A Corpus-Based Study of Human Impersonal Constructions in Russian

Olga Rudolf
University of Jena

1. Introduction

Human impersonals are constructions whose subject (sometimes object) is human and non-referential (cf. e.g. Siewierska’s (2012) ‘R-impersonals’), and which ranges over a group of individuals (cf. Gast & van der Auwera 2013). In Russian such constructions usually lack overt subject pronouns and impersonal reference is marked on specific forms of verbs, such as a 3rd-person plural form as in (1a), or a modal impersonal form (without a dative argument) as in (1b).

(1a) Segodnâ v Ukrainâ prazdnuût den’ nezavisimosti.
‘Today they celebrate the independence day in Ukraine.’

(1b) Nužno byt’ čutkim k drugim lûdâm.
‘One should be sensitive to other people.’

Human impersonal pronouns and their equivalents have been extensively discussed in the literature, especially with a focus on particular languages (cf. Zifonun 2001 and Linthe 2010 on Germ. man and du, Creissel 2008 on Fr. on, etc.), sometimes dealing primarily with a single strategy (e.g. Moltmann 2010 on Engl. one, Paducheva 2012 on Russ. 3pl). There are also some typological and comparative studies, such as Cabredo Hofherr’s (2003) and Siewierska & Papastathi’s (2011) investigation of 3rd-person plural forms in a number of European languages, or van der Auwera et al.’s (2012) study of impersonal pronouns in English, Dutch and German. This study deals with the translation equivalents of the German human impersonal pronoun man in Russian with a focus on their semantic properties.

The current study is part of a larger typological project entitled Towards a typology of human impersonal pronouns (Typoimp) conducted at the English

---

1 Abbreviations are made according to the Leipzig Glossing Rules.
department of Jena University and a research group from the CNRS. The terminology developed jointly by the participants of the project, partly adopting concepts from earlier work on impersonals, will also be used in this study. Therefore, only a short overview (with some personal remarks) of the theoretical background underlying the study is presented in section 2. Section 3 provides a description of the data, a discussion of the methods of data processing, and a presentation of the results of the study, followed by some conclusions in section 4.

2. Human impersonal strategies

2.1. The scope of the study

Many studies on impersonalization have been conducted on languages that exhibit pronominal impersonal forms (e.g. man in some Germanic languages), or that use personal pronouns to express impersonal meanings (e.g. English you and they). Such forms are commonly termed ‘impersonal pronouns’ or also ‘arbitrary’ elements (cf. Malamud 2012), especially in languages with zero subject pronouns. Since one of the target languages of this study is Russian, which has mostly non-pronominal impersonal strategies (cf. ex. 1), it is more convenient to talk about ‘impersonal constructions’. In grammars and some of the theoretical literature on Russian, such impersonal strategies are approached at the sentential level, e.g. односоставные предложения (‘subjectless clauses’), and are conventionally referred to as обобщенно-личные/неопределенно-личные предложения (kinds of impersonal clauses). By contrast, German uses several pronominal impersonal forms. The most frequently used impersonal in this language is the dedicated impersonal pronoun man, which serves as a basis for my corpus study. Since German and Russian show completely different syntactic behavior with regard to their ways of expressing impersonal meanings, a unified approach and, therefore, an appropriate common term should be used for their comparison. Thus, as a generalization over the German impersonal pronoun man and its Russian equivalents, the terms ‘impersonal constructions’ or ‘impersonal strategies’ will be used.

Both обобщенно-личные and неопределенно-личные предложения are included in the study, but this classification will not be adhered to. Russian grammars (e.g. Valgina 2003) contain a thorough description of the syntactic, semantic and

2 The principal investigators are V. Gast (Jena) and P. Cabredo Hofherr (Paris); cf. http://www.personal.uni-jena.de/~mu65qev/improtoype/
pragmatic properties of these clauses, and the aim of this study is not to criticize this work or argue against the inclusion of certain constructions in one of these clause types. On the contrary, this description can serve as a basis for the investigation of selected impersonal constructions in Russian. The problem is that *obobšenno-ličnye predloženiâ*, for example, encompass several different constructions according to the grammars, i.e. the 3rd-person plural, the 2nd-person singular and the 1st-person plural as well as the 1st-person singular (referred to as 3pl, 2sg, etc. in the following) forms of the verb in their different tense forms and moods. This study takes a closer look at the semantic differences between these constructions, and at the types of context where they occur.

