CLUSTER SIMPLIFICATION IN A NONWORD REPETITION STUDY OF RUSSIAN CHILDREN WITH SLI

Darya Kavitskaya
UC Berkeley

Sharon Inkelas
UC Berkeley
Outline of the talk

1. Introduction
2. SLI
3. The setting: villages in the Russian north
4. The experimental study
5. Results and discussion
Introduction
Introduction

• We present preliminary results of a nonword repetition study involving two groups of children in several villages of the Russian north

• One group of children is diagnosed with SLI (Specific Language Impairment); the other is typically developing

• The nature of phonological impairment in SLI is complex. Grammar-related? Memory-related? other?

• A controlled comparison may shed some light on this
SLI
Defining SLI

- Atypical language development with no apparent neurobiological pathology
- Most common term is Specific Language Impairment (SLI)
  - “Specific” in a sense that the impairment of language occurs in otherwise normal development (non-verbal IQ within normal range, no obvious sensory-motor deficiencies)
  - Specificity of SLI has been questioned (e.g., Bishop 1994, Hill 2001, Plante 1998, Ullman & Pierpont 2005)
  - Other terms used: developmental language impairment, developmental dysphasia, developmental language disorder (DLD)

Bishop (1992); Leonard (1998); Tomblin et al. (1997); Rakhlin et al. (2013)
Defining SLI

- SLI is a heterogeneous disorder with several different profiles possible that can be responsible for the low performance on the standardized verbal tests.
- The diagnosis is based on broad exclusionary criteria, so individuals can exhibit a wide range of symptoms, e.g.,
  - Smaller receptive and expressive vocabulary
  - Word-finding difficulties
  - Omission or incorrect use of morphological forms
  - Low syntactic complexity
  - Poor performance on nonword repetition tests
The Setting:
Villages in the Russian North
A cluster of SLI

- SLI participants were recruited from a small cluster of villages in Northwestern Russia
  - As of 2012, the number of residents is 861
  - Substantial rate of distant intermarriages
  - More than 31% of the population (children and adults) exhibit atypical language development

Rakhlin et al. (2013)
A cluster of SLI

- This population is characterized by a high degree of genetic and environmental uniformity
  - 120 km to the nearest train station; 600 km from the nearest major city
  - 45 km to the provincial center by a dirt road
- The population has been geographically and culturally isolated since the time of its establishment in the early 15th century
  - Surrounded by forest and swamp; difficult to settle or navigate; harsh climate
- Same socioeconomic class: Russian rural poverty

Rakhlin et al. (2013)
SLI study subjects

- 7 monolingual Russian-speaking children aged 7.3:9.5 years old classified as DLD (SLI)
- 7 Typically Developing (TD) children matched to the SLI group in age
- Non-verbal IQ for all children within normal range: 88-113
- Standard measures, adapted to Russian, used to assess language development
Control group: TD subjects

- Goal: matched comparison populations
- Typically developing (TD) participants were recruited from a demographically similar population (a village in the same administrative region of Russia)
Participants compared

- Same type of schools
- Same types of jobs for parents
- Similar social behavior
- Same peers
- No labeling (SLI not an identified disability)
- No intervention at school
Participants

- All participants’ parents agreed that their child could participate in this and related studies conducted at the same time under guidelines approved by the Yale University Human Subjects Research Review Committee and Northern State Medical University.

- The data were collected as a part of a larger study of familial Disorders of Spoken and Written Language.
The experimental study
Study design

• Nonword imitation task, testing ability to correctly reproduce disyllabic words varying in whether or not they begin or end with consonant clusters

  ptkoka patugmn tabont krata paran

• Why clusters?
  • Consonant clusters are famously difficult in L1 acquisition and in L2 acquisition for L2 learners whose L1 has different cluster restrictions
  • Russian has a variety of different cluster types
Phonology of Russian

- Russian has both onset and coda CC and CCC clusters
- Not all clusters in Russian obey the Sonority Sequencing Generalization (SSG)
  - rta ‘mouth-GEN.SG’
  - lba ‘forehead-GEN.SG’
  - mxa ‘moss-GEN.SG’
- Not all combinations of Russian consonants are attested in CC and CCC clusters
Study design: tokens

- Nonwords
- No palatalization
- Vowels: a, o, u
- Clusters in stressed syllables only
- CC vs. CCC clusters
  - CC: bnapa, dbota, lbuka
  - CCC: gmrota, ptkoka, nzboka
- Onset vs. coda clusters
  - Onset: brupa, pflata
  - Coda: tabolk, takodnl
Study design: tokens

