Cross-linguistic argument-coding predictability and the lexical meaning of the verb

Abstract. While it is widely acknowledged that verbs' valency patterns are partly determined by meaning (predictable) and partly idiosyncratic (unpredictable), systematic empirical measurement of this property remains scarce. This study addresses this gap by assessing predictability through the lens of cross-linguistic consistency. I introduce quantitative metrics capturing the degree to which language-specific verb–valency associations align with the valency class assignments of verbs with the same meaning in other languages, and apply these metrics to data from BivalTyp, a typological database documenting valency patterns of 130 bivalent verbs across 149 languages. Predictability is highest for predicates close to the transitivity prototype, as well as for symmetric predicates and those with concrete meanings, in contrast to more abstract predicates, including those expressing emotions, which often display idiosyncratic valency behavior. Cross-linguistic variation in aggregate predictability values reveals broad areal and genealogical patterns, highlighting gradual yet fundamental differences in how verbs are organized into valency classes across languages. Overall, this study serves as a proof of concept demonstrating that the degree of semantic motivation in valency class assignment can be quantified, revealing significant variation across predicates and languages.

Keywords. Argument coding; typology; database; predictability; valency classes; semantic roles.

1. Introduction

A consensus emerging from research on valency patterns is that they are shaped both by general principles of clausal syntax and by idiosyncratic lexical factors. This perspective reconciles two kinds of observations. On the one hand, languages typically exhibit major valency patterns that are highly frequent and hold a special status in grammar, such as the basic transitive pattern. On the other hand, most, if not all, languages include verbs whose valency patterns cannot be predicted from general principles and must instead be stored individually in the lexicon. The two extremes can be illustrated by (1) and (2), respectively:

German (deu; Indo-European, Germanic;[1] personal knowledge)
(1) *Emma hat ihr-en Bruder geschlagen*
PN[NOM.SG] AUX.PRS.3SG her-M.ACC.SG brother[ACC.SG] hit.PST.PTCP
'Emma hit her brother.'

(2) *Emma wart-et auf ihr-en Bruder*
PN[NOM.SG] wait-PRS.3SG on her-M.ACC.SG brother[ACC.SG]
'Emma is waiting for her brother.'

In (1), the verb *schlagen* 'hit' is transitive. To use it correctly, one only needs to know that in German subjects take the nominative and direct objects the accusative. Other verbs meaning 'hit' typically follow the transitive pattern both in German and cross-linguistically (Haspelmath 2015: 143). More generally, the transitive pattern is the default for bivalent verbs, so its use here can be explained by the absence of conditions that would force a deviation. In contrast, *warten* 'wait' in (2) requires its object to be marked with the preposition *auf* 'on' plus the accusative. This pattern is largely idiosyncratic, as shown by semantically similar verbs with different valency patterns, such as *erwarten* 'expect' (transitive) and *harren* 'await' (object in the genitive).

There is some debate over whether differences between default and idiosyncratic patterns, such as those in (1) and (2), should be considered categorical (see Section 2 for discussion). Even if transitive patterns are treated as distinct from all patterns involving obliques, non-transitive verbs can still exhibit gradual differences in systematicity versus idiosyncrasy in their valency patterns. This parameter, which can be called 'predictability' or 'consistency', can have wide-ranging effects on a verb's behavior. For instance, less predictable valency patterns are expected to be more prone to errors in first and second language acquisition and to be less stable diachronically.

And yet, there is a lack of empirical research on the predictability of lexical valency patterns as a measurable property. The aim of this paper is to address this gap. My approach is typological and assumes that if a verb's choice of valency frame is driven by its meaning, parallel meaning-to-form mappings should be observed in other languages. By contrast, idiosyncratic lexical valency patterns are expected to be typologically more varied. Based on this reasoning, I propose a quantitative technique to measure the predictability of the verb-pattern association and apply it to a large

[1] Here and below, genealogical information (family, genus) follows WALS (Dryer & Haspelmath 2013).

typological dataset of verbs and their valency patterns. The results indicate that predicates with features associated with semantic transitivity systematically display predictable valency behavior, whereas less semantically transitive verbs are not uniform. In particular, symmetric verbs and verbs with more concrete meanings tend to exhibit more predictable valency patterns, while more abstract verbs, including so-called psychological verbs, display less predictable valency patterns.

The remainder of the paper is organized as follows. Section 2 provides an overview of approaches to valency class assignment and outlines the goals of the study. Sections 3 and 4 introduce the dataset and the statistical methods used, respectively. Sections 5 and 6 present the main empirical results, focusing on differences between predicates and between languages. Finally, Section 7 summarizes the study.

