Eötvös Loránd University
Faculty of Humanities

THESES OF PHD DISSERTATION

Nikolett F. Gulyás

IMPERSONAL CONSTRUCTIONS IN FINNO-UGRIC LANGUAGES
A FUNCTIONAL AND TYPOLOGICAL APPROACH

Doctoral School of Linguistics
Head of Doctoral School: Prof. Dr. Gábor Tolcsvai Nagy MHAS

Uralic Linguistics and Languages Doctoral Program
Head of the Doctoral Program: Prof. Dr. Ferenc Havas DSc

Members of the Committee:
Chair: Prof. Dr. Mária Ladányi DSc
Secretary: Dr. Katalin Gugán PhD
Opponents: Dr. habil. Márta Csepregi PhD
Prof. Dr. Edith Moravcsik PhD
Members: Prof Dr. Sándor Csúcs DSc
Dr. habil. Nóra Kugler PhD
Dr. Zsuzsanna Salánki PhD

Supervisor: Prof. Dr. Ferenc Havas DSc

Budapest
2016
1. Introduction

1.1 Topic and aim of the dissertation

The aim of this study is to describe the impersonal constructions of six Finno-Ugric languages: Finnish, Mari, Udmurt, Komi-Permyak, Surgut Khanty and Hungarian. Impersonality is considered in this work as a complex syntactic-semantic category based on functional linguistic theories (e.g. Givón 1997, 2001). In this framework, the main function of impersonals is to defocus the causer, the initiator or the actor, so broadly speaking, the agent of the event (Siewierska 2008). Impersonal constructions are used when the agent of the action is not known, irrelevant to the utterance, or the speaker does not want to specify the identity of the actor of the event.

Comparative analyses of Finno-Ugric impersonal constructions have already been conducted (Stipa 1962, Schiefer 1981), but these approaches ignore typological considerations. On the other hand, the comprehensive typological works have concentrated mainly on weather related expressions (Bartens 1995, Salo 2011). Some of the languages have a vast literature on impersonal constructions but these sources vary in quantity and quality. The Hungarian and the Finnish constructions were described in various linguistic frameworks (cf. Tóth 2000, Kádár 2006, Holmberg 2005, Helasvuo & Vilkuna 2008, Huumo & Helasvuo 2015), and contrastive studies were conducted targeting these two languages (e.g. Keresztes 1995). However, impersonals of smaller Finno-Ugric languages are lesser described, and the available resources are accessible mainly in Russian (for an exception, see Kalinina et al. 2006). Thus, the present research may enrich our knowledge about impersonal constructions of Finno-Ugric languages, and in the case of Surgut Khanty, the present study could be the first systematic analysis of impersonals.

The aim of the dissertation is to answer the following questions:

i) Which constructions suffice as impersonals in Finno-Ugric languages, based on the functional theoretical framework? What morphological, syntactic, and semantic features characterize these constructions?

ii) How are impersonal constructions used in the different Finno-Ugric languages? Are there any pragmatic factors which are specific to the different impersonal types?
iii) What similarities and differences are there among the Finno-Ugric impersonal constructions?

iv) How do the analyzed Finno-Ugric impersonal constructions relate to the impersonal constructions of other languages? Should we assume a special usage pattern or structural construction which is specific to Finno-Ugric languages?

1.2 The language sample

When the language sample was compiled for the analysis, I tried to select languages from different branches of the language family, and I also tried to compile a sample which could enable the detection of areal patterns. The starting point of the analysis is Hungarian, which belongs to the Ugric branch of the Finno-Ugric language family. The Ob-Ugric branch is represented here by Surgut-Khanty. There are two Permic languages in the sample: Udmurt and Komi-Permyak1, additionally, the Mari language (more specifically Meadow Mari) was also selected for detailed analysis. The Finnic branch is represented by Finnish. Some basic information and typological characteristics of the selected languages can be seen in the table below.

