Как разные языки классифицируют предметы быта

How different languages categorize everyday items

Iomdin B. L. (iomdin@ruslang.ru)
V. V. Vinogradov Russian Language Institute,
Russian Academy of Sciences

Piperski A. Ch. (apiperski@gmail.com)
M. V. Lomonosov Moscow State University

Russo M. M. (rousseau@mail.ru)
Institute for Linguistics, Russian Academy of Sciences

Somin A. A. (somin@tut.by)
Russian State University for the Humanities

Исследуются классификации бытовых предметов на материале более 40 языков. Показано, что большинство классов являются «скрытыми» — не имеют нейтральных общепринятых названий (ср. офиц. предметы личной гигиены и разг. мыльно-рыльное). Кроме того, наборы и состав классов в разных языках существенно различаются.

Introduction

The idea of the present study was born at our academic seminar devoted to developing an explanatory and encyclopedic thesaurus of Russian everyday life terminology. In Iomdin 2009, 2010, 2011 this lexicon has been shown to be treated very differently in dictionaries, industrial standards, and usage; uniform lexicographic definitions of such words are very difficult to produce. The thesaurus is being developed by a group of researchers led by Boris L. Iomdin. The group members helped to perform the study at all stages (organizing the survey, recruiting participants, collecting and discussing the results). We would like to especially thank those members who made many valuable contributions: Anna Kadykova, Anastasiya Lopukhina, Varvara Matissen-Rozhkova, Pavel Vasilyev, Fedor Vinokurov, and Anna Vybornova.

1 Under partial financial support by Russian Foundation for Humanities (Project No. 10-04-00273a), Fundamental Research Program of History and Philology Branch of the Russian Academy of Sciences, and a President grant for leading scientific schools of Russia (No. NSh-4019.2010.6).

2 We would also like to thank Julia Khaleeva who calculated all statistics for us; Vladimir Belikov and Aleksandr Berdicevskis who made valuable comments; Elena Muravenko, Elizaveta Kushnir and Hugo Dobbs who promoted the survey among speakers of various languages; Anastasia Zaytseva who commented on Japanese; professors and students from the
1. Classification

We conceive our dictionary as a thesaurus where similar objects are grouped together, which allows an easy search for information on a certain object or group of objects. However, when trying to classify the lexicon, we were faced with problems of different kinds.


1.2. Covert classes. Several classes that obviously exist in speakers’ minds do not have natural names. For example, most travelers pack their toothbrush, toothpaste, soap, shampoo, sponge etc. together, but no good Russian word exists for this class. If asked, or urged, to use a superordinate, people either resort to awkward official phrases like предметы личной гигиены ‘personal hygiene items’, or highly colloquial occasional words like умывалки [from умываться ‘to wash oneself’]. In fact, most superordinates prevail either in official documents (e.g. парфюмерия ‘perfumery’, бытовая химия ‘household chemistry’, пищебумажные принадлежности ‘stationery supplies’, чулочно-носочные изделия ‘hosiery’), or in colloquial texts such as blogs (мыльно-рыльное ≈ ‘soap and stuff’ / ‘phiz wash’, косметика ‘make-up’, аксессуары ‘accessories’, имотки ‘duds’, прибамбасы ‘gismos’). Consider two examples covering similar topics, where игрушки ‘toys’ is the only item named in the same way:

(1) Согласно договору о патронате воспитателю перечисляются заработная плата и денежные средства на содержание ребенка (питание, приобретение предметов хозяйственного обихода, личной гигиены, медикаментов, канцелярских товаров, игрушек и др.) (Russian Tax Courier, 2008, No.13–14)

[≈ ‘According to the patronage agreement, salary and money for upkeeping the child (nutrition, purchase of household objects, medicaments, stationery, toys, etc.) is transferred to the tutor’s account]


Slavistic Institute of Karl-Franzens-Universität (Graz, Austria) and the Department of Foreign Languages of University of Bergen (Norway); subscribers of http://community.livejournal.com/by_mova; and everyone who submitted answers for our survey.

