

Avvakum Im(personal): A Categorization Approach to Information Structure and Coreference in the Historical Russian Text from the 17th Century

Diese Studie analysiert die unpersönlichen Konstruktionen mit dem Verb in der dritten Person Plural im altrussischen Text *Das Leben des Protopopen Avvakum*. Verschiedene Interpretationen dieser Konstruktionen werden basierend auf ihren semantischen Merkmalen, Häufigkeiten und Kontexten klassifiziert. Mögliche Verbindungen zur diachronen Entwicklung des russischen grammatischen Systems und zum narrativen Genre des Textes werden diskutiert.

This study analyzes third-person plural impersonal constructions in the Old Russian text *The Life of Archpriest Avvakum*. It examines various interpretations of these constructions based on their semantic features, frequencies, and contexts. The findings provide insights into the semantic patterns of impersonal constructions. Their connection to the diachronic evolution of the Russian grammatical system and the narrative genre of the text is discussed.

Where is the boundary between personal and impersonal constructions? This question is not as trivial as it may seem. In Russian, there is a human impersonal construction that contains a third-person plural verb and lacks an overt subject (further referred to as 3PL-IMP). While usually being associated with reduced referentiality, 3PL-IMP-structures can also have referential interpretations and characterize temporally anchored events with specific agents (cf. Siewierska 2011: 59–60). This observation, made on modern language material, appears to hold true for 17th-century

Russian as well, as exemplified by the following sentence from *The Life of Archpriest Avvakum* taken from the TOROT treebank (Eckhoff and Berdičevskis 2015): »привели ко мнѣ бѣшанова. фѣѡдором звали«¹. Both sentences contain 3PL-IMP-structures, but their interpretations are clearly different. The first sentence describes a specific event with a closed set of participants (in the sense of Berger 1991), while the second sentence denotes a general state of affairs applicable to a potentially unlimited number of agents.

The present study aims to investigate the range of interpretations of 3PL-IMP-structures in *The Life of Archpriest Avvakum* (further referred to as *The Life*) by analyzing their semantic features defined in previous research, particularly in the typological study by Gast and Auwera (2013). To achieve this goal, all instances of 3PL-IMP-structures in the text were extracted and qualitatively annotated for the relevant features, aiming to identify if there are combinations of features not attested in modern Russian as documented by Gast and Auwera (2013). Subsequently, a descriptive quantitative analysis was conducted on the collected data to examine the distribution of feature combinations in the text.

The paper is organized as follows. The first section is the introduction. Section 2 summarizes the previous classifications of 3PL-IMP-structures with a focus on the study by Gast and Auwera (2013). Section 3 characterizes the analyzed text and formulates the research questions. Section 4 presents the results while commenting on the relevant methodological aspects. In Section 5, the findings are discussed in the context of the research questions and predictions. Section 6 concludes the paper.

Previous Accounts on Different Readings of Impersonal Constructions

Several syntactic and semantic criteria have been proposed in the literature to capture the variation of the referential status of impersonal constructions, which should help differentiate between distinct types of readings (cf., e.g., Cabredo-Hofherr 2006, Siewierska and Papastathi 2011, Malamud 2013). This study primarily relies on the classification put forward by Gast and Auwera (2013) which was designed to account for different ranges of possible interpretations of impersonals

1 Transl. by the author: »They brought the demon-possessed man before me. His name was Feodor (literally: They called him Feodor).«

across European languages. The classification is summarized and commented on in the following paragraphs.

The first characteristic taken into consideration is veridicality. Drawing upon the account of Giannakidou (2011), Gast and Auwera (2013: 137) categorize a proposition denoted by a bare clause as veridical »if it is assumed to be true, in the context in which it is uttered«. In other words, veridicality manifests itself in that if the clause entails or presupposes the truth of the proposition it expresses (cf. Giannakidou 2011: 1674).

