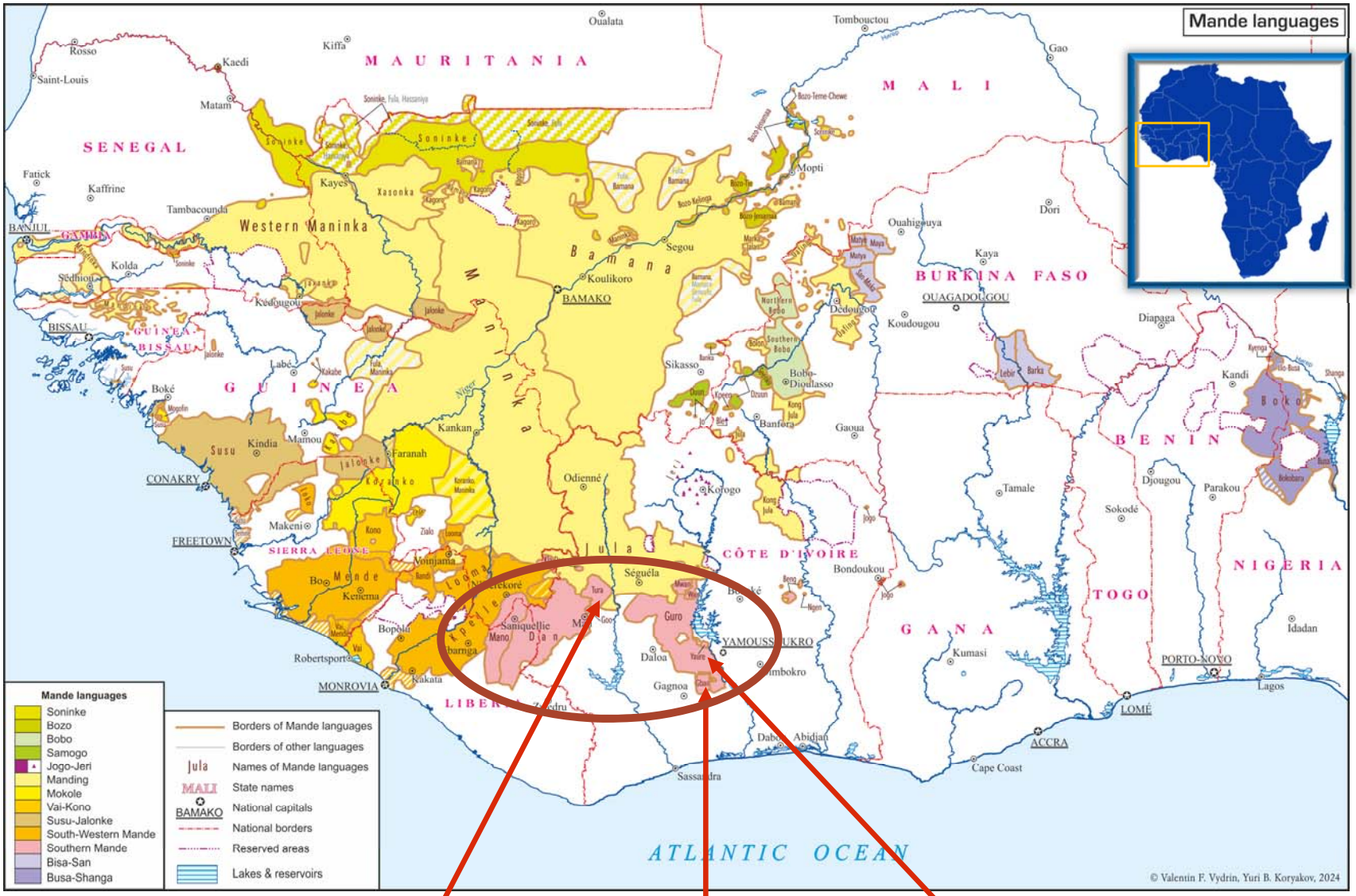
STRANGER THINGS OF WORD (DIS)INTEGRITY IN SOUTHERN MANDE

Dmitry Idiatov

LLACAN (CNRS – INALCO – EPHE)

dmitry.idiatov@cnrs.fr

SOUTHERN MANDE



Tura Gban Yaure

VIOLATION OF WORD INTEGRITY IN SOUTHERN MANDE

- In several Southern Mande languages, the integrity of words can be violated by inserting other bona fide words inside them.
  - Tura (Bearth 1971:171-176; Idiatov 2005, 2008)
  - Gban (Fedotov 2014)
  - Yaure (Kushnir 2016)
- 2 types of word structures whose integrity can be violated:
  - complex words: $[\text{Root} + \text{Suffix}]_{\text{Word1}} > [\text{Root}]_{\text{Word1}} + \text{Word}_2 + [\text{Suffix}]_{\text{Word1}}$
  - simplex words: $[\text{Root}]_{\text{Word1}} > [\text{Root}_{\text{Part1}}]_{\text{Word1}} + \text{Word}_2 + [\text{Root}_{\text{Part2}}]_{\text{Word1}}$
- Phonological restriction: only words of the shape C…C… can be subject to violation of word integrity