3 Data obtained via several searches in Russian blogs (blogs.yandex.ru) and in the Consultant-Plus juridical information system (base.consultant.ru).
[≈ ‘2000 roubles for the whole year as flat payment by the parent org: water, soap and stuff, toys, prezzies and gz, office stuff, etc.’, a highly colloquial forum message].

1.3. Vague classes. Some other classes with relatively established names are too fuzzy: for instance, Russian галантерея ≈ ‘haberdashery’ for different speakers might refer to handkerchiefs, ties, gloves, belts, bags, purses, threads, needles, pins, umbrellas, combs, hair rollers, beads, costume jewellery, mirrors, clothes hangers, etc. Similar phenomena were discussed in semantic literature in the 1970s, consider e. g. Kempton 1978. Cruse 1995 reports on a study where some 200 American college students were asked to estimate sixty household items as good or bad examples of furniture.

For some classes, standard dictionaries often provide vague, hardly translatable explications of doubtful usability, e.g. ширпотреб ‘Товары широкого спроса и массового производства’ ['mass demand and mass production goods'], утварь ‘Совокупность предметов, необходимых в обиходе, в какой-л. области жизни’ ['a range of items needed in common use, in one of life spheres']; аксессуары ‘1. Мелкие предметы сценической обстановки, бутафория. 2. Принадлежность чего-л.; сопутствующие предметы’ ['1. Small items of stage set, props. 2. Accessories of something, accompanying items'] (Kuznetsov 1998) (see below for more on Accessories).

Considering all this, we decided to investigate into the subject with the purpose to find out whether unclassifiable items, covert and vague classes are universal or language specific. Semantics of category words was studied a lot (see e.g. Wierzbicka 1985 or Taylor 1995); some studies are also under way in the domain of natural ontologies, cf. Mihatsch 2007. However, this topic has not received much attention cross-linguistically. With this aim in view, we launched a survey, to be outlined below.

2. Survey

Under http://www.lingling.ru/useful/pics-survey-en.php, we posted 33 clickable images depicting the following items: suitcase, pot, notepad, toothbrush, receipts, toy blocks, eraser, sock, glasses, pencil, blanket, passport, gloves, tacks, umbrella, ruler, make-up bag, ear plugs, handkerchief, CD, vase, barrette, charger, keys, spoon, soap, slippers, teapot, lipstick, table cloth, high heels, comb, glass. The following task was given:

Please add two headings for each image according to the examples below:
[image of a chair] chair furniture
[image of a bed] bed furniture
[image of an iron] iron appliances

Write in your native language. Choose words that you use yourself when speaking with your family members. If you find it difficult to add headings to some of the pictures, leave those fields blank.

563 participants aged 12 to 64 (mean age 30) submitted their results in 41 languages: Albanian, Arabic, Azerbaijani, Belorussian, Bulu, Catala, Chinese, Croatian,
How different languages categorize everyday items

Czech, Dutch, English, Finnish, French, Georgian, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Luxembourgish, Norwegian, Occitan, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovenian, Spanish, Swedish, Tagalog, Turkish, Ukrainian.

3. Results

3.1. Categorization difficulties in different languages. Our survey allowed making preliminary cross-linguistic observations on how easily speakers of a given language can use generic terms. The subjects were instructed to leave a field blank if they had difficulties filling it in. If speakers of language $L_1$ fail to provide generic terms more often than speakers of language $L_2$, then $L_2$ is probably richer in generic terms than $L_1$ and uses them more frequently.

However, when analyzing the data, certain precautions had to be taken. First of all, samples must of course be rather large: it is not enough to count the mean number of blank fields in the responses of e.g. three speakers of a language. Second, it is well understandable that not all participants of the survey invest as much zeal and enthusiasm into this work as the researchers would wish. Each response had to fulfil two criteria in order to be counted: (1) all 33 specific terms should be provided; (2) at least 10 generic terms should be provided (in other words, no more than 23 gaps are allowed).

