

Building a Database of Conversational Routines

Polina Bychkova, Alyaxey Yaskevich, Serafima Gyulasaryan,
Ekaterina Rakhilina

University of Ljubljana, University of Warsaw, Ca' Foscari University of Venice,
Vinogradov Russian Language Institute
Ljubljana, Warsaw, Venice, Moscow
{poljabickova, rakhilina}@gmail.com, a.yaskevich@uw.edu.pl, 899236@stud.unive.it

Abstract

This paper discusses the Routinicon, a new constructicographic resource for the description of conversational routines. Conversational routines are defined as conventional formulaic expressions that language speakers use in standard extralinguistic situations (cf. *Bless you!* as a reaction to sneezing or *Who's there?* as a typical answer to a knock on the door). The Routinicon's goal is to accumulate the routines that constitute the inventory of conventional expressions in Russian language and systematically describe them in a way that would enable future cross-linguistic comparison and typological research. Conceptually, the Routinicon is a natural extension of such projects as the Russian Constructicon and Pragmaticon. It inherits their approach to the systematization of phraseological units as well as to the data collection. At the same time, the new project focuses on a fundamentally different domain of units and hence offers a radically new structure of linguistic annotation. Its principles and challenges are addressed in the paper.

Keywords: linguistic databases, formulaic units, pragmatics, conversational routines, constructicography
main scope of most works on computational phraseology¹.

1. Introduction

This paper concerns a relatively new field in linguistic documentation, namely, constructicography (see Lyngfelt et al., 2018). Constructicography aims at systematization of the phraseological level of language, which has been known as one of the most challenging both for structured linguistic description and for automatic processing (see, e.g., Dunn, 2017).

As a cognitive inventory, the constructicon is highly heterogeneous. To better understand its structure, it seems practical to divide it into manageable and relatively homogeneous groups and create separate databases, geared to account for the peculiarities of each group. The linguistic research on the distinctive features of a group of constructions is an essential part of the development of the corresponding database. The wider the scope of the project is, the less distinctions are captured.

The project presented in the current paper inherits from the constructicographic projects of Russian Constructicon (2021) and Pragmaticon (2021), as well as the typological Multilingual Pragmaticon (see Buzanov et al., 2022). Continuing the constructicographic resources naming tradition, the current project is called Routinicon. Its goal is to collect and systematically describe the so-called **conversational routines** (see Coulmas, 1981; Aijmer, 1997), sometimes also referred to as situation-bound utterances (Kecskes, 2010). Note that this type of expressions is not in the

The term “conversational routines”, coined in Coulmas, 1981, refers to conventional clichés that language speakers use in standard extralinguistic situations. Cf. *Bless you!* as a reaction to sneezing, or *Who's there?* as a typical answer to a knock on the door. This kind of units are essentially different from the ones that are collected in the Constructicon and Pragmaticon and require a separate description framework. Russian Constructicon collects and classifies semi-formulaic multiword schemas that contain open syntactic slots (this restriction corresponds to the definition of a construction recently proposed in Haspelmath, 2023). The description of these units' usage rests on defining the semantic/syntactic constraints on their slots. Cf. the classical example of [X], *let alone* [Y] in (Fillmore et al., 1988), wherein X and Y are clauses representing events of different likelihood (X is more probable than Y).

Unlike the Russian Construction, Pragmaticon focuses on fully formulaic items without overt slots. Those are pre-formed clichéd answers that occur in a dialog as NO-like or YES-like answers to a verbal stimulus, i.e. the interlocutor's speech act. Accordingly, the illocutionary type of the stimulus determines their usage and serves as the basis for the discourse formulas description in the database.

Although the Constructicon and Pragmaticon differ in their annotation principles, both

1 E.g., Dunn 2017 explicitly focuses on schematic constructions (ibid: p. 5). The computational works that deal with any type of MWE, like the ones described in Constant et al, 2017, do not distinguish them as a

subclass with its own parameters. It is thus unclear to what extent their methods are effective in application to routines.

databases contain units that require verbal context. Hence, the classification of the contexts and, through the contexts, the units themselves can rely upon well-developed linguistic categories. For the constructions, these are grammatical/syntactic features (like VP, past tense, accusative case, etc., cf. *ne smej VP-lpfv.Inf!* = 'neg dare-imp.2sg'). For the discourse formulas, the contextual constraints can be accounted for with Searle's speech act classification (e.g., certain formulas react to polar questions or opinions, other to offers or requests). Meanwhile, the usage of the routines is mainly conditioned by the aspects of the extralinguistic situation. Therefore, the descriptive inventory for their contextual constraints is yet to be defined. In that respect, constructing a database of routines appears to be the most challenging compared to the other two resources.