2.2. Parameters describing the readings of the human referent and the contexts of their occurrence

German *man* can be found in various types of context, with different interpretations of the human referent (cf. Malamud 2012), as in (2a) (universal reading) and (2b) (existential reading):

(2a) Auf dieser Insel spricht man Englisch.
    ‘On this island, they speak English.’
Roughly ‘Everybody on the island speaks English.’

(2b) Man hat hier alles umgeräumt.
    ‘Someone has rearranged everything here.’

In Russian, there is no such polysemous construction, and the speakers have to choose a more specific strategy from a range of different ways of expressing impersonalization. On the basis of the German-Russian corpus data, which will be presented in the next section, I have distinguished eleven (groups of) equivalents of *man* in Russian.

In the literature on impersonals, several claims have been made about the trigger conditions of distinct impersonal meanings, e.g. that generic contexts are a prerequisite for a universal interpretation (cf. Moltmann 2010; Malamud 2006, 2012). They will be taken into account in working out a classification of the potential licensing conditions for impersonal uses. But the primary starting point will be parameters (i) characterizing the state of affairs (such as veridical vs. non-veridical propositions, episodic vs. generic contexts) and (ii) the set of human participants (such as quantification, perspective, etc.) as described by Gast and van der Auwera (2013) in their typological study of human impersonal pronouns in European languages. These categories seem to capture the major
semantic and grammatical meanings of human impersonal referents and contextual conditions, so they will be partly adopted in this study. However, some modifications are required, since Gast and van der Auwera’s (ibid) classification of the categories is appropriate for the arrangement of semantic maps (which is the focus of their study), but not to the same extent for a (quantitative) description of the corpus data.

Let us first consider the parameters describing the state of affairs or the context, as it will be termed in this paper. The definition of ‘veridicality’ will be adopted as it is described in the paper (Gast & van der Auwera 2013, p. 23), i.e. “[t]he proposition denoted by a (bare) clause is veridical if it is assumed to be true, in the context in which it is uttered […].” According to the authors, the bare clause consists of the predicate and its arguments plus any adjuncts and internal negation.

At this point, I would like to emphasize that not only a sentential but also a wider discourse setting should be considered, as sometimes a sentence may figure as non-veridical in a situation going beyond the immediate context. This is taken into account in the analysis of the data – the interfaces of the corpora chosen for this study allow for resorting to a broader context of any given sentence.

The further pairs of context features, singled out by the authors (ibid.), are ‘generic’ vs. ‘episodic’ and ‘modal’ vs. ‘non-modal.’ Before elaborating on these categories, let us look at their relation to (non)veridicality. In the current study veridicality and genericity will be considered as two independent categories: they are not redundant and can presumably influence the choice of impersonal strategies in different degrees and ways. The relationship between veridicality and modality seems to be different, since almost all modal contexts are non-veridical. An important observation should be mentioned: modal contexts constitute a favorable condition for the use of impersonal strategies, and modal contexts are among the most frequently occurring ones in the German and Russian data. Therefore, it is tempting to take (non)modal contexts as a separate variable, along with (non)veridicality. However, not all non-veridical sentences are modal: they also include conditional clauses, sentences with predicates expressing wanting or wishing, direct questions etc. Crucially, in Russian these subtypes of non-veridical sentences – and not only modal ones – may be at least partially responsible for the choice of certain impersonal strategies. Thus, the category ‘non-veridical’ is more general and descriptive than ‘modal’, and therefore only the former will be taken as an independent parameter.

