How learnable is Russian aspect?

Laura A. Janda
UiT The Arctic University of Norway

with: Hanne M. Eckhoff, Olga Lyashevskaia and Robert J. Reynolds
How learnable is Russian aspect?

- Use and meaning of Russian aspect is topic of long-standing debate (cf. Janda 2004 and Janda et al. 2013 and references therein)
- It is unclear how children acquire Russian aspect in L1
  - Generativist theory would assume that aspect is part of UG
  - Gvozdev (1961), based on his diary of son Ženja, claimed Russian aspect was fully acquired early on, but re-analysis of his and other data (Stoll 2001, Gagarina 2004) has shown that L1 acquisition is far from complete even at age 6
- It is clear that L2 learners struggle with Russian aspect
  - Russian aspect is considered the most difficult grammatical feature for L1 English speakers (Offord 1996, Andrews et al. 1997, Cubberly 2002); it is not clear how L2 acquisition takes place (Martelle 2011)
  - “Rules” offered in textbooks for when to use perfective vs. imperfective are relevant for only 2% of verb forms in a corpus (Reynolds 2016)
Aspect in Russian (a crash course)

• All forms of all verbs obligatorily express **perfective** vs. **imperfective** aspect

• **Perfective aspect**: unique, complete events with crisp boundaries
  – *Pisatel’ na-pisal/na-pišet roman ‘The writer has written/will write a novel’*

• **Imperfective aspect**: ongoing or repeated events without crisp boundaries
  – *Pisatel’ pisal/pišet roman ‘The writer was writing/is writing a novel’*

• Morphological marking is not entirely reliable:
  – **bare verb**: usually imperfective (*pisat’ ‘write’*), some biaspectual (*ženit’sja ‘marry’*), a few perfective (*dat’ ‘give’*)
  – **prefix + verb**: usually perfective (*pere-pisat’ ‘rewrite’*), some imperfective (*pre-obladat’ ‘prevail’, *pere-xodit’ ‘walk across’*)
  – **prefix + verb + suffix**: imperfective (*pere-pis-yva-t’ ‘rewrite’*)
Where the aspects do and do not compete

- **Paradigmatically competing:**
  - Non-past (future if perfective, present if imperfective)
  - Past
  - Imperative
  - Infinitive

- **Syntagmatically competing:**
  - In some contexts, either aspect is grammatical

- **Paradigmatically non-competing:**
  - Past gerunds and participles are perfective
  - Present gerunds and participles are imperfective

- **Syntagmatically non-competing:**
  - In some contexts only one aspect is allowed

Study 1 takes a **paradigmatic perspective**
Study 2 takes a **syntagmatic perspective**
Research Questions

Study 1 Paradigmatic Perspective:
To what extent can the aspect of a verb be figured out based on the distribution of its grammatical forms (grammatical profiling)?
Can this type of learning be modeled by means of corpus data?

Study 2 Syntagmatic Perspective:
To what extent can the aspect of a verb be figured out based on the context in which it appears?
Can this type of learning be modeled by means of experiments with native speakers, L2 learners, and machine learning?
Study 1 Paradigmatic Perspective: Aspect via Grammatical Profiles

• Janda & Lyashevskaya (2011) showed that, for paired verbs, perfective and imperfective verbs have in aggregate different grammatical profiles
  – This was a top-down approach (we started out by segregating perfective from imperfective verbs) and was limited to paired verbs
• Can aspect be approached bottom-up?
• Is it possible to figure out the aspect of individual verbs of all types (not just paired verbs) based only on the distribution of their grammatical forms in a corpus?
• Goldberg (2006) gives evidence that children are sensitive to statistical tendencies in L1 acquisition
• Could children learn to distinguish between perfective and imperfective verbs based solely on the distributions of their forms?
What is a grammatical profile?

Verbs have different forms:

- eat 749 M
- eats 121 M
- eating 514 M
- eaten 88.8 M
- ate 258 M

The grammatical profile of eat
Janda & Lyashevskaya 2011
Grammatical Profiles of Russian Verbs **Top-Down**

<table>
<thead>
<tr>
<th></th>
<th>Nonpast</th>
<th>Past</th>
<th>Infinitive</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfective</td>
<td>1,330,016</td>
<td>915,374</td>
<td>482,860</td>
<td>75,717</td>
</tr>
<tr>
<td>Perfective</td>
<td>375,170</td>
<td>1,972,287</td>
<td>688,317</td>
<td>111,509</td>
</tr>
</tbody>
</table>

**chi-squared**
= 947756
**df** = 3
**p-value** < 2.2e-16
**effect size** (Cramer’s V) = 0.399 (medium-large)
Janda & Lyashevskaya 2011
Grammatical Profiles of Russian Verbs **Top-Down**

<table>
<thead>
<tr>
<th></th>
<th>Nonpast</th>
<th>Past</th>
<th>Infinitive</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imperfective</td>
<td>1,330,016</td>
<td>915,374</td>
<td>482,860</td>
<td>75,717</td>
</tr>
<tr>
<td>Perfective</td>
<td>375,170</td>
<td>1,972,287</td>
<td>688,317</td>
<td>111,509</td>
</tr>
</tbody>
</table>

