Launch of BivalTyp

Typological database of bivalent verbs and their encoding frames (www.bivaltyp.info)

Sergey Say
Institute for Linguistic Studies, RAS
serjozhka@yahoo.com

Dmitry Nikolaev
Stockholm University
mail@dnikolaev.com
Preliminaries

Prominent role of valency for linguistic typology

- transitivity
- alignment
- voice and related phenomena: passive, reflexive, ...
Preliminaries

Typology is mainly focused on major clause types

- monovalent: ‘sleep’, ‘run’, ...
- transitive: ‘kill’, ‘break’, ...
- ditransitive: ‘give’, ...
Preliminaries

- All (?) languages have minor (a.k.a. non-canonical) valency patterns
- (Until recently) underrepresented in typological research
- Goal: to fill this gap for bivalent verbs
Preliminaries

Why bivalent verbs?

- they are especially prone to show deviant valency behaviour [Bickel et al. 2014]

(1) *The boy looked at the clouds*

(2) *Мне нравится эта рубашка*

- they often form relatively large classes, unlike non-canonical trivalent verbs
Project: goals

- Which factors determine valency class assignment in individual languages?
- To what extent are valency classes similar across languages?
- What is the role of genealogical and areal factors?
Project: major design features

- First-hand data
  - St. Petersburg-style typology

- Questionnaire with 130 verbs given in context
Project: major design features

#21 (Peter was crossing the river in a boat)

‘Peter reached the bank’

#22 (The wall was covered with fresh paint)

‘Peter touched the wall’ (and got dirty)

=> Two pre-defined arguments (X, Y) for each predicate
Project: major design features

• The valency of a verb = “the list of its arguments with their coding properties”

• Coding properties
  – flagging (cases & adpositions)
  – indexing (agreement, cross-referencing)
  – word order (rarely)
Project: major design features

• Problem: coding devices (e.g. ‘Instrumental case’, ‘post-verbal agreement slot’, etc.) are language specific

• How to typologize behavior of verbs like ‘be afraid’, ‘follow’, ‘listen’, ‘touch’, etc.?

• Solution: use the lexical lists as a tertium comparationis = set partition variable
## Eastern Armenian

<table>
<thead>
<tr>
<th>#</th>
<th>Predicate</th>
<th>Translation</th>
<th>Valency Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>reach</td>
<td>Petros-ə hasav ap'-i-n</td>
<td>NOM_DAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Petros[NOM]-DEF reach:AOR:3SG bank-DAT-DEF</td>
<td>NOM_DAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘P. reached the bank’</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>touch</td>
<td>Petros-ə dipav pat-i-n</td>
<td>NOM_DAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Petros[NOM]-DEF touch:AOR:3SG wall-DAT-DEF</td>
<td>NOM_DAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘Petros touched the wall’</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>attack</td>
<td>Arj-ə harjakvec' jknors-i vra</td>
<td>NOM_DATvra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bear[NOM]-DEF attack:AOR:3SG fisherman-DAT on</td>
<td>NOM_DATvra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘A bear attacked a fisherman’</td>
<td></td>
</tr>
</tbody>
</table>

=> Eastern Armenian equivalents of ‘reach’ and ‘touch’ belong to the same class; the equivalent of ‘attack’ is different
Project: a brief history

- 2010: first version of the questionnaire
- Laboratory for the typological study of languages, Institute for Linguistic Studies, RAS
- Team members
  - Sergey Dmitrenko
  - Dmitry Gerasimov
  - Maria Kholodilova
  - Viktor Khrakovskij
  - Elena Kordi
  - Olga Kuznecova
  - Daria Mischenko
  - Arseniy Vydrin
  - Natalia Zaika
Project: a brief history

- 2009-2013: two consecutive grants from the Russian Foundation for Humanities

- 2010-present: collection and annotation of data contributed by a large group of supportive language experts without whose help the project would have never got off the ground

- Further contributions are very welcome!
Project: a brief history

- A big big THANK YOU to language experts

Project: a brief history

Sample size

Year

Languages

20 40 60 80

2012 2014 2016 2018 2020
Project: a brief history

- 2018: edited volume (in Russian)
Project: a brief history

- 2020: Building the web-site

https://www.bivaltyp.info/

- All credit for this phase goes to Dmitry Nikolaev
Intermission

a virtual tour of BivalTyp [www.bivaltyp.info]
Potential applications

- transitivity ratio of verbs
- (dis)similarity distances between verbs
- typologically informed analysis of language-specific valency class systems
- transitivity profiles for languages
- (dis)similarity distances between languages
- comparison with genealogical and areal data
- comparison with structural data: case, WO, etc.
- predictability of valency patterns
- and many more
Transitivity ratio of verbs

- Tsunoda’s implicational transitivity hierarchy
  1a) direct effect (kill, break subtype) >
  1b) direct effect (hit, shoot subtype) >
  2a) perception (see subtype) >
  2b) perception (look subtype) >
  3) pursuit (search, wait) >
  4) knowledge (know, understand, remember, forget) >
  5) feeling (love, like, want, need) >
  6) possession (have) [Tsunoda 1981, modified in 1985].

