INDUCING VERB CLASSES FROM FRAMES IN RUSSIAN: MORPHO-SYNTAX AND SEMANTIC ROLES

Olga Lyashevskaia & Egor Kashkin
olesar@yandex.ru & egorkashkin@rambler.ru
INTRODUCTION

- The systematic lexicography approach: lexical classes ⇔ similar trends in grammatical, syntactic and lexical co-occurrence behaviour.
- Computational linguistics: lexical classes ⇔ reliable patterns for machine learning.
- Problem: lack of large-scale lexical classifications as open-access resources.
- In future: well, we have lots of open-access classifications but how to assess them?
INTRODUCTION

- Manual builds of verb classes:
  - English (Levin 1993, VerbNet, Palmer 2009; Baker et al. 1998),
  - Spanish (Vázquez et al. 2000),
  - Russian (Kuznetsova 1989; Shvedova 1998-2007; Babenko 2007; Apressjan lavori in corso).

- Automatically built classifications (Dorr and Jones 1996; Lapata 1999; Korhonen 2002; Schulte im Walde 2006; etc.)

- Lenci 2014: ontology-based vs. distribution-based

  * eat – devour *

  eat – devour

  share the same frame Ingestion

  *She devours (obj drop)

  'an Ingestor consumes Ingestibles'

  *She devoured at the cake

  (conative construction)
WHY NOT TRANSFORMATION-BASED CLASSIFICATIONS?

- Syntactic transformations/alternations: verb in two states
  
  *She knocked the door. -- She knocked on the door.*

  (Adams 2001)

- Distributive hypothesis: the verbs are similar semantically if their context patterns are similar

- Corollary (corpus data): ...if the distribution of their constructions is similar

  - VERB1: meaning A cx A1 27
  - VERB2: meaning A cx A1 27
    
    cx A2 6
    meaning B cx B1 19
    cx B2 12
    cx B3 7
    meaning C cx C1 15
  - cx A2 6
    meaning B cx B1 19
    cx B2 12
    cx B3 7
    meaning C cx C1 15

- Strong preference for cx (Stefanowitch, Gries 2004)
WHY NOT TRANSFORMATION-BASED CLASSIFICATIONS?

- Syntactic transformations/alternations: verb in two states
  
  _She knocked the door._ -- _She knocked on the door._

  preference: 1572 21481 (Adams 2001)

- Distributive hypothesis: the verbs are similar semantically if their context patterns are similar

- Corollary (corpus data): ...if the distribution of their constructions is similar

<table>
<thead>
<tr>
<th>VERB1: meaning A</th>
<th>cx A1</th>
<th>27</th>
<th>cx A2</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERB2: meaning A</td>
<td>cx A1</td>
<td>27</td>
<td>cx A2</td>
<td>6</td>
</tr>
<tr>
<td>meaning B</td>
<td>cx B1</td>
<td>19</td>
<td>cx B2</td>
<td>12</td>
</tr>
<tr>
<td>meaning B</td>
<td>cx B1</td>
<td>19</td>
<td>cx B2</td>
<td>12</td>
</tr>
<tr>
<td>meaning C</td>
<td>cx C1</td>
<td>15</td>
<td>cx B3</td>
<td>7</td>
</tr>
<tr>
<td>meaning C</td>
<td>cx C1</td>
<td>15</td>
<td>cx B3</td>
<td>7</td>
</tr>
</tbody>
</table>

- Alternations of peripheral uses (Kuznetsova, Lyashevskaya 2009)
OUR TASK

- Grouping Russian verbs within a lexical field + on corpus vectors
- Observable morpho-syntactic cues
- Latent frame-based cues
- Which cues are more powerful?
DATA

- A dictionary of lexical constructions + a corpus of up to 100 manually annotated examples for each verb.
- Construction: morpho-syntactic pattern + semantic roles.
- Flexible hierarchy of semantic roles, their correlation with the main semantic classes of verbs.
Паспорт конструкции: 1.2. Погонщик легонько ударил верблюда по спине.

Схема: Snom V Sacc по + Sdat
Пример: Погонщик легонько ударил верблюда по спине [по шее].