Can we turn this upside-down and go **Bottom-Up**?
Grammatical Profiles of Russian Verbs Bottom-Up

Data extracted from the manually disambiguated Morphological Standard of the Russian National Corpus (approx. 6M words), 1991-2012
Stratified by genre, 0.4M word sample for each

<table>
<thead>
<tr>
<th>Genre</th>
<th># Verb Tokens</th>
<th># Verb Lemmas</th>
<th># Verb Lemmas Frequency &gt;50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journalistic</td>
<td>52 716</td>
<td>5 940</td>
<td>185</td>
</tr>
<tr>
<td>Scientific-technical</td>
<td>43 528</td>
<td>4 494</td>
<td>74</td>
</tr>
<tr>
<td>Fiction</td>
<td>78 084</td>
<td>8 665</td>
<td></td>
</tr>
</tbody>
</table>

Study 1 focuses on **Journalistic** data
Correspondence Analysis of Journalistic Data

**Input:** 185 vectors (1 for each verb) of frequencies for verb forms
Each vector tells how many forms were found for each verbal category: indicative non-past, indicative past, indicative future, imperative, infinitive, non-past gerund, past gerund, non-past participle, past participle

*rows are verbs, columns are verbal categories*

**Process:**
Matrices of distances are calculated for rows and columns and represented in a multidimensional space defined by factors that are mathematical constructs. Factor 1 is the mathematical dimension that accounts for the largest amount of variance in the data, followed by Factor 2, etc.

Plot of the first two (most significant) Factors, with Factor 1 as x-axis and Factor 2 as the y-axis

You can think of **Factor 1** as the **strongest parameter** that splits the data into **two groups** (negative vs. positive values on the x-axis)
On the Following Slide…

- Results of correspondence analysis for Journalistic data
- Perfective verbs represented as “p”
- Imperfective verbs represented as “i”
- Remember that the program was not told the aspect of the verbs
- All it was told was the frequency distributions of grammatical forms
- All it was asked to do was to construct the strongest mathematical Factor that separates the data along a continuum from negative to positive (x-axis)
Factor 1 looks like aspect
Factor 1 correctly predicts aspect 91.5% 
(negative = perfective vs. positive = imperfective)

Of the 185 verbs:

- 87 perfectives
  - 84 negative values, 3 positive values, so 96.6% correct
    - 3 deviations are: obojtis’ ‘make do without’, smoč’ ‘manage’, prijtis’ ‘be necessary’

- 96 imperfectives
  - 83 positive values, 13 negative values, so 86.5% correct
    - 13 deviations are: ezdit’ ‘ride’, rešat’ ‘decide’, xodit’ ‘walk’,
      prinimat’ ‘receive’, iskat’ ‘seek’, rassčityvat’ ‘estimate’,
      provodit’ ‘carry out’, ožidat’ ‘expect’, borot’sja ‘struggle’,
      platit’ ‘pay’, čitat’ ‘read’, učastvovat’ ‘participate’, smotret’ ‘look’

- 2 biaspectuals with low negative values
  - obeščat’ ‘promise’, ispol’zovat’ ‘use’
Summary of Study 1: Paradigmatic Perspective

• When we look at the distribution of verb forms, aspect (or a close approximation) emerges as the most important factor distinguishing verbs.

• It is possible to sort most high-frequency verbs as perfective vs. imperfective based only on the distribution of their forms.

• Need to study the exceptions more carefully.
Study 2 Syntagmatic Perspective: Aspect via Context

• This study is still underway!!
• Given contexts where both aspects are morphologically possible, what happens when you offer a choice of a perfective vs. an imperfective form to:
  – L1 native speakers of Russian
  – L2 learners of Russian
  – machine learning algorithms
• Source material: texts representing three written genres (journalistic, scientific-technical, fiction) and two spoken genres (monologue, dialogue)
• All texts represent authentic Russian (produced by native speakers) and plenty of context (1100-1500 words)
The contexts where both aspects are morphologically possible in Russian: example verb *na-pisat’*(p) vs. *pisat’*(i) ‘write’

<table>
<thead>
<tr>
<th></th>
<th>Perfective</th>
<th>Imperfective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past</td>
<td><em>na-pisal</em> ‘he wrote’</td>
<td><em>pisal</em> ‘he wrote’</td>
</tr>
<tr>
<td>Future</td>
<td><em>na-pišet</em> ‘s/he will write’</td>
<td><em>budet pisat’</em> ‘s/he will write’</td>
</tr>
<tr>
<td>Infinitive</td>
<td><em>na-pisat’</em> ‘write’</td>
<td><em>pisat’</em> ‘write’</td>
</tr>
<tr>
<td>Imperative</td>
<td><em>na-piši</em> ‘write!’</td>
<td><em>piši</em> ‘write!’</td>
</tr>
</tbody>
</table>

Excluded:
- Present tense (imperfective only)
- Gerunds & participles (specific to one aspect or the other)
- Biaspectual verbs
- Verbs not paired for aspect (Aktionsarten, -*sja* passives)
- Forms of verb *byt’* ‘be’
В Соединенных Штатах наготове дожидается своего часа огромное число буровых установок и комплектов оборудования для гидроразрыва пласта, чтобы [возобновить / возобновлять] работу, как только оставленные на время промыслы [выйдут / будут выходить] на уровень рентабельности.

