The Semantics of Split-Intransitive Alignment Systems: A Multidimensional Scaling Analysis

Jens Van Gysel – University of New Mexico
Syntax of the World’s Languages 8, 3-5 September 2018, Paris
jelvangysel@unm.edu
Introduction: split-intransitive alignment

Previous cross-linguistic research

Data and methodology: Multidimensional Scaling

Results and discussion

Further research
Introduction: Split-Intransitive Alignment

- Split-intransitive alignment (also known as active-inactive, split-S)
  - (Usually) two large classes of intransitive verbs, whose arguments are marked in different ways
  - One class usually has the same marking as A arguments, the other as O arguments
- Example: Tupí-Guaranían languages (e.g. Tupinambá, Jensen 1990: 117):
  - a-só
  - 1sg-go
  - ‘I go’
  - a-i-nupã
  - 1sg-3sg-hit
  - ‘I hit it’
Introduction: Split-Intransitive Alignment

- Split-intransitive alignment (also known as active-inactive, split-S)
  - (Usually) two large classes of intransitive verbs, whose arguments are marked in different ways
  - One class usually has the same marking as A arguments, the other as O arguments
- Example: Tupí-Guaranían languages (e.g. Tupinambá, Jensen 1990: 117):
  - syé katú
  - 1sg be.good
  - ‘I am good’
  - syé nupã
  - 1sg hit
  - ‘He/she/it/they hit me’
Introduction: Split-Intransitive Alignment

- Within functional frameworks: essentially a semantic phenomenon
  - Van Valin (1990): splits based on Aktionsart (e.g. Georgian) or on agentivity (e.g. Acehnese)
  - Many studies on factors conditioning the split in individual languages/families (e.g. Klamer 2008)

- Lexicalisation and grammaticalisation can obscure semantic motivation (Mithun 1991)
Introduction: Split-Intransitive Alignment

- Even considering language-specific “weighting” of factors and idiosyncrasies: significant systematicities have been found

- This work builds on cross-linguistic accounts by Mithun (1991) and Croft (1998, 2012)

- Use Multidimensional Scaling to give a preliminary quantitative account of the factors cross-linguistically conditioning split-intransitive alignment
Previous Cross-Linguistic Research

- Mithun (1991): sample of five languages: Paraguayan Guaraní, Lakhota, Caddo, Central Pomo, Mohawk
  - Guaraní: Aktionsart is main conditioning factor

- Events (achievements, accomplishments, activities, see Vendler 1967) take Sa marking
  - ‘get up’, ‘walk’, ‘fire a gun’, ‘work’, ‘fall’, ‘die’

- States take So marking
  - ‘be sick’, ‘be sleepy’, ‘be short’, ‘have a cramp’
Previous Cross-Linguistic Research

- Mithun (1991): sample of five languages: Paraguayan Guaraní, Lakhota, Caddo, Central Pomo, Mohawk
  - North American languages in the sample: mostly force-dynamic factors
  - Lakhota: agency, instigation
    - ‘be patient’, ‘walk’ vs. ‘die’, ‘be cold’
  - Central Pomo (and Caddo): control + affectedness
    - ‘sneeze’ would be Sa in Lakhota, but So in Pomo

=> interaction of Aktionsart, instigation, control and affectedness, which have a different weight in different languages
Croft (1998, 2012): same sample as Mithun + Tsova-Tush (Caucasian, Georgia)

- MDS analysis => one-dimensional model based on only causal factors explains the data well:
  - controlled activities
  - inactive actions
  - inherent properties/dispositions
  - bodily actions
  - inchoatives
  - uncontrolled activities/transitory states

More Sa-like
More So-like
Data and Methodology

- The samples of aforementioned studies are small and rather concentrated geographically
  - E.g. importance of causal factors in Croft (2012) is not surprising, since for most of the sample, Mithun (1991) already shows this

- This preliminary study: six languages which have been described as showing active-inactive/split-intransitive alignment, geographically and genetically balanced
Data and Methodology

- Twenty languages were sampled, based on the WALS data for active-inactive alignment (Comrie 2013; Siewierska 2013), and other references in the literature.

- Six were chosen, based on maximal geographical and genetic distance, and availability of sources, for this initial study.
## Data and Methodology

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Region</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acehnese</td>
<td>Chamic (Austronesian)</td>
<td>Indonesia (Sumatra)</td>
<td>Durie (1985)</td>
</tr>
<tr>
<td>Beria</td>
<td>Saharan</td>
<td>Chad, Sudan</td>
<td>Jakobi &amp; Crass (2004); Jakobi (2011)</td>
</tr>
<tr>
<td>Creek</td>
<td>Muskogean</td>
<td>USA (Oklahoma, Florida)</td>
<td>Martin &amp; Mauldin (2000)</td>
</tr>
<tr>
<td>Pilagá</td>
<td>Guaykuruuan</td>
<td>Argentina (Formosa)</td>
<td>Vidal (2001)</td>
</tr>
<tr>
<td>Rotokas</td>
<td>East Papuan</td>
<td>Papua New Guinea (Bougainville)</td>
<td>Robinson (2011)</td>
</tr>
<tr>
<td>Tsova-Tush</td>
<td>Caucasian</td>
<td>Georgia</td>
<td>Holisky (1987)</td>
</tr>
</tbody>
</table>
Data and Methodology

- For each language, the type of S-marking with as many intransitive predicates as possible was coded.
- Twenty-four were predicates were then chosen for analysis based on:
  - Their semantic characteristics
  - Number of languages for which data for this predicate was found
## Data and Methodology