<table>
<thead>
<tr>
<th>Буква</th>
<th>Вершина</th>
<th>Экспликация</th>
<th>Ранг</th>
<th>Сем. ограничения</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Snom</td>
<td>агенс</td>
<td>Субъект</td>
<td>одушевленный</td>
</tr>
<tr>
<td></td>
<td>ударить</td>
<td></td>
<td>Предикат</td>
<td>-</td>
</tr>
<tr>
<td>Y</td>
<td>Sacc</td>
<td>пациентс</td>
<td>Объект</td>
<td>одушевленный</td>
</tr>
<tr>
<td>T</td>
<td>по + Sdat</td>
<td>подвергающаяся воздействию часть пациенса</td>
<td>Периферия</td>
<td>часть тела</td>
</tr>
</tbody>
</table>
CASE STUDIES

- Two case studies with different clustering procedures:
  - Speech verbs – automatic clustering, vector models
  - Possessive verbs – manual rule-based clustering
SPEECH VERBS

- 80 verbs which refer to speech in their primary meaning: "govorit' 'say, speak', "povtorit' 'repeat', "khvalit' 'praise'.
- 3000 uses of constructions that refer to speech.
- Three types of vector models:
  - **F:** morpho-syntactic tags, eg. o + NPloc 'about smth'
  - **R:** semantic roles, eg. Speaker, Message-as-Topic
  - **R+F:** both, eg. o + NPloc|Message-as-Topic

Dimensionality of vector space:
- **F:** 34 dimensions (tags that occur less than 5 times removed, cf. ot imeni + NPgen)
- **R:** 20 dimensions
- **R+F:** 71 dimensions ‘on behalf of’, Result, etc.)
HOW OFTEN DO WE GET THE SLOTS FILLED?
IN SEARCH OF GOLD... STANDARD

- There cannot be an ideal gold classification due to different principles that could be applied to data (e.g. thematic proximity, event structure, pragmatic goals, etc.,
- 4 classifications taken:
  1) Babenko 2007: six speech classes + ...
  2) Role2: the role of the 2nd participant: addressee-like, interlocutor-like, patient-like, benefactor-like verbs...;
  3) SpeakerGoals: the goals of the speaker (7 classes);
  4) Goals19: more detailed version of (3) with 19 classes of sharing information, getting information, symmetric communication + speech affect like asking, abusing, etc. ..
## SIZE OF CLASSES

<table>
<thead>
<tr>
<th>Class</th>
<th>26</th>
<th>10</th>
<th>9</th>
<th>7</th>
<th>4</th>
<th>4</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>1 in each other class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babenko</td>
<td>30</td>
<td>20</td>
<td>16</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role2</td>
<td>33</td>
<td>25</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpeakerGoals</td>
<td>26</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1 in each other class</td>
</tr>
<tr>
<td>Goals19</td>
<td>26</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1 in each other class</td>
</tr>
</tbody>
</table>
k-means
# k-means: RESULTS

## Babenko (k(gold)=23, elements in the largest class = 26)

<table>
<thead>
<tr>
<th>Roles &amp; Forms</th>
<th>Roles</th>
<th>Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>PU</td>
<td>CO</td>
</tr>
<tr>
<td>4</td>
<td>0.3418</td>
<td><strong>0.9114</strong></td>
</tr>
<tr>
<td>5</td>
<td>0.3544</td>
<td><strong>0.9114</strong></td>
</tr>
<tr>
<td>13</td>
<td>0.4430</td>
<td>0.6329</td>
</tr>
<tr>
<td>20</td>
<td>0.4810</td>
<td>0.4937</td>
</tr>
<tr>
<td>21</td>
<td><strong>0.5063</strong></td>
<td>0.5190</td>
</tr>
<tr>
<td>23</td>
<td>0.4810</td>
<td>0.4684</td>
</tr>
<tr>
<td>25</td>
<td>0.4304</td>
<td>0.4684</td>
</tr>
</tbody>
</table>

## Role2 (k(gold)=7, elements in the largest class = 33)

<table>
<thead>
<tr>
<th>Roles &amp; Forms</th>
<th>Roles</th>
<th>Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>PU</td>
<td>CO</td>
</tr>
<tr>
<td>4</td>
<td>0.4051</td>
<td><strong>0.9114</strong></td>
</tr>
<tr>
<td>5</td>
<td>0.4177</td>
<td><strong>0.9114</strong></td>
</tr>
<tr>
<td>7</td>
<td>0.5570</td>
<td>0.6582</td>
</tr>
<tr>
<td>10</td>
<td>0.4937</td>
<td>0.6076</td>
</tr>
</tbody>
</table>
## k-means: RESULTS

### SpeakerGoals ((k(gold)=7, elements in the largest class = 30))

<table>
<thead>
<tr>
<th>Roles &amp; Forms</th>
<th>Roles</th>
<th>Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>PU</td>
<td>CO</td>
</tr>
<tr>
<td>4</td>
<td>0.4304</td>
<td>0.9114</td>
</tr>
<tr>
<td>5</td>
<td>0.4430</td>
<td>0.9114</td>
</tr>
<tr>
<td>7</td>
<td>0.5823</td>
<td>0.6962</td>
</tr>
<tr>
<td>8</td>
<td>0.5696</td>
<td>0.5570</td>
</tr>
<tr>
<td>9</td>
<td>0.5696</td>
<td>0.5570</td>
</tr>
</tbody>
</table>