✋ At least in Tura and Gban, the concept of syllable is not relevant in the phonology (compare Hyman 1983, 1985, 2011, 2015 on a comparable situation in Gokana)

INTEGRITY: MORPHOLOGY OF THE BASE

- The morphologically **complex** verb ***dɔ́-dɔ̋*** 'stop' in Tura (from *dɔ́* 'stand' plus the deobliquative suffix *-dɑ̋*)

(1) Ɔ̀ɔ́ dɔ́ dɛ̋ɛ́ ké dɔ̋-ɔ́

3SG.NEG.PFV stand$_{P1}$ new a.certain DEOBL$_{P2}$\PFV-PFV

'He did not stop again.' (Idiatov 2008:162)

- The morphologically **simplex** verb ***gbɑ́dɑ̋*** 'thunder' in Tura:

(2) Dá=á gbɑ́ dɛ̋ɛ́ ké dɑ̋-á

rain=AUX thunder$_{P1}$ new a.certain thunder$_{P2}$\PFV-PFV

'It thundered again.' (Idiatov 2008:164)

- The morphologically **simplex** numeral ***pɪ̀ɪ̀dɛ̋*** 'two' in Tura:

(3) Wàá pɪ̀ɪ̀ dɛ̀-fɪ̀ɪ̀ dɛ́

3SG.NEG.COP two$_{P1}$ RESTR-empty two$_{P2}$

'They are not even two/ they are not two at all [but just one].' (Idiatov 2005:32)

INTEGRITY: MORPHOSYNTAX OF THE SPLIT FORM

(1) Ồó dɔ́ dɛ́ɛ́ ké dɔ̌-ɔ́

3SG.NEG.PFV stand$_{P1}$ new a.certain DEOBL$_{P2}$\PFV-PFV

'He did not stop again.' (Idiatov 2008:162)

(2) Dá=á gbā́ dɛ́ɛ́ ké dā́-á

rain=AUX thunder$_{P1}$ new a.certain thunder$_{P2}$\PFV-PFV

'It thundered again.' (Idiatov 2008:164)

- Part2 behaves as a **regular verb** in taking the TAM morphology: tone change + suffix
- Within the rigid SOVX framework of the Mande clausal syntax, Part1 behaves as the **nominal Object** of Part2.
- Additional evidence for its nominal status is provided by the fact that it takes **adnominal modifiers**, such as adjectives (not adverbs), focus markers, plural markers, etc., that neither the base form, nor Part2 can take.

INTEGRITY: MORPHOSYNTAX OF THE SPLIT FORM

- Part2 functions as the **syntactic head** governing Part1 of the original word, and this irrespective of whether it is an **actual morpheme itself** (2) or just **a meaningless part of the word**
- Idiatov (2005) refers to these elements as ***quasi-words*** (or ***pseudo-words***)
- This is **typologically highly unusual** and different from:
  - endoclisis (Harris 2002)
  - syllabemes in languages like Vietnamese (cf. Nhàn 1984, Bickel et al. 2007) in those cases when polysyllabic words happen to be split by other words

    cà phê 'coffee' + *với* 'and' vs. *cà với phê* 'coffee and the like'
  - expletive insertion in English, recently argued by Zingler (2024) to be an example of discontinuous compounding

    bloody + *absolutely* vs. *abso-bloody-lutely*

INTEGRITY: CATEGORY OF THE BASE

- Tura
  - 19 verbs derived with the deobliquative suffix *-dɑ̃́*, such as:

    dɔ́-dɔ̃́ ‘stop; wait’ < dɔ́ ‘stand; stop; wait’
    gɩ̃̀ɩ̃-nɑ̃́ ‘roll (up)’ < gɩ̃̀nɩ̃̀ ‘roll’
    gbíé-dɑ̃́ ‘drag, trail’ < gbíé ‘pull, draw’
    yà-dɑ̃́ ‘sit down’ < yàà ‘sit down (somewhere)’
    púó-dɑ̃́ ‘turn, coil, wring’, sèè-dɑ̃́ ‘move; turn’, *yùù-dɑ̃́* ‘crash down’…