Another distinction pertains to the relation of the described event to space and time. States of affairs that are »thought of as taking place at a specific time and place« are identified as episodic, whereas the ones that »generalize over time and space« are called generic (ibid.: 137). Although Gast and Auwera only apply this distinction to veridical contexts, the present study follows the approach of Rudolf (2014) in annotating these features independently to get the full picture of the proportion of two levels of this parameter.

Non-veridical states of affairs are further categorized into modal and non-modal based on the presence or absence of expressions indicating »possibility or necessity« within the clause (Gast and Auwera 2013: 137). Whereas Rudolf excludes this property from her analysis due to its close connection with veridicality (cf. 2014: 332), the present study treats it as a separate parameter.

All the above-mentioned features apply to the event described by the clause containing the impersonal. The following categories, in contrast, pertain to the type of quantification over the semantic subject. Once again, while acknowledging the rationales of applying these in a hierarchical manner and excluding some improbable combinations, this study treats them as independent features. Firstly, expressions with existential quantifiers are distinguished from those with universal quantifiers (Gast and Auwera 2013: 138). Secondly, the perspective of the speaker is investigated. If »the speaker identifies him/herself with the set of referents«, the perspective of the expression is classified as internal, otherwise, it is categorized as external (Rudolf 2014: 333). Additionally, the difference between definite or indefinite referents is considered relevant (cf. Gast and Auwera 2013: 138). A final criterion is the presence or absence of number ambiguity which distinguishes between vague and plural expressions, respectively (cf. ibid.).

Gast and Auwera (2013: 140–149) have identified four bundles of features that are attested for modern Russian 3PL-IMP constructions: 1) Veridical, episodic, existential, indefinite, vague (diagnostic sentence: »They're knocking on the door«); 2) Veridical, episodic, existential, indefinite, plural (»They've surrounded us«); 3) Veridical, ep-

isodic, existential, definite («They've raised the taxes again»); 4) Veridical, generic, universal, external («They eat dragonflies in Bali»). However, it has also been suggested that in certain generic readings, the speaker can be included in the set of referents (cf. Berger 1991: 74, citing Guiraud-Weber 1984: 67; Malamud 2013: 5).

The notion of generic, speaker-inclusive interpretation of 3PL-IMP constructions is particularly important from a diachronic perspective. Meyer (2011: 200) links the possibility of such interpretations to changes in the subject realization system in Russian. This idea builds upon the claim by Holmberg (2009) that languages with consistent null referential subjects do not allow for these readings. In contrast, languages where null referential subjects occur only in certain configurations are said to have generic, speaker-inclusive 3PL-IMP constructions. According to Meyer (2011), Russian has transitioned from the first to the second type. Meyer also cites Borkovskiy (1968: 83–100) to support the argument that this shift may have been accompanied by the emergence of generic, speaker-inclusive interpretations not documented in earlier texts. However, the results of the corpus study presented by Meyer (2011: 225–226) indicate that aside from diachronic development there is also a correlation with the content of texts. In the case of *The Life*, occasional instances of generic readings were identified. While Meyer used only a short sample from *The Life*, the present study analyzes the usage of 3PL-IMP constructions in the entire text to provide further insights into its position in the described process.

Research Materials and Questions

The Life is an Old Russian text dating back to the seventeenth century. It takes the form of an autobiography and revolves around the experiences of archpriest Avvakum, a prominent clergyman who opposed the Nikonian reforms within the Russian Orthodox Church. The main focus of *The Life* is the repression and persecution faced by the author, his family, and other adherents of the old belief system. The text exhibits a combination of religious and vernacular registers, with narrative passages intermingled with theological discussions. These discussions often rely on the citation and reinterpretation of sacred texts, demonstrating the author's deep engagement with religious scriptures (cf. Martin 2019: 1478–1483, and the references therein).

The text version used for the present study is sourced from the TOROT treebank (cf. Eckhoff and Berdičevskis 2015). All examples cited in this paper originate from this source. This text version is based on the Pustozerskiy Sbornik (Demkova et al. 1975) that is considered to contain the most recent edition of *The Life*, written in 1675

(cf. Demkova 1970: 219). The TOROT treebank version of the text is digitalized, tokenized (divided into individual words), segmented into sentences, lemmatized (words reduced to their base form), tagged for parts of speech, annotated for morphological features, and parsed for syntactic dependencies (cf. Eckhoff and Berdičevskis 2015).