- Only partially supported by the data from BivalTyp, see next slide
## Transitivity ratio of verbs

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a) direct effect</td>
<td>break</td>
<td>1,00</td>
</tr>
<tr>
<td>1b) direct effect</td>
<td>hit</td>
<td>0,77</td>
</tr>
<tr>
<td></td>
<td>shoot</td>
<td>0,25</td>
</tr>
<tr>
<td>2a) perception</td>
<td>see</td>
<td>0,88</td>
</tr>
<tr>
<td></td>
<td>hear</td>
<td>0,86</td>
</tr>
<tr>
<td>2b) perception</td>
<td>look</td>
<td>0,30</td>
</tr>
<tr>
<td></td>
<td>listen</td>
<td>0,67</td>
</tr>
<tr>
<td>3) pursuit</td>
<td>search</td>
<td>0,81</td>
</tr>
<tr>
<td></td>
<td>wait</td>
<td>0,65</td>
</tr>
<tr>
<td>4) knowledge</td>
<td>know</td>
<td>0,88</td>
</tr>
<tr>
<td></td>
<td>understand</td>
<td>0,84</td>
</tr>
<tr>
<td></td>
<td>remember</td>
<td>0,71</td>
</tr>
<tr>
<td></td>
<td>forget</td>
<td>0,41</td>
</tr>
<tr>
<td>5) feeling</td>
<td>love</td>
<td>0,76</td>
</tr>
<tr>
<td></td>
<td>like</td>
<td>0,22</td>
</tr>
<tr>
<td></td>
<td>need</td>
<td>0,25</td>
</tr>
<tr>
<td>6) possession</td>
<td>have</td>
<td>0,40</td>
</tr>
</tbody>
</table>
(Dis)similarity distances between verbs

- Which verbs tend to pattern together?

- Distance metric $D(V_1, V_2) =$ the ratio of languages where the two verbs $V_1$ and $V_2$ belong to different classes

- NeighborNet visualization of the distance matrix
74 predicates favoring intransitivity; intransitive patterns only; 37 languages

“Motion from”
- be_ashamed, forfeit, go_out, dismount, avoid, be_afraid, be_squeamish, fall_behind, be_different, depend
- become_upset, get_stuck, reach, shoot_at, hit@target@

“Attained goal”
- touch, get_stuck, reach, shoot_at, hit@target@

Interaction
- help, answer, flatter, trust, believe, be_similar, envy, obey
- lose_to_follow, beat

Comitative
- influence, attack, bear_a_grudge, look

Possession
- have, have_left, have_a_ache, lack, have_enough, love@inanimate@

Instrument/Cause
- need, dream@sleeping@

Some emotions
- miss, forget, dream, think
- have, have_a_ache, lack, have_enough, love@inanimate@
- need, dream@sleeping@

Instrument/Cause
- get_mixed, be_friends, get_to_know, fight, speak_with, agree, meet_with, have_a_quarrel
- be_content, enjoy
- have@illness@
- cut_oneself, wave

“Motion towards”
Typologically informed analysis of language-specific valency class systems

- E.g., build a distance matrix of verbs based on their typological behaviour

- And explore the ways in which individual language’s valency classes carve out sections of that space
Basic valency classes in North Eastern Neo-Aramaic against a typological background (MDS-visualization)
The ratio of intransitive verbs
(Dis)similarity distances between languages

- DistValPat: a dissimilarity metric based on Mutual Information
  - Captures dissimilarities between the ways in which the verbal lexicon is partitioned into valency classes
Neighbotnet visualization based on DistValPat
Comparison with genealogical and areal data

- Similarities in valency class organization, including minor classes
  - no family-level genetic effects
  - strong areal effects

- Similarities in transitivity profiles:
  - family-deep genealogical effects
  - no large-scale areal effects
$\text{DistValPat} = \text{Entropy-based distance between valency class systems}$

Cumulative average z-scores: DistValPat

Geographic distance (km)

- Same genus
- Same family
- Unrelated
DistTrProf is the only metric that showed significant genetic signal not only on the level of individual genera, but also on the family-size level. This indicates that languages are relatively stable in terms of those semantic features that are relevant for the assignment of the [+/-] transitivity values to individual verbs. The data imply that verb hierarchies of transitivity prominence are partially family-specific: given a certain level of DistTrRat, genetically related languages show lower DistTrProf, which would not be expected if the transitivity-prominence scale of verbs were universal. Finally, DistValPat shows the strongest low-level areal signal for both genetically related and unrelated languages, everywhere apart from the Caucasus.
THANK YOU!
References


