were clarified with a speaker in Berlin
- From 41 of these kind of texts, I **selected 8 for this presentation** (for the dissertation, I selected 14 texts which are fully given in the annex of the dissertation)
  - ○ **6 Monologues:** 2 narrations from picture series (texts A + B in Table 4), 1 event report (D), 1 folktale (E), 2 narrations from short movies (G + H)
  - ○ **2 Dialogues:** 1 argumentative discussion after short movies (L), 1 free discussion (M)
- The texts were annotated and broken down according to **clause type**:

(6)    jooni  ɗe-ya    le-ɗɗe    (7)    ɗe    waaw-ay          rim-ugol
        now    3-other   tree-3              3     be.able-IPFV3   bear.fruit-INF

(7)    lutt-u = ɗe
        remain-PFV2 = 3.REL
        'Now the other trees [which are remaining]$_{Rel}$ they can bear fruits.'
        (Text A-31/32)

| Type and Name of the text | # Sentences | # Clauses | Length (min.) |
|---|---|---|---|
| **Monologues: Narrations** | | | |
| A The Giant Tree Story | 24 | 33 | 1:04 |
| B The Tomato Story | 17 | 25 | 1:00 |
| D Naming Ceremony | 19 | 22 | 0:47 |
| E The Two Brothers | 139 | 271 | 5:57 |
| G The Stolen Watch I | 18 | 34 | 0:43 |
| H The Pear Story | 38 | 62 | 1:57 |
| *Subtotal* | *255* | *447* | *11:28* |

| Dialogues | | | |
|---|---|---|---|
| L The Stolen Watch III | 146 | 213 | 3:19 |
| M Girlfriend Problems | 158 | 247 | 4:49 |
| *Subtotal* | *304* | *460* | *17:29* |
| **Total** | **559** | **907** | **28:57** |

**Table 4:** Text types, total number of sentences and clauses, and text length (in min.)

- For each text, each clause was classified into either **main clause**, **subordinate clause**, **exclamation**, **imperative** or **question**, and then more finely according to the construction type (verb paradigm, type of subordinate clause, etc.)
- Clauses containing dislocation were classified separately
- Over all texts, the clauses have the following distribution (this distribution is quite consistent across monologues and dialogues):



Main clause verbal 47%
Main clause non-verbal 8%
Subordinate clause 21%
Exclamation 9%
Imperative 8%
Question 3%
Unidentified 4%
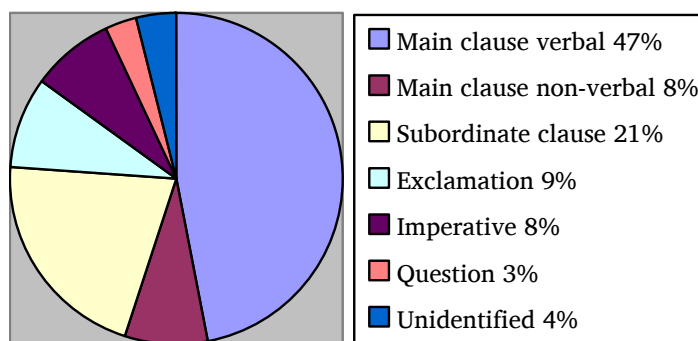
**Figure 1:** Distribution all texts

- There is a clear difference between monologues and dialogues in the use of the marked term focus constructions (see also Fiedler, in press)
  o From 447 clauses in monologues, 1% are focal
  o From 460 clauses in dialogues, 8% are focal
  o From 40 clauses with focus (questions excluded), 13% occur in monologues and 87% in dialogues (direct speech and rhetorical answers in monologues are counted as dialogues)

# 3    Research Questions and Results

## 3.1    In Which Kind of Clauses Does the Cleft Construction Occur?

The cleft construction is the most frequent construction for marking term focus in Fula. It consists of two parts: the focus part and the background part. The focal term stands in the

sentence-initial position and is preceded by the term focus marker *ko*[3]. If a pronoun is focalized, it stands in its emphatic form. The verb in the background part is always in a backgrounding verb form, i.e. in Perfective 2 or Imperfective 4. Additionally, there is prosodic prominence on the focal term (Anyanwu & Diallo 2007, Bao 2012 on Pulaar, Senegal).

