Language Resources and Evaluation (2026) 60:51
https://doi.org/10.1007/s10579-026-09922-y

ORIGINAL PAPER

Studying valency patterns of bivalent verbs with BivalTyp

Dmitry Nikolaev[1] · Sergey Say[2]

Received: 27 October 2025 / Accepted: 15 April 2026

Abstract
This paper presents BivalTyp, a typological database of bivalent verbs and their argument coding patterns across 140+ languages. Based on a questionnaire comprising 130 contextualized sentences with bivalent predicates, BivalTyp aims to facilitate cross-lingual comparison by combining broad language coverage with rich descriptive detail. The online database provides multiple query and visualization tools, including customizable maps, as well as downloadable datasets. We outline diverse research applications, from identifying transitivity hierarchies and semantic-role clustering to analyzing language-specific valency systems and their complexity. A case study demonstrates how BivalTyp data can disentangle genealogical inheritance from areal convergence in shaping argument-coding patterns, revealing that geographical proximity significantly influences valency pattern similarities even after controlling for phylogenetic relatedness.

Keywords Bivalent verbs · Valency classes · Dataset · Online resource

1 Introduction

The aim of this paper is to present BivalTyp, a typological database of bivalent verbs and their argument coding. The database is designed as a tool for the typological study of valency classes of verbs, i.e. classes of verbs defined by shared patterns of argument coding (e.g. case marking or adpositional marking) that are specified at the level of individual lexical verbs, as illustrated by English *look at* and *listen to*, which

✉ Dmitry Nikolaev
dmitry.nikolaev@manchester.ac.uk

Sergey Say
serjozhka@yahoo.com

1 Department of Linguistics and English Language, University of Manchester, Oxford Rd., Manchester M13 9PL, UK

2 Department of Slavic Linguistics, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Published online: 20 June 2026

differ in their choice of preposition. It is based on a detailed questionnaire and covers a continuously growing sample of languages (145 as of June 2, 2025). The entries in the database consist of translations of the questionnaire sentences into the target languages, with valency patterns annotated according to the principles explained below.

The project originated in 2009 in an offline form.[1] Interim results were summarized in the edited volume (Say 2018), which also included a published version of the questionnaire (in Russian). The online version of BivalTyp was launched in 2020 (Say 2020a).

In the remainder of the paper, we provide an overview of current research on valency patterns and the motivation for BivalTyp (Sect. 2), describe the database design (Sect. 3), outline possible applications (Sect. 4), and present a BivalTyp-based case study disentangling genealogical and areal signals in distributions of valency pattern (Sect. 5), before concluding in Sect. 6.

2 State of the art

In a standard definition, the valency of a verb is understood as "the list of its arguments with their coding properties" (Malchukov et al., 2015, 30). Cross-lingually, argument coding properties include case markers and adpositions (flags), verbal indexes, and, more rarely, word order (Malchukov et al., 2015; Sinnemäki, 2008). Valency plays a central role in typological research, underpinning discussions of transitivity, alignment, grammatical relations, differential argument marking, voice, valency-changing operations, and valency orientation (Hopper & Thompson, 1980; Lazard, 1994; Nichols et al., 2004; Andrews, 2007; Dixon, 2010; Bahrt, 2020).

Most typological studies of valency have focused on a small set of patterns: monovalent (e.g. *The dog barked*), basic transitive (*The child opened the cupboard*), and basic ditransitive (*The teacher gave the book to the girl*). Additional valency patterns are often grouped under the label *non-canonical*, with a large body of work explaining them in terms of referential, semantic, discourse-pragmatic, and diachronic factors (Tsunoda, 1985; Aikhenvald & Dixon, 2001; Bhaskararao & Subbarao, 2004; Næss, 2007; Barðdal et al., 2009; Grossman et al., 2019). While the term *non-canonical* suggests that such patterns constitute a form of deviation, it may be pointed out that lexically determined non-basic patterns, such as English *She is looking at the clouds* vs. *She is listening to the music*, are probably found in all languages. Accordingly, verbs in every language can be grouped into valency classes, defined by the coding devices they assign to their arguments. From this perspective, the so-called canonical patterns correspond to just three large valency classes, coexisting with many others.

The core idea in the study of valency classes is that they are at least partly semantically motivated: verbs with similar meanings tend to pattern together (Fillmore, 1968; Jackendoff, 1990; Lehmann, 1991; Levin, 1993; Van Valin, 1999; Levin & Rappaport Hovav, 2005; Malchukov, 2005; see also Cysouw [2014] on the broader

[1] From 2009 to 2013, the project was funded by two consecutive grants from the Russian Foundation for Humanities and remained coordinated at the Institute for Linguistic Studies, Russian Academy of Sciences, until 2022.

"isomorphism hypothesis" underlying much typological research). This hypothesis has been widely supported by dictionaries and lexical databases (Apresjan, 1982; Helbig & Schenkel, 1983; Fillmore et al., 2003; Ruppenhofer et al., 2016), which collectively stimulated much of the modern work on valency Haspelmath (2014). At the same time, it is precisely in non-canonical classes that cross-lingual differences become most visible (Bickel et al., 2014; Gaszewski, 2020). Despite some important contributions (see below), we still lack a typological understanding of the extent to which languages are similar or different in the principles linking verbs to (non-canonical) valency patterns. Two main obstacles are responsible for this gap – one conceptual and one empirical.

The conceptual problem is the absence of a clear basis for cross-lingual comparison. Valency patterns are defined in terms of language-specific coding devices – such as cases or adpositions – that cannot be convincingly equated across languages, even if they share the same label, such as *dative* or *instrumental* (Haspelmath, 2009, 510). Typological valency research has mainly avoided this problem by relying on arguably more robust comparative concepts such as generalized argument roles (S, A, P), grammatical relations (subject, object), or semantic roles (agent, patient). These concepts, however, also pose challenges as aprioristic tools in cross-lingual comparison (Croft, 2001; Haspelmath, 2011; Cysouw, 2014). More importantly for the present study, they cannot capture fine-grained contrasts, such as that between *look* and *listen*. In English, these verbs display different valency patterns, even though their second arguments would be likely labeled with the same grammatical relation and semantic role.

The empirical problem is the lack of resources that are both broad in language coverage and rich in lexical detail. Dictionaries may describe hundreds of verbs with their valency patterns, but such resources are available only for a few major languages and employ diverse terminologies and annotation schemes. Large typological databases, by contrast, usually cover only a limited set of diagnostic meanings (e.g., 'have' or 'give').

Several typological projects have recently addressed these gaps. Bickel et al. (2014) identify non-canonical valency classes across a wide sample but rely on grammars and dictionaries, which differ in depth and detail. Gaszewski (2020) surveys a large number of verbal meanings, but his sample is geographically restricted to Central Europe. The ValPal project (Malchukov & Comrie 2015) provides the most extensive openly available cross-lingual dataset, covering 36 languages from all macro-areas.[2] It is based on a questionnaire including 70 predefined verb meanings spanning all major numeric valency types (0–3 arguments). While this suffices to identify major classes and alternations, it is not detailed enough to adequately capture non-canonical valency classes and, consequently, to trace recurrent patterns in their distributions.[3]

[2] See also the Pavia Verbs Database Paveda; https://paveda.unipv.it/; (Zanchi et al., 2022), which is an extension of the original ValPal database.

