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 (or 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): 4,833 words (Danielsen et al. 2009)
  - Bora (Borom): 6,701 words (Seifart 2009)
  - Chintang (Sino-Tibetan): 44,488 words (Bickel et al. 2009)
  - Njua (Tibetan): 31,691 words (Güldemann et al. 2010)
- 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: \[
  \text{Ratio} = \frac{\text{Nouns}}{\text{Nouns} + \text{Pronouns} + \text{Verbs}}
  \]
- Analyzed (i) per text — and (ii) per corpus:
  - Analyzed (i) per text
  - Analyzed (ii) per corpus

RESULTS STUDY 1: VERB MORPHOLOGY

- Ratios in different texts follow a "normal" distribution
- Effect of genre negligible
- Except for smallest corpus (Baure)
- Drastic differences, e.g.,
  - Chintang \( \approx 1.1 \) (\(1/1+1=0.5\)) vs. S. L. Malay \( \approx 2.1 \) (\(2/2+1=0.667\))
- 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 6-10, 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