Due to the complexity of this material, and the difficulties related to data collection (see Data Sources), Routinicon is not merely a database for accumulation and presentation of already gained knowledge, but rather an instrument of active linguistic research. The pilot data fragment that is in the database comprises a limited set of items, which was balanced to maximize the representation of the functional diversity of routines within a single language. The core annotation principles developed within the project framework are set by the database structure (see Annotation Structure). They are expected to be scalable to a much larger sample of routines, as well as cross-linguistically. However, within the fixed annotation fields, the classification is flexible because it is being developed inductively, bottom-up. The initial sample is used to develop preliminary classification (annotation tags). Once new data becomes available, this classification is revisited and updated so that it accommodates for the novelties discovered. If needed, new classes may be added, or narrower classes may be merged under one umbrella class.

The following section discusses the data collection process and the established annotation structure in the database.

2. Data Sources

Our current goal is to develop a framework of machine-readable pragmatic description that would be appropriate for conversational routines as a peculiar type of multiword expressions. For this task, we were looking for a sample of a manageable size, and at the same time representative in terms of the diversity of the items.

Since the Routinicon project emerged in the context of constructicographic research and after the creation of the two large datasets of constructions and discourse formulas, it was

possible to take advantage of their "remainders", i.e. the list of items that were discovered during data collection for the Constructicon and Pragmaticon but did not fit the inclusion criteria of either of the two resources.

The initial list of constructions for the Russian Constructicon was gathered through manual annotation of literary texts (Janda et al., 2020). This method was efficient since constructions as a functional class are both numerous and highly frequent in language use, so even the manual search of relatively short texts could yield many results. Later, this list was updated based on the data provided in phraseological dictionaries and language coursebooks for foreign students. For the discourse formulas, manual annotation did not produce as many results, because they do not occur as often, especially not in written texts. This is why the data collection for the Pragmaticon was partially automatized. The detailed description of the automatic extraction that was conducted for the Russian Pragmaticon can be found in Gerasimenko et al, 2019. Briefly, it was based on initial manual annotation and targeted relatively short sentences, frequently reoccurring at the onset of the character lines in the corpus of dramatic texts DraCor (Fischer et al. 2019). After the automatic extraction, a Russian dictionary containing about 650 so-called "communiquemes", defined as fixed expressions that convey a communicative intent, (Melikyan 2001) was manually processed, which extended the list of expressions.

The raw lists of constructions and discourse formulas that were subsequently reviewed and filtered based on the established semantic and formal requirements for each of the resources. In the residual data, conversational routines constitute the most salient functional group. Thus, we could take advantage of the sizeable list of linguistic items that were filtered out from the Russian Pragmaticon and Constructicon and use it as a pilot sample to model the Routinicon resource. We consciously chose this shortcut because we believe that the subsequent extraction of routines from texts (drama and movie subtitles) could benefit from a more in-depth understanding of this subclass and its properties, achieved by the analysis of a pilot dataset.

3. Annotation

Since the work implied developing a new annotation convention, it consisted of regular group discussions by the annotators (the authors) on the tags to be introduced. When there was significant disagreement or ambivalence on the use of a particular tag, a consensus was reached by adjusting the tag set in a way that it did not cause controversy. Currently, a little over 200

routines are manually annotated by all pragmatic parameters.

Every routine in the database is provided with examples, glosses, an approximate English translation, and a list of formal variations. The formal variation of an expression can be based on (1) grammatical alternations (e.g. in number or gender), (2) synonymity-alike relations between lexical constituents (*Možno vas na **paru** slov / dva slova?* ~ ‘Can we have a word? lit. possible.PRED you.ACC on **couple** word-GEN.PL / **two**-ACC word-GEN.SG’), or (3) pragmatic particles that are consistently added to the basic form (*Čto vy stoíte?* Vs. *Čto **že** vy stoíte?* ~ ‘What are you waiting for? lit. what you (PART) stand-prs.2pl’).