(8)     **[Ko X]     [(S) V$_{BG}$ (O)]**
        'It is X     (that/who) do'
        Focus      Background

- In total, 6% of all clauses contain the cleft construction
- Of 52 clauses containing the cleft, 60% are main clauses, 31% are subordinate clauses and 9% are questions

The cleft in questions and in main clauses are kind of echo-like:

(9)  ko        hombo     wujj-i      *montre*     on?
     T.FOC    who        steal-PFV2   watch.1   DEF.1
     [It is who        ]     [Background]
     'Who stole the watch?'

(10) ko        an          wujj-i      *montre*     on
     T.FOC    2S.EMPH   steal-PFV2   watch.1   DEF.1
     [It is X          ]     [Background]
     'It is you (who) stole the watch.' (Text L-52)

The data show that clefts are also used as **complement clauses** (e.g. after the verbs to find/say/know/make, etc.) (ex. 11) and in an **adverbial clause** after "because" (ex. 12):

(11) a.    **S V [complement clause]**

     b.    a    waaw-ataa            wi'-ude
           2S   be.able-IPFV.NEG    say-PROG
           'You cannot say'

           ko       min          wujj-i＝on
           T.FOC   1S.EMPH     steal-PFV2＝2P
           'it is me (who) stole from you.' (Text L62/63)

---

[3]    In other dialects, the focus marker may be ∅ or *ɗum*.

(12) a.　**[Main clause] [Subordinate clause]**

b.　ko　　　an　　　　ƴett-ata　　　*montre*　on …
　　T.FOC　2S.EMPH　take-IPFV4　watch.1　DEF.1
　　'It is you (who) will take the watch,'

　　ɓaa　　　ko　　　an　　　　won-no　　takko　*montre*　on .
　　because　T.FOC　2S.EMPH　be-PRET　beside　watch.1　DEF.1
　　'because it is you (who) were next to the watch.' (Text L16/17)

⇒ The latter two uses have not been described before.

## 3.2 How is the Canonical Bipartiteness Cancelled in Thetic Statements?

As stated above, thetic statements consist of one information unit. This means that the bipartite "sentence-internal information structure that is induced by the morphosyntax of an unmarked sentence" needs to be cancelled (Güldemann 2010: 86). In order to achieve this, one can a) manipulate the subject and lift it from its "low" saliency position to a higher position, b) lower or lift the predicate (see Section 3.3) or c) cancel either the topic or the comment.

The comparison of all thetic statements which are used at the beginning of each story in the narratives showed:

- In two cases, there was no predicate. Example (13) is an identificational sentence:

(13)　————————
　　(Subject　　　　　　　　　)
　　ko　siinaa-ɓe　ɗiɗ-o　　non
　　ID　co.wife-2　two-2　　now
　　'(It is said now) there are now two co-wifes.' (Text E-8)

- In one case, the subject was marked as focus (= entity-central) (i.e. the subject is marked as being more salient than the predicate):

(14)　————————　　　　　　————————————————
　　(Subject　　　　　　　)　(Predicate　　　　　　　　　　　)
　　ko　　　gork-o　maw-ɗo　　yeh-i　　tett-ugol　　e　　nges-a　　mun
　　T.FOC　man-1　old.person-1　go-PFV2　pick-INF　　LOC　field-15　POSS.24
　　'(This film here, what I saw,) it is an old man (who) came pick in his field (, where he planted trees).' (Text H-1)

- In three cases, the first sentence of the narration was unmarked, i.e. there was no formal marking in order to cancel the bipartite structure. The unitary character of the first sentence is only pragmatical (which means that neither the subject nor the predicate is more salient):