In the USA there stands at the ready an enormous number of drills and sets of equipment for the fracking of rock layers in order [RESUME] work as soon as the temporarily held up businesses [ACHIEVE] the level of profitability.


Thousands of holes have been bored where there [REMAIN] only [INITIATE] periodic fracking operations.
The interface we are developing

Click to write the question text

Беспризорник Жук

Most correct  Possible  Impossible

Право выбора и определенного судьбу. У Василия этого права не было. Он безропотно [принял / принимал] выбор, который за него [сделала / делала] судьба, и это был великий шаг.


Вскоре мать [умерла / умирала] от отравления, и по деревне поползли слухи. Но доказательств вины в смерти женщины не было, да и искать было некому.

Отец Василия женился на работнице, которая уже была на сносях. Дети у мачехи Василия [посели / шли] один за другим.
What a pilot revealed
(NB: based on a small sample with a Wikipedia text, which will not be part of the experiment)

Native Speaker Responses

- Unambiguous with trigger
- Unambiguous without trigger
- Ambiguous

2%  83%  15%
Examples of triggers cited in Russian textbooks  
(Only available 2% of the time but 96% reliable)

<table>
<thead>
<tr>
<th>Adverbials</th>
<th>Complements of verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Imperfective</strong></td>
<td>Phasal verbs: stat’ ‘start’, načat’/načinat’ ‘begin’, prodolžit’/prodolžat’ ‘continue’, končit’/končat’ ‘stop’</td>
</tr>
<tr>
<td>categorical negation: ne nado ‘should not’, ne stoit ‘not worth’, ne razrešaetsja ‘not allowed’</td>
<td>Others: učit’sja ‘learn’, umet’ ‘know how’, ljubit’ ‘love’</td>
</tr>
</tbody>
</table>
**Examples of the three types**
(sentences with triggers from Fiction; others from the Wikipedia text we will not be using)

<table>
<thead>
<tr>
<th></th>
<th>Unambiguous with trigger ($T$): 2%</th>
<th>Unambiguous without trigger: 83%</th>
<th>Ambiguous: 15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfective</td>
<td>neobyknovennoe vezenie nakonec($T$) prišlo k nemu ‘unusual luck finally($T$) came to him’</td>
<td>V 1909 godu brat’ja otkryli pervuju v Kolorado gostinicu ljus-klassa ‘In 1909 the brothers opened the first luxury hotel in Colorado’</td>
<td>Pojavlenie legkogo, kompaktnogo i dostatočno moščnogo dvigatelja vnutrennego sgoranija otkrylo širokie vozmožnosti dlja razvitija avtomobilja. ‘The appearance of a lightweight, compact and rather powerful internal combustion motor opened up extensive possibilities for the development of the automobile.’</td>
</tr>
<tr>
<td>Imperfective</td>
<td>Za bol’šim stolom na terrase často($T$) sobiralis’ gosti. ‘Guests often($T$) gathered at the big table on the terrace.’</td>
<td>Brat’ja Stènli proizvodili okolo 1000 avtomobilej v god. ‘The Stanley brothers produced about 1000 automobiles per year.’</td>
<td>Legkij i vosokoproizvoditel’nyj dvigatel’ dostigal 75km/č i mog nabirat’ ot 300 do 1000 oborotov v minutu. ‘The lightweight high-capacity motor reached the speed of 75m’km/h and could achieve from 300 to 2000 revolutions per minute.’</td>
</tr>
</tbody>
</table>
Summary of Study 2: Syntagmatic Perspective

• For 2% of verb forms in a corpus the choice of aspect is clearly marked by a “trigger” in the context: here everyone (L1, L2, machine learning) should know what to do and be correct 96% of the time
  – But what about the exceptions to these rules?
• For about 83% of verb forms in a corpus L1 speaker knows what to do, but is it possible to deduce tendencies and guidelines?
  – Maybe machine learning can find patterns for us?
• For about 15% of verb forms in a corpus, L1 has a “free” choice
  – What makes free construal possible?
  – Are some choices freer than others? Why?
Conclusions

• In the overwhelming majority of cases, the aspect of a verb can be determined either solely on the basis of the distribution of forms, or solely on the basis of context.

• It is likely that L1 learners use both cues in acquisition.

• But we don’t know enough about the cues.

• More study could tell us more about the role of construal in language.

• And we could learn things that can be applied to pedagogy.
References