Let us now turn to the dichotomy ‘episodic’ vs. ‘generic.’ Generic contexts can be defined as those contexts in which the speaker “abstract[s] away from particular events and facts” (cf. ‘characterizing sentences’ in Krifka et al. 1995, p. 4)
and “ascribe[s] a general property to all members of a class” (Langacker 1997, p. 191) as in (3a). Episodic expressions, on the contrary, describe a specific event anchored in time (cf. (3b)).

(3a) Tam živut v užasnoj nišete.  
‘They live in terrible poverty there.’

(3b) Včera otkryli novuû stanciû metro.  
‘Yesterday a new metro station was opened.’

(3c) V ètih nomerah ubiraût každyj den’.  
‘These rooms are cleaned every day.’

Generic contexts are rare, however: they are restricted to the sentences containing generic verbs, which do not occur often in impersonal constructions. There is another type of context that falls under the above definition of generic contexts, namely ‘habitual’: habitual sentences do not refer to particular occurrences of an event (cf. Langacker 1997), but generalize over occurrences of it (cf. (3c)). Therefore, instead of ‘generic’, the category ‘generalizing’ context will be used, which subsumes both generic and habitual contexts.

As to the interpretation of the referent, the first dichotomy describing it concerns the type of quantification over the domain of human individuals. A referent with a universal interpretation can be roughly paraphrased as everyone, whereas an existential interpretation means someone. These readings have been exemplified with the two German sentences in (2a) and (2b) above, respectively.

The second pair of interpretation features, ‘internal’ vs. ‘external’ ‘perspective’ tells us whether the speaker identifies him/herself with the set of referents or not (Gast & van der Auwera 2013, pp. 24-25, cf. also Moltmann’s (2010) ‘generalized detached self-reference’). Consider the examples in (4).

(4a) One should have fun in life.  
(4b) They may declare a state of emergency.

In (4a), unlike in (4b), the speaker makes a generalization over the referent and obviously identifies him/herself with this group. In expressing empathy towards the referent (cf. Malamud 2006, p. 5), the speaker usually simultaneously appeals to the addressee to take this perspective. Whereas German man can express both perspectives, Russian impersonal strategies such as 2sg and 3pl are expected to be sensitive to this distinction.

---

3 This category was born in the project team’s discussions.
The implication made by Gast and van der Auwera (2013, p. 25) that only the universal interpretation can take either internal or external perspective, and that existential instances are almost always external, is also intuitively true for the German-Russian data. However, I will assume that the two pairs ‘universal/existential’ and ‘internal/external’ are two independent sets of parameters and test them for independence with a statistical test.

All the dichotomous categories described in this section can figure as binary predictors of the Russian translation strategies of German *man*. The following section will briefly describe the data and the statistical methods of their evaluation as well as present the results of the analysis.

3. Data, methods of their analysis, results

3.1. Data

This study was conducted on literary texts extracted from two parallel German-Russian corpora, i.e. the parallel subsection of the Russian National Corpus with original German texts and their translations into Russian, and the Parallel Corpus of Slavic and Other Languages (ParaSol) containing both German and Russian original texts.

The procedure of the search for the relevant data was as follows: the German specialized impersonal pronoun can be easily searched for without additional checking of its impersonal status. This method does not take into account other, though less frequent, impersonal strategies in German (e.g. the personal pronoun *du* ‘you’ used impersonally, impersonal passives, etc.); moreover, a reverse procedure, i.e. searching for Russian impersonal strategies and identifying their German equivalents, may provide fruitful results for both Russian and German. However, filtering out non-impersonal occurrences of 2sg or 3pl forms would have to be done manually, normally resorting to a broader context for each sentence. This is very time-consuming and will be reserved for future research.

The inventory of Russian translation strategies corresponding to *man* identified on the basis of the data is quite diverse. Figure 1 presents eleven translation strategies with their frequencies.

The three major impersonal strategies are 3pl, modal impersonal and 2sg (also predominantly discussed in Russian grammars, e.g. Valgina 2003), followed by reflexive verbs, infinitives, participial constructions, and adverbs. Not as frequent are nominalizations, the group of “non-impersonal pronouns” (a heterogeneous category which contains different kinds of pronouns that are not impersonal per se but express impersonal meanings), some instances of the pronoun *my* “we”, and finally the noun *čelovek* (“a man”) with a universal interpretation.
This paper only presents the results of the statistical analysis of the three most frequent strategies, which together, as can be seen in Figure 1, cover about 70 percent of the whole data, and thus can be regarded as quite representative.