<table>
<thead>
<tr>
<th></th>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of speakers</td>
<td>13–14 million</td>
<td>2800</td>
<td>61,000</td>
<td>324,000</td>
<td>365,000</td>
<td>5,5 million</td>
</tr>
<tr>
<td>EGIDS type2</td>
<td>1</td>
<td>6b</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>canonical case for subject</td>
<td>Nom</td>
<td>Nom</td>
<td>Nom</td>
<td>Nom</td>
<td>Nom</td>
<td>Nom</td>
</tr>
<tr>
<td>marking of object nouns3</td>
<td>Acc (Nom)</td>
<td>Nom</td>
<td>Nom/Acc</td>
<td>Nom/Acc</td>
<td>Acc</td>
<td>Part/Nom/ Acc</td>
</tr>
<tr>
<td>marking of object pronouns</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc/Part</td>
</tr>
<tr>
<td>pro-drop in 1st and 2nd person</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>pro-drop in 3rd person</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>

1 The Komi-Zyrian language was also analyzed in the study, however, the results are almost identical to the Komi-Permyak data, thus Zyrian examples were only indicated as additional information.

2 Meaning of the codes: 1 – National, 4 – Educational, 5 – Developing, 6b – Endangered (Lewis et al. 2015).

3 Optional constructions are in brackets.
copula in 1st and 2nd person

<table>
<thead>
<tr>
<th>Basic word order</th>
<th>SOV</th>
<th>SOV</th>
<th>SVO</th>
<th>SOV/SVO</th>
<th>SOV</th>
<th>SVO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st person</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2nd person</td>
<td></td>
<td></td>
<td></td>
<td>(+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd person</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

1.3 Structure of the dissertation

The dissertation has seven chapters. The introductory chapter specifies the aims of the dissertation, states the research questions, describes the socio-linguistic and typological features of the selected languages, and the transcription and glossing principles. Chapter 2 gives the theoretical background of the dissertation: this chapter addresses the terminological issues, and it also reviews the literature on Finno-Ugric languages. The different types of impersonals (R-impersonals, A-impersonals, and T-impersonals) are introduced and their morphological, syntactic, and semantic features are discussed in detail. Chapter 3 describes the data sources and the methodology of the research. The chapters 4–6 introduce the results of the study. Chapter 4 shows the two types of non-referential impersonals (R-impersonals): the indefinite and non-referential subject constructions. Chapter 5 discusses constructions sensitive to agentivity properties of the subject (A-impersonals). One group of these consists of constructions with inanimate subjects, the other type includes constructions which have non-volitional subjects. Chapter 6 discusses the impersonal constructions with non-topical subjects. Chapter 7 summarizes the results of the research.

2. Theoretical framework of the study

There are various ways of encoding impersonal constructions cross-linguistically, but this topic emerged in typological studies only recently (cf. Sansò 2006, Malchukov & Siewierska 2011). The typological classification of Malchukov & Ogawa (2011) is based on Keenan’s (1976) subject definition, which treats the subject as a universal phenomenon with prototypical syntactic and functional properties. Malchukov & Ogawa (2011) use 5 criteria of the subject prototype in order to gain a universal definition of impersonals. According to Keenan (1976), a subject is prototypical if it is

a) a referential argument
b) a definite NP
c) topical
d) animate
e) agentive (cf. Malchukov – Ogawa 2011: 23)

Impersonality can be defined and characterized in terms of deviations from the prototypical subject properties. Depending on which prototypical property the subject lacks in a given impersonal construction, the impersonals can be divided into three main categories: 1) those sensitive to reference and definiteness (R-impersonals), 2) agentivity and animacy (A-impersonals) 3) topicality of the subject (T-impersonals). The latter type marks as a transition between the two other groups (Malchukov & Ogawa 44–45).

Those constructions that have indefinite or non-referential subject fall into the group of R-impersonals. Among others, this group contains 3Pl, generic pronoun (or generic noun) constructions, impersonal passives, as well as weather verbs. A-impersonals sensitive to agentivity make up two groups based on their semantic features: constructions containing inanimate and non-volitional subjects. The former, for example, can be a natural force. A-impersonals have a common feature that their subject marking is usually non-canonical, dative, genitive or oblique cases are often used within these constructions. A common feature of T-impersonals, those sensitive to topical features, is the lack of topicality of the subject. Constructions that fall in to this category can have indefinite or generic and new subject (e. g. a subject introduced as a new participant of the event, cf. Malchukov & Ogawa 2011: 28–31). The dissertation
analysis the Finno-Ugric languages mentioned above through the lenses of these five subject properties.