The functional annotation of the routines in the database includes the following general parameters:

- situation structure
- pragmatics
- usage domain
- usage conditions
- situation type
- summary

Since conversational routines always involve a certain extralinguistic situation linked to their usage, the main parameter that underlies the classification of the items in Routinicon is **situation structure**. Situation structure is characterized by one of four classes: **reaction**, **prompt**, **reaction + prompt**, or **verbal accompaniment**. These classes are defined in terms of the chronological order between the use of the routine and the event in the extralinguistic context to which this routine is directly linked. In reactions, this event (“trigger”) precedes the routine. In prompts, the event (“intended effect”) follows the routine². In verbal accompaniments, the routine is used during the event (“accompanied action”). Reaction + prompt situation structure involves both the trigger, and the intended effect, so the routines of this type are directly connected both to the preceding, and following event.

Every event type (**trigger**, **intended effect**, **speaker’s action**) corresponds to a separate annotation field. For the annotation of the routines of different situation structures, only the fields that are relevant for their structure are filled. For example, for **reactions**, only the field **trigger** is annotated (see Table 1). Within these fields, the extralinguistic events are classified with the use of a system of tags that broadly describe what is happening: e.g., “encounter”, “transferring an object”, “serving a meal”, etc.

² Theoretically speaking, every verbal expression has a certain intended (or expected) effect. We consider the routine to be a prompt only when the main goal of the

Situation structure	Trigger	Intended effect	Accompanied action
Reaction	+	–	–
Prompt	–	+	–
Reaction + Prompt	+	–	–
Accompaniment	(+)	(+)	+

Table 1. Events (columns) relevant for different types of situation structure (rows)

The classification of the routines functions is represented in the field **pragmatics**. The values of this parameter mostly correlate with the contextual settings represented by the **situation structure**, and the event tags in the fields **trigger**, **intended effect**, or **accompanied action**. For example, a greeting (pragmatics) is always a reaction (structure type) to an encounter (class of triggers). However, multiple types of pragmatics can correspond to the same contextual settings. E.g., a sudden encounter can trigger either a joyful (cf. *What a surprise!*) or irritated (cf. *Not you again*) reaction.

The overall characteristics of the “pragmatics” field can be summarized as following.

- The tags are not mutually exclusive. E.g., for the routine *What the hell is going on?* the tags <request for information> and <indignation> are both applicable simultaneously.
- The tags can refer to different levels of specificity. Each routine should be assigned with at least one “higher-level” tag that generalizes over at least 4-5 routines.
- The cases when a routine has different meanings depending on the context are considered polysemy. For a polysemous routine, several pragmatic profiles are created, each assigned with its own set of pragmatics tags.

The following paragraphs provide illustrations of the annotation for the routines of each situation structure type.

3.1 Reaction

The usage of reaction routines is always motivated by a non-verbal stimulus (**trigger**). For instance, the expression *Long time no see!* is typically used as a reaction to an encounter. Moreover, there is a distinctive extra-linguistic condition licensing its usage (introduced in the

speaker is to elicit a specific action or a verbal response from the addressee).

field **usage condition**), namely, a long separation preceding the meeting. In terms of **pragmatics**, the routine combines the functions of an informal greeting and an expression of surprise. See Table 2 for an illustration of the Routinicon annotation of the Russian routine *Skol'ko let skol'ko zim!* 'how_many summer-gen.pl how_many winter-gen.pl' which is an approximate equivalent of *Long time no see!*

Many routines that belong to the class "reaction" are associated with assessment towards the addressee or the stimulus event. The pragmatic tags for the routines in this class include disapproval (cf. *Shame on you!*), shock (cf. *You've got to be kidding me!*), fascination (cf. *Bravo*), excitement (cf. *Someone pinch me!*), and the like.