  - 5 underived verbs that however match the deobliquative derivation formally

    gɑ́dɑ́ ‘change, transform’, gbɑ́dɑ́ ‘thunder’, sɑ́dɑ́ ‘sweep’, tɑ̃́nɑ̃́ ‘lean’, wɑ́dɑ́ ‘collapse’

  - 3 underived verbs that do not match the deobliquative derivation formally

    kɔ́dɔ́ ‘change, transform’, tɑ́dɑ́ ‘become black, blacken’, zɑ̀nɑ̀ ‘wake up’

  - numerals (except the CV(V) forms *dó* ‘1’ and *bùù* ’10’)

    pɩ̀ɩ̀dɛ́ ‘2’ sɑ̂à dó ‘6’
    yàká ‘3’ sɑ̂à pɩ̀ɩ̀dɛ́ ‘7’
    yɩ̀sɛ́ ‘4’ sɑ̂àká ‘8’
    sʊ́dʊ́ (~ arhaic: sʊ́ʊ́dʊ́) ‘5’ sɔ̂ɩ̀sɛ́ ‘9’

INTEGRITY: CATEGORY OF THE BASE

- Gban
  - any C…C… obliques: SOV**X**
- Yaure
  - locative nominal C…C… obliques (except non-integrated foreign toponyms)

(4) Ã́ tà Áɓî lɛ́ jà̀

1SG.PFV go\PFV Abijan$_{P1}$ CONTR Abijan$_{P2}$

'As for Abijan, I went there.' (Kushnir 2016:111)

INTEGRITY: INSERTED MATERIAL

- Tura
  - any adnominal modifier that matches semantically: adjectives, various focus markers, (especially with numerals) restrictors
- Gban
  - focus marker *lì*
- Yaure
  - contrastive topic marker *lé́*

INTEGRITY: SYNTACTIC RESTRICTRIONS ON INSERTION

Syntactic restriction: The violation of word integrity is only possible in a limited number of syntactic positions.

- Gban & Yaure
  - Obliques: SOVX
- Tura
  - (both verbs and numerals) predicate
  - (numerals) transposed oblique

TRANSPOSITION CONSTRUCTIONS

- As I argued in Idiatov (2005, 2008) for Tura, the transposition constructions served as a **model for the reanalysis** that eventually resulted in the observed cases of word integrity violation
  - In addition, the **word phonotactic patterns** of Tura have played a crucial role in this reanalysis scenario
- Tura has 2 basic types of transposition constructions:
  - Verbal transposition construction: [V_{NMLZ} + (verb) *wó* 'accomplish']
  - Oblique transposition construction: [X_{NMLZ} + (postposition) *wó*]
- The transposed element is nominalized by conversion and becomes a complement of the transposer, viz. the verb *wó* 'accomplish' or the postposition *wó* (derived from the verb)
- The function of the transposition constructions is to give the transposed element combinatorial possibilities of a noun: [N ADJ], [N PL], [N **FOC**]

SPLIT TRANSPOSITION CONSTRUCTIONS

(5) Ò**ó** **dɔ́-dɔ̋** **dɛ́ɛ́** **ké** **wó**-ó

3SG.NEG.PFV stand-DEOBL new a.certain TR\PFV-PFV

'He did not stop again.' (Idiatov 2008:162)

(1) Ò**ó** **dɔ́** **dɛ́ɛ́** **ké** **dɔ̋**-ɔ́

3SG.NEG.PFV stand_{P1} new a.certain DEOBL_{P2}\PFV-PFV

'He did not stop again.' (Idiatov 2008:162)

☞ The construction in (1) where the integrity of the verb is violated is equally a Verbal transposition construction, that we can refer to as the **Verbal split transposition construction**.

☞ Similarly, the construction in (3) where the integrity of the numeral is violated is equally a Oblique transposition construction, that we can refer to as the **Oblique split transposition construction**.

LINGUISTIC UNITS AS STATISTICAL ABSTRACTION

- Constructive vs **abstractive** perspective on the morphosyntax (cf. Geertzen et al. 2016):
  - "words and other linguistic units are not independent components from which larger expressions are 'constructed' but are, instead, abstracted from larger utterances"
  - "the units that 'emerge' in different languages are abstracted on the basis of recurrent statistical patterns, specifically patterns of syntagmatic and paradigmatic interpredictability"
  - "word boundaries are the most informative boundary type [as compared to morpheme or utterances boundaries]" and "words are optimal-sized units for describing the regularities" of the language structure
- Both the **recurrent presence and absence of linguistic boundaries** are important