2. Background and goals

Argument coding is a well-studied topic at the interface of syntax and semantics. While the literature employs a wide range of terms and frameworks, one of the central issues concerns the extent to which valency patterns are predictable from meaning (regular) as opposed to being lexically specified (idiosyncratic).

For several decades, lexicographic approaches have played a central role in the analysis of valency (Haspelmath 2014: 4), as reflected in valency dictionaries for a number of well-studied languages, including German (Helbig & Schenkel 1983), Hungarian and Russian (Apresjan & Páll 1982), English (Herbst et al. 2004), and French (van den Eynde & Mertens 2006). These dictionaries, often designed with second-language learners in mind, specify the argument structure of individual verbs, providing a baseline for valency research, with broader generalizations built on top.

Lexicographic approaches share several key ideas. First, for both theoretical and practical reasons, they specify a discrete set of arguments, rather than adjuncts, for each verb (Helbig & Schenkel 1983: 35ff.; Apresjan & Páll 1982: 38–46). Arguments are taken to represent participants inherent in the verb's meaning (Van Valin 2001: 93), and their presence and form are assumed to be largely lexical properties. Second, these approaches typically assume a systematic relationship between meaning and form (Fillmore 1968; Jackendoff 1990: 246–251; Levin 1993: 1–10): verbs with similar semantic properties tend to exhibit similar argument-coding patterns (Apresjan & Páll 1982: 42; Helbig & Schenkel 1983: 60–61).

This same idea underlies approaches that focus on verb classes defined by shared valency behavior (Levin 1993; Zolotova 2006), including the FrameNet project for English (Fillmore et al. 2003) and its derivatives for other languages (Lyashevskaya & Kashkin 2015). Collectively, these projects pursue the attractive view that a verb's valency pattern can "be predicted by some general principles related to the semantics of verbs" (Jolly 1993: 275).

Despite their descriptive power, such approaches imply a certain degree of redundancy: valency information is stored lexically even when it may, in principle, be predictable. This tension has motivated more reductionist accounts that seek to derive argument coding from general principles. In the generative tradition, a common strategy is to distinguish between structural and lexical case (Chomsky 1981; Yip et al. 1987). Structural case typically covers only a few options, such as nominative and accusative in German, as in (1), whereas oblique devices like auf 'on' + accusative (2) are treated as lexical case.

Empirically, the distinction between "structural" and "lexical" largely mirrors the contrast between "default" and "non-default" case-marking (Zaenen et al. 1985: 466), a distinction also used outside generative frameworks. Canonical valency patterns—monovalent (e.g., *sleep*, *run*), transitive (*kill*, *break*), and ditransitive (*give*)—play a central role, especially in studies of alignment and voice. The semantics and syntax of transitivity are well understood (Hopper & Thompson 1980; Tsunoda 1985; Dowty 1991; Lehmann 1991; Lazard 1994; Malchukov 2005), while other bivalent patterns are treated as extensions of or deviations from the canonical transitive pattern.

Non-default case frames are usually assumed to result from mapping rules linking semantic roles to argument positions (for variations on this theme, see Fillmore 1968; Levin & Rappaport Hovav 2005; Van Valin 1999). However, evidence suggests that non-default valency patterns are not uniform. For instance, the mechanism of case assignment in (2) likely differs from that in (3).

German (personal knowledge)
(3) *Emma hat dem Lehrer geantwortet*
PN[NOM.SG] AUX.PRS.3SG DEF.M.DAT.SG teacher[DAT.SG] answer.PST.PTCP
'Emma answered the teacher.'

The dative marking of *der Lehrer* 'the teacher' in (3) is predictable from its semantic role, typically analyzed as 'addressee'. By contrast, the preposition *auf* 'on' in (2) must

be learned as a lexical property of *warten* ‘wait’ and cannot be derived from general German grammar rules. Such contrasts are often framed as “thematic” vs. “idiosyncratic” case (Yip et al. 1987; Woolford 2006; see also Blume 1998). Case is “thematic” when its choice is tied to generalized roles such as agent, patient, or instrument, whereas “idiosyncratic” case is specific to individual verbs. However, a strict division is difficult to maintain: semantic roles are not standardized, and their inventories vary across approaches, which limits the predictive power of role-based accounts.

As an alternative to semantic annotation, it has been proposed to infer roles bottom-up from the prevalence of coexpression patterns in typological data (Bickel et al. 2014; Hartmann et al. 2014; Cysouw 2014; Say 2018: 566–572). The idea is simple: if semantic roles drive argument-coding choices but are not directly observable, similarities in coding across verbs can reveal these roles.