[3] Further studies that are related to BivalTyp's main goals to varying degrees but do not fully align with them and are therefore not discussed here in detail include Cysouw (2014), Hartmann et al. (2014), the World Atlas of Transitivity Pairs (WATP 2014), and Seržant et al. (2022).

The BivalTyp project is designed to address both the conceptual and empirical issues outlined above, as discussed in the following section.

3 Database design

3.1 The questionnaire

BivalTyp is based on a questionnaire comprising 130 sentences with bivalent predicates. In the following, a "predicate" is an item in the questionnaire that links database entries across languages, while a "verb" refers to a language-specific lexical item. The focus on bivalent predicates ('see', 'find', 'wait', 'touch', 'be afraid', 'need', etc.) reflects a compromise between descriptive richness and manageability. On the one hand, investigating how individual verbs are assigned to valency classes requires a sufficiently dense sample of predicates. On the other hand, covering predicates accommodating any number of arguments would produce an unmanageably large questionnaire. Bivalent verbs strike a balance. They often exhibit diverse non-canonical patterns, unlike the more uniform monovalent verbs. At the same time, they still form sizable language-specific classes, unlike trivalent verbs, which are typically rare and idiosyncratic (Bickel et al. 2014).

The BivalTyp questionnaire focuses mainly on predicates from broad semantic classes that the literature identifies as prone to non-canonical coding, such as psychological ('like'), contact ('touch'), or interaction ('help') predicates. Only a small subset of the questionnaire predicates involves a volitional agent acting on a fully affected patient—the configuration that cross-lingually forms the core of the transitive class (Hopper & Thompson, 1980; Tsunoda, 1985, Kittilä, 2002, Malchukov, 2005, Næss, 2007, Malchukov & Comrie, 2015).

Each questionnaire item presents a clause with two target arguments, labeled X and Y, sometimes with additional context in parentheses, e.g., no. 79 'Lightning (X) hit the house (Y)'; no. 62 '(The teacher asked P.[4] what his parents' names are.) P. (X) answered the teacher (Y)'. Arguments are labeled according to Dowty's (1991) lexical entailments, with X corresponding to the more agent-like participant.

The questionnaire follows the "exemplar semantics" approach, which treats contextualized, grounded meanings as more tangible than generalized meanings derived as coarse summaries of individual uses (Cysouw, 2014, 24–25). In practice, this means that translations need not involve specialized verbs in the target language. Thus, item no. 79 above may yield verbs with broad meanings like 'hit', 'attain', or 'fall', while no. 62 may produce a trivalent structure such as, literally, 'P. gave an answer to the teacher'. While translations may not map directly onto language-specific lexical classes, they are expected to convey comparable meanings across languages, thereby enhancing cross-lingual comparability.

In sum, the questionnaire consists of full situationalized contexts, while English verbs like 'touch', 'wait', or 'answer' serve as convenient labels.

[4] Placeholders like this are used for any appropriate person name in the target language.

3.2 Data collection and language sample

The database relies entirely on contributions from language experts ("contributors"). By default, contributors elicit translations from native speakers using one of the available questionnaire versions (Russian, English, French, or German). If multiple translations are offered, contributors select one according to four criteria specified in the guidelines. The criterion overriding all others is *naturalness*, with contributors instructed to use all available means, such as multiple consultants and corpus frequencies, if available, to ensure that the selected translation could be used in naturalistic contexts. In this way, the BivalTyp design aims at mitigating the possible effects of using translated sentences as database entries. Contributors also identify the coding devices used for the X and Y arguments – typically cases, adpositions, or verbal indices. If a translation does not allow for unambiguous identification of these devices (e.g. due to syncretism), contributors try to obtain modified translations of the same stimulus to ensure clarity. Some translations fail to meet acceptance criteria, such as the requirement that both X and Y be overtly expressed as clause-level dependents in the same clause, excluding, for example, second-order adnominal modifiers in constructions like *The boy* (X) *answered the teacher's* (Y) *question*. These cases result in gaps in the dataset. In addition, contributors provide metadata on their languages, including genealogical classification at two levels ("family" and "genus," following The World Atlas of Language Structures [Dryer & Haspelmath, 2013]), geographical coordinates,[5] and several typological parameters (basic word order, number of morphological cases).

BivalTyp's sample is a convenience sample. As of June 2, 2025, it included 145 languages, the vast majority from Eurasia. Even within Eurasia, coverage is uneven: some regions (Europe, Northern Asia, and especially the Caucasus) are represented much more densely than others. Nevertheless, the sample is genealogically diverse, spanning 20 families and 48 genera. The strong genealogical and geographical bias of the sample limits the interpretation of BivalTyp observations in terms of universal tendencies; however, recent approaches separate data collection from statistical modelling and offer tools to control for such biases (see Guzmán Naranjo & Becker [2022] and Sect. 5 below).

3.3 Entry structure

Each database entry that meets the acceptance criteria contains four obligatory fields: "X", "Y", "Valency pattern", and "Locus". The X and Y fields record schematic representations of the coding devices used for the two arguments. The valency pattern, the main response variable, is defined as the ordered (X+Y) combination of these devices. For example, the Budugh entry for no. 9 'see', shown in (1), has the

[5] By default, contributors use the coordinates provided by Glottolog, overriding them if there are errors in Glottolog or if the variety studied differs from the one registered there.

valency pattern schematically represented as DAT_NOM, meaning that X ('the seer') is marked with the dative case and Y ('the seen') with the nominative.[6]

(1)	aslan-ız		k'ul	ürqhü-rü
	PN(M)-DAT		house(III)	III.see.IPFV-NONPST
	'Aslan	sees a	house.'	

Valency-pattern tags are language-particular and sometimes include specific markers. For instance, the German tag NOM_aufACC represents a pattern whereX is nominative and Y is coded by auf 'on' plus the accusative.

For each language, one pattern is designated as transitive ("TR"), namely the pattern where X and Y are coded by the same devices as the killer and victim in the entry for 'kill' (cf. Haspel math, 2015, 136). For an easily operationalizable definition of transitivity, we needed a verb that (i) displays the usual semantic features of transitivity – with a volitional agent acting on a patient that undergoes a significant change as a result of the action – and (ii) is expected to have equivalents in all languages of the sample. 'Kill' met both criteria. In Budugh, the transitive pattern, illustrated in

(2), involves ergative case for X and nominative case for Y.

(2)	aslan-ır		rosul	at'u-ci
	PN(M)-ERG		PN(I)	I.kill.PFV-PRF
	'Aslan	killed	Rasul.'	

The devices used in a language's transitive pattern are termed "direct", while all others are "oblique". This distinction underlies the parameter of "locus of non-transitivity" ("locus" for short). By definition, both arguments are direct in transitive patterns. Non-transitive patterns are classified as X-locus, Y-locus, or XY-locus depending on which argument(s) are coded by oblique devices. The Budugh pattern in (1), for example, displays the X-locus because only X is marked by an oblique device (dative).