(15)

| | (Subject | Predicate | | ) |
|---|---|---|---|---|
| ɗoo | miɗo | mar-i | kot-iraa-ɓe | ɗiɗ-o |
| here | 1S.LOC | have-STAT | older.brother-kinship-2 | two-2 |

'Here, I have two older brothers.' (Text A-1)

⇒ The last two strategies confirm the theoretical assumptions about a formal cancellation of the bipartite structure. The first strategy (the unmarked cancellation) proves that the pragmatical information status of a sentence does not have to be presented by morpho-syntactic means (cf. in-situ focus). Nevertheless, the role of prosody needs to be checked.

## 3.3   How Often Does the Verb Form Perfective 1 Appear?

In Fula, there is one verb paradigm in the perfective which is only used for state-of-affairs focus: the Perfective 1.

(16)  O     yi'-aali＝mo,      o    nodd-**u**＝mo.
      3S    see-PFV.NEG＝3S   3S   call-PFV1＝3S
      '{The woman said that she saw Ibrahima.} She did not see him, she CALLED him.'
      (Apel, f.n.)

With little variation with respect to whether the TAM suffix is optional or not in active and middle voice, it is found in all dialects. This verb form is also used for thetic statements:

(17)  Ɓe    war-**u**      lan-ɗo   on.
      3P    kill-PFV1    chief-1   DEF.1
      '{What happened?} They KILLED THE KING.' (Apel, f.n.)

State-of-affairs focus and event-central thetic statements have in common that they cancel the canonical reading (topic-comment) of a categorical sentence by raising the predicate.

(18)  Categorical statement

      [Topic ≈ Subject]  [Comment ≈ Predicate]

(19) State-of-affairs focus/Thetic statement

[Background = Subject] [Focus = Verb]
[Subject             Predicate]
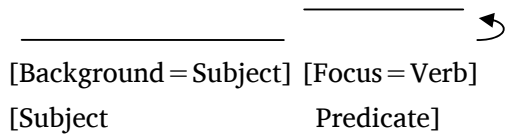
⇒ In the corpus, this verb form, which is widely described elsewhere in the literature, is not found at all. This hints to a very specific use.

## 3.4 Is Left-Dislocation More Frequently Used for Subjects and Right-Dislocation for Objects?

Being an SVO language, one could assume that left-dislocation has the tendency to refer to subject and right-dislocation to objects (the dislocated element always refers to a pronoun in the clause):

(20) a. [min]ᵢ,    mi  ar-i       e     hakkill-e       an       taho
         1S.EMPH  1S  come-PFV2  with  intelligence-3  1S.POSS  also
         'Me, I came also with my intelligence.' (Text A-22)

     b. ɓe  land-ii = lan,       min       tolaajon
        3P  ask-MIDD.PFV2 = 1S  1S.EMPH   last.born.1:DEF.1
        'He (my father) asked me, me the last born.' (Text A-22)

- Dislocation occurs more often in dialogues (8%) than in monologues (4%)
- From 52 clauses, 83%  refer to a topic-comment structure (above), 11% to an imperative (ex. 21) and 6% to a cleft construction (ex. 22)

(21) **an**       kadi  [sel-**u**        on       samakala  ɗon]ₗₘₚₑᵣₐₜᵢᵥₑ
     **2S.EMPH**  also  leave-**IPFV1:2S**  DEM.1   joke.1    there
     'You too, leave this joke there!' (Text L-38)

(22) **rew-ɓe**    **ɓen**  [ko    non       ɓe  wa'-i]꜀ₗₑfₜ
     woman-2    DEF.2  T.FOC  like.that  3P  be.like-PFV2
     'The women, it is like that they are.' (Text M-49)

- The data show 43 clauses with a dislocation in a topic-comment structure: (90% left-dislocation and 10% right-dislocation)
- Right-dislocation in a topic-comment structure is not attested in dialogues at all

- From 39 clauses with left-dislocation in a topic-comment structure, 85% refer to a subject, 15% refer to an object in the clause

- From 4 clauses with right-dislocation in a topic-comment structure, 50% refer to a subject and 50% to an object

⇒ The assumption holds for left-dislocation, but not for right-dislocation.