Grosso modo, this paper continues this line of research with one key modification. Previous studies, notably Bickel et al. (2014), acknowledge that semantic roles are inherently fuzzy and focus on areas where their organization is relatively clear. Typically, this fuzziness is treated as a nuisance that obscures meaning–form mappings. By contrast, I treat it as an object of study. Specifically, I assume that the regularity of argument coding is gradable and measurable: some verbs show highly consistent patterns across languages, while others vary considerably. This variation reflects differences in the degree to which argument coding is semantically motivated.

My approach seeks a middle ground between two extremes. One treats (non-default) case assignments as lexical—descriptively useful but lacking predictive power. The other treats all case assignments as semantically determined and fully predictable, which becomes empirically problematic when applied to broad datasets rather than selective examples. Rather than dividing cases into “thematic” versus “idiosyncratic”, I evaluate them along a **continuum** of regularity versus idiomaticity. In this respect, I propose a quantitative implementation of the regularity continuum, as occasionally noted in the literature (Huddleston & Pullum 2002: 1240–1241; Haspelmath 2014: 5).

In my approach, examples (1)–(3) illustrate a continuum from more regular (predictable, grammatical; see (1)) to more irregular (unpredictable, lexical; see (2)) verbs (cf. Barðdal 2011, passim, on the link between regularity and predictability of case assignment). I propose measuring a verb’s valency regularity via its **cross-linguistic consistency**, as detailed in Section 4.

Below, I focus exclusively on **bivalent** verbs, which are better suited to my purposes than other valency types. Bivalent verbs typically center on a large transitive class, while non-canonical bivalent classes can also be sizable and often include frequent verbs—unlike trivalent verbs, which are more lexically limited. Bivalent verbs are also more likely to exhibit non-default case assignments than other valency types (Bickel et al. 2014).

Non-default case assignment in bivalent patterns varies in its locus. It often occurs on the argument lower on the agentivity scale, typically reflecting low affectedness, as in (4), where the person waited for is flagged by a preposition. In some languages, it appears on the argument higher on the agentivity scale, typically reflecting low control, as illustrated by the use of the dative case for the 'finder' in (5). Finally, bivalent patterns in which both arguments are coded by non-default devices are also possible, as in (6) from Svan.[2]

Czech (ces; Indo-European, Slavic; BivalTyp)[3]
(4) *Petr čeká na Michal-a*
PN(M)[NOM.SG] wait(IPFV).PRS.3SG on PN(M)-ACC.SG
'Petr is waiting for Michal.'

Telugu (tel; Dravidian; BivalTyp)
(5) *Pravīṇ-ki tana tāḷaṁcēvu-lu dorikāyi*
PN(M).SG.OBL-DAT own key(N)-PL.NOM be_found.PST.3PL.N
'Praveen found his keys.'

Svan (sva; Kartvelian; BivalTyp)
(6) *Maizer-s žeɣ-išd x-a-q'luːn-i*
PN-DAT dog-BEN IO3-VER:SUP-be_afraid-PRS
'Maizer is afraid of the dog.'

[2] Like other Kartvelian languages, Svan displays complex TAM-conditioned argument coding. While the details cannot be discussed here, neither argument of 'to be afraid' in (6) matches canonical transitive coding.

[3] Here and below, "BivalTyp" signals examples taken from the BivalTyp database (Say (ed.) 2020-). For more information, see Section 3.

In short, bivalent verbs strike an ideal balance between regularity and irregularity across languages, making them an excellent testing ground for my study.

3. Dataset

The data come from BivalTyp, an online database of bivalent verbs and their coding frames (Say 2020–). The database is based on a questionnaire of 130 sentences with bivalent predicates, often presented in short contexts 'Peter sees a house', 'Lightning struck the house', '(Peter played chess with Mary). Peter lost to Mary'.

Each database entry includes a translation of a stimulus sentence into a target language, typically provided by native speakers and annotated by experts. To ensure cross-linguistic comparability, the two arguments in each sentence are labeled "X" and "Y",[4] where X corresponds to the argument bearing more agentive entailments (Dowty 1991), regardless of syntactic structure. For instance, in translations of *Peter likes this shirt*, "Peter" is marked as X even if the language uses constructions such as *This shirt pleases Peter* or *To Peter, this shirt is pleasing*.

Entries are annotated for the morphosyntactic devices used to code "X" and "Y" in the target languages. Flagging—through case marking or adpositions—is common in the data, as shown in (4)–(6). In some cases, however, indexing plays the primary role, as in (7) from Abaza, where the Y argument of 'believe' is indexed in the slot preceding and grammatically linked to the benefactive preverb *ʒ-*, while no flagging occurs.

Abaza (abq; Northwest Caucasian; BivalTyp)
(7) *Fatíma Murád jə-ʒ-qá-l-ç-əj-ṭ*
PN PN [3SG.M.IO-BEN]-LOC-[3SG.F.ERG]-believe-PRS-DCL
'Fatima trusts Murad.'

