WHAT IS A POSSIBLE WORD? EVIDENCE FROM RUSSIAN FACTITIVE VERBS

Anna Endresen
Laura A. Janda

CLEAR (Cognitive Linguistics: Empirical Approaches to Russian)
University of Tromsø
What happens in the gap between the actual and the impossible in a language?

ACTUAL  IMPOSSIBLE
What happens in the gap between the actual and the impossible in a language?
What happens in the gap between the actual and the impossible in a language?

- **ACTUAL**
  - **STANDARD WORDS**
    - Words that are standard and conventionalized, might be stored in memory rather than generated on the fly
  - **NONCE WORDS**
    - Words that cannot be generated and do not exist

- **POSSIBLE**
  - **MARGINAL WORDS**
    - Words that are generated by some speakers and can be understood / accepted by some speakers

- **IMPOSSIBLE**
What happens in the gap between the actual and the impossible in a language?

**ACTUAL**

**STANDARD WORDS**
Words that are standard and conventionalized, might be stored in memory rather than generated on the fly

- e.g.: *undo*

**POSSIBLE**

**MARGINAL WORDS**
Words that are generated by some speakers and can be understood / accepted by some speakers

- e.g.: *unworry*

**IMPOSSIBLE**

**NONCE WORDS**
Words that cannot be generated and do not exist

- e.g.: *unblick*
What is a possible (marginal) word?

A word which

• is attested at least once
• is not established in standard language
• is a spontaneous creation generated on the fly, on a certain occasion
• is generated on the basis of a productive morphological pattern
• is analyzable and semantically transparent

• It exists and at the same time it does not exist.
Case study: Russian factitive verbs

- *Ob’jasnit’* ‘clarify, make X be clear’ < *jasnyj* ‘clear\(_{\text{ADJ}}\)’

- **Two most productive patterns:** o-…-it’ and u-…-it’

- We are interested in **new coinages** like
  - *omuzykatalit’* ‘musicalize’ (< *muzykal’nyj* ‘musical\(_{\text{ADJ}}\)’)
  - *ukonkretit’* ‘concretize’ (< *konkretnyj* ‘concrete\(_{\text{ADJ}}\)’)
  - *ovnešnit’* ‘externalize’ (< *vnešnij* ‘external\(_{\text{ADJ}}\)’)
Productivity: competence or performance?

• Usually newly coined words are examined from the perspective of linguistic performance (Haspelmath 2002: 112).

• “A widespread view among linguists is that linguistic competence and linguistic performance are conceptually quite distinct and should therefore be studied separately. The different degrees of productivity that we observe in word-formation are a problem for this view, because rule productivity is not clearly a property of either competence or performance.” (Haspelmath 2002: 110)

• We offer a study of productivity that combines both perspectives: performance and competence. Frequency reflects performance, whereas knowledge of productive patterns reflects competence.
**Performance:**
We look at novel marginal factitive verbs generated by some native speakers and attested in the corpus.

**Competence:**
We want to know how these possible words are perceived by other native speakers?

<table>
<thead>
<tr>
<th>Hypothesis 1:</th>
<th>Standard</th>
<th>Marginal</th>
<th>Nonce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 2:</td>
<td>Standard</td>
<td>Marginal</td>
<td>Nonce</td>
</tr>
<tr>
<td>Hypothesis 3:</td>
<td>Standard</td>
<td>Marginal</td>
<td>Nonce</td>
</tr>
<tr>
<td>Hypothesis 4:</td>
<td>Standard</td>
<td>gradual transition</td>
<td>Marginal</td>
</tr>
</tbody>
</table>
Experimental design

STANDARD WORDS
Words that are standard and conventionalized, might be stored in memory rather than generated on the fly

e.g.: undo

MARGINAL WORDS
Words that are generated by some speakers and can be understood / accepted by some speakers

e.g.: unworry

NONCE WORDS
Words that cannot be generated and do not exist

e.g.: unblick
Experimental design

STANDARD WORDS

10 o + 10 u

osurit’
usaglit’

NONCE WORDS

10 o + 10 u

osurit’
usaglit’

MARGINAL WORDS

10 o + 10 u

oser’ëznit’
uvkusnit’

10 o + 10 u

osurit’
usaglit’

10 o + 10 u

oser’ëznit’
uvkusnit’

ob”jasnit’
uskorit’
60 stimuli

• All factitives used in the experiment are **deadjectival**. This decision is made in order to reduce the number of valuables.