The rationale for this four-way distinction follows Malchukov's (2006) relevance principle: deviations from syntactic transitivity are typically marked on the participant that departs from high semantic transitivity in Hopper & Thompson's (1980) sense. Thus, X-locus is expected for predicates with less agentive X arguments, and Y-locus for those with less affected Y arguments. That said, there is a high degree of cross-lingual variability in how participants are coded, and the locus parameter offers a cross-lingually applicable annotation grid.

Most entries in the online database also include the translation of the questionnaire sentence into the target language, glosses, an English back-translation, the citation form of the verbal expression, and comments where necessary (for 36 languages, these fields are not yet available online but should be made publicly available by

[6] In Budugh, the verb normally agrees with the nominative argument, so indexing does not provide additional information for identifying valency classes. Indexing is taken into account when identifying valency patterns if there are indexing contrasts between arguments that are not predictable from flagging contrasts, as in, e.g., Adyghe.

the end of 2027). Verbal expressions are usually simplex verbs but may also be non-verbal predicates (e.g. lit. ‘be envious’ for no. 116 ‘envy’), or complex verbs (e.g. lit. ‘throw paint’ for no. 40 ‘paint’).[7]

3.4 Online tools

In addition to several sections with meta-information (e.g. questionnaire, contributor instructions, user guidelines), the database website offers five tabs for exploring and analyzing the data: “Languages”, “Predicates”, “Data overview”, “Maps”, and “Download”. The “Languages” tab provides two views: a map with dots for the languages and an interactive spreadsheet. The spreadsheet combines contributor-supplied metadata (genealogical affiliation, coordinates, Glottocode, basic word order, etc.) with summary statistics derived from the database, including counts of entries for the four locus types, the total number of distinct valency patterns, and entropy-based variation metrics. Both the map and the spreadsheet provide links to individual language pages, where all entries for that language are displayed together, preceded by an introduction with data collection details, a short grammar sketch, and a list of glossing abbreviations.

The “Predicates” tab combines datapoints by predicate, so that, for example, all entries for ‘be afraid’, ‘touch’, or ‘wait’ can be displayed together.

The “Data overview” tab offers two ways to view the database as a whole. The “Patterns overview” is a table where rows are predicates, columns are languages (with tools for selecting customized language groups), and cells contain language-specific valency pattern tags. The “Data in one spreadsheet” mode presents a large table where each row corresponds to a single entry with all its fields. This mode is especially useful for creating customized subsets.

The “Maps” tab provides three types of maps. The “Transitivity and locus ratio maps” represent the prevalence of each locus type as percentage ranges. For example, Fig. 1 illustrates the prevalence of X-locus patterns, revealing areal trends such as the concentration of X-locus in the Caucasus. The “Customizable transitivity and locus maps” show locus distributions for individual predicates. Figure 2 presents such a map for no. 95 ‘look’, with patterns that are either transitive or display the Y-locus, as in English *The boy is looking at the clouds*. The third type, “Customizable predicate-pair maps”, shows whether a selected pair of predicates (e.g. no. 3 ‘be afraid’ and no. 35 ‘avoid’) share the same valency pattern across languages.

Finally, the “Download” tab provides access to (i) numbered database releases stored at Zenodo and (ii) a link to a GitHub repository with the latest development version, which may be of use given that the database is continuously expanding.[8]

[7] Argument-encoding information in the four obligatory fields for languages that lack this additional information can be found in the downloadable dataset tables and in the “Patterns overview” section of the website.

[8] The data and code used for the analyses in this paper can be found at https://github.com/macleginn/bivaltyp-paper

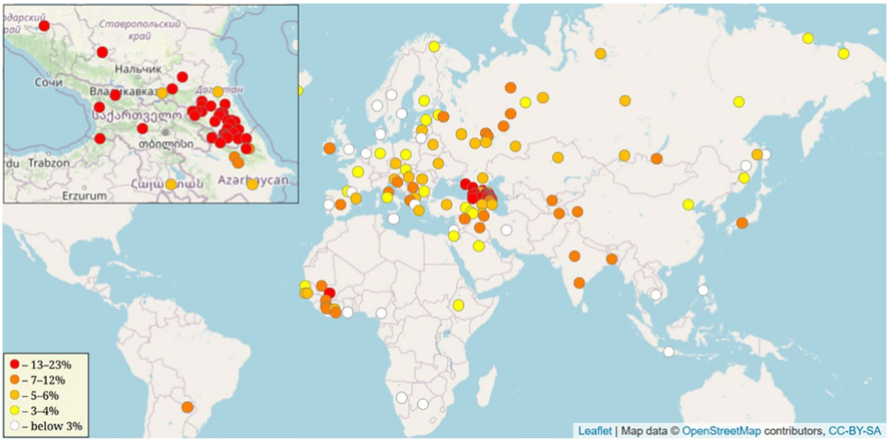

Fig. 1 Prevalence of X-locus patterns across the languages of the sample. The inset zooms in on the languages of the Caucasus

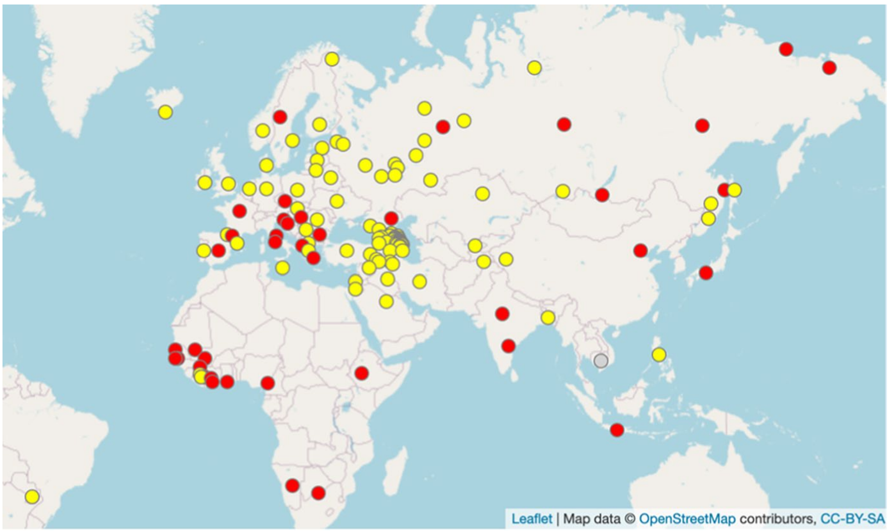

Fig. 2 Locus in the entries for 'look': TR (red) vs. Y (yellow)

4 Research topics with BivalTyp

In this section, we outline research topics explored or potentially addressable with BivalTyp, with selected references to studies based on the project. If the project's main output is imagined as a table with predicates as rows and languages as columns, the data can be approached from two complementary perspectives. The first perspective collates observations across languages to identify cross-lingually recurrent phenomena (Sect. 4.1). The second perspective compares languages by pooling observations from individual predicates, highlighting inter-language similarities and differences (Sect. 4.2).