In rarer cases, word order is the main device for argument coding. English past-tense clauses with transitive verbs and nominal arguments, such as *Peter saw Mary*, illustrate this.

[4] These labels are inspired by Lazard's framework (e.g. 2002) but do not fully correspond to his definitions.

The main response variable in the database is the **valency pattern** (see Malchukov et al. 2015: 30 for a definition) observed in the translation, defined as the ordered combination of argument-coding devices for X and Y in the target language. For example, the valency pattern of (4) is schematically represented as "NOM_naACC", indicating that the X argument of the Czech verb *čekat* 'wait' is coded by the nominative, while the Y argument is marked by the preposition *na* 'on' plus the accusative.

Two verbs—or other "valency carriers" (Gaszewski 2020), such as adjectives, serial verb constructions—found in two entries from the same language are considered to belong to the same language-specific **valency class** if their arguments use the same valency pattern. BivalTyp's valency labels are applied consistently within each language and define the lexical range of valency classes. For example, besides *čekat* 'wait', the Czech class "NOM_naACC" includes the equivalents of 'reach', 'forget', 'call', 'play (instrument)', 'shoot (at)', 'be angry', and 'have a grudge'. The use of the same label within the Czech dataset in BivalTyp reflects the objective fact that all these verbs exhibit the same argument-coding behavior.

By contrast, BivalTyp's labels cannot equate argument-coding devices or valency patterns across languages. For example, both the Svan pattern in (6) and the Abaza pattern in (7) include a device labeled "BEN," but in Svan it denotes the benefactive case, whereas in Abaza it marks a specific preverb hosting the index of an argument. Even when two languages use identical traditional case labels (dative, instrumental, locative, etc.), these devices cannot be directly equated (Haspelmath 2009: 510). Thus, more sophisticated methods are needed for cross-linguistic comparison of valency classes (see Section 4; Haspelmath, Hartmann 2015; Say 2014).

The only exception to the language-specific nature of valency class annotations in the BivalTyp design is the transitive class, consistently labeled "TR" in the database. Unlike all other classes, which are defined exclusively based on language-specific coding devices, this class is defined as the one that includes the equivalent of 'kill' (see Haspelmath 2015: 136 for a similar approach). In this respect, the transitive class is treated as a comparative concept suitable for cross-linguistic analysis, following a long-standing tradition of typological research, as outlined in Section 2.

Methodologically, BivalTyp builds upon wordlist-based studies of verbs and their categories (Nedjalkov 1969; Bossong 1998; Nichols et al. 2004; Nichols 2008; Malchukov & Comrie 2015). To interpret the analysis that follows, however, two disclaimers on the notion of 'bivalent verbs' are necessary.

First, the database focuses on bivalent predicates in the sense that each entry is annotated for the coding of exactly two predefined participants, X and Y. It does not assume that these noun phrases necessarily meet argumenthood criteria in the target languages, nor that the elicited verbs or verb-like expressions do not have additional arguments.

Second, while English verbs are used to reference questionnaire items, the actual translation units are whole sentences. This approach facilitates cross-linguistic alignment of entries and allows control over properties such as animacy, volitionality, and aspect inherent in the stimulus sentences. Similarly, language-specific verbs, such as the Czech *čekat* 'wait', sometimes serve as shorthand for database entries, each illustrating a single sentence with a specific valency pattern, even though such verbs can occur in highly variable contexts. Thus, the 130 sentences are treated as probes in an infinite semantic space, without implying fixed boundaries between language-specific lexical items.

For the analysis below, I used the latest development version of BivalTyp available as of October 3, 2025. The sample comprises 149 languages, with a total of 17,791 entries (calculated as 130 entries per language across 149 languages, minus 1,579 gaps, averaging about 8% or 10–11 missing entries per language). The dataset used in this study is available in the Supplementary materials on the GitHub repository: https://github.com/serjozhka/Argument_coding_predictability_LTC/tree/main.

The sample size is substantial for a lexically based typological database. However, its areal and genealogical composition is heavily skewed: it is a convenience sample, mainly comprising Eurasian languages, with disproportionate representation across families and genera. The absence of many genealogical taxa and the overrepresentation of others pose challenges for the analysis. To address these issues, specific techniques were employed when introducing statistical metrics for capturing typological consistency in argument coding, as described in the next section.

4. Statistical metrics

The major premise of this study is the idea that the argument-coding pattern of a verb is partly determined by its meaning, though the extent varies across cases. This section introduces a statistical technique to assess that extent empirically. To illustrate the approach, consider the Azerbaijani sentence in (8).

Azerbaijani (azj; Altaic, Turkic; BivalTyp)
(8) *mənim köynəyim sən-in-ki-ndən fərqlən-ir*
1.GEN shirt.POSS1SG 2-GEN-that-ABL be_different-IPFV(3)
'My shirt is different from yours.'