This left us with representative samples ($\geq 15$ responses) for five languages: Belorussian, English, German, Norwegian, and Russian. Even though this sample is obviously ill-balanced genealogically as well as geographically, we can see that the mean number of gaps varies significantly even within this sample:

<table>
<thead>
<tr>
<th>Language</th>
<th>Responses (total)</th>
<th>Responses (suitable)</th>
<th>Mean number of gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norwegian</td>
<td>21</td>
<td>18</td>
<td>8.2</td>
</tr>
<tr>
<td>German</td>
<td>44</td>
<td>37</td>
<td>5.1</td>
</tr>
<tr>
<td>Russian</td>
<td>230</td>
<td>188</td>
<td>4.2</td>
</tr>
<tr>
<td>Belorussian</td>
<td>32</td>
<td>17</td>
<td>3.2</td>
</tr>
<tr>
<td>English</td>
<td>21</td>
<td>18</td>
<td>1.7</td>
</tr>
</tbody>
</table>

This shows that speakers of Norwegian had the greatest difficulty finding generic terms, while speakers of English had the fewest problems with this$^4$. The fact that such closely related and similar languages as Russian and Belorussian pattern similarly

$^4$ The hypothesis that English develops category words more easily than other languages could be supported by the following funny observation. In current English usage, suffixes -wear and -ware tend to be mixed: Google search yields thousands of occurrences for footwear, eyeware, outerware etc. as well as cookwear, silverware, glasswear etc. This might mean that speakers of English start to consider this a single suffix with a general meaning ‘category of artifacts’.
in respect to the number of gaps (they occupy neighbouring rows in the table above) supports the assumption that the amount of gaps in generic terms is not arbitrary but constitutes an important characteristic of a language.

One of the important consequences of this fact is that generic terms cannot serve as markers of linguistic identity because of high fluctuation in their use. For example, it is shown elsewhere in this volume (Piperski 2011) that Serbian and Croatian speakers make less notice of differences between their languages that concern generic terms than of differences in specific terms even though the sociolinguistic situation in the Balkans favours language awareness.

Furthermore, as we will show below, languages vary considerably in how much speakers agree when using generic terms for the same items.

### 3.2. Unclassifiable items.

In total, 11 items were not classified at all by 20% or more respondents: keys (48.4%), ear plugs (42.4%), table cloth (32.1%), umbrella (30.8%), glasses (29.5%), handkerchief (25.5%), make-up bag (24.8%), CD (24.0%), comb (23.5%), receipts (22.4%), barrette (21.5%). The answers of respondents who did submit generic terms for these items exhibit great variation. E.g. in Russian, 230 respondents submitted 63 different unique superordinates for ear plugs (with a maximum of 7 answers (3.0%) reached in vague личные вещи ‘personal belongings’) and 53 for keys (a maximum of 11 answers (4.8%) reached in helpless ключи ‘keys’), while e.g. for blocks, only 8 different superordinates were offered, other than игрушки ‘toys’ used by 207 (90.0%) respondents. For some of these items, these numbers were comparable in all languages, but there are exceptions.

The umbrella was classified by more than half of Japanese speakers as 雨具 [amagu]; web search confirms that this category is indeed well established in the Japanese language and includes umbrellas, raincoats, rubber boots, tents etc (see e.g. http://shopping.yahoo.co.jp/category/2585). Rain gear in English was used by 25% respondents, Regenschutz in German by 17.5% respondents, regnustyr/regntøy in Norwegian by 10.5% respondents each. According to the submitted results, this category is virtually non-existent in other languages participating in the survey: e.g. in Russian only one respondent (0.5%) used защита от дождя ‘rain protection’ and another one used средство от дождя ‘aid against rain’.

An interesting tendency we observed in many languages is categorizing unclassifiable items like these (but not only them!) under a special extremely nebulous class called Accessories. This word was borrowed into and widely used in almost half of the languages studied, and is the most frequent category word used by the survey participants.