### Goals19 (k(gold)=19, elements in the largest class = 26)

<table>
<thead>
<tr>
<th>Roles &amp; Forms</th>
<th>Roles</th>
<th>Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>PU</td>
<td>CO</td>
</tr>
<tr>
<td>4</td>
<td>0.3418</td>
<td>0.9241</td>
</tr>
<tr>
<td>5</td>
<td>0.3544</td>
<td>0.9241</td>
</tr>
<tr>
<td>11</td>
<td>0.3924</td>
<td>0.6076</td>
</tr>
<tr>
<td>19</td>
<td>0.5316</td>
<td>0.6076</td>
</tr>
<tr>
<td>22</td>
<td>0.5696</td>
<td>0.4684</td>
</tr>
<tr>
<td>23</td>
<td>0.5443</td>
<td>0.4430</td>
</tr>
<tr>
<td>24</td>
<td>0.5443</td>
<td>0.4810</td>
</tr>
<tr>
<td>25</td>
<td>0.5696</td>
<td>0.4557</td>
</tr>
</tbody>
</table>
**k-means:** Roles & Forms >= Roles > Forms

- metrics: Purity (PU), Collocation (CO), and F1 (Lang, Lapata 2011)
- Roles > Forms except Goals19
- ! Forms perform better at smaller k-s -- while Roles work better at larger k values
- Roles & Forms is the best in 3 of 4 but at certain k's Roles > Roles & Forms

- Are k-means results the best at max k's? interestingly, no!
HIERARCHICAL CLUSTERING: CONSISTENCY IN ERRORS

- Verbs that group together under all three conditions M, R, M+R (at k=7, 63 verbs in eight clusters of size 4 to 14)
  
  \{besedovat', zdorovat'sja, obschat'sja, prostit'sja\},
  
  \{blagodarit', informirovat', pozdravit', privatstvovat', khvalit', zvat', klikat', oprosit'\}

- manual assessment

- obvinit' and ugoverit' ('accuse' and 'persuade'),
  - similar R-vectors (Speaker - Addressee - Message-as-content), partly different M-vectors: NPnom NPacc
    v+NPloc vs NPnom NPacc VPinf

- prokl'ast' and obosnovat' ('imprecate' and 'justify'),

- podkhvatit' and ugrozhat' ('play along' and 'threaten')
HIERARCHICAL CLUSTERING: CONSISTENCY IN ERRORS

- Orthogonality effects:

<table>
<thead>
<tr>
<th>dimensions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>vector 1</td>
<td>a</td>
<td>b</td>
<td>-</td>
<td>c</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>vector 2</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- obvinit' and ugovorit' ('accuse' and 'persuade'),
  - v+NLoc |55| – VPInf |54|
  - Message-as-content|Conj + CL |1|, Reason|Sins |1|

- Unique context patterns

- objasnit' (explain'): v+NLoc
INTERIM CONCLUSIONS: SPEECH VERBS

- Roles+Morph slightly overperforms Roles and Morph
- one-to-one correspondence VS one-to-many
- In contrast to rules based on a standard valency dictionary (Apresjan 1982) statistics reveals:
  - 'core' frame elements are often not expressed
  - some 'non-core' frame elements are frequently expressed in a certain domain-anchored patterns
- Larger data sets could improve the results.
POSSESSIVE VERBS: DATA

- 128 Verbs having the arguments with the roles of Initial Possessor or Eventual Possessor.
- Different meanings are treated separately, e. g. *vz’at’* 1 ‘to take’ ≠ *vz’at’* 9 ‘to buy’.
- Morpho-syntactic tags, semantic roles and their pairings.
- E. g. *krast’* 1 ‘to steal’:
  - NPnom V NPacc (Eventual Possessor – Patient).
  - NPnom V NPacc *u* + NPgen (Eventual Possessor – Patient – Initial Possessor).
  - …
- Morpho-syntactic patterns & semantic roles => manual clustering.
- Verification: agglomerative clustering method.
POSSESSIVE VERBS: RESULTS

- Two main classes:
  - Taking (34 verbs): NPnom is an Eventual Posessor.
  - Giving (90 verbs): NPnom is an Initial Posessor.
- Further classification.
POSSESSIVE VERBS: TAKING

- **Buying** (4 verbs): NPnom V NPacc za + NPacc (Eventual Possessor – Patient – Price), e. g. *kupit’* 1 ‘to buy’, *ar’endovat’* 1 ‘to rent’.