4.1 Cross-lingually recurrent tendencies

i. *Transitivity and locus-based prominence scales.* Predicates are not equally likely to display different valency patterns across languages. Such generalizations can be captured in cross-lingual hierarchies, for example, based on a predicate's *transitivity prominence* – the ratio of transitive to all bivalent patterns in a language sample (Haspelmath & Hartmann,2015, 142–144). BivalTyp allows for testing these hierarchies as well as similar ones reflecting the prevalence of different locus types (Say, 2018, 560–566). It can also shed light on inherent asymmetries between X- and Y-locus.
ii. *Argument co-expression likelihood and semantic microrole clustering.* Systematic sharing of coding devices across predicates suggests semantic similarity, thus leading to more empirically substantiated definitions for semantic roles. At a minimum, BivalTyp refines our understanding of such co-expression tendencies (Say, 2018, 566–572), complementing findings from other datasets (Bickel et al. 2014; Cysouw 2014; Hartmann et al. 2014).
iii. *Argument coding consistency.* Argument coding in individual languages is partly semantically motivated (predictable) and partly lexically assigned (idiosyncratic). BivalTyp's data allow assessing to what extent valency patterns are cross-lingually consistent, thus providing empirical support for earlier observations that predicates differ in their degree of idiosyncrasy (Huddleston & Pullum, 2002; Woolford, 2006); (Haspelmath, 2014). For example, the so-called psychological predicates often display more idiosyncratic patterns than symmetric predicates such as 'fight' or 'agree', while direct-effect verbs like break systematically display the most predictable patterns (Say, 2018, 572–575).
iv. *Verbal expression types and argument coding.* BivalTyp records not only simplex verbs but also non-verbal predicates and complex verbs. This enables identifying which meanings are most likely to attract these expression types and testing hypotheses about their relation to argument coding (Say, 2018, 602–604), e.g., that non-verbal predicates are rarely transitive.

4.2 (Dis)similarities between languages

i. *Typologically informed analyses of individual languages and taxa.* BivalTyp provides a solid basis for analyzing the valency patterns of a specific language (or genealogical group) against a typological background, helping to distinguish rare and common traits and detect cases of areal convergence (cf. Ovsjannikova [2018] on Forest Enets; Blinova [2020] on Saami; Say [2020b] on Neo-Aramaic; Studenikina [2023] on Hill Mari).
ii. *Analyses within genealogical taxa.* For related languages, language-specific valency patterns can often be equated through shared marker etymologies. In

such cases, BivalTyp is a useful tool for analyzing divergence within genealogical groups, ideally combined with verb cognacy tracing (Rostovtsev-Popiel, 2024) on Kartvelian; (Kozhanov & Say, 2025) on Romani; (Say, 2025a) on Slavic.

iii. *Transitivity prominence.* Languages differ in *transitivity prominence*, calculated as the ratio of transitive to all bivalent patterns in the language-specific dataset (Haspelmath, 2015, 139–142). To illustrate, languages of the Caucasus tend to use non-transitive patterns for many meanings typically conveyed by transitive verbs in the languages of Western Europe; this is captured by transitivity prominence, cf. 0.28 for Avar (Nakh-Daghestanian) vs. 0.60 for Dutch. Languages with similar transitivity prominence scores may still vary in which predicate meanings are coded by transitive verbs. BivalTyp enables such comparisons across the entire language sample (Say, 2018, 576-582) or its subsets (Say, 2014, 2017; Kozhanov & Say 2025).

iv. *Valency class system complexity.* Some languages make finer distinctions in their valency class systems, yielding greater complexity (cf. Hawkins [1986] on German vs. English). BivalTyp allows systematic exploration of such differences, e.g. with entropy-based metrics (Say, 2018; 2025b).

v. *Locus type prevalence.* Like transitivity, locus types of non-transitivity (X-locus vs. Y-locus vs. XY-locus) vary in prevalence across languages, potentially reflecting competing principles shaping deviations from transitivity, as well as phylogenetic and areal tendencies Say (2018).

vi. *Organization of verbs into valency classes.* Languages differ in how they structure their verbal lexica across (minor) valency classes. BivalTyp enables tracing these (dis)similarities with quantitative methods (Say, 2014; 2018).

The (dis)similarities described above can be captured as indices, hierarchies, maps, or distance matrices. Ultimately, valency-related typological variation can be linked to other structural dimensions (e.g. word order and morphological-case richness; see Say [2025b] on valency-class-system complexity and its structural correlations) or to genealogical and areal relations (Alfimova, 2024). Our case study in Section 5 focuses on this latter task, examining argument-coding differences discussed in 4.2(v) and 4.2(vi) above.

5 Case study: genealogical vs. areal factors in shaping argument-coding differences

Disentangling the differential effects of genealogical inheritance and areal convergence is one of the key challenges in current typological research (Murawaki & Yamauchi, 2018; Ranacher et al., 2021; Skirgård et al., 2023; Sinnemäki et al., 2024). The BivalTyp data make it possible to approach this problem for certain valency-related phenomena.

To do so, the analysis begins with a distance matrix capturing cross-lingual differences. In this case study, we consider two types of linguistic distances derived from the BivalTyp dataset, with individual languages represented as 130-item vectors, each item corresponding to a predicate and coded as a categorical variable.

The first distance measure, DistValLoc, originally introduced by Say (2014; see also Say [2025a]), captures cross-lingual differences in the assignment of individual verbs to four locus types: “TR”, “X”, “Y”, and “XY”. Since locus is defined as a comparative concept with a universally applicable value set, DistValLoc is calculated simply as the normalized Hamming distance, i.e. the number of positions in two language vectors at which the corresponding locus annotations are different, restricted to the predicates annotated in both languages and divided by the number of such predicates. More formally:

$$\mathrm{DistValLoc}(L_1, L_2) = \frac{|i : L_1[i] \neq \mathrm{NA} \wedge L_2[i] \neq \mathrm{NA} \wedge L_1[i] \neq L_2[i], 1 \leq i \leq n|}{|i : L_1[i] \neq \mathrm{NA} \wedge L_2[i] \neq \mathrm{NA}, 1 \leq i \leq n|},$$

where n is the number of surveyed predicates. To illustrate, DistValLoc(Armenian, Azerbaijani) equals 0.176, calculated as 22 (the number of predicates with different locus annotations in the two languages) divided by 125 (the total number of predicates with non-missing datapoints in both languages).