The text in the TOROT treebank consists of 3,237 sentences and 26,630 tokens, making it a substantial corpus for analysis. In the context of the present study, further annotation steps were carried out as part of the DFG/AHRC project »The History of Pronominal Subjects in the Languages of Northern Europe (Pronoun History)«. This project focuses on the diachronic analysis of subject realization patterns in Slavic, Celtic, and Germanic languages, particularly examining the opposition between empty subjects and overt subjects. The project involves the creation of historical corpora, where texts are annotated for syntactic and information structure, coreference relations, and specific categories related to empty subjects. The analysis presented in this study specifically focuses on the indefinite human category in one specific text, *The Life*.

The research questions for this study are as follows: 1) Which combinations of the semantic features of 3PL-IMP-structures described above occur in *The Life*? 2) What are the proportions between these combinations? 3) What are the proportions inside individual pairs of opposed features? Based on previous research conducted by Gast and Auwera (2013) in combination with the evidence presented by Meyer (2011), it was expected that five types of readings would be found in the text: four combinations listed above with the addition of the generic speaker inclusive interpretation. Considering the narrative nature of *The Life*, it was anticipated that there would be a higher occurrence of episodic and existential contexts, as these are characteristic features of this text type (cf. Borkovskiy 1968: 97). The analysis of the data and the results obtained from it are described in the following chapter, preceded by providing relevant methodological details to ensure a comprehensive understanding of the study's approach.

Research Methods and Results

To prepare the material for the analysis, two rounds of annotation were performed. The first round involved using INCEpTION, an interactive tool for multilayer annotation (cf. Klie et al. 2018). During this stage, 3PL-IMP-structures were identified and distinguished from other personal or impersonal null subject constructions. Most of the other parameters used in the general schema of the Pronoun History project were found to be irrelevant to the research questions of this study. An exception is the cate-

gory of givenness, the extent to which the entity is present in the Common Ground (cf. Krifka 2008: 37). The Common Ground refers to »information that is mutually known to be shared« by the participants of the discourse (ibid.: 15). This notion of givenness appears to be very close to the definition provided by Gast and Auwera (2013: 140) for definiteness as the property of an entity to »be accessible in the discourse environment«. Therefore, all 3PL-IMP-structures marked as given were considered definite.

The second annotation step required extracting all instances of 3PL-IMP-structures. To accomplish this, the annotated text was converted to a format suitable for such extraction using the Pepper converter (cf. Zipser and Romary 2010). The converted text was then processed using ANNIS, a software tool for corpus search and visualization (cf. Krause and Zeldes, 2016), to locate and export the 3PL-IMP-structures along with the givenness annotation layer. The exported data was organized into a table format for further analysis. This involved identifying the veridicality, episodic or generic state of affairs, modality, existential or universal quantification, definite or indefinite referents, and vague or plural expressions associated with each instance of the 3PL-IMP-structures. The manual annotation allowed for a detailed analysis of the combinations of these features and their proportions within the text.

In total, 142 instances of 3PL-IMP construction were attested for in *The Life*. The overwhelming majority of them (135 or 95%) belong to five categories which were expected to cover all 3PL-IMP-structures found in the text. The overview of these with examples and frequencies of occurrence is presented in Table 1. It is evident from it that the most frequent feature combination is veridical, episodic, existential, and definite, with 66 instances. By treating all parameters as independent, a variation inside this category was discovered. While Gast and Auwera (2013) only differentiate between plural and vague expressions by indefinite referents, this feature was annotated also for definite ones. As a result, 54 instances of unambiguously plural were found. 12 3PL-IMP-structures have the vague number feature. While the example from Table 1 illustrates the plural variety, the following passage was attributed to the vague variety: »ѿ ѿсѣкоша четырьѣ перѣста.«²