Figure 1: Frequencies of the Russian translation strategies

3.2. Statistical analysis of the data

The main goal of this study is to predict the occurrence of the three major Russian translation strategies of German *man* under given context conditions. The statistical procedure applied is quite complex and comprises several steps. It includes the testing of the hypotheses as well as choosing predictor variables for the multinomial logistic regression analysis. The task can be formulated in terms of the following questions:

I. Do all of the context and interpretation variables mentioned in section 2.2 have a significant influence on the translation strategies? This question is crucial for sorting out the variables that appear to have no impact on the translation strategies.

II. Do the context variables GENERALIZING and VERIDICAL⁴ have a significant influence on the interpretations of the referent, represented by the variables UNIVERSAL and INTERNAL? This step includes testing of the hypothesis that a specific type of context triggers a specific interpretation of the human agent. If any such correlations are identified, taking their combinations for further statistical analysis can be considered. Otherwise, a variable showing no

---

⁴ It should be kept in mind that all the variables are binary (having the values TRUE/FALSE).
correlation with the others will be considered as an independent variable (a predictor) in the next step.

III. What kind (positive or negative) and degree of influence do the independent context and interpretation variables (or their combinations) have on the translation strategies of German man in Russian? This is the final step which involves making predictions about the choice of the translation strategies in Russian under specific semantic conditions.

The statistical tests and methods applied in the study include Pearson’s Chi-squared test for independence, which shows whether two categorical variables are independent of each other (cf. Gries 2013), and a multinomial logistic regression analysis, which allows for presenting the results as a hierarchy of probabilities predicting the occurrence of each strategy under specific conditions.

Let me briefly explain how the tests were used with regard to the three tasks described above. First, each context and interpretation variable was paired with the set of translation strategies to detect any variable that has no influence on them. The results showed that all the context and interpretation variables do have a significant (positive or negative) impact on the translation strategies in Russian. Then all the variables were tested pairwise for independence. The variable GENERALIZING has a much stronger association with the interpretation variables (especially with UNIVERSAL, less markedly with INTERNAL) than the variable VERIDICAL. Table 1 provides Cramer’s V values for measuring the strength of an association between each pair of variables and its verbal interpretation. These values can range from 0 to 1, signifying no association to an absolutely perfect association (Gries 2013, p. 186).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Level of association</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal + universal</td>
<td>0.91</td>
<td>very strong assoc.</td>
</tr>
<tr>
<td>Universal + generalizing</td>
<td>0.363</td>
<td>strong assoc.</td>
</tr>
<tr>
<td>Veridical and the other</td>
<td>0.105 - 0.198</td>
<td>very weak or weak</td>
</tr>
</tbody>
</table>

As can be seen from the table, the only variable that has minimal or weak association with the other variables is VERIDICAL. It can be concluded that, on the one hand, this variable is not as related to (or is not as good a predictor for) the interpretation of the human referent as the variable GENERALIZING (e.g. as a predictor for a universal reading). On the other hand, VERIDICAL can be considered as a relatively good candidate for an ‘independent’ predictor.
in the regression analysis. The other variables show a strong degree of association between one another, which means that generalizing contexts indeed tend to trigger universal interpretations of the referent, and universal referents, in turn, take an internal perspective. Hence, these variables should not be taken as independent variables for the multinomial logistic regression but rather grouped into one, “generalizing + universal + internal”, and a representative of this group should be chosen. From these three categories, internal has the most precise application, that is, if the perspective is internal, then the quantification is very likely to be universal, and this kind of referent interpretation is expected to occur in a generalizing sentence. If we consider the reverse logical chain, some problems arise: universal reference does by no means regularly take an internal interpretation (such as in ‘Did you know that in India they don’t eat beef?’); generalizing contexts, in turn, especially their subtype ‘habitual’, readily take existential referents. Therefore, a more reliable indicator of the group is ‘internal’. In the regression analysis, wherever the results for internal are shown, it should be kept in mind that internal also stands for universal and generalizing.