TURA: PRIMARY LINGUISTIC BOUNDARIES

- There is never a linguistic boundary in the sequences:
  - non-syllabic C + V
  - V̰ + C̰ + V̰ (the ultra-short realization is possible only when C̰ is [ɗ̰, d̰, l̰, ɾ, n̰], V̰ and their tones are identical)
- There is normally no boundary in the sequences of V1 + V2 when:
  - Both V1 and V2 are both oral or both nasal
  - Tone2 is the same or lower than T1
  - V2 is followed by C or pause (not by V3)
- There is rarely a boundary between V + syllabic N when followed by pause or C.
- There is always a linguistic boundary in the sequences:
  - V + non-syllabic C, where is C is not /ɗ/ or /n/
  - syllabic N + V
  - syllabic N + C

TURA: EMERGING UNITS

- The following **major types of linguistic units** emerge from the recurrent presence or absence of linguistic boundaries in Tura:
  - CV ; (with additional restrictions on the combination vowel qualities and tones) CVV, CV̰C̰V̰ → typical for lexical units (words)
  - CVN̰ → less common shape of lexical units (words)
  - (with additional restrictions on the combination vowel qualities and tones) CVVN̰, CV̰C̰V̰N̰, CVVV → exceptional shape of lexical units (words)
  - V; (with additional restrictions on the combination vowel qualities and tones) VV̰ → typical for functional units
- The functional units can be further subdivided into several types that we can refer to as words, clitics or affixes based on whether or not they are subject to a number phonological process that may span linguistic boundaries (assimilation in quality, nasalization, vowel elision) and the morphosyntactic criteria of selectivity

TIPITS & WIPITS

- These major types of linguistic units reflect the typical phonotactic templates of Tura, that we can refer to as ***tipits***
- Those tipits that are typical for lexical units can be referred to as ***wipits*** (from word + tipit)
- We can continue in the same spirit and speak of:
  - *sipit* < suffix + tipit
  - *aipit* < affix + tipit
  - *fipit* < functional morpheme + tipit
  - *clipit* < clitic + tipit
  - …

- Wipits have shaped **the restrictions on the split transposition of numerals** in Tura observed, but not explained by Idiatov (2005):

pɪ̀ɪ̀dɛ́ '2'	sáà dó '6'
yàká '3'	sáà pɪ̀ɪ̀dɛ́ '7'
yɪ̀sɛ́ '4'	sáàká '8'
sʊ̋dʊ̋ (~ arhaic: sʊ̋ʊ̋dʊ̋) '5'	sɔ̃ɪ̀sɛ́ '9'

- With respect **to the verbal split transposition** in Tura:
  - Wipits can help us reconstruct the specific TAM construction in which the verbal split transposition is most likely to have emerged.

    ☞ **PFV**

    gálá 'thunder' [gálá ~ gárá] is 1 wipit → gálá-á thunder\PFV-PFV [gáráá ~ gáláá ~ gáɗáá] is not 1 wipit

  - Wipits can help us explain why the split transposition could extend both on the verbs that do not contain, and especially on the verbs that cannot contain, the deobliquative suffix, such as *kʊ́dʊ́* 'educate'

WIPITS IN ACTION

- Wipits can explain the general phonological restriction on the violation of word integrity that only words of the shape C…C… can be split
- Most obviously, wipits and tipits in general drive and shape various formal erosion and reduction processes when morphological boundaries erode and when units change their status in the grammar, most noticeably from lexical to functional.

WIPITS ARE NOT…

- Wipits are **phonotactic units** that emerge from the generalizations on **the most informative** linguistic boundaries
- They need **not** correlate with (natural) units of pronunciation, such as syllables, so that we can have C̥V̥C̥V̥(N̥) wipits

  ☹ syllabemes, syllabomorphemes
- They need **not** correlate with (natural) units of rhythm, such as feet, so that:
  - we can have monomoraic CV wipits
  - the bimoraic CVV wipits would not need to behave like other "feet" with respect to the possibility of their split
  - in the heavy C̥V̥C̥V̥(N̥) wipits, both vowels can be reduced similarly

  ☹ (featural) feet
- They need **not** correlate with morphemes, so that we can have wipits that are quasi-words

LANGUAGES WITH VIOLATION OF WORD INTEGRITY

- Languages reported with **the true violation of word integrity** (that is, not including endoclisis of the Udi type) :
  - Southern Mande: Tura, Gban, Yaure
  - Southeast Asian languages described with "syllabemes", such as Vietnamese
  - English with its expletive insertion
- What all these languages appear to **have in common** (to various degrees):
  - their native lexicon is arranged around a limited number of short and structurally simple wipits
  - while they have also acquired a sizeable amount words of longer and more diverse structure
  - limited to no bound morphology
  - rigid word order
- As a result, they have reunited **conditions propitious for the reanalysis** that may lead to a violation of word integrity

  ☞ Compare Zingler's (2024) hypothesis on the source of the expletive insertion in English in a "kind of 'prosodic reanalysis'".