The second distance measure, also proposed by Say (2014; 2025a), which we refer to as DistValPat, is aimed at capturing cross-lingual differences in the ways individual verbs are assigned to specific valency classes, as reflected in BivalTyp’s valency-pattern annotations. These annotations, unlike locus annotations, are fully language specific and cannot be directly equated across languages. DistValPat therefore focuses on the ways individual languages partition their verbs into valency classes and quantifies how well we can predict if two verbs are in the same valency class in language B if we know how verbs are divided into valency classes in language A. Similarly to DistValLoc, when computing this measure we only consider the predicates that have non-missing values in both language vectors. DistValPat can be equivalently operationalized as the complement of the average of mutual information between the two clusterings weighted by the entropies of the clusterings or as the average of conditional entropies of the two clusterings vis-à-vis each other weighted by the same entropies:

$$\begin{aligned}\mathrm{DistValPat}(L_1, L_2) &= 1 - \frac{1}{2}\left(\frac{MI(L_1, L_2)}{H(L_1)} + \frac{MI(L_1, L_2)}{H(L_2)}\right)\\ &= \frac{1}{2}\left(\frac{H(L_1|L_2)}{H(L_1)} + \frac{H(L_2, L_1)}{H(L_2)}\right)\end{aligned}$$

Both measures range from 0 to 1.

For statistical analyses to work properly, our distance measures should be metrics, i.e. obey non-negativity, symmetry, and the triangle inequality.[9] The first two conditions are trivially satisfied, but as both measures depend on the size of the intersec-

[9] Which states, in our case, that given three languages A, B, and C it can never be the case that the sum of DistValLoc or DistValPat distances between A and B, on one hand, and B and C, on the other hand, can be smaller than the distance between A and C, similarly to how the sum of the lengths of two sides of a triangle can never be smaller than the length of the third side.

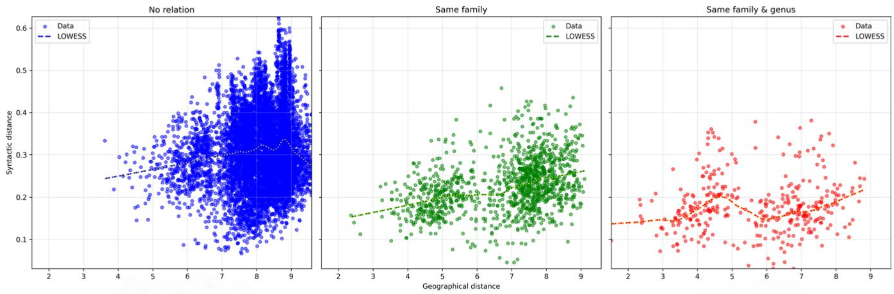

Fig. 3 The empirical dependency of DistValLoc on the log-transformed geographical distance with LOWESS smoothing curves

tion of the set of annotated predicates in a language pair, which can vary between language pairs, the triangle inequality can be easily violated. We safeguard against that by excluding the four languages with less than 90 annotated predicates – this is enough to ensure that triangle inequality, with some tolerance for numerical errors, is never violated in the computed distance matrices. We also exclude Latin and Ancient Greek, the two ancient languages in the dataset, thus reducing the size of the sample analyzed in this case study to 139.[10]

The upper-right halves of the distance matrices computed using the two measures are the response variables that we target in our statistical modelling. We regress the values, corresponding to different language pairs, on two genealogical predictors and a single predictor representing geographical distance between languages.

We encode genealogy as a single categorical variable with three values: *no relation*, *same family & different genus*, *same genus*. Following the standard practice in regression modelling, which is well suited to our case, *no relation* is treated as the base case, which does not have its own variable, while *same family & different genus* and *same genus* each get their own binary values, zero or one of which can be equal to 1 at any given time.[11]

As for modelling geography, we use the traditional approach of directly regressing linguistic distances on geographical distances. We compute the latter based on latitudes and longitudes using the Vincenty algorithm[12] and log-transform the values for modelling stability since pairwise distances between our languages are distributed roughly lognormally.

The empirical relation between the geographical distances and the two structural distance metrics, DistValLoc and DistValPat, is shown in Figs. 3 and 4, respectively.

[10] Extinct languages were excluded because geographic distance, operationalized via present-day coordinates, does not reliably reflect potential contact or areal interaction in their case.

[11] This setup is a compromise between, on the one hand, having a single ordered "relation" variable (*no relation* = 0, *same family & different genus* = 1, *same genus* = 2), which, in the standard OLS setup, would imply that the step from *no relation* to *same family & different genus* and from *same family & different genus* to *same genus* is of the same magnitude, and on the other hand, introducing two correlated binary predictors *same family* and *same genus*.

[12] An iterative algorithm for computing geodesic distances on the surface of a spheroid, which provides more accurate estimates than the traditional great-circle method. See Vincenty (1975) for details.

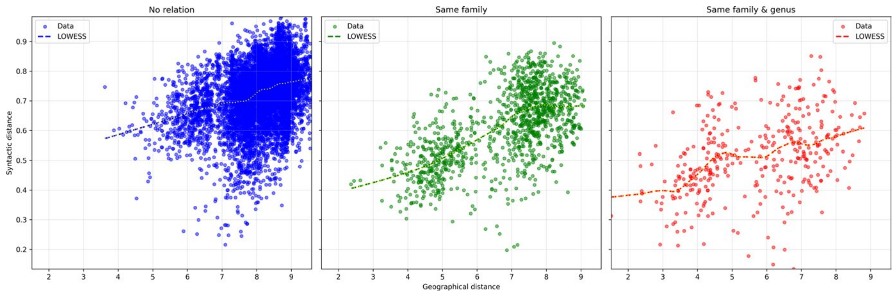

Fig. 4 The empirical dependency of DistValPat on the log-transformed geographical distance with LOWESS smoothing curves

The plots demonstrate that, unsurprisingly, geographical distances are not independent from genealogy (pairs of related languages correspond, on average, to lower geographical distances) and, more surprisingly, that the effect of geographical distance on argument-coding similarity is much more pronounced if we measure the latter using DistValPat, i.e. the measure that disregards any specific morphosyntactic details focusing exclusively on the lexical composition of language-specific valency classes. The plots also demonstrate that in all cases the relationships are roughly linear, which motivates the use of a straightforward linear regression to model them.

In order to create uncorrelated predictors for our model, we treat the genealogical predictors as primary and decompose the geographical distance into a component that can be predicted from genealogy and an independent component. We achieve this by fitting an auxiliary regression model that predicts log-transformed Vincenty distances using the two genealogical binary variables. After fitting the model, we treat the so-called "residuals" (differences between the model's predictions and the true values) as the component of geographical distances that cannot be explained by phylogeny and include that part as a predictor in the main model.

We then consider all subsets of the individual predictors and their pairwise interactions and test the performance of each subset using two approaches: by checking the Akaike Information Criterion and by considering the predictive performance (R-squared) of the model on held-out data, using 5-fold cross validation. Both methods agreed on selecting the same set of predictors for DistValLoc and another one for DistValPat. We then fitted two models with selected predictors and additionally fitted 10,000 bootstrap replications of the models with the number of language pairs from each phylum-phylum combination capped at 500 to compute.95 confidence intervals for the coefficients in a stratified-sampling setting. The model summaries are shown in Table 1.