The valency pattern of (8) is labeled "NOM_ABL" in BivalTyp, indicating the use of the nominative and ablative cases to code the X and Y arguments. The key question is to what extent this pattern is semantically motivated. One appealing approach is to identify the semantic roles of the arguments in (8) and test whether other Azerbaijani constructions with the same roles show the same valency pattern. Greater parallelism across such constructions would suggest higher predictability and, consequently, stronger semantic motivation.

However, this approach is impractical. Its main drawback is that discrete semantic roles cannot be assigned on a priori semantic grounds to all possible constructions, making straightforward calculations impossible or arbitrary. As an empirical alternative, I propose using **data from other languages** as a proxy for the fuzzy, unobservable semantic roles. This approach builds on the idea that each valency-class system reflects a language-specific grouping of argument *microroles* (cf. Haspelmath 2014; Hartmann et al. 2014), where microroles are predicate-specific argument roles tied to individual verb meanings, also called "verb-specific semantic roles" (Van Valin 2001: 29). Assuming that "verbs with similar patterns have similar meanings" (Klotz 2007: 118; see also Helbig & Schenkel 1983: 62–63), the degree of semantic affinity among verb-specific arguments can be estimated from the cross-linguistic prevalence of their co-expression (Bickel et al. 2014; Hartmann et al. 2014; Cysouw 2014; Say 2018: 566–572). Although no individual language can be viewed as an ideal representative of "meso-roles," the valency-class system of any language reflects semantic affinities and can serve as a reference point for analyzing another language. As an illustration, French is used below, with the equivalent of (8) shown in (9).

French (fra; Indo-European, Romance; BivalTyp)
(9) *ma chemise se distingue de la tienne*
my.SG.F shirt(F) REFL distinguish.PRS.3SG of DEF.SG.F your.SG.F
'My shirt is different from yours.'

The valency pattern of (9) is represented schematically as “SBJ_de”. As with any valency pattern, French verbs[5] showing this pattern are assumed to share some semantic similarity. The empirical question, then, is whether the correspondence between the French “SBJ_de” and the Azerbaijani “NOM_ABL” patterns is regular. To test this, we examine all BivalTyp entries with the “SBJ_de” pattern in French and inspect the valency patterns of their Azerbaijani counterparts. The relevant data are shown in Table 1.[6]

Table 1. French entries with the “SBJ_de” pattern and their Azerbaijani counterparts

predicate label	French verb	pattern	Azerbaijani verb	pattern	PredLib3
be afraid	*avoir peur*	SBJ_de	*qorxmaq*	NOM_ABL	0.47
go out	*sortir*	SBJ_de	*çıxmaq*	NOM_ABL	0.47
depend	*dependre*	SBJ_de	*asılı* + COP	NOM_ABL	0.47
play (instrument)	*jouer*	SBJ_de	*çalmaq*	TR	0.13
make fun	*se moquer*	SBJ_de	*sataşmaq*	NOM_DATLAT	0.20
fill (intr)	*se remplir*	SBJ_de	*dolmaq*	NOM_ile[7]	0.07
need	*avoir besoin*	SBJ_de	*lazım* + COP	DATLAT_NOM	0.07
be different	*se distinguer*	SBJ_de	*fərqlənmək*	NOM_ABL	0.47
remember	*se souvenir*	SBJ_de	*yadında qalmaq*	GEN_NOM	0.07
be glad	*se réjouir*	SBJ_de	*sevinmək*	NOM_DATLAT	0.20
dismount	*descendre*	SBJ_de	*düşmək*	NOM_ABL	0.47
dream (sleeping)	*rêver*	SBJ_de	*yuxuda görmək*	TR	0.13
be content	*être content*	SBJ_de	*razı* + COP	NOM_ABL	0.47
fall in love	*tomber amoureux*	SBJ_de	*vurulmaq*	NOM_DATLAT	0.20
be_shy	*avoir honte*	SBJ_de	*utanmaq*	NOM_ABL	0.47

The data in Table 1 show that the correspondence between French “SBJ_de” and Azerbaijani “NOM_ABL” is fairly regular, accounting for 7 of the 15 French “SBJ_de” cases. To capture the predictability of the Azerbaijani pattern given the French one, I

[5] For the sake of simplicity, I use the term *verb* to refer to all types of valency carriers in the database, including non-verbal predicates (e.g. *être content* ‘be content’) and complex verbs (e.g. *avoir peur* ‘be afraid’).

[6] Table 1 includes only entries with equivalents in both languages; cells with NAs are omitted and excluded from later calculations.