---

5 Of course we should bear in mind that all Belorussian speakers live in Russian language environment, so that they are inevitably strongly influenced by Russian.

6 Note also the Tagalog word gamit, which was used by our surveyees as the superordinate for a considerable number of different objects. Cf. “The term gamit means several things. Its definition as a Filipino word is legion. In Tagalog colloquial term, it means an object that has several utilitarian purposes or simply a utilitarian object with specific usage in a particular space and time” (Ruston Ocampo Banal Jr. Gamit: subjectifying objectivity).
3.3. Well-established classes. Surprisingly, only one class appears to have a distinct name in most languages, namely Toys. For almost all languages, there is one word with this meaning that gets more than 60% of answers (and usually much more, cf. игрушки (91%) in Russian mentioned in the previous subsection) and little variation. An interesting example is presented by the Documents class: it appears to be well-established in most languages, but not quite so in some others (mostly Germanic ones7). For passport, we got документы in Belorussian (100%), dokumenty in Polish (100%), документы in Russian (92%), documentos in Spanish (67%), etc. However, in Norwegian, the best result was reisedokument (21%), with 17% of respondents who couldn’t provide any answer and the rest using various other words (документ, идентификационный документ, документ, статут, документ, документ, документ, документ, документ). Similar situation happened in Dutch: document (25%) and many other answers (reisdocument, identiteitspapier, identificatiebewijs, officiële papieren, reisperken, reisdocumenten), English, Swedish and (outside the Germanic branch) also Japanese, where 36% of respondents could not come up with any answer and different words were offered by the rest.

We also discovered some language-specific classes. These include Rain gear in Japanese described above and Luggage in English and Polish. For the suitcase, 66% used bagaż in Polish and 65% used luggage in English, while the next closest result was 32% for German Gepäck, with similar or lower numbers for other languages. 22% of Russians did not come up with any answer, and the leading one was сумки (29%).

3.4. Covert classes. Several other classes have no good names in most analyzed languages. These are: (a) personal hygiene, (b) appliances, (c) stationery, and (d) data storage. For all items belonging to these classes, the variation (calculated as the total number of different answers in all languages divided by the number of non-empty answers) was twice as high as for the well-established class of toys. Most answers here are compounds or word combinations8, and no answer was given by 50% of respondents or more. The highest results for all languages in these groups are (a) Hygieneartikel in German (45%), (b) denkaseihin in Japanese (45%), (c) skrivesaker in Norwegian (42%), and (d) носі́бі́ інформаці́ї (32%) in Belorussian. Interestingly, in many cases there is one leading word or root (depending on the morphological structure of the language) which occurs in various compounds or word combinations. E.g. generic names for items from group (c) in Russian mostly contain the root канц-, and counting together all answers containing this root (канцелярские товары, канцтовары, канцелярские принадлежности, канцелярская продукция, канцелярские

---

7 One of the striking results of our survey that requires much more data to be confirmed is that cognate languages often use similar categorization patterns even when words they choose are not related.

8 According to Mihatsch 2007, superordinates are often morphologically more complex than subordinates; our data clearly support this hypothesis.
мешочки, канцелярские предметы, канцелярское изделие, канцелярия, канцелярка, канцелярщина, etc.) we get as much as 81%. The same applies to roots toilet- and hygien- for group (a) or to technic- and electro- for group (b), in various forms depending on the language.

3.5. Overlapping classes. Some classes exist in many languages but include different items. Let us give two examples. For four images that are grouped together as посуда in Russian, different languages have several classes. Cf. the summarization table, characteristic for lexical typology (cf. Hjelmslev 1957, Haspelmath 2001, Koch 2005), which only uses data from languages where 50% or more respondents agree on certain generic term:

