



# Cross-linguistic variation in the noun-to-verb ratio: the role of verb morphology and narrative strategies

Frank Seifart

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany  
frank\_seifart@eva.mpg.de

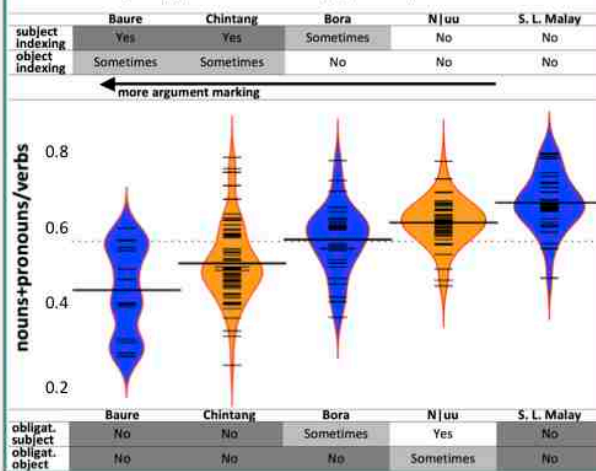
## AIMS & LANGUAGES INVESTIGATED

- Does the ratio of nouns(+pronouns) to verbs in texts depend on
  - The language's system of argument indexing on verbs (study 1)?
  - Narrative strategies, i.e. the position within a text (study 2)?
- Five corpora of spoken language, tagged for parts of speech by experts:
  - Baure (Arawakan): 4,839 words (Danielsen et al. 2009)
  - Bora (Boran): 6,701 words (Seifart 2009)
  - Chintang(Sino-Tibetan): 44,488 words (Bickel et al. 2009)
  - N|juu (Tuu): 31,691 words (Güldemann et al. 2010)
  - Sri Lanka Malay (Austronesian): 12,900 words (Ansaldo et al. 2009)
- Exploration of new methods and some preliminary results



## METHODS STUDY 1: VERB MORPHOLOGY

- Languages coded for typological characteristics
  - presence and extensiveness of argument indexing
  - obligatoriness of argument realization
- Corpora analyzed for the ratio of nouns+pronouns to verbs
  - Standard formula:  $\frac{N(\text{nouns}) + N(\text{pronouns})}{N(\text{nouns}) + N(\text{pronouns}) + N(\text{verbs})}$
  - Analyzed (i) per text: — and (ii) per corpus: —

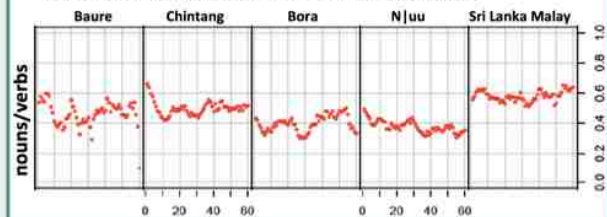


## RESULTS STUDY 1: VERB MORPHOLOGY

- Ratios in different texts follow a “normal” distribution
  - effect of genres negligible
  - except for smallest corpus (Baure)
- Drastic differences, e.g.
  - Chintang ~1:1 (1/1+1=0.5) vs. S. L. Malay ~2:1 (2/2+1=0.667)
- Extensive argument indexing correlates with a low nouns+pronouns to verbs ratio
- Argument indexing on verbs makes argument realization unnecessary, this effect prevails in the overall number of nouns and pronouns
- Obligatory argument realization (+/- pro drop) does not predict the number of nouns-and-pronouns in text well

## METHODS STUDY 2: NARRATIVES

- Ratio of nouns (excluding pronouns) to verbs in narratives
- Progressive windows of five annotation units (window 1: units 1-5, window 2: units 2-6, etc.)
- Means for all narrative texts for each language



## RESULTS STUDY 2: NARRATIVES

- High noun to verb ratio at the beginning, reflecting the introduction of new discourse participants, then decrease
- Sinusoidal alternations as narrative texts unfold, with peaks at about 10-15 annotation units that may reflect the decay of activation of discourse participants

## CONCLUSION

- A promising method to uncover usage patterns linked to typological characteristics and narrative strategies
- Outlook:
  - Include more languages to test for effects of areality
  - Include more genres per language to test for genre effects

## REFERENCES

- Ansaldo, U., L. Lim, S. Nordhoff. 2009. *Sri Lanka Malay Documentation*. Nijmegen: DOBES-MPIP.
- Bickel, B., G. Banjade, T. N. Bhatta, M. Gaenzle, N. P. Paudyal, M. Raj, N. K. Raj, I. P. Raj, S. Stoll. 2009. *Audiovisual Chintang corpus*. Nijmegen: DOBES-MPIP.
- Danielsen, S., F. Riedel, F. Admiraal, L. Terhart. 2009. *Baure Documentation*. Nijmegen: DOBES-MPIP.
- Güldemann, T., A. Witzlack-Makarevich, M. Ernst, S. Siegmund. 2010. *A text documentation of N|juu*. Leipzig, London: MPI-EVA, ELDP.
- Seifart, F. 2009. Bora documentation. *A multimedia documentation of the languages of the People of the Center*, ed. by F. Seifart, D. Fagua, J. Gasché, J. A. Echeverri. Nijmegen: DOBES-MPIP.