Although the assumption that a definite discourse referent can nonetheless be vague concerning its number might appear controversial, this approach corresponds to the specifics of the analyzed text. Due to the challenges of fully reconstructing the discourse situation and the Common Ground in a historical text, slightly different criteria were applied to definiteness and number. In the case of definiteness, the speaker-oriented

2 Transl. by the author: »And they cut off his four fingers.«

perspective was chosen, meaning that referents accessible to the speaker, or narrator, in the given context were considered definite. In contrast, the number feature was annotated from the point of view of the reader, since the knowledge of the narrator about the number of referents was in many cases considered inaccessible to the reader and annotator, making it difficult to reliably determine.

| Type | Features | Diagnostic sentence | Example from <i>The Life</i> | n |
|------|--|---------------------------------|---|---------------|
| 1 | Veridical, episodic, existential, indefinite, vague | They're knocking on the door. | пята слова гдрвы сказывали на меня в полтара годы. »I had been denounced five times in the course of one year and a half.« | 24 |
| 2 | Veridical, episodic, existential, indefinite, plural | They've surrounded us. | вся по дробну расказал. какъ без ѡстатку войско побили у него. »He told us everything in detail about how his whole army was crushed.« | 34 |
| 3 | Veridical, episodic, existential, definite | They've raised the taxes again. | и ругавъ много столкали с судна. »And after a lot of cursing, they pushed me out of the ship.« | 66 (54+12) |
| 4 | Veridical, generic, universal, external | They eat dragonflies in Bali. | до василия пояху во цркви англския рчи. »Before Vasilij, they sang angelic words in the church.« | 3 |
| 5? | Veridical, generic, universal, internal? | One only lives once. | в тоболске была двѣца у меня. анною звали. »I had a maid in Tobol'sk. Her name was Anna.« | 8 |

Tab. 1: Distribution of the main types of 3PL-IMP-structures in *The Life*. All examples are translated by the author.

The next two most frequent types of the 3PL-IMP construction in *The Life* are veridical, episodic, existential, indefinite, plural, and veridical, episodic, existential, indefinite, vague. Generic interpretations, on the other hand, are relatively rare in the text. It is interesting to observe that the combination of a generic reading with an internal perspective occurs more frequently than with an external perspective, with eight instances compared to three. However, it is important to note that this type of combination only

occurs with the verb *зъвати* (<call>) in the construction introducing the name of a person. Therefore, these instances cannot be described as building a separate class in a strict sense (and are hence marked with a question sign in Table 1). This phenomenon in *The Life* was previously identified by Borkovskiy (1968: 34), highlighting its significance within the text.

Alongside the predicted bundles of features, seven combinations not attested for by Gast and Auwera (2013) were found in *The Life*. These can be divided into two types. The first group consists of three instances of 3PL-IMP-structures whose readings are veridical, episodic, non-modal, universal, external, definite, and plural. Gast and Auwera exclude the possibility of an episodic event with universal quantification. However, the following example from *The Life* shows that, in certain configurations, this combination can occur: «ѿ давали мнѣ мѣсто гдѣ бы я захотѣл.»³ On the one hand, the expression refers to a specific and limited period of time and does not generalize over time, indicating an episodic reading. On the other hand, it allows for an interpretation where not only the place but also the subjects of the predicate are universally quantified over.

The second group of unexpected readings is represented by four instances where the 3PL-IMP constructions combine the following features: non-veridical, non-modal, episodic, existential, external, indefinite, vague. These instances all occur in the same passage where Avvakum is praying for salvation in the event that he is executed. The passage consists of conditionals and is structured very homogeneously, so one example is sufficient to illustrate it: «аще меня задушатъ. причти мя с митрополитом московским.»⁴ According to Gast and Auwera (2013), non-veridical contexts which are explicitly claimed to include conditionals (cf. *ibid.*: 137), only have generic interpretations. However, in the context of this passage, it is evident that the conditional statement refers to a certain point in time. The narrator expects execution in the near future as a consequence of a specific situation. Therefore, this type of reading should be labeled as episodic rather than generic.