<table>
<thead>
<tr>
<th>Item</th>
<th>Russian</th>
<th>Belorussian</th>
<th>English</th>
<th>Norwegian</th>
<th>German</th>
<th>Japanese</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pot</td>
<td>посуда</td>
<td>посуд</td>
<td>cook-ware</td>
<td>kjøkkenutstyr</td>
<td>?</td>
<td>調理器具 [chor-ikigu]</td>
<td>أدوات (ال) مطبخ [adwa:t (al)-ma[ba]</td>
</tr>
<tr>
<td>Teapot</td>
<td></td>
<td></td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wineglass</td>
<td></td>
<td></td>
<td></td>
<td>Geschirr</td>
<td></td>
<td>食器 [shokki]</td>
<td></td>
</tr>
<tr>
<td>Spoon</td>
<td></td>
<td></td>
<td></td>
<td>bestikk</td>
<td>besteck</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Russian word посуда is indeed rarely translated as a similar generic term in English: interpreters use various strategies to avoid direct translation. Cf. examples from parallel corpora:

(3) Проходивший в это время по коридору старший доктор, услыхав звон разбитой посуды и увидав выбежавшую раскрасневшуюся Маслову, сердито крикнул на нее (Л. Н. Толстой, Воскресение).

(4) The head doctor, who was passing at that moment, heard the sound of breaking glass, and saw Maslova run out, quite red, and shouted to her (Lev Tolstoi, Resurrection, translated by Louise Maude).

(5) Хохлушка в платке внесла поднос с посудой, потом самовар (А. П. Чехов, Красавица).

(6) A Little Russian peasant woman in a kerchief brought in a tray of tea-things, then the samovar (Anton Chekhov, The Beauties, translated by Constance Garnett).

A less clear situation takes place with the Clothing class. In Russian, it distinctly falls into two subclasses: одежда ‘garments’ and обувь ‘shoes’. In most other languages, the respondents disagree as to which of the four items in the survey fall
into which class. The following table summarizes the results (words in brackets correspond to answers that received 40% to 50% votes).

<table>
<thead>
<tr>
<th>Item</th>
<th>Russian</th>
<th>Belorussian</th>
<th>English</th>
<th>Norwegian</th>
<th>German</th>
<th>Japanese</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>High heels</td>
<td>обувь</td>
<td>абутак</td>
<td>?</td>
<td>?</td>
<td>Schuhe</td>
<td>靴 [kutsu]</td>
<td>?</td>
</tr>
<tr>
<td>Slippers</td>
<td>(footwear)</td>
<td>(sko)</td>
<td>(Schuhe)</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socks</td>
<td>одежда</td>
<td>(вопратка / адзенье)</td>
<td>(clothing)</td>
<td>klær</td>
<td>Kleidung</td>
<td>衣類 [irui]</td>
<td>?</td>
</tr>
</tbody>
</table>

Google search results\(^{10}\) seem to confirm that clothes/clothing and shoes/footwear do not constitute same-level classes in English, as do одежда and обувь in Russian:

<table>
<thead>
<tr>
<th>Search Term</th>
<th>Google Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>clothing site:uk</td>
<td>48,500,000</td>
</tr>
<tr>
<td>shoes site:uk</td>
<td>47,800,000</td>
</tr>
<tr>
<td>footwear site:uk</td>
<td>2,190,000</td>
</tr>
<tr>
<td>“clothing and shoes” site:uk</td>
<td>33,400</td>
</tr>
<tr>
<td>“clothing and footwear” site:uk</td>
<td>105,000</td>
</tr>
<tr>
<td>“clothing such as shoes” site:uk</td>
<td>2,050</td>
</tr>
<tr>
<td>одежда site:ru</td>
<td>21,700,000</td>
</tr>
<tr>
<td>обувь site:ru</td>
<td>14,800,000</td>
</tr>
<tr>
<td>“одежда и обувь” site:ru</td>
<td>16,700,000</td>
</tr>
</tbody>
</table>

The same might be the case in Arabic, where ملابس [mala:bis] ‘clothing’ is used much more frequently than أحذية ['aḥḍi:ya] ‘shoes’. In the survey, no Arabic speakers used أحذية:ya, and two of them even referred high hills to ملابس.