3. Data and methods

For the data collection, a usual methodology of language typology was used (cf. Croft 2003: 28–30). The data collection started with the reviewing of the previous literature of impersonal constructions in the selected languages, as well as with reviewing of descriptive grammars (e.g. Kenesei et al. 1998, Tolcsvai Nagy 2013, Honti 1984, Csepregi 2011, Batalova 1975, Kalina & Raspopova 1983, Alatyrev 1970, Csúcs 1990, Bartens 2000, Vasikova 1987, Bereczki 1990, Alhoniemi 2010, Vilkuna 1996, Karlsson 1999, VISK). After that the described phenomena were put into the typological framework of the study, following the work of Siewierska (2008), and Malchukov & Ogawa (2011) respectively.

During the research, five questionnaire-based studies were conducted between 2011 and 2015. The survey data was collected with the help of native speakers of the selected languages. Four questionnaire contained 60-100 sentences each which had to be translated to the target languages, and the fifth survey targeted acceptability judgements and pragmatic features. The surveys were filled out by 6 Surgut Khanty, 6 Udmurt, 7 Komi-Permyak, 7 Mari, and 8 Finnish informants. The primary language of the survey was Hungarian, however, as controls the test were taken in Russian too in order to see the effects of the different source languages. All of the studies were conducted with the help of native professionals. They helped throughout the study, from the data collection, through the processing of the results and determining the pragmatic factors of the impersonal constructions, to explain the differences of the Hungarian and the Russian survey results.

The sentences of the first survey were compiled based on the classification of Siewierska (2008). It contained weather verbs, non-canonical subjects, indefinite pronouns, generic nouns in 2Sg, 1Pl, and 3Pl forms, presentative and locative constructions, and impersonal passives. The second survey targeted non-canonical subjects, the third survey contained reflexive and causative verbal constructions. The last survey was conducted in order to clarify the details of the constructions retrieved from the previous surveys. The native consultants filled out a pragmatic questionnaire.
to determine the functional characteristics of the constructions according to referentiality, topicality, agentivity. On the one hand, the main results of the present dissertation come from the systematizing the results of previous studies, and more significantly, from my own survey. Additionally, a small corpus analysis was also conducted for 3Pl and generic noun constructions following the example of Siewierska and Papastathi (2011). The corpus was an originally Russian literary text (a propagandistic fable about the life of Pavlik Morozov, PM) which has parallel translations for many Finno-Ugric (and some Turkic) languages compiled by the Research Unit for Volgaic Languages at the University of Turku. These parallel translations cannot be considered as corpora in the traditional sense, because they are not annotated and they are quite small in number (containing approx. 13,000 tokens per language).

4. Results

4.1 R–impersonals

The first type of impersonals contain indefinite subjects. From R-impersonals, I analyzed three constructions in detail: 1) the 3Pl impersonal construction, 2) the construction containing the generic noun meaning ‘man’, and 3) impersonal passives.

3Pl impersonals do not show structural variation, since they have fixed morphosyntactic features. The verbal predicate is always in 3Pl form agreeing with the pronominal subject. In the study, I investigated which constructions are the functional equivalents of the Hungarian 3Pl impersonals expressed by intransitive and transitive verbs. The results are summarized in the table below.

<table>
<thead>
<tr>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Pl intr.</td>
<td>3Pl intr.</td>
<td>3Pl intr.</td>
<td>3Pl intr.</td>
<td>3Pl intr.</td>
<td>passive</td>
</tr>
<tr>
<td>3Pl tr.</td>
<td>passive</td>
<td>3Pl tr.</td>
<td>3Pl tr.</td>
<td>3Pl tr.</td>
<td>passive</td>
</tr>
<tr>
<td>(3Pl tr.)</td>
<td></td>
<td></td>
<td></td>
<td>('man')</td>
<td>(3Pl non-impersonal)</td>
</tr>
</tbody>
</table>

Table 2

Equivalents of Hungarian 3Pl impersonals according transitivity

4 The fable does not have a Surgut Khanty translation.
Semantically, the construction has 3 different usages: generic, episodic, and specific. 3Pl impersonals in generic readings typically mark habitual actions, or they code other irrealis contexts as probability, negation or modality. 3Pl impersonals with episodic usage describe events anchored in time, usually with perfective aspect, but sometimes the verb can refer to a present tense event. Temporal anchoring narrows the possible referents of the subject. In generic usage, the referent could be any individual, but in the episodic usage, the referents could be people in a given time (space or situations). The subject of constructions with specific usage can be understood from the context, which typically have special local and temporal settings. The subject here is a specified but non-determined, concrete entity (Siewierska 2011: 61–65). In the related languages, the following constructions are used as compared to Hungarian:

<table>
<thead>
<tr>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>generic</strong></td>
<td>3Pl</td>
<td>3Pl</td>
<td>3Pl</td>
<td>3Pl</td>
<td>impersonal passive</td>
</tr>
<tr>
<td>passive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>episodic</strong></td>
<td>passive</td>
<td>3Pl</td>
<td>3Pl</td>
<td>3Pl</td>
<td>impersonal passive</td>
</tr>
<tr>
<td>(passive participle)</td>
<td></td>
<td></td>
<td>(passive participle)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>specific</strong></td>
<td>3Pl</td>
<td>3Pl</td>
<td>3Pl</td>
<td>3Pl</td>
<td>impersonal passive</td>
</tr>
</tbody>
</table>

Table 3
Equivalents of Hungarian 3Pl impersonals according usage

The second type\(^5\) of R-impersonals with an indefinite subject consists of constructions using generic nouns meaning ‘man’. In this chapter, I examined the statement of Siewierska (2011), stating Hungarian is the only Finno-Ugric language in her sample that uses the generic nouns as impersonals. The equivalents of Hungarian noun ‘man’ can be seen in the following table:

---

\(^5\) I tangentially overviewed those other constructions that differ from the prototypic subject in terms of definiteness, namely the 1Pl and 2Sg forms. However, these did not show such systematic correspondences as the 3Pl impersonals do. Pragmatic tests showed that informants found these sentences deficient (as opposed to impersonals which are complete expressions), so the classification of this construction as impersonal requires further examination.
I examined the properties of these nouns in terms of agreement, and the results partly confirmed the statements of previous literature stating that singular generic words meaning ‘folk’ in Khanty, Komi-Permyak, Udmurt and Mari do not show agreement in number with the verb. In Hungarian and Finnish, such lack of grammatical agreement cannot be found. Furthermore, I tested the three types of usage I described in connection with 3Pl impersonals, and came to the conclusion that in the examined languages, generic usage is the only option. So we can see that impersonal generic noun constructions can be found, besides Hungarian, in all tested languages, however, their usage is strongly limited in Finnish.

<table>
<thead>
<tr>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>ember</td>
<td>qo</td>
<td>mort</td>
<td>aďami</td>
<td>ajdeme</td>
<td>ihminen</td>
</tr>
<tr>
<td>nép(ek)²</td>
<td>jāy</td>
<td>jōz</td>
<td>kalyk</td>
<td>kalāk, jeñ-βlak⁷</td>
<td>vāki</td>
</tr>
</tbody>
</table>

Table 4
Lexemes meaning ’man’ and ’folk’ in the examined languages

In this study, I also have investigated the so-called impersonal passive domain. A consider a language has impersonal passives here if the following criteria apply to given constructions in the language:

a) there is no overt grammatical subject in the construction

⁶ In Hungarian, the plural form of the word nép is used in the meaning of ‘an undefined group of people’. ⁷ Jeñ means ‘folk’, -βlak indicates plural form.
b) the only argument of the verb is an object
c) passivization applies also to intransitive verbs
d) the construction is lacking an overt (oblique) agent

Results can be seen in the table below:

<table>
<thead>
<tr>
<th>criteria</th>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>no overt subject</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>object argumentum</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>intransitive verb</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>overt agent is ungrammatical</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6
Fulfillment of the criteria for impersonal passive

I continue the description of R-impersonals with constructions having a non-referential subject. This type contains various construction types expressing weather phenomena. Following the work of Bartens (1995), Salo (2011) and Eriksen et al (2015), I classified the constructions according to predicates and arguments. In the predicate type, the verbal predicate stands alone without a subject, expressing the meteorological circumstance on its own. In the argument type, a nominal represents the meaning concerning the weather, or the construction does contain a verbal predicate, but it does not have a semantic feature related to the weather phenomena. In the predicate-argument type both components of the construction contain the semantic momentum concerning weather. Here we can distinguish figura etymologica (e.g. Esik az eső ‘It is raining’) and structures originating from different word stems (e.g. Fúj a szél ‘The wind is blowing’). In these cases, expressions with non-subject-predicate setup can also be observed. Finnish usually uses the partitive, sometimes the accusative case for marking a non-subject participant of the event, while Surgut Khanty, Komi-Permyak and Udmurt uses the instrumental case marking.
4.2 A-impersonals

Constructions with subjects sensitive to agentivity consist of a smaller, but not less heterogenic group compared to R-impersonals. A-impersonals can differ from the prototypic subject specified by Keenan (1976) in two ways: their subject referent can either be inanimate or non-volitional.