The present study not only examined the combinations of features in the 3PL-IMP constructions but also analyzed the proportions within each dichotomy. The relationships between the parameters that characterize the state of affairs are illustrated in Figure 1, while those related to the type of quantification are depicted in Figure 2. In both figures, the light-grey bars represent the occurrences of the features on the left

3 Transl. by the author: «And they offered me a position wherever I wanted.»

4 Transl. by the author: «If they strangle me, give me Holy Communion with Metropolitan of Moscow.»

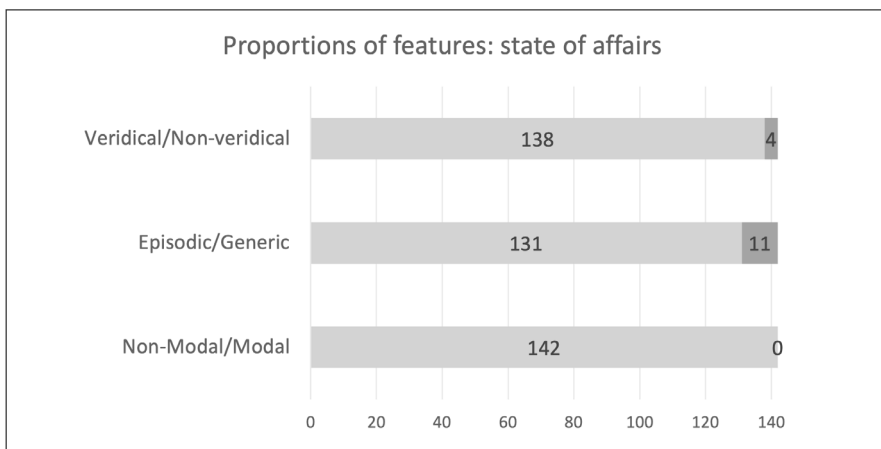


Fig. 1: Relations between the event features of 3PL-IMP-structures in *The Life*.

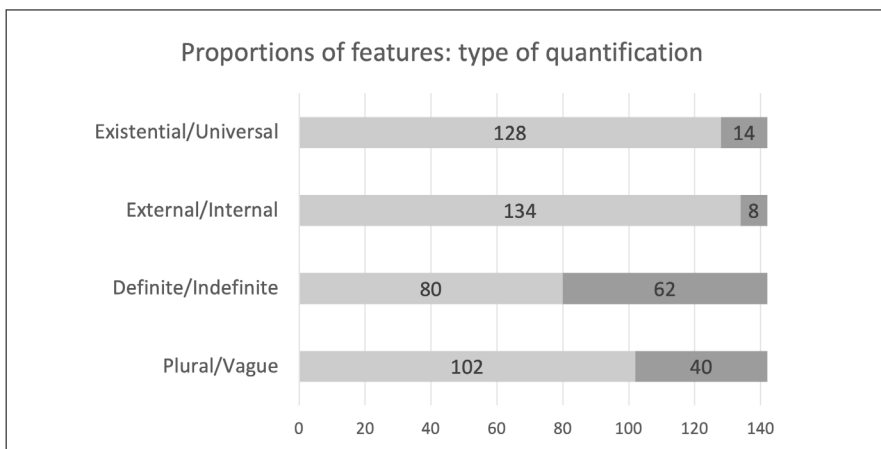


Fig. 2: Relations between the quantification features of 3PL-IMP-structures in *The Life*.

side of the slash symbol, while the dark-grey bars represent the features on the right side. The analysis of *The Life* revealed that modal readings were never observed in the 3PL-IMP constructions. Veridical, episodic, and external interpretations were found to significantly dominate over non-veridical, generic, and universal interpretations, respectively. In terms of definiteness, the readings were relatively balanced, with a slight tendency towards definite interpretations. Additionally, plural interpretations occurred more frequently than vague interpretations.