All selected predictors are significant, in that their confidence intervals never include zero (although the coefficient for the *same genus*: *geographical distance* interaction has a very small magnitude). Furthermore, both models demonstrate the same tendencies: different degrees of genealogical and geographical relatedness correspond to reduced DistValLoc and DistValLoc values, with the effect of geography clearly noticeable over and above that of genealogy. What is remarkable, however, is that the interaction between structural distance and geography has opposite effects at the inter-genera and intra-genus level. At the intra-genus level, both models have

Table 1 The intercepts and coefficients for the linear-regression models predicting pairwise DistValLoc and DistValPat distances with .95 confidence intervals. SFDG – same family, different genus; SG – same genus; GeoDRes – residuals of regressing log-transformed geographical distances on SFDG and SG

	DistValLoc	Conf. int	DistValPat	Conf. int
Intercept	0.312	(0.311, 0.314)	0.72	(0.718, 0.721)
SFDG	− 0.086	(− 0.09, −0.082)	− 0.11	(− 0.117, − 0.104)
SG	− 0.13	(− 0.137, −0.123)	− 0.206	(− 0.219, − 0.194)
GeoDRes	0.016	(0.015, 0.017)	0.04	(0.038, 0.043)
SFDG:GeoDRes			0.011	(0.007, 0.015)
SG:GeoDRes	− 0.012	(− 0.016, −0.008)	− 0.009	(− 0.016, − 0.0017)

a negative coefficient for the *same genus*: *geographical distance* interaction, which means that the slope of the dependence of the response on distance is less steep in the same-genus group than on average. Conversely, although this effect is reported only for DistValPat, the same slope is steeper for language pairs from the same family but different genera.

This finding may be tentatively interpreted as follows: in our data, most languages in close contact are genealogically related (cf. Figs. 3 and 4). Contact-induced convergence is therefore expected to be most pronounced among genealogically related languages. However, languages from the same genus already start out as highly similar due to common inheritance and then diverge, seemingly rather slowly and mainly through geographical drift. In contrast, languages from the same family but different genera share few inherited traits and start out less similar; yet, when in in contact, they tend to converge rapidly. As a result, the role of the geographical distance becomes more pronounced. A similar pattern was reported, based on phonological-inventory data, by Nikolaev (2018). These findings suggest that even though over a wide range of morphosyntactic features the phylogenetic signal is arguably stronger (Skirgård et al., 2023), there is a principled way to stratify different linguistic phenomena by their affinity to geographical spread, as argued by Moran et al. (2024). More specifically, the lexical distribution of individual verbs across valency patterns is arguably a phenomenon that is more likely to converge areally than more basic morphosyntactic phenomena that are mainly discussed by Skirgård et al. (2023).

One caveat in this analysis is that while residuals of log-transformed distances are uncorrelated with genealogical variables, they are not perfectly uncorrelated with the actual unobserved relatedness. This confound arises when dealing with "unrelated" languages (cf. Indo-European vs. Uralic and Indo-European vs. Niger-Congo) as well as inside a given family (different branches of Indo-European are not all pairwise phylogenetically equidistant), and even inside a given genus (different pairs of Slavic languages are certainly not pairwise phylogenetically equidistant). Thus, there may

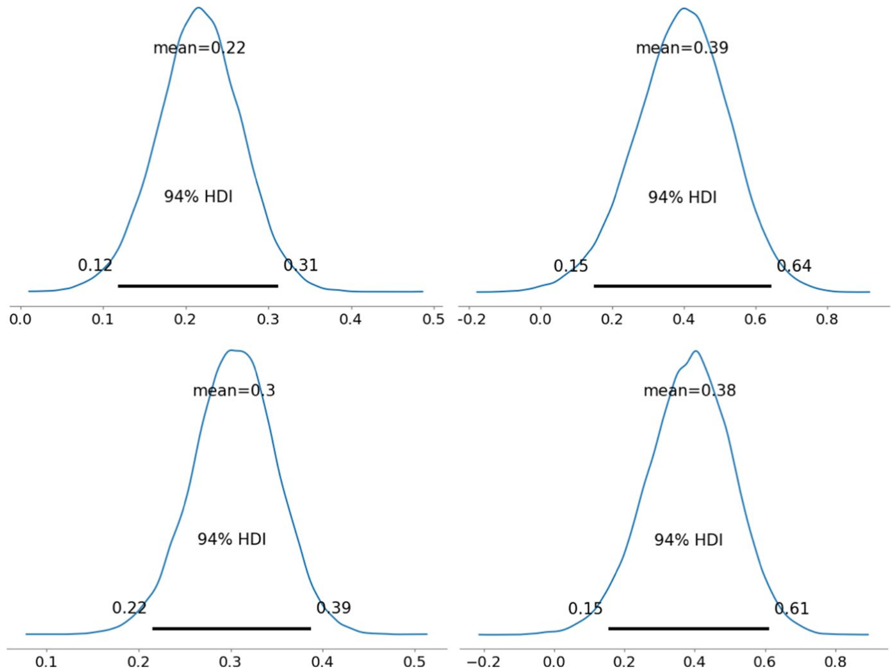

Fig. 5 Draws from the posterior distribution of the slope of the regression of linguistic distance (top: DistValLoc; bottom: DistValPat) on the log-transformed geographical distance in the different-family (left) and same-family-different-genera (right) subsets

be cases of genealogy in disguise of geography, where language pairs with same notional relatedness level are in fact more closely related when they are separated by shorter geographical distances than more geographically distant pairs. It is therefore, unfortunately, impossible to estimate the precise effect of geography *per se* without very detailed phylogenetic knowledge, which is generally unavailable.

The best option to control for these effects, given available data, seems to be to construct hierarchical models for data subsets at different levels of relatedness, with family pairs (for unrelated languages), genus pairs (for *same family* & *different genera*), and genera (for *same genus*) as random effects. The situation inside genera remains as above, but when looking, for example, at all pairs of languages where one is Indo-European and the other Uralic, or likewise Slavic and Germanic, we can safely assume that the degree of genealogical relatedness in these pairs is the same. (In this way we can also avoid a potential case of Simpson's paradox.) Unfortunately, due to small sizes of individual genera, models fitted in this way have convergence issues, which make the results suspect, but with filtering out small ($N < 10$) genera, choosing very conservative priors, and carefully tuning the sampling procedure we can just make the Bayesian approach work.

Sampled posterior slopes for the dependence of the linguistic distances on the log-transformed geographical distance are shown in Fig. 5. Our observations from Figs. 3 and 4 and Table 1 above hold, with slopes for the different-families and different-genera settings being robustly positive. In other words, even after our best attempt to control for phylogenetic relatedness, pairwise similarities in the coding patterns of bivalent predicates remain largely driven by geography. Given the Eurasian bias of the BivalTyp sample, however, it is still an open question whether this effect is as strong in other macro-areas.

6 Conclusion

In this article, we presented BivalTyp, a database of argument coding in bivalent predicates, which provides fine-grained data on individual languages and at the same time can be used for statistical analyses. Already in its offline stage, BivalTyp has provided interesting findings in typological morphosyntax, and we hope that its availability as a downloadable dataset with rich query mechanisms and mapping functionality will make it a useful tool for linguistics research.