- **Stealing** (10 verbs): NPnom V NPacc {ADV / PRfrom_where + NPx} (Eventual Possessor – Patient – Starting Point), e. g. *vorovat’* 1 ‘to steal’, *krast’* 1 ‘to steal’, *pohitit’* 1 ‘to steal, to kidnap’, *taskat’* 4 ‘to pinch’.

- **Receiving** (3 verbs): NPnom V NPacc ot + NPgen (Eventual Possessor – Patient – Initial Possessor) – *polučit’* 1 ‘to receive’, *prin’at’* 1 ‘to accept (e. g. a present)’, *prin’at’* 4 ‘to accept (e. g., an advertisement)’.

- **Earning** (2 verbs): NPacc is Price – *zarabotat’* 1, *polučit’* 2 ‘to earn (money)’.

- **Borrowing** (3 verbs): verbs of taking which have Time Period among their participants – *ar’endovat’* 1 ‘to lease sth. from sb.’, *zan’at’* 8 ‘to borrow’, *sn’at’* 13 ‘to rent’.
POSSESSIVE VERBS: GIVING

- **Selling** (6 verbs): NPnom V NPacc za + NPacc (Initial Possessor – Patient – Price), e.g. *prodat*’ 1 ‘to sell’, *sdat*’ 4 ‘to lease sth. to sb.’

- **Paying** (2 verbs): verbs of giving which have a direct object expressing Price – *platit*’ 1 ‘to pay’, *ustupit*’ 3 ‘to take off (a price)’

- **Giving somewhere** (8 verbs): NPnom V NPacc {ADV / PRwhere(to) + NPx} (Initial Possessor – Patient – Point of destination).
  - *Vernut*’ 1, *vozvratit*’ 1 ‘to return sth. to sb.’
  - *Vyslat*’ 1, *poslat*’ 2 ‘to send’, *p’er’edat*’ 1 ‘to pass sth. to sb.’
  - *Podat*’ 2 ‘to submit’, *sdat*’ 1 ‘to return, to surrender’, and *sdat*’ 2 ‘to submit, to hand in’
POSSESSIVE VERBS: GIVING

- **Giving with some goal** (12 verbs): NPnom V NPacc na + NPacc (Initial Possessor – Patient – Goal) or NPnom V NPacc NPdat na + NPacc (Initial Possessor – Patient – Eventual Possessor – Goal), e.g. assignovat’ 1 ‘to allocate’, žertvovat’ 1 ‘to donate’, tratit’ 1, 2 ‘to spend’, darit’ 1 ‘to make a present’, pr’epodn’esti 1 ‘to present (a gift)’.

- **Supplying** (8 verbs): verbs (balovat’ 2 ‘to make sb. glad by giving sth.’, vooružit’ 1 ‘to arm’, nagradit’ 1 ‘to award’, ob’esp’ečit’ 1 ‘to provide’, (n’e) obid’et’ 2 ‘not to stint sb. of sth.’, obogatit’ 2 ‘to enrich’, ssudit’ 1 ‘to loan’) with a pattern NPnom V NPacc NPins (Initial Possessor – Eventual Possessor – Patient), a verb od’et’ 2 ‘to provide clothes for sb’ with a pattern NPnom V NPacc (Initial Possessor – Eventual Possessor – Patient), and a verb obogatit’ 1 ‘to enrich’ with a pattern NPnom V NPacc (Method – Eventual Possessor).

- **Lending** (2 verbs): verbs of giving which have Time Period among their participants (odolžit’ 1 ‘to lend’, sdat’ 4 ‘to rent out’).
Various possible cues for clustering.

Semantic roles => inadequate, conversives!

Morpho-syntactic patterns => better, but often too broad classes, e.g. Snom V Sacc za + Sacc – both Bying and Selling, Snom V Sacc na + Sacc – 7 possible roles of PP.

Semantic roles & morpho-syntactic patterns => the best, few cases of too broad classes (some peripheral verbs in the class of Stealing, a possibly heterogeneous class of Giving somewhere).
POSSESSIVE VERBS: FRAMEBANK VS. FRAMENET

- Berkeley FrameNet as the gold standard.
- The classes appear to be mostly the same.
- The main differences:
  - Surrendering_possession (‘A Surrenderer is compelled to transfer a Theme to a Recipient’, e.g. *Shortly after the boy surrendered the gun, the three remaining warriors made a rush for liberty*): missed in our clustering.
  - Submitting_documents, Supply: there are more peripheral verbs in our study.
  - Giving with some goal – missed in FrameNet, but found in our experiment.
CONCLUSIONS

- Lexical classifications: input data structure!
- Clustering is better if semantic roles are taken into account, either as the only input (in the case of speech verbs) or together with the morpho-syntactic patterns.
- In the future: large corpora with SRL annotations?