3.3. Discussion of the results

Table 2 presents the most important results of the logistic regression analysis.

Table 2: Results of the logistic regression analysis conducted on strategies 3pl, 2sg, and modal impersonal

<table>
<thead>
<tr>
<th>Alt. category: predictor</th>
<th>B (logistic coefficient)</th>
<th>P-values (the cutoff = 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3pl:verTRUE</td>
<td>1.32749</td>
<td>2.811e-05</td>
</tr>
<tr>
<td>Modal:verTRUE</td>
<td>-1.36637</td>
<td>2.414e-05</td>
</tr>
<tr>
<td>3pl:intTRUE</td>
<td>-4.50129</td>
<td>8.145e-10</td>
</tr>
<tr>
<td>Modal:intTRUE</td>
<td>-1.20211</td>
<td>0.122516</td>
</tr>
</tbody>
</table>

The first column shows the translation strategies, except for 2sg, which was chosen as a ‘reference level’, and the values of each binary predictor, namely ‘true’ in all cases. The reference level strategy (2sg) is the one for which the coefficients are set to zero, and to which each result for any alternative strategy is compared. The coefficient B (log-odds) shows the degree of the influence of the variables ‘veridical.true’ and ‘internal.true’ on the translation alternatives: the more the coefficient deviates from 0, the stronger is the influence of the variable on the translation. Positive and negative signs indicate positive and negative influence of the predictor on the response variable. The p-value
in the last column indicates whether each predictor contributes significantly to the model. With the cut-off level of 0.05, the coefficient for the impact of ‘internal.true’ on the strategy modal is not significant and can be ignored. This result is in accordance with linguistic observations and intuitions about the behavior of the Russian impersonal modals and modal contexts in general: most modal contexts have a flavor of a rule of general applicability which involves both the speaker and the addressee. Therefore, it can be assumed that internal does play a significant role for the choice of the modal strategy. All the other results in Table 2 are of high significance. Now the probabilities can be hierarchically ordered and formulated in the form of the following claims:

(5a) Veridicality has a significantly positive impact on the strategy 3pl, but a moderate negative one on the strategy 2sg and a significantly negative on the modal.

(5b) The strategies modal and 2sg, rather than 3pl, are expected to be found in generalizing sentences with internal and universal reference.

The conclusions based on the figures can now be compared with the hypotheses and the relevant theoretical claims made in the literature on impersonals. For example, with regard to 3pl, Malamud (2006, p. 5) observes both “generic or almost-universal interpretation” as well as “(seemingly) existential interpretations”. This compatibility of 3pl with both universal and existential interpretations, which tend to be triggered by generalizing and episodic contexts respectively, is regularly observed in the data, as in (6a,b):

(6a) Ne znaâ, kak postupaût v takih slučââh, Not knowing how behave.3PL.PRS in such cases
Steþa podnâlsâ na trásuþiesâ nogi i poþel po molu k beregu.
‘Not knowing how to behave in such a case, Styopa got up on his trembling legs and walked along the jetty towards the shore.’ [ParaSol]

(6b) poþemu vas, sobstvenno, dostavili k nam?
And why you.ACC actually bring.3PL.PST to us?
‘And why, actually, were you brought here?’ [ParaSol]