• All standard and marginal factitives chosen for experiment are **morphologically transparent** and **analyzable** and have a clear existing adjectival base.

• All factitives are given as **perfective infinitives**.

• Factitive verbs are presented **in contexts**.
  • For standard and marginal factitives we are using real contexts from the corpus, often shortened.
  • The contexts of nonce factitive verbs are based on corpus contexts of real verbs with meanings similar to those that are assumed for nonce factitives.
# Standard factitives

- All chosen standard factitives are highly frequent in the RNC.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ob”’jasnit’</td>
<td>18,149</td>
<td>utočnit’</td>
<td>2,860</td>
</tr>
<tr>
<td>2</td>
<td>oblegčit’</td>
<td>1,802</td>
<td>umen’šit’</td>
<td>2,010</td>
</tr>
<tr>
<td>3</td>
<td>oslabit’</td>
<td>1,401</td>
<td>uskorit’</td>
<td>2,008</td>
</tr>
<tr>
<td>4</td>
<td>okruglit’</td>
<td>939</td>
<td>ulučšit’</td>
<td>1,899</td>
</tr>
<tr>
<td>5</td>
<td>obogatit’</td>
<td>800</td>
<td>uprostit’</td>
<td>1,350</td>
</tr>
<tr>
<td>6</td>
<td>ožestočit’</td>
<td>686</td>
<td>ukorotit’</td>
<td>787</td>
</tr>
<tr>
<td>7</td>
<td>osložnit’</td>
<td>410</td>
<td>usložnit’</td>
<td>311</td>
</tr>
<tr>
<td>8</td>
<td>ogolit’</td>
<td>387</td>
<td>uteplit’</td>
<td>205</td>
</tr>
<tr>
<td>9</td>
<td>osčastlivit’</td>
<td>343</td>
<td>uplotnit’</td>
<td>201</td>
</tr>
<tr>
<td>10</td>
<td>osvežit’</td>
<td>280</td>
<td>uxudšit’</td>
<td>199</td>
</tr>
</tbody>
</table>
Marginal factitives – POSSIBLE WORDS

• All marginal factitives have low token frequency in the RNC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>omeždunaroditi’</td>
<td>1</td>
<td>uvkusniti’</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>opoxabit’</td>
<td>1</td>
<td>umedliti’</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>opriličiti’</td>
<td>1</td>
<td>ukrasiviti’</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>oser’ezniti’</td>
<td>1</td>
<td>user’ezniti’</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>ostekljaniti’</td>
<td>1</td>
<td>usovremenenti’</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>oržavit’</td>
<td>2</td>
<td>ukonkretiti’</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>osuroviti’</td>
<td>2</td>
<td>ustrožiti’</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>obytoviti’</td>
<td>3</td>
<td>ucelomudrati’</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>ovnešniti’</td>
<td>4</td>
<td>uprozračiti’</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>omuzykaliti’</td>
<td>4</td>
<td>udorožiti’</td>
<td>8</td>
</tr>
</tbody>
</table>
Nonce factitives

- O- and U- nonce factitives correlate with respect to the initial consonant of the stem:

- Variety of initial stem consonants

- Nonce words should not be similar to standard or possible factitives so that the speaker should not hesitate about them.