Generally speaking, no grammatical marker is specified to mark inanimate subjects in the examined languages, however, in some cases it is possible to encode the inertness of the subject. In the passive construction of Surgut Khanty, the marking of the agent always goes with a special marker, the locative suffix (Kulonen 1989), which, based on my data, can also be applied when the notional subject of the event is inanimate:

(1) Surgut Khanty

\[wôt-na \; qôt \; owti \; īa \; kōs-i.\]

wind-LOC house roof away break-PASS.PST.3SG

'The wind blew the roof away.' (intended meaning)

Supposedly by Russian influence, the instrumental coding of the inanimate notional subject is possible in Udmurt and Komi-Permyak:
(2) Komi-Permyak

\[Va-\text{ön} \quad kyrö̈t-i-s \quad bereg.\]

water-INS  wash\_away-PST\_3SG  shore

'The water washed the shore away.'

(3) Udmurt

\[Töl-en \quad nu-i-z \quad l'ipet-ez.\]

wind-INS  carry-PST\_3SG  roof-ACC

'The wind carried away the roof.'

No such a specific marking of inanimate subject can be found in Hungarian, Mari and Finnish. The second group of A-impersonals is constructions coding involuntary subject. This is a widely interpretable semantic category that can consist of constructions with various meanings, of which I describe four:

a) unintentional actions
b) expressing of will
c) quasi-causative event
d) description of physical experience

The notional subject in all these constructions are expressed by very diverse encoding strategies, which are summarized in the following table.

<table>
<thead>
<tr>
<th>type of occurrence</th>
<th>Hungarian</th>
<th>Surgut Khanty</th>
<th>Komi-Permyak</th>
<th>Udmurt</th>
<th>Mari</th>
<th>Finnish</th>
</tr>
</thead>
<tbody>
<tr>
<td>unintentional event</td>
<td>–</td>
<td>–</td>
<td>Gen+reflexive</td>
<td>reflexive</td>
<td>–</td>
<td>Abl</td>
</tr>
<tr>
<td>expression of will</td>
<td>Dat</td>
<td>specific construction</td>
<td>Gen+specific construction</td>
<td>Gen+specific construction</td>
<td>Gen+specific construction</td>
<td>Gen</td>
</tr>
<tr>
<td>quasi-causative physical experience</td>
<td>–</td>
<td>–</td>
<td>Acc</td>
<td>Acc</td>
<td>Acc</td>
<td>Part</td>
</tr>
<tr>
<td>physical experience</td>
<td>Dat</td>
<td>Akk/Dat/Lat</td>
<td>Dat/Gen</td>
<td>Dat</td>
<td>Dat</td>
<td>Ade/Gen</td>
</tr>
</tbody>
</table>

Table 8
Grammatical marking of involuntary subject
4.3 T-impersonals

In Malchukov and Ogawa’s (2011) classification system, T-impersonals, e.g. constructions sensitive to the topicality of the subject, count as the most problematic part of the classification. These structures usually contain a non-topical subject, which may be indefinite, generic, and new (in the sense being a new participant of the described event, cf. Malchukov & Ogawa 2011: 28–31). Grammatical marking of T-impersonals can occur via verbal agreement or word order. Since the examination of correlation between topicality and word order would have exceeded the frames of this study, I only included those constructions to T-impersonals that mark the non-topical nature of the subject by case marking or the lack of agreement. In this sense, T-impersonals can only be found in Finnish.

5. Summary

In this study, I provided an overview on the equivalents of the Hungarian R-, A- and T-impersonals in its cognate languages. It can be stated that impersonal constructions of these examined languages are very heterogenic in all three semantic groups. In summary, Finnish encoding strategies are the most distant, as compared to Hungarian, followed by Khanty, then Mari and the Permic languages. The Finno-Ugric languages I examined mainly differ from other European languages in the usage of R-impersonals – this confirms the results of previous studies (Siewierska 2011) – however, they do not use constructions that can’t be detected cross-linguistically. However, the impersonal constructions of the Finno-Ugric languages form a various category, it can be seen from the study, that Komi-Permyak, Udmurt and Mari show much more concord in terms of structure and usage than the other examined languages.
References


**Source**


**Articles related to the dissertation**

2014


2013


2012


2011

2010