Discussion

Based on the results obtained, the research questions can be answered as follows. The majority (95%) of the instances of 3PL-IMP constructions in *The Life* align with the interpretations proposed by Gast and Auwera (2013) and the notion of speaker-inclusivity based on Berger (1991) and Meyer (2011). Additionally, two readings of the construction were identified that were not accounted for by Gast and Auwera (2013). These involve combinations of episodic events with universal quantification and non-veridicality with an episodic interpretation. These findings highlight additional nuances and complexities in the interpretation of 3PL-IMP constructions in *The Life*.

Although cases, where the speaker is included in the set of referents, are accounted for, they are not very frequent and are limited to one specific verb. Therefore, the role of *The Life* in the spread of the speaker-inclusive generic reading of 3PL-IMP-structures and its connection with the shift from a consistent null subject language to a partial null subject language cannot be conclusively determined. The three most frequent types share common features of veridicality, episodic interpretation, and existential quantification. However, they differ in terms of definiteness and number. This indicates that 3PL-IMP-structures in *The Life* consistently exhibit a veridical and episodic nature.

The analysis of proportions within each feature dichotomy supports the predicted patterns, showing a prevalence of episodic and existential readings over generic and universal interpretations. This aligns with the narrative genre of the text, as previously suggested by Borkovskiy (1968: 97). Furthermore, the tendency towards definite and plural interpretations suggests that the 3PL-IMP constructions are frequently employed to describe events in which contextual information about the referents is accessible, while overt personal reference is intentionally avoided. This indicates a specific narrative strategy employed in *The Life*.

Conclusion

The present study examined the third-person plural impersonal constructions (3PL-IMP-structures) in an Old Russian text, *The Life of Archpriest Avvakum*. The analysis aimed to explore the various interpretations of these constructions based on their semantic features. By utilizing a classification framework proposed by Gast and Auwera (2013), the 3PL-IMP constructions were categorized based on veridicality, episodic or

generic reading, modality, quantification type, definiteness, and number. Through a comprehensive analysis of the text, it was found that the most prevalent feature combination was veridical, episodic, existential, and definite. Within this category, further variations were observed, distinguishing between unambiguously plural and vague expressions. These results align with the narrative nature of the text, where specific events involving a closed set of participants are described.

Furthermore, other combinations of features were discovered that, although infrequent, can be considered significant for the formulated research questions. On one hand, generic and speaker-inclusive interpretations of 3PL-IMP constructions are relevant from the point of view of the history of the Russian language. On the other hand, instances of combinations of features excluded by Gast and Auwera (2013) might indicate the developmental potential of this classification framework. Future research could build upon these findings by expanding the analysis to other Old Russian texts or exploring additional features and factors that influence the interpretation of impersonal constructions. Additionally, comparative studies across different languages and time periods could provide a broader understanding of the evolution and variation of impersonal constructions.

References

- Berger, T. (1991): »Zum denotativen Status ›unbestimmt-persönlicher‹ Konstruktionen im Tschechischen«. In: Grochowski / Weiss (eds.): »*Words are Physicians*« for an Ailing Mind. München, pp. 71–81.
- Borkovskiy, V. I. (ed.) (1968): *Sravnitel'no-istoricheskij sintaksis vostochnoslavjanskikh yazykov. 1. Tipy prostogo predlozheniya*. Moskva.
- Cabredo-Hofherr, P. (2006): »Arbitrary: Pro and the Theory of Pro-Drop«. In: Ackema / Brandt / Schoorlemmer / Weerman (eds.): *Arguments and Agreement*. Oxford / New York, pp. 230–258, <https://doi.org/10.1093/oso/9780199285730.003.0008>.
- Demkova, N. S. (1970): »Tvorcheskaya istoriya Zhitiya protopopa Avvakuma«. In: *Trudy otdela drevnerusskoj literatury* (25), pp. 197–219.
- Demkova, N. S. / Droblenkova, N. F. / Sazonova, L. I. (eds.) (1975): *Pustozerskiy sbornik. Avtografy sochineniy Avvakuma i Epifaniya*. Leningrad.
- Eckhoff, H. / Berdičevskis, A. (2015): »Linguistics vs. Digital Editions: The Tromsø Old Russian and OCS Treebank«. In: *Scripta & e-Scripta* (14–15), pp. 9–25.
- Gast, V. / Auwera, J. V. D. (2013): »Towards a Distributional Typology of Human Impersonal Pronouns, Based on Data from European Languages«. In: Bakker / Haspelmath (eds.): *Languages Across Boundaries*. Berlin / Boston, pp. 119–158, <https://doi.org/10.1515/9783110331127.119>.