[7] The actual form of the relevant Azerbaijani postposition is *ilə* ‘with’. For technical reasons, however, BivalTyp valency class labels often simplify such items, omitting non-standard symbols and diacritics. The same applies to several cases mentioned below.

use a metric called *Predictability (liberal estimate)*, or *PredLib*, whose formula is introduced in two steps. First, define a two-argument function *ValPat* ('valency pattern') that identifies the valency pattern of a verb in a given language. For example, the following equation shows that the Azerbaijani verb 'be different' has the "NOM_ABL" pattern:

(10) *ValPat*('Azerbaijani', 'be_different') = "NOM_ABL"

Second, define the *PredLib3* function, which takes three arguments—a predicate and two languages—and calculates the predictability of the pattern observed in $lang_2$ based on the data from $lang_1$.

$$(11)\quad PredLib3(\mathrm{v_n}, \mathrm{lang_1} \rightarrow \mathrm{lang_2}) = \frac{|\{i \mid \mathrm{ValPat}(\mathrm{lang_1}, v_i) = \mathrm{ValPat}(\mathrm{lang_1}, v_n) \wedge \mathrm{ValPat}(\mathrm{lang_2}, v_i) = \mathrm{ValPat}(\mathrm{lang_2}, v_n)\}|}{|\{i \mid \mathrm{ValPat}(\mathrm{lang_1}, v_i) = \mathrm{ValPat}(\mathrm{lang_1}, v_n)\}|}$$

The numerator in (11) represents the number of verbs that share the same pattern as v_n in both $lang_1$ (the predictor language) and $lang_2$ (the predicted language). The denominator represents the number of verbs that share the same pattern as v_n in $lang_1$. Applying this function to the data in Table 1 yields the equation in (12).

$$(12)\ PredLib3(\text{'be\_different', 'French'} \rightarrow \text{'Azerbaijani'}) = \frac{7}{15} = 0.47$$

Both the numerator and denominator in (11) are natural numbers equal to or greater than 1, since every verb belongs to the same class as itself in both languages. The numerator is always equal to or smaller than the denominator, so *PredLib3* is a positive value equal to or smaller than 1.

PredLib3 partly models a second-language learner: a French speaker acquiring Azerbaijani might generalize that verbs in the "SBJ_de" pattern correspond to Azerbaijani "NOM_ABL" verbs. This generalization is imperfect—reflected by *PredLib3* of 0.47 here—but more reliable than for less regular correspondences.

To illustrate *PredLib3*, let's compare more and less transparent cross-linguistic correspondences. Table 2 shows all Kumyk (kum; Altaic, Turkic) entries with the

"NOM_NOMbulan" valency pattern[8] alongside their Albanian (als; Indo-European) counterparts, illustrating a scenario with high predictability values.

Table 2. Kumyk entries with the "NOM_NOMbulan" pattern and their Albanian counterparts

predicate label	Kumyk verb	pattern	Albanian verb	pattern	PredLib3
fight	*tutušmaq*	NOM_NOMbulan	*përleshem*	NOM_meACC	0.9
be friends	*qurdaš* + COP	NOM_NOMbulan	*kam miqësi*	NOM_meACC	0.9
get to know	*taniš* + COP	NOM_NOMbulan	*njihem*	NOM_meACC	0.9
play (instrument)	*ojnamaq*	NOM_NOMbulan	*bie*	NOM_DAT	0.1
cut oneself	*gesmek*	NOM_NOMbulan	*pritem*	NOM_meACC	0.9
speak	*söjlemek*	NOM_NOMbulan	*flas*	NOM_meACC	0.9
mix	*bulʁanmaq*	NOM_NOMbulan	*përzihem*	NOM_meACC	0.9
agree	*razi* + COP	NOM_NOMbulan	*bie dakord*	NOM_meACC	0.9
have a quarrel	*erišmek*	NOM_NOMbulan	*grindem*	NOM_meACC	0.9
be content	*razi* + COP	NOM_NOMbulan	*kënaqur* + *COP*	NOM_meACC	0.9

Table 2 shows that the Kumyk "NOM_NOMbulan" pattern maps almost perfectly onto the Albanian "NOM_meACC," reflected in very high *PredLib3* values. Importantly, these patterns are not related etymologically or via direct contact. The high predictability of the Albanian pattern likely stems from semantic similarities among the predicates in Table 2 (see Section 5 for substantive details).

At the opposite extreme, Table 3 presents Hungarian (hun; Uralic, Ugric) entries with the "NOM_SUB" pattern alongside their Shinaz Rutul (rut; Nakh-Daghestanian, Lezgic) counterparts.