This dual feature of 3pl is not rejected or contradicted by claim (5b). The focus of the statistical analysis is rather on the comparison of the strategies, i.e. on the relative likelihood of the occurrence of a certain strategy under a given condition in relation to the other strategies. Thus, from this perspective, if we have an episodic sentence with an existential and external interpretation of the referent, 3pl is indeed the most likely choice from the three strategies. 2sg loses its impersonal meaning as soon as it occurs in an episodic sentence (as in (7b)):
As to the interpretation of the referent of 2sg constructions Malamud (2006, p. 5) writes: “Arbitrary interpretation of 2nd-person pronoun always has a sense of addressee and speaker inclusion […]. The sense of inclusion of conversational participants in the impersonal 2nd-person pronouns stems from an appeal (on speaker’s behalf) for (addressee’s) empathy”. This claim is supported in this paper, though I have used the term ‘(internal) perspective’ (as discussed in section (2.2)) instead of ‘inclusiveness’. It can be rightly claimed that 2sg always implies inviting the addressee to identify with the referents in question, i.e. take an internal perspective, even if the addressee seems to be excluded from the range of the referents (cf. (7a)). This may be due to the deictic characteristic of the 2sg form.

A peculiar feature of Russian 3pl (as opposed to, for example, English they) is that it can also take an internal perspective and even be explicitly inclusive:

(8) Sovr̄senno estestvenno, čto vas prinimaût za sumasšedšego.
quite natural that you take.3PL.PRST for madman
‘It’s quite natural that you’re taken for a madman.’ [ParaSol]

This should be kept in mind when interpreting the statistical figures, but again, figures show tendencies, and the tendency of 3pl (especially in contrast to 2sg) is to take external perspective in sentences like (9).

(9) Trebuû, čtoby menâ nemedlenno vypustili.
I demand that me immediately release.3PL.PFV
‘I demand to be released immediately!’ [ParaSol]

Modal impersonals are found primarily in generalizing sentences with a universal and internal referent. The strategy does occur in a few episodic sentences, as in (10), but the existential interpretation of the referent with external perspective is questionable:

(10a) Čerez samoe korotkoe vremâ možno bylo uvidet’ Ivana
after most short time can.IMP be.N.PST see.INF Ivan
(10b) Nikolaeviča na granitnyh stupenâh amfiteatra Moskvy-reki.
Nikolaevich.ACC on granite steps of amphitheatre of Moscow River
‘In the very shortest time, Ivan Nikolaevich could be seen on the granite steps of the Moscow River amphitheatre.’ [ParaSol]
Here the author explicitly refers to the specific situation anchored in time, and the described experience obviously stems from the author’s or character’s personal perception of the situation (in this use modal impersonals resemble some uses of 2sg as in (7)). However, the personal experience seems to be generalized and ‘transferred’ to the range of potential individuals, who would be in the characters’ place, in other words saying “anyone who would be in that place at that particular time could experience the same”. Thus, even if a modal impersonal occurs in an episodic sentence, it features a generalization over the referent set, perhaps due to its non-veridical nature.

Some non-veridical instances of 2sg also have a modal flavor (e.g. expressing unavoidability or necessity as in (11)) or occur in conditional sentences.

(11) Ničего не podelaеš! Tak už vyšlo.
nothing NEG do.2SG.FUT
‘There’s no help for it, it just happened.’ [ParaSol]

3pl does not seem to be as sensitive to (non-)veridicality as modal impersonals and 2sg, but in comparison to them, it is the one that tends to be used in veridical propositions.

Finally, it can be concluded that both veridicality and genericity are important context predictors for the use of impersonal strategies. They are independent from each other; moreover, veridicality has minimal influence on the interpretation of the human referent, whereas genericity is strongly correlated with the interpretation and perspective.

4. Conclusions and prospects for future research

The distribution of the major translation equivalents of German man and impersonalization strategies in Russian, 3rd-person plural, modal impersonal and 2nd-person singular forms of the verb, has been described on the basis of cross-linguistic semantic features of referent interpretation and context. These features served as parameters for the qualitative and quantitative evaluation of the data from the parallel German-Russian corpora. The statistical methods of the data evaluation used in this study seem to be appropriate for making predictions about the occurrence of the major translation strategies in Russian, and the results obtained have been compared and supported with linguistic observations. It has been concluded that the two context variables, i.e. ‘generalizing’, together with the relevant interpretation of the human referent, and ‘veridical’, play a central role in the choice of impersonal strategies in Russian.
References


Sources: ParaSol (A parallel corpus of Slavic and other languages): http://parasol.unibe.ch/