An obvious desideratum for the future development of BivalTyp is the inclusion of more languages from macro-areas that are currently underrepresented or not represented at all. Based on our experience so far, we believe that the growth of BivalTyp can play a beneficial role in the documentation of underdescribed languages, and we therefore strongly hope to attract more contributions from descriptive linguists worldwide.

That said, even with its areal bias, we hope to have shown that BivalTyp is a large and carefully designed database that can serve as a valuable tool for analyzing the ways in which languages map verbal meaning onto valency patterns, thereby establishing both robust typological generalizations and cross-lingual differences that can be interpreted in genealogical and areal terms.

Author contributions D.N. designed and implemented the case study and contributed the bulk of the analysis section, in addition to technical work on the underlying database. S.S. led the development of the described resource and contributed the description of the database, the overview of the research applications, and the literature review. The final text of the article was written jointly by the two authors.

Funding Sergey Say's contribution was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317633480 – SFB 1287.

Data availability A link to a repository with data and code for the analyses is provided in the manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Aikhenvald, A., Dixon, R.M.W., & Onishi, M. (Eds.). (2001). *Non-canonical marking of subjects and objects*. John Benjamins.

Alfimova, D. (2024). Object marking in Western Eurasia. The Circum-Baltic area dissolves into the broader areal background. *Journal of Language Contact, 17*, 281–314.

Andrews, D.A. (2007). The major functions of the noun phrase. T. Shopen (Ed.), *Language typology and syntactic description, vol. i, clause structure* (2nd ed., pp. 132–223). Cambridge: Cambridge University Press.

Apresjan, Y.D., & Páll, E. (1982). *Russkij glagol–vengerskij glagol. upravlenie i sočetaemost'* (Vols. 1–2). Budapest: Tankönyvkiadó.

Bahrt, N. (2020). *Towards a typology of voice syncretism* (Unpublished doctoral dissertation)University of Helsinki.

Barðdal, J., & Eythorsson, T. (2009). The origin of the oblique subject construction: An Indo-European comparison. V. Bubenik et al. (eds.), *Grammatical change in Indo-European languages* (pp. 179–193). John Benjamins.

Bhaskararao, P., & Subbarao, K.V. (Eds.). (2004). *Non-nominative subjects*. John Benjamins.

Bickel, B., Zakharko, T., Bierkandt, L., & Witzlack-Makarevich, A. (2014). Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. *Studies in Language, 38*(3), 485–511. https://doi.org/10.1075/sl.38.3.03bic

Blinova, A., & Shagal, K. (2020). Dvuxmestnye predikaty v saamskix jazykax v areal'noj perspective [bivalent predicates in saami languages from an areal perspective]. K.P. Semenova (Ed.), *Malye jazyki v bol'šoj lingvistike: Sbornik trudov konferencii [minority languages in big linguistics: Conference proceedings]* (pp. 25–33). Moscow: Buki Vedi. (In Russian)

Croft, W. (2001). *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cysouw, M. (2014). Inducing semantic roles. S. Luraghi and H. Narrog (Eds.), *Perspectives on semantic roles* (pp. 23–68). Amsterdam Philadelphia: John Benjamins.

Dixon, R.M.W. (2010). *Basic linguistic theory. Volume 2: Grammatical topics*. Oxford: Oxford University Press.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*, 547–619.

Dryer, M.S., and Haspelmath, M. (Eds.). (2013). *Wals online (v2020.4)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13950591

Fillmore, C. (1968). The case for case. E. Bach and R. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart and Winston.

Fillmore, C. J., Johnson, C. R., & Petruck, M. R. (2003). Background to FrameNet. *International Journal of Lexicography, 16*(3), 235–250.

Gaszewski, J. (2020). Does verb valency pattern areally in Central Europe? A first look. L. Szucsich, A. Kim, and U. Yazhinova (Eds.), *Areal convergence in Eastern Central European languages and beyond* (pp. 13–53). Berlin: Peter Lang.

Grossman, E., Seržant, I., & Witzlack-Makarevich, A. (2019). Valency and transitivity in contact: Theoretical and empirical issues. *Journal of Language Contact,12*(1), 1–26.(Editorial introduction)

Guzmán Naranjo, M., & Becker, L. (2022). Statistical bias control in typology. *Linguistic Typology, 26*(3), 605–670. https://doi.org/10.1515/lingty-2021-0002

Hartmann, I., Haspelmath, M., & Cysouw, M. (2014). Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language, 38*(3), 463–484.

Haspelmath, M. (2009). Terminology of case. A. Malchukov and A. Spencer (Eds.), *The Oxford handbook of case* (pp. 505–517). Oxford: Oxford University Press.

Haspelmath, M. (2015). Transitivity prominence. A. Malchukov and B. Comrie (Eds.), *Valency classes in the world's languages. Vol. 1. Introducing the framework, and case studies from Africa and Eurasia* (pp. 131–147). Berlin & Boston: De Gruyter Mouton.

Haspelmath, M., & Hartmann, I. (2015). Comparing verbal valency across languages. A.L. Malchukov and B. Comrie (Eds.), *Valency classes in the world's languages. vol. 1. introducing the framework, and case studies from Africa and Eurasia* (pp. 41–71). De Gruyter Mouton.

Haspelmath, M. (2011). On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology, 15*(3), 535–567. https://doi.org/10.1515/LITY.2011.035