<table>
<thead>
<tr>
<th>#</th>
<th>O- factitive</th>
<th>U- factitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>osurit’</td>
<td>usaglit’</td>
</tr>
<tr>
<td>2</td>
<td>otovit’</td>
<td>utulit’</td>
</tr>
<tr>
<td>3</td>
<td>oduktit’</td>
<td>udamlit’</td>
</tr>
<tr>
<td>4</td>
<td>ogabit’</td>
<td>uguzvit’</td>
</tr>
<tr>
<td>5</td>
<td>okočlit’</td>
<td>ukampit’</td>
</tr>
<tr>
<td>6</td>
<td>ošaklit’</td>
<td>ušadrit’</td>
</tr>
<tr>
<td>7</td>
<td>očavit’</td>
<td>učopit’</td>
</tr>
<tr>
<td>8</td>
<td>oblusit’</td>
<td>uloprit’</td>
</tr>
<tr>
<td>9</td>
<td>onomit’</td>
<td>unokrit’</td>
</tr>
<tr>
<td>10</td>
<td>obmomlit’</td>
<td>umarvit’</td>
</tr>
</tbody>
</table>
Experiment: score-assignment test

The task: Evaluate the marked word using one of the statements.

Давно пора как-то оприличить наше общение более мягкими выражениями. ‘It’s high time we made our interaction respectable by using kinder statements.’

☐ 5 points - Это совершенно нормальное слово русского языка.
   ‘This is an absolutely normal Russian word’

☐ 4 points - Это слово нормальное, но его мало используют.
   ‘This word is normal, but it is rarely used’

☐ 3 points - Это слово звучит странно, но, может быть, его кто-то использует.
   ‘This word sounds strange, but someone might use it’

☐ 2 points - Это слово звучит странно, и его вряд ли кто-то использует.
   ‘This word sounds strange and it is unlikely that anyone uses it’

☐ 1 point - Этого слова в русском языке нет.
   ‘This word does not exist in the Russian language.’
A few more issues

Order of stimuli:
- Semi-random
- The same for all

Administration:
- Questionnaire
- No limits on time

Age groups:
- School age children aged 14 – 17: 70 participants
- Adults: 51 participants

Website
http://ansatte.uit.no/laura.janda/opyt/opyt.html
Лингвистический эксперимент

Добро пожаловать!

В апреле 2013 мы проводим интересный лингвистический эксперимент, посвященный изучению особенностей языковой интуиции.

Если Вы носитель русского языка, выросли в России и хотели бы принять участие в эксперименте, который занимает 15 – 20 минут, пройдите, пожалуйста, по ссылке http://surveys.questionpro.com/a/tAJSEZPVbR, или заполните анкету в текстовом редакторе и пошлите ее по адресу: anna.endresen@gmail.com.

Если Вы уже приняли участие в нашем эксперименте, мы Вам очень благодарны! В общей сложности, мы рассчитываем на около 200 участников. После окончания сбора анализ ответов результаты эксперимента будут представлены на этой странице. До времени окончания проведения эксперимента лингвистические задачи исследования не могут быть раскрыты. Мы приглашаем Вас вернуться на этот сайт в начале июня 2013 г.

Если у Вас есть друзья, родственники или знакомые, которым было бы интересно принять участие в этом эксперименте, Вы можете порекомендовать им этот сайт.
Questionnaire on-line:
http://surveys.questionpro.com/a/t/AJ5SEZPVbR
A few more issues

Order of stimuli:
- Semi-random
- The same for all

Administration:
- Questionnaire
- No limits on time

Age groups:
- School age children aged 14 – 17: 70 participants
- Adults: 51 participants

Age and prefix turned out to be non-significant
The remaining variables are **acceptability** and **type of word** (standard vs. marginal vs. nonce)
Standard Verbs
MAX = 605
MEAN = 595
MIN = 549
stand dev = 15
variance = 235