- Giannakidou, A. (2011): »Negative and Positive Polarity Items«. In: Maienborn / von Stechow / Portner (eds.): *Semantics: An International Handbook of Natural Language Meaning*. Berlin / New York, pp. 1660–1712, <https://doi.org/10.1515/9783110255072.1660>.
- Guiraud-Weber, M. (1984): *Les propositions sans nominatif en russe moderne*. Paris.
- Holmberg, A. (2009): »Null Subject Parameters«. In: Biberauer / Holmberg / Roberts / Sheehan (eds.): *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge, pp. 88–124, <https://doi.org/10.1017/CBO9780511770784.003>.
- Klie, J.-C. / Bugert, M. / Boullosa, B. / de Castilho, R. E. / Gurevych, I. (2018): »The INCEption Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation«. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Santa Fe, pp. 5–9.
- Krause, T. / Zeldes, A. (2016): »ANNIS3: A New Architecture for Generic Corpus Query and Visualization«. In: *Digital Scholarship Humanities* (31), pp. 118–139, <https://doi.org/10.1093/lc/fqu057>.
- Krifka, M. (2008): »Basic Notions of Information Structure«. In: *Acta Linguistica Hungarica* (55), pp. 243–276, <https://doi.org/10.1556/ALing.55.2008.3-4.2>.
- Malamud, S. A. (2013): »(In)definiteness-Driven Typology of Arbitrary Items«. In: *Lingua* (126), pp. 1–31, <https://doi.org/10.1016/j.lingua.2012.11.003>.
- Martin, E. (2019): »Avvakum Petrov: Zhitiye protopopa Avvakuma, im samim napisannoe (17th Century) [Life of Avvakum]«. In: Wagner-Egelhaaf (ed.), *Handbook of Autobiography/Autofiction. Vol. III: Exemplary Autobiographical/Autofictional Texts*. Berlin / Boston, pp. 1470–1484, <https://doi.org/10.1515/9783110279818-125>.
- Meyer, R. (2011): *The History of Null Subjects in North Slavonic. A Corpus-based Diachronic Investigation* (Habilitationsschrift). Regensburg.
- Rudolf, O. (2014): »A Corpus-Based Study of Human Impersonal Constructions in Russian«. In: Witkoś / Jaworski (eds.): *New Insights into Slavic Linguistics*. Frankfurt a. M. / New York, pp. 329–342.
- Siewierska, A. (2011): »Overlap and Complementarity in Reference Impersonals: Man-Constructions vs. Third Person Plural-Impersonals in the Languages of Europe«. In: Malchukov / Siewierska (eds.): *Impersonal Constructions: A Cross-Linguistic Perspective*. Amsterdam, pp. 57–90, <https://doi.org/10.1075/slcs.124.03sie>.
- Siewierska, A. / Papastathi, M. (2011): »Towards a Typology of Third Person Plural Impersonals«. In: *Linguistics* (49), pp. 575–610, <https://doi.org/10.1515/ling.2011.018>.
- Zipser, F. / Romary, L. (2010): »A Model Oriented Approach to the Mapping of Annotation Formats Using Standards«. Presented at the Workshop on Language Resource and Language Technology Standards, LREC 2010, La Valette, Malta.

Open Access

This paper is published under the Creative Commons Attribution-ShareAlike 4.0 International license (<https://creativecommons.org/licenses/by-sa/4.0>). Please note that individual, appropriately marked parts of the paper may be excluded from the license mentioned or may be subject to other copyright conditions. If such third party material is not under the Creative Commons license, any copying, editing or public reproduction is only permitted with the prior consent of the respective copyright owner or on the basis of relevant legal authorization regulations.