## 3.5 *What is the Relationship Between Narratives and Subordinate Clauses?*

In Pular, like in other West African Niger-Congo and Chadic (<Afro-Asiatic) languages[4] (Bearth 1993, Frajzyngier 2004), one and the same verb form (here: the Perfective 2) has two main uses in asserted sentences: narration and subordination.

- From 360 clauses containing the Perfective 2 verb form, 73% are narratives and 27% are subordinate clauses

In **narratives** the predicate is part of the **salient** information (example 23). This is not the case in **subordinate** clauses (relative, temporal, out-of-focus) which usually serve as background (thus less salient) information (example 24):

(23) a.      O    yah-**i**     ka    kuri.
            3S   go-PFV2   LOC   kitchen.1
            'She went to the kitchen.'

     b.      O    sayƴit-**i**        taasi-hoy koy.
            3S   rinse:REP-PFV2   cup-22    DEF.22
            'She rinsed the cups.'

     c.      O    wowl-**i**     ka    *téléphone.*
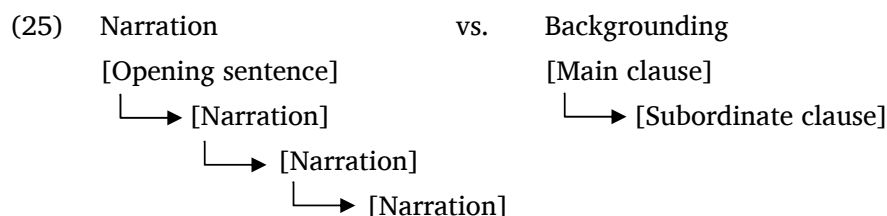            3S   talk-PFV2   LOC   telephone.1
            'She talked on the telephone.' (Text G-15/16/17)

(24) a.      Keku-wal   ngal   (...)    find-in-i           pay-kun
            turtle-11    DEF.11      wake.up-CAUS-PFV2   child-21

            [moyƴ-**u**(<i)=kun]       kun.
            be.good-PFV2=21.REL    DEF.21
            'The turtle woke up the child who was good.' (Text E-220/222/223)

     b.      Ko      an          [jog-it-**i**       *montre*    on].
            T.FOC   2S.EMPH   hold-REP-PFV2   watch.1    DEF.1
            'It is you (who) has the watch.' (Text L-86)

---

[4]   E.g. Tura (Mande<Niger-Congo), Wobe (Kru<Niger-Congo) or Hausa (Chadic<Afro-Asiatic).

Bearth (1993) explains that narratives are also kind of dependent from the opening sentence (= stage setting). As the sentences that develop the storyline represent a temporal sequence, they are pragmatically dependent. In fact, in monologues, stage setting before the "real story line" is often realized in other verb forms or constructions, e.g. a stative construction, with Perfective 3, an identificational or focus construction (see Section 3.2).

(25)  Narration                vs.    Backgrounding
      [Opening sentence]              [Main clause]
         └──→ [Narration]               └──→ [Subordinate clause]
               └──→ [Narration]
                     └──→ [Narration]

In Fula, the setting is not always expressed overtly. A story can also start with the Perfective 2 (cf. the folktales published by Salvaing 1985). This may be an innovation.

⇒ The relation between narration and subordination can be established by dependency.

# 4    Summary

1) Clefts are not only used in main clauses, but also as complement clauses and in adverbial clauses

2) Thetic statements do not always morphosyntactially mark the cancellation of the bipartiteness in a canonical sentence

3) Perfective 1 is not found at all in the corpus, which hints at its specific and restricted use (at least in Pular)

4) Left-dislocation refers indeed mostly to subjects in the topic-comment structure, whereas right-dislocation refers to subjects and objects alike

5) The verb form in narratives and background clauses have in common that they are dependent, the narrative is pragmatically dependent on the opening sentence, the subordinate clause is syntactically dependent on the main clause

It's important to have natural discourse AND elicitation for an analysis.