Routine	<i>Skol'ko let skol'ko zim!</i>
Glosses	how_many summer-GEN.PL how_many winter-GEN.PL
Situation structure	reaction
Trigger	encounter: pleasant encounter
Pragmatics	greeting, surprise
Usage conditions	long separation prior to the encounter

Table 2. Reaction routine. An example of the annotation

3.2 Prompt

The main characteristics of the class "prompt" can be illustrated with a routine *Could I have a word?* This routine does not require any specific event that would trigger its usage. However, by using the routine, the speaker expects to achieve a particular effect: to get an opportunity to speak to the addressee in private. The usage condition is that the addressee is busy, usually talking to other people. The intended effect of this routine is that the addressee and the speaker retire to some place where they could have a private conversation. Thus, the pragmatics of the routine falls into two general classes: communication initiation (along with the routines like *May I ask you something?*) and summoning (along with the routines like *Come on!* or *Help!*). See in Table 3 the annotation fragment of the Russian routine *Na paru slov?* which an approximate equivalent of *Could I have a word?*

Prompts generally tend to express imperative semantics over prohibitive since the situation associated with prohibition usually includes both an expected effect and an event that triggers the usage of the routine. In our framework, this contextual configuration corresponds to the situation structure reaction + prompt.

Routine	<i>Na paru slov?</i>
Glosses	on couple-ACC.SG word-GEN.PL
Situation structure	prompt
Intended effect	private conversation
Pragmatics	communication initiation, summoning
Usage conditions	the addressee is engaged in some activity, third parties are present

Table 3: Prompt routine. An example of the annotation

3.3 Reaction + Prompt

One example of a routine that combines the functions of a reaction and a prompt is *Where do you think you are going?* (Cf. the annotation of its Russian equivalent *Kuda sobral'sja?* in Table 4). It is used to command the addressee to stay where they are (intended effect) after they have attempted to leave (trigger). The speaker's intention is to warn the addressee that an attempt to escape will entail negative consequences and therefore force them to stay.

Routine	<i>Kuda sobral'sja?</i>
Glosses	where_to get_ready-PST.M.SG
Situation structure	reaction + prompt
Trigger	the addressee attempts to leave
Intended effect	the addressee stays in place
Pragmatics	prohibition, threat
Usage conditions	the social status of the addressee is not higher than the status of the speaker

Table 4: Reaction + prompt routine. An example of the annotation

Another example of a reaction + prompt is the routine *Get lost!* (cf. Russian *Ubirajsja!* in Table 5). In contrast to *Where do you think you are going*, it is used to drive the addressee away because their actions or their presence itself offend the speaker. Therefore, the routines *Get lost!* and *Where do you think you are going?* imply effects of the opposite types (the addressee leaving Vs. the addressee staying), but the trigger in both cases is the unwanted behavior of the addressee. Negative assessment of the trigger is a very common characteristic of the reaction + prompt routines.

Routine	<i>Ubirajsja!</i>
Glosses	remove-IMP.SG-REFL
Situation structure	reaction + prompt
Trigger	unwanted presence of the addressee / unwanted behavior of the addressee
Intended effect	the addressee leaves
Pragmatics	driving away, anger
Usage conditions	the social status of the addressee is not higher than the status of the speaker

Table 5: Reaction + prompt routine.
An example of the annotation

The intended effect does not always imply an action. The speaker's intention can also be to prompt a verbal response from the addressee. Cf. the routine *To what do I owe the pleasure?* (see the annotation of its approximate Russian equivalent *Čem objazan?* in Table 6). On the one hand, it functions as a reaction to the speaker's unexpected arrival, and on the other, as a request for explanation. The addressee is expected to reveal the reason of their visit. An essential usage condition for this routine is that the situation takes place in a space that belongs to the speaker (like their home or office).

Routine	<i>Čem objazan?</i>
Glosses	what-INS oblige-PTCP.M.SG
Situation structure	reaction + prompt
Trigger	unexpected visit
Intended effect	explanation of the reason for the visit
Pragmatics	request for information
Usage conditions	the addressee appears in the speaker's personal space; the conversation takes place in person; the social status of the addressee is not higher than the status of the speaker

Table 6: Reaction + prompt routine.
An example of the annotation

3.4 Verbal Accompaniment

Essentially, the routines that function as verbal accompaniments are performative speech acts (cf. Eckardt, 2012). The verbalization serves as an announcement and explanation of the action that the speaker is simultaneously performing. For example, the routine *I'll be right back!* and its approximate Russian equivalent *Sejčas pridu!* (see Table 7) is used by the speaker while taking leave. With this routine, they notify the addressee

that they are leaving and at the same time specify that they are planning to be back soon.