Table 3. Hungarian entries with the "NOM_SUB" pattern and their Shinaz Rutul counterparts

predicate label	Hungarian verb	pattern	Shinaz Rutul verb	pattern	PredLib3
resemble	*hasonlít*	NOM_SUB	*kejkɨn*	NOM_CONT	0.08
think	*gondol*	NOM_SUB	*fikɨr havʔɨn*	ERG_CONTEL	0.08
wait	*vár*	NOM_SUB	*güzlemiš hiʔɨn*	TR	0.17

[8] In this pattern, the Y argument is marked by the Kumyk postposition *bulan*, roughly corresponding to 'with'.

remember	*emlékszik*	NOM_SUB	*hik'ij ʁatkɨn*	DAT_NOM	0.08
obey	*hallgat*	NOM_SUB	*qhacun*	NOM_APUD	0.08
shoot at	*rálő*	NOM_SUB	*vihin*	ERG_APUD	0.08
be squeamish	*kényes*	NOM_SUB	*k'vač' hɨʔɨn*	TR	0.17
envy	*irigy*	NOM_SUB	*ul livxin*	ATTR_SUPER	0.08
be angry	*mérges*	NOM_SUB	*qhaˤl biʔin*	ERG_ATTR	0.08
have a grudge	*haragszik*	NOM_SUB	*qhaˤl libq'ɨn*	DAT_EL	0.17
take offence	*megsértődik*	NOM_SUB	*χatir atkɨn*	DAT_EL	0.17
get irritated	*haragszik*	NOM_SUB	*bizar hišin*	NOM_EL	0.08

Table 3 shows that the Hungarian NOM_SUB pattern has no regular counterpart in Shinaz Rutul: ten different patterns appear, none covering more than two entries. Thus, in this semantic zone, Shinaz Rutul valency patterns are largely unpredictable (captured by low *PredLib3* values), likely requiring a hypothetical Hungarian-speaking second-language learner of Shinaz Rutul to acquire them verb by verb.

The *PredLib3* measure, as defined in (11), is conceptually similar to Jaccard similarity, a well-known metric operating on sets. The key difference is that Jaccard's denominator represents the size of the **union** of two sets, whereas *PredLib3* uses only the size of the set of verbs belonging to the same class as the target verb v_n in the **predictor** language ($lang_1$). This makes *PredLib3* an asymmetric measure: it compares the intersection of two sets to one of them, while Jaccard similarity compares the intersection to the union.

The need for an asymmetric function becomes clear in the next step of the analysis. There is no reason to assume that French—or any other language—is inherently better suited for predicting valency patterns across languages. The main aim here is not to measure (dis)similarities between language pairs but to assess how predictable verb-specific valency patterns are within each language against the background of all others. A natural next step, then, is to use **all** available languages as predictors. Mathematically, this is expressed by the *PredLib2* function, which takes two arguments and calculates the mean predictability of the valency pattern of the n^{th} verb in the m^{th} language, as shown in (13), where s is the size of the language sample.

(13) $\text{PredLib2}(V_n, lang_m) = \frac{\sum_{i=1}^{s} PredLib3(v_n, lang_i \rightarrow lang_m)}{s}$

PredLib2 can be used to assess whether the valency pattern of a given verb in one language is typologically consistent with the behavior of verbs with the same meaning

across other languages. It yields high values when the verb belongs to a language-specific valency class whose members' cross-linguistic equivalents tend to fall into the same valency classes as those of the verb in question. As elsewhere in the paper, the typological predictability of a verb-specific valency pattern is taken to reflect its degree of semantic motivation.

However, a straightforward application of *PredLib2* to the entire sample is problematic. The paper's argument rests on the assumption that similarities in the lexical composition of valency classes primarily reflect semantic similarities between verbs. Yet, as with other typological phenomena, distributions in individual languages are influenced by genealogical and areal factors. For instance, the fact that verbs v_n and v_m belong to the same valency class in both Danish (dan; Indo-European, Germanic) and Swedish (swe; Indo-European, Germanic) is less indicative of inherent semantic similarity than if the same pattern were found in Danish and Telugu, since Danish and Swedish may share cognate valency patterns inherited from their common ancestor. In this respect, the BivalTyp sample is not ideal for directly computing *PredLib2* across all languages. As noted in Section 3, some genealogical taxa are overrepresented. For example, the eleven Slavic languages would be advantaged if *PredLib2* were calculated over the full sample, since any one of them would almost inevitably show high predictability in Slavic–Slavic pairs, inflating the average values in equation (13).

In response to this problem, I used a subsampling procedure to calculate the actual *PredLib2* values. Rather than using all sample languages as predictors, I drew random subsamples of 48 languages, each containing exactly one language per genus represented in BivalTyp. Sixteen such subsamples were created, ensuring that every language served as a predictor at least once (16 being the maximum number of languages within a single genus, Lezgic). The final *PredLib2* values were then obtained as the mean of the 16 averages calculated from these genealogically controlled subsamples.