\(^{9}\) In Belorussian, two different superordinates are used for socks, both of which have received more than 40% votes. Вопратка is explained in dictionaries as outerwear, but is used to describe all clothing as can be seen from the survey and confirmed by web searches. This could have happened either under the influence of Russian that does not have a separate word for outerwear (only word combinations like верхняя одежда) and does not specially name it unless needed, using the neutral word одежда for all types of clothing (so вопратка develops the same meaning as одежда and starts competing with адзенье), or by following the tendency of lexical differentiation of closely related Russian and Belorussian languages (the word адзенье is akin to the Russian одежда).

\(^{10}\) Data obtained on January 31, 2011. The total number of hits given by Google varies considerably and may only serve as a very approximate estimate.
4. Possible areas of further research

4.1. Language acquisition and bilingualism studies. Superordinate categories play an important role in language acquisition. Reportedly small children master well enough many category names that are well established in a language, including those of everyday objects they use. This might be a way of finding out which categories play a key role in a language. Consider e.g. a characteristic quotation about Russian children: “К 3 годам среди родовых наименований появляются более «книжные» термины: фрукты, овощи, животные, посуда, насекомые, обувь, одежда, транспорт и т.п. <…> В речи ребенка появляются конструкции, <…> соотносящие видовое и родовое <…>: Каструля — это посуда. Чашка — это посуда” (Yeliseeva 2006) [‘3-year-olds start using more “bookish” generic terms: fruits, vegetables, animals, dishes, insects, footwear, clothing, transport, etc. The child starts producing constructions correlating specific and generic terms: Pot is dishes. Cup is dishes’]. It is clear from the awkward translation we provided that Russian-speaking kids learn other hierarchies that English-speaking ones. Bilinguals are especially interesting here, since they might mix up different classification strategies (see e.g. Malt & Pavlenko 2009 who report a study of English-Russian bilinguals naming cups, mugs and glasses of different types).

4.2. Folk biology. Further research of everyday items classification in different languages might use the experience of folk biology, which studies linguistic classifications of animals and plants (Berlin 1992, Atran 1990). It also describes covert categories that have no special names in languages but apparently exist in speakers’ minds. Often these even include the highest taxa, which are animals and plants (Berlin 1973: 266–267). Latin started to use plantae for all plants only in the 13th century, English and French accepted this term only in the 16th century (Kupriyanov 2005: 14). It is suggested that generic terms for animals and plants appear when a language becomes a written one (Slaughter 1982). Covert classes in folk biology and in everyday items are evidently similar. In folk biology, several techniques for revealing covert classes through speaker surveys have been developed (cf. Hays 1976), which could be used in deeper studies of everyday items classification. Probably their names appear in professional sublanguages before progressing into standard language and then into colloquial speech; this is subject to further investigation.

Scholars of folk biology believe that folk taxa in world languages are organized into a hierarchical system of levels, or ranks: folk kingdom (e.g. animal, plant), life form (e.g. bug, fish, bird, mammal/animal, tree, herb/grass, bush), folk species (gnat, shark, robin, dog, oak, clover, holly), folk specific (poodle, white oak), and folk varietal (Berlin 1992: 15–25). The levels are thought to be universal, unlike the taxa. Folk species can unite into folk series: chains of species that look similar to the speakers. These chains only rarely have names (see Merkulova 1967) and obey several universal tendencies. For instance, longer series less frequently get names. This might be explained by the fact that speakers cease to consider remote elements of a long chain as similar enough (Kupriyanov 2005:15). Such phenomena might be present in our case, too. S. Atran (Atran 1990) mentions that linguistic
classifications of artifacts provide much more freedom for intersecting classes and several alternative groupings: e.g. a piano could be considered a musical instrument or a piece of furniture. However this might vary for different items and in different languages.

4.3. Folksonomics. This is another domain thoroughly investigated in recent years. It studies classification emerging from the collective action of users who tag resources with an unrestricted set of key terms, such as flickr.com (Veres 2006). Since on many websites like these objects of everyday use are discussed and tagged, it would be interesting to compare these tag sets with the categories we describe.

References