Routine	<i>Sejčas pridu!</i>
Glosses	now come-FUT.1SG
Situation structure	verbal accompaniment
Accompanied action	departure
Pragmatics	departure announcement, promise to return

Table 7: Verbal accompaniment routine.
An example of the annotation

One of the most numerous subclasses of the verbal accompaniments is related to basic manipulations with objects (cf. *Here you are!* → handing over an object, *You shouldn't have!* → accepting a present).

3.5 Additional Annotation Fields

Besides the parameters directly related to the situation structure and pragmatic classes, the annotation includes the fields "usage conditions", "usage domain", "situation type", and "summary".

The field **usage conditions** represent the prerequisites for the usage of the routine (i.e. presuppositions). They often concern the relationships between the situation participants, their relative hierarchy, the temporal and spatial localization of the situation, etc.

The field **usage domain** is annotated for the specialized routines that are only involved in the situations within a certain extralinguistic domain (e.g., dog training, sports, restaurant, theater, playground games, etc.). For instance, the Russian routine of verbal accompaniment *Za ščet zavedenija* ('for account-acc.sg establishment-gen.sg', ~ *On the house*) carries the label "restaurant".

The field **situation types** assigns each routine to a general category. Unlike the parameter "pragmatics", which marks the meaning of the routine itself, the parameter "situation type" refers to the entire communicative situation to which it is bound. For example, a situation can be a <fight> (*Take that!*), an <encounter> (*Look who's here!*), a performance (*Bravo!*), queue (*Who's the last? or I'll be after you!*), taking a picture (*Say cheeze!*), etc. The labels used in this field consist of common lexical terms (cf. encounter, commercial deal, theft, accident). They are primarily meant to simplify the search for the database users. At the same time, these categories are also useful for the annotators. When a new set of routines is introduced in the database, they are first distributed between the major situation types. This way, before starting the detailed annotation of the new routine, the annotator can compare it

to the similar ones. Such comparison helps with identifying the routine's distinctive features.

The field **summary** aims to generalize the content of other fields in a simple and comprehensive way. It is written as a coherent textual description once the routine is analyzed based on all the parameters.

4. Database

4.1 General Properties

The database contains several tables that reflect the internal structure of the project data. The main table, "units", stores the relations between the linguistic items (conversational routines) and the sets of their contextual characteristics (pragmatics, usage domain, intended effect, etc.). Those characteristics that are not represented with free-form text fields but are chosen by the annotator from a pre-defined list are stored in the "features" table and referenced by their identifiers. The relation between the "units" records and the routines verbalizations is not straightforward. Every *unit* is related to a *phrase* (table "phrases"). *Phrase* aggregates the *expressions* (fixed word sequences) stored in the "expressions" table. The *expressions* represent the formal variation of the same routine (*unit*) which was discussed in the beginning of this Section. One formal *expression* can belong to multiple routines. The lexical components of the *expressions* are stored in the "tokens" table. The relations between *units*, *phrases*, and *expressions* are stored as references between tables via their unique identifiers.