<table>
<thead>
<tr>
<th>Predicates studied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get up</td>
</tr>
<tr>
<td>Enter</td>
</tr>
<tr>
<td>Work</td>
</tr>
<tr>
<td>Be in a lying position</td>
</tr>
<tr>
<td>Be in a standing position</td>
</tr>
<tr>
<td>Be in a sitting position</td>
</tr>
<tr>
<td>Spit</td>
</tr>
<tr>
<td>Vomit</td>
</tr>
<tr>
<td>Breathe</td>
</tr>
<tr>
<td>Sneeze</td>
</tr>
<tr>
<td>Die</td>
</tr>
<tr>
<td>Fall</td>
</tr>
<tr>
<td>Spill</td>
</tr>
<tr>
<td>be bad</td>
</tr>
<tr>
<td>be quiet</td>
</tr>
<tr>
<td>be good</td>
</tr>
<tr>
<td>be short</td>
</tr>
<tr>
<td>be tall</td>
</tr>
<tr>
<td>Be big</td>
</tr>
<tr>
<td>Dry up</td>
</tr>
<tr>
<td>Get startled</td>
</tr>
<tr>
<td>Be cold</td>
</tr>
<tr>
<td>Be hungry</td>
</tr>
<tr>
<td>Be hot</td>
</tr>
</tbody>
</table>
Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)
  - A methodology for measuring (dis)similarity between entities
  - Dissimilarities between entities are represented as distances in geometric space
    => In this case, the further away from each other two predicates are in the plot, the bigger the difference between them with regards to argument marking
    => A mathematical implementation of the semantic map methodology (Croft & Poole 2008)
Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)

<table>
<thead>
<tr>
<th></th>
<th>Creek Set I</th>
<th>Creek Set II</th>
<th>Creek Dative S</th>
<th>Tsova-Tush NOM</th>
<th>Tsova-Tush ERG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Be lying</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Be standing</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sit</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Spit</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>
Data and Methodology

- Multidimensional scaling (see Poole 2005 for the maths)
  - For each construction, a cutting point, line, or plane is fitted in space so that it divides the predicates into yea vs nay groups with as good a fit as possible
    - 23 constructions
    - 24 predicates
  - W-NOMINATE algorithm (Poole & Rosenthal 1985), implemented in a linguistics-friendly format by Timm (2017) for R (R Core Team 2018), calculates an ideal location for every point (= predicate) in space
    - I.e. where it is on the correct side of as many cutting points/lines as possible
Data and Methodology

- Dimensionality of the model?

<table>
<thead>
<tr>
<th></th>
<th>Correct classification</th>
<th>Average proportional reduction of error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>86,1%</td>
<td>0,46</td>
</tr>
<tr>
<td>2D</td>
<td>95,1%</td>
<td>0,81</td>
</tr>
<tr>
<td>3D</td>
<td>96,3%</td>
<td>0,86</td>
</tr>
</tbody>
</table>

- A 2D model provides a significant increase in correct classification, the increase from 2D to 3D does not warrant the use of extra statistical power
Results and Discussion
Results and Discussion

- The first dimension seems to correspond to causal factors:
  - On the left: non-agentive predicates, S is not in control
    - Non-controlled states, both inherent and transitory
    - S argument is significantly affected
Results and Discussion
Results and Discussion

- The first dimension seems to correspond to causal factors:
  - On the right: agentive predicates, S is in control
    - Controlled activities
    - Controlled states: inactive actions, ‘be quiet’
Results and Discussion
Results and Discussion

- The first dimension seems to correspond to causal factors:
  - In the middle: predicates where S can be either volitional or non-volitional
    - Uncontrolled activities
    - Bodily actions
  - Predicates with typically inanimate S, so less affected
    - ‘spill’, ‘dry up’
Results and Discussion
The first dimension seems to correspond to causal factors:

Unexpected: inchoative-like predicate ‘get startled’ is found in the middle, even though the S is not in control and is affected
Results and Discussion
Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - Higher up: directed predicates, result in a change of state
    - Directed achievements
    - Directed accomplishments
Results and Discussion
The second dimension seems to correspond to aspectual factors:
- Lower down: predicates that do not result in a change of state
  - Cyclic achievements
  - Undirected activity
Results and Discussion
Results and Discussion
Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - In the middle
    - Transitory states
    - Permanent states
    - Inactive actions (can be construed as transitory states, see Croft 1998)
  - Does not seem to correspond to Mithun’s (1991) dichotomy between event-like Sa predicates and state-like So predicates
  - Croft (2012: 170-1): many states can be analysed as having the same base as a directed achievement, but with a different profile
Results and Discussion

- The second dimension seems to correspond to aspectual factors:
  - In the middle: temporary and permanent states

From Croft (2008)
The second dimension seems to correspond to aspectual factors:

- Unexpected: ‘get up’ and ‘get startled’ would be expected higher up with the directed predicates.
Results and Discussion
Conclusions and Further Research

- MDS plot indicates general cross-linguistic trends in the semantics of agentive-inagentive alignment systems
  - Quite clear aspectual and force-dynamic dimensions
  - Directed predicates with controlling S => more prototypically Sa
  - Note: Aktionsart and agentivity are not always independent

- More geographically balanced sample shows the presence of aspectual factors, next to the causal factors found by Croft (2012)
- However, the most important aspectual factor seems to be directedness, rather than eventhood (as proposed by Mithun 1991)
Conclusions and Further Research

- Croft (2012: 166): “the general effect of MDS is to detect broad patterns in complex and messy data”
  - We always expect noise, especially with relatively small datasets

- To reduce noise: increase number of languages under study
  - More datapoints for the predicates studied here improves their location
  - More cutting lines allows the inclusion of more predicates in order to find subtler distinctions

- Bayesian MDS models are not yet very widespread (see e.g. Okada & Shigematsu 2009) but could improve the model
  - Provide more accurate estimates of uncertainty in the data
  - Gives a direct indication of optimal dimensionality
References


References


References