Marginal Verbs
MAX = 479
MEAN = 286.4
MIN = 169
stand dev = 67
variance = 4446

Nonce Verbs
MAX = 223
MEAN = 183.4
MIN = 150
stand dev = 19
variance = 360
ANOVA RESULTS overall:
F = 546, df = 2, p-value < 2.2e-16

**STANDARD WORDS**
Words that are standard and conventionalized, might be stored in memory rather than generated on the fly

**MARGINAL WORDS**
Words that are generated by some speakers and can be understood / accepted by some speakers

**NONCE WORDS**
Words that cannot be generated and do not exist (because they do not conform phonotactic laws and / or are not based on productive morphological patterns)

- ob”jasnit’
- uskorit’
- oser’ëznit’
- uvkusnit’
- osurit’
- usaglit’
T-test RESULTS for **standard vs. marginal words:**

**STANDARD WORDS**
Words that are standard and conventionalized, might be stored in memory rather than generated on the fly

**MARGINAL WORDS**
Words that are generated by some speakers and can be understood / accepted by some speakers

$t = 20$
$df = 21$
$p$-value $= 3.173e-15$
95% confidence interval is 277 - 340

*Russian examples:
  - **ob”jasnit’**
  - **uskorit’**
  - **oser’ëznit’**
  - **uvkusnit’**
T-test RESULTS for marginal vs. nonce words:

\[ t = 7 \]
\[ df = 22 \]
\[ p-value = 1.098 \times 10^{-6} \]

95% confidence interval is 71 - 135

**MARGINAL WORDS**
Words that are generated by some speakers and can be understood / accepted by some speakers

- oser’ëznit’
- uvkusnit’

**NONCE WORDS**
Words that cannot be generated and do not exist (because they do not conform phonotactic laws and / or are not based on productive morphological patterns)

- osurit’
- usaglit’
Hypothesis 3 is confirmed, but...

<table>
<thead>
<tr>
<th>Hypothesis 1:</th>
<th>Standard</th>
<th>Marginal</th>
<th>Nonce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 2:</td>
<td>Standard</td>
<td>Marginal</td>
<td>Nonce</td>
</tr>
<tr>
<td>Hypothesis 3:</td>
<td>Standard</td>
<td>Marginal</td>
<td>Nonce</td>
</tr>
<tr>
<td>Hypothesis 4:</td>
<td>Standard</td>
<td>gradual transition</td>
<td>Marginal</td>
</tr>
</tbody>
</table>

Marginal words are much closer to nonce words than to standard words (compromise between Hypothesis 2 and Hypothesis 3)
## Variation across stimuli

<table>
<thead>
<tr>
<th>Word category</th>
<th>Total score</th>
<th>% of maximal score</th>
<th>Highest score</th>
<th>Lowest score</th>
<th>Interval of variation in scores across stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>11,900</td>
<td>98 %</td>
<td>605</td>
<td>549</td>
<td>56</td>
</tr>
<tr>
<td>Marginal</td>
<td>5,729</td>
<td>47 %</td>
<td>479</td>
<td>169</td>
<td>310</td>
</tr>
<tr>
<td>Nonce</td>
<td>3,669</td>
<td>30 %</td>
<td>223</td>
<td>150</td>
<td>73</td>
</tr>
</tbody>
</table>

## Variation across subjects

<table>
<thead>
<tr>
<th>Marginal factitive</th>
<th>Gloss</th>
<th>Number of subjects who gave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 points (“normal word”)</td>
<td>4 points</td>
</tr>
<tr>
<td>usovremenit’</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>opriličit’</td>
<td>9</td>
<td>25</td>
</tr>
</tbody>
</table>
What do these results mean?

- Each type of word has a different behavior
- Marginal words are semantically transparent, but nonce words are not
- **Marginal words** are rated more like nonce words than like standard words
- Speakers are more sensitive to frequency than to semantic transparency
- **Frequency**, which is related to **performance**, is a stronger factor than competence (ability to unpack morphological patterns)
- **Memory** may be a stronger factor than use of productive rules