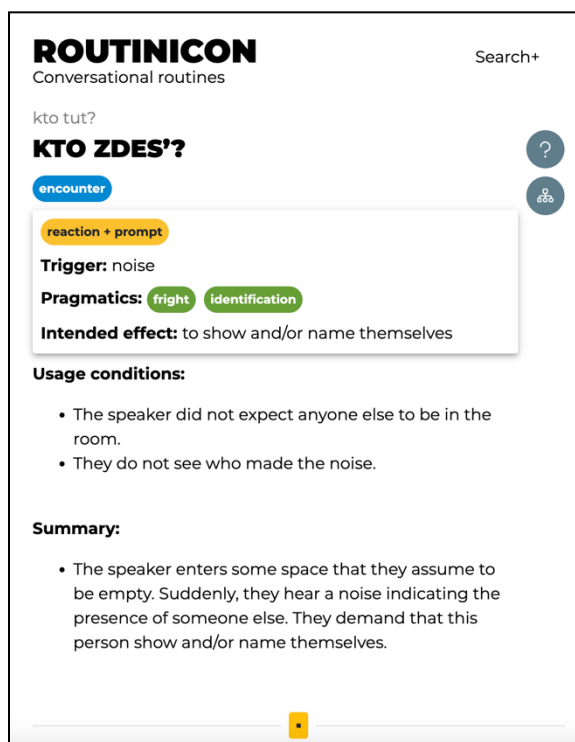
4.2 Preprocessing

The source data is collected in a flat-table spreadsheet. The application (or, to be more precise, the set of applications) is designed in a way to minimize the requirements for data processing and maintaining online presence. Hence, the source data is converted into several related tables of a PostgreSQL database. Afterwards, a client application is built with the processed dataset embedded into it. The technical requirements for the actual hosting platform are quite modest: it is a single page application that only requires several megabytes of space, while the database backend is not needed. It makes the web-platform easy to maintain and host, which is crucial for such projects, especially in terms of its sustainability in the long-term perspective.

4.3 Web-Interface

The web-interface of the project is a single-page application written in JavaScript (with Vue.js library)³. The design of the application is two-fold,

providing capabilities for accessing the routines both via entering a segment of a text form (main page) and filtering the full list by categorical parameters of a routine (search page). The reason for this is that the project is targeting two user groups. The first group is represented by teachers and students of Russian as foreign language who can use the resource as a sort of a dictionary. The second group are the linguists interested in exploring the phenomenon of conversational routines.



The screenshot shows the Routinicon web interface. At the top, the title "ROUTINICON" is displayed in large, bold, black letters, with the subtitle "Conversational routines" below it. A search bar with a magnifying glass icon is in the top right corner. The main content area shows a routine entry for "KTO ZDES?". The routine is categorized as "encounter" (indicated by a blue tag). Below this, a "reaction + prompt" box contains the text "Trigger: noise", "Pragmatics: fright identification" (with "fright" and "identification" in green tags), and "Intended effect: to show and/or name themselves". Underneath, the "Usage conditions" section lists two bullet points: "The speaker did not expect anyone else to be in the room." and "They do not see who made the noise.". The "Summary" section contains one bullet point: "The speaker enters some space that they assume to be empty. Suddenly, they hear a noise indicating the presence of someone else. They demand that this person show and/or name themselves." The interface is clean and modern, with a white background and blue accents.

Figure 1: Representation of a routine in the web-interface

The full annotation of every routine includes fixed-list properties as well as free-text descriptions and is presented in a separate page (see Figure 1). There are three components of the user interface in this page that are developed especially for the language learners: an audio file with the correct pronunciation of the routine, a video fragment from the RNC Multimedia corpus which illustrates the extralinguistic setting, and a button of navigation to the units that are similar to the currently shown routine. The similarity is established automatically: the routines are treated as similar if they share at least one tag in the "pragmatics" field.

For some entries in the Routinicon, the description also includes the links to the formally related entries in the Pragmaticon and Russian Constructicon. In the long run, it is planned to

³ The test version with a limited dataset is available at <https://r.congram.org>. The data in the web-application are downloadable from the interface.

cross-link all the three projects. This would enhance the navigation capabilities and thus help better understand the relations between the linguistic classes of conversational routines, discourse formulas, and schematic constructions.

5. Conclusion

The primary practical purpose of a resource like Routinicon is the application to the language learning. In a way, it serves as a phrasebook, providing the users with an inventory of ready-made linguistic expressions reserved for specific occasions. Regular phrasebooks are made for foreigners that do not speak the language and need the ready-made expressions as the only means to be understood. Meanwhile, native speakers also tend to turn to prefabricated expressions in the typically recurring situations (Barron, 2003; House, 2013). In those cases, they do not need to parse or modify the formula to adjust it to the context and can reproduce it as it is, just like the foreigners that use a phrasebook solution. Ironically, the formulas actually used by the language speakers (except for the basic expressions of politeness like *Thank you!*) are unlikely to end up in the regular phrasebooks since the latter tend to include the most simple and compositional ways to convey the speaker's needs. Meanwhile, the Routinicon offers a collection of linguistic means that can be of interest for the advanced learners that would like to enhance their communicative skills.