- Rising, falling, level sonority clusters
  - Rising: brupa, gmruka
  - Falling: lbata, rskupa
  - Level (obstruent and sonorant): dbota, ptkoka, mnota

- Lexically attested vs. unattested clusters
  - Attested in Russian: lbata
  - Unattested in Russian: nbota

- Fillers without clusters: dopa, kalus
Study design

- Total number of tokens
  - 144 tokens in a list
  - 2 lists (2 testing blocks)
- Each word repeated once after pronounced by experimenter
- Recording done in the villages’ schools
- Experimenters spoke the same Northern Russian dialect as the children
Transcription and coding

- Transcription in Berkeley
- Random cross-checking for accuracy
- Some tokens too noisy or inaudible
Results and discussion
The focus of this talk

- Word and cluster repetition accuracy as a function of
  - SLI vs. TD
  - Cluster size
  - Typological markedness of cluster (by syllable position)
  - Lexical attestedness of cluster type
## Examples of errors

<table>
<thead>
<tr>
<th>Repair to cluster</th>
<th># tokens</th>
<th>Example target word</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion</td>
<td>624</td>
<td>ptkoka</td>
<td>ptkoka</td>
</tr>
<tr>
<td>Segmental change</td>
<td>436</td>
<td>patubml</td>
<td>patugmn</td>
</tr>
<tr>
<td>Epenthesis (C or V)</td>
<td>128</td>
<td>pmota</td>
<td>pmota</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mtupa</td>
<td>mutupa</td>
</tr>
<tr>
<td>Assimilation</td>
<td>129</td>
<td>mnota</td>
<td>n:ota</td>
</tr>
<tr>
<td>Metathesis</td>
<td>124</td>
<td>pakatp</td>
<td>pakapt</td>
</tr>
</tbody>
</table>
Word accuracy, SLI vs. TD

- SLI subjects produce fewer words correctly than TD subjects do (p < .008)
Cluster accuracy, SLI vs. TD

- SLI subjects make more errors in clusters than TD subjects do (p > .003)
Cluster size: CC vs. CCC

• Both groups make more errors with bigger clusters than with smaller ones (for TD, p < .001; for SLI, p < .001)
Typological markedness

- From a cross-linguistic perspective, clusters obeying the Sonority-Sequencing Generalization are unmarked.
- Hypothesis: performance should be better on unmarked clusters.

Rising onsets, Falling codas…

BEST

Falling onsets, Rising codas

WORST
Typological markedness

- Russian exhibits unmarked and marked clusters:
  - rising sonority onsets  brat  ‘brother’
  - level sonority onsets  kto  ‘who’
  - falling sonority codas  volk  ‘wolf’
  - level sonority codas  glimn  ‘anthem’
  - falling sonority onsets  lba  ‘forehead-GEN.SG’
  - rising sonority codas  bobr  ‘beaver’

- Marked clusters are less frequent than unmarked clusters
For codas, both SLI and TD do better on unmarked clusters (e.g., kabukr, patabm), as expected. Significant for SLI ($p < .02$)
But: For onsets, TD subjects do better with marked clusters (e.g., *lbuka*, *nbota*) than with unmarked (p < .001)
Lexical attestedness of clusters

• 50% of the CC clusters are attested in Russian
• 0% of the CCC clusters are attested in Russian (though Russian does have other CCC clusters)
• If the preference for marked onsets is due to a lexical effect (the existence of clusters like /lb/ in experimental tokens and actual Russian words), then the preference should only be exhibited in experimental tokens with CC clusters.
• **Hypothesis:** subjects do better at attested clusters
Lexical attestedness of clusters

- TD subjects do better at attested clusters
Lexical attestedness of clusters

- SLI subjects are indifferent to cluster attestedness

Effect of lexical and typological markedness on cluster accuracy
Falling onsets: TD

Cluster accuracy, TD only

- Rising attested onsets (kr, br, dl)
- Falling attested onsets (rv, rt, lb)
- Falling attested codas (nt, rk, lk)
- Rising attested codas (dr, br)
- Falling unattested onsets (rb, rt, lb)
- Rising unattested codas (nt, rk, lk)
- Falling unattested codas (dr, br)
- Rising unattested onsets (rb, rt, lb)
- Rising unattested codas (nt, rk, lk)