Haspelmath, M. (2014). Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery, 12*(2), 3–11.
Hawkins, J. A. (1986). *A comparative typology of English and German: Unifying the contrasts*. Austin: University of Texas Press.
Helbig, G., & Schenkel, W. (1983). *Wörterbuch zur Valenz und Distribution deutscher Verben*. Tübingen: Max Niemeyer.
Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251–350.
Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Jackendoff, R.S. (1990). *Semantic structures* (No. 18). The MIT Press.
Kittilä, S. (2002). *Transitivity: Towards a comprehensive typology*. Turku: Åbo Akademis Tryckeri.
Kozhanov, K., & Say, S. (2025). Variation in valency patterns across Romani dialects is primarily shaped by contact languages. *Studies in Language, online,*. https://doi.org/10.1075/sl.23067.koz
Lazard, G. (1994). *L'actance*. Paris: Presses Universitaire de France.
Lehmann, C. (1991). Predicate classes and participation. H. Seiler and W. Premper (Eds.), *Partizipation. Das sprachliche Erfassen von Sachverhalten* (pp. 183–239). Tübingen: G. Narr.
Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.
Malchukov, A. (2005). Case pattern splits, verb types and construction competition. M. Amberber and H. de Hoop (Eds.), *Competition and variation in natural languages: The case for Case* (pp. 73–117). Oxford: Elsevier.
Malchukov, A. (2006). Transitivity parameters and transitivity alternations: constraining co-variation. L. Kulikov, A. Malchukov, and P. de Swart (Eds.), *Studies on case, valency and transitivity* (pp. 329–359). Amsterdam: Benjamins.
Malchukov, A., and Comrie, B. (Eds.). (2015). *Valency classes in the world's languages*. Mouton de Gruyter.
Malchukov, A., et al. (2015). Leipzig questionnaire on valency classes. A. Malchukov and B. Comrie (Eds.), *Valency classes in the world's languages. Vol. I. Introducing the framework, and case studies from Africa and Eurasia* (pp. 27–40). Berlin: Mouton de Gruyter.
Moran, S., Eisen, E., Nikolaev, D., & Grossman, E. (2024). Operationalizing borrowability: Phonological segments as a case study. *Language, 100*(4), 671–698. https://doi.org/10.1353/lan.2024.a947038
Murawaki, Y., & Yamauchi, K. (2018). A statistical model for the joint inference of vertical stability and horizontal diffusibility of typological features. *Journal of Language Evolution*. https://doi.org/10.1093/jole/lzx022
Næss, Å. (2007). *Prototypical transitivity*. Amsterdam & Philadelphia: John Benjamins.
Nichols, J., Peterson, D. A., & Barnes, J. (2004). Transitivizing and detransitivizing languages. *Linguistic Typology, 8*(2), 149–211.
Nikolaev, D. (2018). Areal dependency of consonant inventories. *Language Dynamics and Change, 9*(1), 104–126.
Ovsjannikova, M. A. (2018). Kodirovanie u?astnikov pri dvuxmestnyx predikatax v lesnom dialekte èneckogo jazyka [bivalent predicates' argument encoding in the forest dialect of enets]. *Uralo-altajskie issledovanija, 1*(28), 50–69. (In Russian).
Ranacher, P., Neureiter, N., van Gijn, R., Sonnenhauser, B., Escher, A., Weibel, R., & Bickel, B. (2021). Contact-tracing in cultural evolution: A Bayesian mixture model to detect geographic areas of language contact. *Journal of The Royal Society Interface*. https://doi.org/10.1098/rsif.2020.1031
Rostovtsev-Popiel, A.A. (2024). Odnorodnost, raznorodnost i pereosmyslenie: pervye zame?ania po obrabotke kartvel'skogo materiala v baze bivaltyp [homogeneity, heterogeneity, and reanalysis: First remarks on the processing of kartvelian data for the bivaltyp database]. *Izvestija Rossijskoj akademii nauk. Serija literatury i jazyka [Bulletin of the Russian Academy of Sciences: Studies in Literature and Language]*, *83*(2), 64–71,
Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., & Scheffczyk, J. (2016). *Framenet ii: Extended theory and practice*. Berkeley, CA: International Computer Science Institute.

Say, S. (2017). Leksi?eskoe raspredelenie perexodnyx i neperexodnyx dvuxmestnyx glagolov v indo-evropejskix jazykax: kvantitativno-tipologi?eskoe issledovanie [lexical distribution of transitive and intransitive bivalent verbs in the Indo-European languages: a quantitative-typological study]. N.N. Kazanskij (Ed.), *Indoevropejskoe jazykoznanie i klassi?eskaja filologija-xxi. materialy ?tenij, posvjaš?ennyx pamjati professora iosifa moiseevi?a tronskogo. 26–28 ijunja, 2017* (pp. 714–747). Saint Petersburg: Nauka.
Say, S. (2018) Markirovanie aktantov dvuxmestnyx predikatov: predvaritel'nye itogi tipologičeskogo issledovanija [argument encoding of bivalent predicates: preliminary results of a typological inquiry]. S. Say (Ed.), *Valentnostnye klassy dvuxmestnyx predikatov v raznostrukturnyx jazykax [valency classes of bivalent predicates in languages of various types]* (pp. 557–616). St. Petersburg: ILI RAN.
Say, S. (2020–). *BivalTyp: Typological database of bivalent verbs and their encoding frames.* https://bivaltyp.info (Accessed on 14 October 2025)
Say, S. (2025). Degrees of complexity in valency class systems: implications for efficiency. *STUF—Language Typology and Universals*, *4*, , (Forthcoming)
Say, S. (Ed.). (2018). *Valentnostnye klassy dvuxmestnyx predikatov v raznostrukturnyx jazykax [bivalent valency classes in structurally diverse languages]*. Saint Petersburg: ILS RAN. (In Russian)
Say, S. (2014). Bivalent verb classes in the languages of Europe: A quantitative typological study. *Language Dynamics and Change, 4*(1), 116–166.
Say, S. (2020). Markirovanie aktantov dvuxmestnyx predikatov v novoaramejskix idiomax sela urmiya [bivalent predicates' argument encoding in neo-aramaic varieties spoken in urmiya, krasnodar krai]. *Acta Linguistica Petropolitana, 16*(2), 654–689. (In Russian).
Say, S. (2025). Bivalent verb classes across Slavic: areal and genealogical patterns. *Russian Linguistics*. https://doi.org/10.1007/s11185-025-09322-w
Seržant, I.A., Wiemer, B., Bužarovska, E., Ivanová, M., Makartsev, M., Savić, S.. Uhlik, M. (2022). Areal and diachronic trends in argument flagging across slavic. E. Dahl (Ed.), *Alignment and alignment change in the Indo-European family* (pp. 300–327). Oxford: Oxford University Press.
Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. M. Miestamo, K. Sinnemäki, and F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 67–88). Amsterdam, Philadelphia: John Benjamins.
Sinnemäki, K., Di Garbo, F., Napoleão de Souza, R., & Ellison, T. M. (2024). A typological approach to language change in contact situations. *Diachronica, 41*(3), 379–413. https://doi.org/10.1075/dia.23029.sin
Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., & Gray, R. D. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances, 9*(16), 6175. https://doi.org/10.1126/sciadv.adg6175
Studenikina, K.A. (2023). Glava 22. valentnostnye klassy dvuxmestnyx predicatov [bivalent valency classes]. E.V. Kashkin (Ed.), *Elementy gornomarijskogo jazyka v tipologicheskom osveshchenii [elements of the hill mari language in a typological perspective]* (pp. 478–490). Moscow: Buki Vedi. (In Russian)
Tsunoda, T. (1985). Remarks on transitivity. *Journal of Linguistics, 21*, 385–396.
Van Valin, J., Robert D. (1999). Generalized semantic roles and the syntax-semantics interface. Corblin, Dobrovie-Sorin, and Marandin (Eds.), *Empirical issues in formal syntax and semantics 2* (pp. 373–389). The Hague: Thesus.
Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review, 23*(176), 88–93. https://doi.org/10.1179/sre.1975.23.176.88
WATP (2014). *The world atlas of transitivity pairs (2014).* Available online at https://watp.ninjal.ac.jp/en/. Tokyo: National Institute for Japanese Language and Linguistics. (Last access 28 August 2025)
Woolford, E. (2006). Lexical case, inherent case, and argument structure. *Linguistic Inquiry, 37*(1), 111–130.
Zanchi, C., Combei, C.R., Luraghi, S. (2022). Paveda - pavia verbs database: Challenges and perspectives. *Proceedings of the 4th workshop on research in computational linguistic typology and multilingual nlp* (pp. 99–102). Seattle, Washington: Association for Computational Linguistics. (Available at https://aclanthology.org/2022.sigtyp-1.14/)