# 5    Abbreviations

| | | | | | |
|---|---|---|---|---|---|
| BG | background clause | DEM | demonstrative article | i | co-reference |
| CAUS | causative suffix | | | ID | identificational marker |
| COP | copula | EMPH | emphatic pronoun | | |
| DEF | definite article | f.n. | field note | INF | infinitive |
| | | FOC | focus | IPFV | imperfective |

| | | | | | |
|---|---|---|---|---|---|
| LOC | locative | POSS | possessive pronoun | REP | repetitive suffix |
| MIDD | middle voice | PP | prepositional phrase | S | singular |
| NEG | negation | PRET | preterite | STAT | stative |
| P | plural | PROG | progressive | T | term |
| PFV | perfective | Q | question particle | | |
| PN | proper name | REL | relative pronoun | | |

# 6    References

Anyanwu, Rose-Juliet & Abdourahmane Diallo. 2007. Rhythmic units and stress: Word accent mobility in Fula. *Frankfurter Afrikanistische Blätter* 17(5). 65-89.

Bao, Mingzhen. 2012. *Prosody of Focus among Information Structures in Pulaar*. Afrikanistik online, Vol. 2012. http://www.afrikanistik-aegyptologie-online.de/archiv/2012/ 3528/. (23 June, 2014.)

Bearth, Thomas. 1993. Satztyp und Situation in einigen Sprachen Westafrikas. In Möhlig, Wilhelm J.G., Siegmund Brauner & Hermann Jungraithmayr*, IX. Afrikanistentag: Beiträge zur afrikanischen Sprach- und Literaturwissenschaft*, 91-104. Köln: Rüdiger Köppe.

Dijk, Teun A. van. 1977. Sentence topic and discourse topic. *Papers in Slavic Philology* 1. 49-61.

Dik, Simon C. 1997. *The theory of functional grammar, Part 1: The structure of the clause*. Berlin: Mouton de Gruyter.

Fiedler, Ines. in press. Ex situ and in situ focus in Kwa - A text-based study on Anii. *Afrika und Übersee*.

Frajzyngier, Zygmunt. 2004. Tense and aspect as coding means for information structure: A potential areal feature. *The Journal of West African Languages* XXX(2). 53-67.

Güldemann, Tom. 2009. *Prädikatszentrierte Fokustypen: Eine sample-basierte typologische Studie zu afrikanischen Sprachen,* Allgemeine Angaben zum Teilprojekt B7.

Güldemann, Tom. 2010. The relation between focus and theticity in the Tuu family. In Fiedler, Ines and Anne Schwarz (eds.), The expression of information structure: a documentation of its diversity across Africa. *Typological Studies in Language* 91, 69-93. Amsterdam: John Benjamins.

Harrison, Annette. 2003. *Fulfulde language family report,* SIL Electronic Survey Reports 2003-009. http://www.sil.org/silesr/2003/silesr2003-009.html. (23 June, 2014.)

Lambrecht, Knud. 1994. Information structure and sentence form: topic, focus and the mental representations of discourse referents. *Cambridge Studies in Linguistics* 72. Cambridge: Cambridge University Press.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International. http://www.ethnologue.com. (12 June, 2014.)

Salvaing, Bernard. 1985. *Contes et récits peuls du Fouta Djalon*, Paris: Conseil international de la langue française.

Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25. 511-580.

Skopeteas, Stavros, Fiedler, Ines, Hellmuth, Sam, Schwarz, Anne, Stoel, Ruben, Fanselow, Gisbert, Féry, Caroline and Krifka, Manfred. 2006. *Questionnaire on information structure (QUIS)*. Interdisciplinary Studies on Information Structure 4. Working Papers of the SFB 632. Potsdam: University of Potsdam.