At the same time, this resource is designed to enable further theoretical, and above all, typological research on conversational routines. Some phraseological approaches hold the view that the phraseological units are too culturally specific and irregular across languages to fit into a single comparative framework (cf. Dobrovolskij & Piirainen, 2009). On the contrary, we believe that phraseological units can and should be a subject of linguistic typology. However, to make it possible, there is need for a large and thoroughly systematized inventory of these units. In the nearest perspective, the Russian Routinicon will be provided with rough English counterparts to the Russian routines. However, they will merely serve as an approximate representation of the Russian routines' semantics. A multilingual resource requires more thorough research for every included language and should be built on different principles. Cf. the structural differences between the Russian Pragmaticon and the Multilingual Pragmaticon databases discussed in Buzanov et al., 2022. We are already developing a framework for multilingual comparison of routines, starting with the data of Slavic languages.

The third dimension of the Routinicon project is making a step towards a full-cycle digital platform for linguistic data gathering, curation and

exploration. Our digital resource is designed to be a platform for continuous linguistic research and not merely for storage and dissemination of the research results. The fundamental principle of such a platform is to provide means for multi-level comparison of the entries. This is why we try to formalize the descriptions as much as possible. The formalization sets a strict framework for the research. Instead of providing traditional textual definitions, we divide the semantic description into multiple parameters and annotate these parameters using recurring tags. These tags represent generalizations developed to reflect the significant similarities and differences identified between the data entries. Consequently, we do not manually establish synonymy relationships between the linguistic units. These relationships can be found by searching the database.

Thus, on the one hand, the data structurization serves practical needs, and on the other, it provides grounds for theoretical insights into this linguistic domain.

Our long-term goal is to equip the interface with various functionality that facilitates the analysis of the data on the side of the annotator and allows for adjustments in the annotation structure. Then it could serve as a workspace for catalogization of different types of constructions.

6. Bibliographical References

- Aijmer, K. (1997). *Conversational Routines in English*. Routledge.
- Barron, A. (2003). *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. John Benjamins Publishing.
- Buzanov, A., Bychkova, P., Molchanova, A., Postnikova, A., & Ryzhova, D. (2022). Multilingual Pragmaticon: Database of Discourse Formulae. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 3331–3336.
- Constant, M., Eryigit, G., Monti, J., Van Der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics*, 43(4), 837–892.
- Coulmas, F. (Ed.). (1981). *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*. De Gruyter Mouton.
- Dunn, J. (2017). Computational learning of construction grammars. *Language and Cognition*, 9(2), 254–292. Cambridge Core.
- Eckardt, R. (2012). Hereby explained: An event-based account of performative utterances. *Linguistics and Philosophy*, 35(1), 21–55.
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501–538.

- Gerasimenko, E., Puzhaeva, S., Zakharova, E., & Rakhilina, E. (2019). Defining discourse formulae: Computational approach. In *Proceedings of Third Workshop "Computational linguistics and language science"* (Vol. 4, pp. 61–69).
- Haspelmath, M. (2023). On what a construction is. *Constructions*, Bd. 15 Nr. 1 (2023): Special Issue "35 Years of Constructions".
- House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics*, 59, 57–67.
- Janda, L. A., Endresen, A., Zhukova, V., Mordashova, D., & Rakhilina, E. (2020). How to build a constructicon in five years: The Russian example. In *Belgian Journal of Linguistics* (Vol. 34, Issue 1, pp. 161–173). John Benjamins.
- Kecskes, I. (2010). Situation-bound utterances as pragmatic acts. *Journal of Pragmatics*, 42(11), 2889–2897.
- Lyngfelt, B., Borin, L., Ohara, K., & Torrent, T. T. (Eds.). (2018). *Constructicography: Constructicon development across languages*. John Benjamins Publishing Company.

7. Language Resource References

- Pragmaticon. (2021). The Russian Pragmaticon. An electronic database of the Russian pragmatic constructions. <https://pragmaticon.ruscorpora.ru/>
- Russian Constructicon. (2021). An electronic database of the Russian grammatical constructions. <https://constructicon.github.io/russian/>