Onsets
Codas
Onsets
Codas
Falling onsets: SLI

Cluster accuracy, SLI only

Falling onsets

Unattested

Attested

Unattested

Attested

Unattested

Attested

Unattested

Attested

Rising codas

Rising unattested codas (dn, tl, bm)

Rising attested codas (kr, dr, br)

Rising unattested onsets (pm, codas (mk, nf, mt)

Rising unattested codas (nt, bn, bv)

Falling unattested codas (rk, lk)

Falling attested onsets (rv, rt, lb)

Falling unattested onsets (rb, mt, nb)
Summary, word & cluster accuracy

- Cluster size: CC better than CCC for both SLI & TD
- Lexical attestedness of cluster type
  - TD perform better on lexically attested clusters overall
  - SLI subjects are generally insensitive to lexical attestedness
- Typological markedness of cluster (by syllable position)
  - Both groups do better on onset clusters than on coda clusters
  - Codas: as expected, both groups do better on rising than on falling codas
  - Onsets: For SLI subjects, and for TD subjects on unattested clusters, **falling onsets** do better than rising onsets
Ctree
(conditional inference tree)
Ctree

Black bar = P(incorrect)
Gray bar = P(correct)
Ctree

CC cluster

unattested

attested in Russian

level sonority

changing sonority

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position

sonority

position
Deletion
## Deletion is most common error type

<table>
<thead>
<tr>
<th>Repair to cluster</th>
<th># tokens</th>
<th>Example target word</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion</td>
<td><strong>624</strong></td>
<td>ptkoka</td>
<td>ptoka</td>
</tr>
<tr>
<td>Segmental change</td>
<td>436</td>
<td>patubml</td>
<td>patugmn</td>
</tr>
<tr>
<td>Epenthesis (C or V)</td>
<td>128</td>
<td>pmota</td>
<td>pmota</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mtupa</td>
<td>mutupa</td>
</tr>
<tr>
<td>Assimilation</td>
<td>129</td>
<td>mnota</td>
<td>n:ota</td>
</tr>
<tr>
<td>Metathesis</td>
<td>124</td>
<td>pakatp</td>
<td>pakapt</td>
</tr>
</tbody>
</table>
Deletion by group

Clusters with deletion
Clusters without deletion

Deletion

TD
SLI
Ctree

Black bar = P(preservation)
Gray bar = P(deletion)
Deletion inhibited adjacent to V

Deletion of V-adjacent C, all clusters

- V-adjacent C deletes
- Peripheral C deletes

TD
SLI
V-adjacent C preservation

- A similar effect is observed in L1 acquisition (e.g. Gerlach 2010) ('snow' → [no]) and in a study of onset cluster production in the speech of two individuals with impaired speech secondary to aphasia and apraxia of speech (Miozzo & Buchwald 2013) ('flow' → [lo])

- The opposite effect is observed in those rare situations in adult language in which clusters are reduced; e.g. in reduplication, sonority tends to determine the outcome (Tagalog trabaho → ta-trabaho, etc.)
The two groups differ in the effects of position and sonority profile for falling codas.
With SLI subjects only, the preference for preserving V-adjacent C is not observed in falling codas (e.g., rk).
Exception: Falling codas for SLI

- The effect is observed mainly with liquid-obstruent clusters, not with nasal-obstruent clusters

<table>
<thead>
<tr>
<th></th>
<th>V-adjacent C-deletion</th>
<th>Other C deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial liquid (4/7 items)</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Initial nasal (3/7 items)</td>
<td>5</td>
<td>12</td>
</tr>
</tbody>
</table>
Wrapup

- This talk is a first attempt at mapping the terrain of the results

- Key initial landmarks deserving further inspection:
  - TD subjects are very influenced by lexical attestedness; SLI subjects apparently are not
  - SLI and TD subjects generally preserve the V-adjacent C if there is any deletion involved in a cluster
  - Cross-linguistic sonority-sequencing preferences are not always respected: Both groups are more accurate with falling (e.g. lb) onsets than rising (e.g. kr) onsets
Acknowledgments

We thank:

- Elena Grigorenko, Sergey A. Kornilov, and Natalia Rakhlin (Child Study Center, Yale University)
- Antony d’Avirro (UC Berkeley)
- Jem Orgun (CU Boulder)
Thank you!

dkavitskaya@berkeley.edu
inkelas@berkeley.edu