The procedure outlined above does not fully eliminate non-independence in the data (see Guzmán Naranjo and Becker 2022 for a more systematic approach to this issue, which is, however, best suited to samples much larger than BivalTyp). Areal convergences across genera and, at least hypothetically, deep genealogical signals may still bias between-verb similarities that are ideally shaped by semantic factors. Nevertheless, the subsampling procedure presumably mitigated the most evident cases of near-congruent valency class systems, primarily found in closely related

language pairs such as Danish–Swedish, Erzya–Moksha (myv and mdf; both Uralic, Mordvin), and Tsakhur–Southern Rutul (tkr and rut; both Nakh-Daghestanian, Lezgic).

The final step in assessing valency pattern predictability in a typological context is to abstract away from language-specific patterns and calculate an index reflecting each verb's predictability across the entire sample. For this purpose, I introduced the *PredLib* function, which takes a single argument (the verb), as shown in (14).

(14) $PredLib(v_n) = \frac{\sum_{i=1}^{s} PredLib2(v_n, lang_i)}{s}$

For the same reasons discussed for *PredLib2*, PredLib values were first calculated within the same random genealogically controlled subsamples and then averaged across them. The resulting PredLib values for individual verbs represent one of the main empirical outcomes of this study, discussed in Section 5.

However, *PredLib* and its variants sometimes produce values that differ intuitively from what these metrics are meant to capture. This is illustrated in Table 4, which lists all Standard Arabic (arb; Afro-Asiatic, Semitic) entries with the "NOM_anGEN" pattern and their Icelandic (isl; Indo-European, Germanic) counterparts.

Table 4. Standard Arabic entries with the "NOM_anGEN" pattern and their Icelandic counterparts

predicate label	Standard Arabic verb	pattern	Icelandic verb	pattern	PredLib3
look for	*baḥaṯa*	NOM_anGEN	*að leita*	NOM_adDAT	0.33
be different	*ʔiḫtalafa*	NOM_anGEN	*að vera ólík*	NOM_DAT	0.33
fall behind	*taḫallafa*	NOM_anGEN	*að dragast aftur*	NOM_urDAT	0.33

The *PredLib3* values in Table 4 are all 0.33—a relatively high score that correctly reflects that each Icelandic pattern accounts for one third of the patterns found in Icelandic counterparts of Standard Arabic verbs using the "NOM_anGEN" pattern. Mathematically, these values are only slightly lower than the 0.47 *PredLib3* value in Table 1, where the Azerbaijani "NOM_ABL" pattern covered 7 of 15 French "SBJ_de" entries. Substantively, however, the situations differ sharply: while the French–Azerbaijani correspondence is non-random (if imperfect), the Arabic–Icelandic pairs in Table 4 show no regular correspondence at all. The relatively high *PredLib3* scores

arise simply because the relevant class in the predictor language (here, Standard Arabic) is small. An extreme case of this low-base effect occurs when a class in the predictor language contains only one entry, as shown in Table 5.

Table 5. Basque entries with the "ABS_kontra" pattern and their Wolof counterparts

predicate label	Basque verb	pattern	Wolof verb	pattern	PredLib3
fight	*borrokatu*	ABS_kontra	*xeex*	SBJ_ak	1

In short, the *PredLib3* value in Table 5 captures the strength of a potential "rule" stating that whenever a Basque (eus; isolate, Spain and France) verb displays the "ABS_kontra" pattern, its Wolof (wol; Niger-Congo, Wolof) counterpart should show "ABJ_ak." Although this rule is technically exceptionless in BivalTyp, its predictive power is vacuous: a learner would need exposure to the single instantiating entry, leaving no scope for generalization.

To address this limitation of *PredLib3*, I introduce a more conservative metric for evaluating cross-linguistic pattern correspondences, called *Predictability (conservative estimate)*, or *PredCons*. The formula for the three-argument version, *PredCons3*, is given in (15).[9]

$$\text{(15) PredCons3}(v_n, \text{lang}_1 \rightarrow \text{lang}_2) = \frac{|\{i \mid \text{ValPat}(\text{lang}_1, v_i) = \text{ValPat}(\text{lang}_1, v_n) \wedge \text{ValPat}(\text{lang}_2, v_i) = \text{ValPat}(\text{lang}_2, v_n)\}|-1}{|\{i \mid \text{ValPat}(\text{lang}_1, v_i) = \text{ValPat}(\text{lang}_1, v_n)\}|-1}$$

PredCons differs from *PredLib* in that it does not consider the specific verb's pattern itself when assessing its predictability in a target language based on a predictor language (as reflected in the subtraction of one in both the numerator and the denominator). In this sense, *PredCons* models the scenario of guessing a verb's valency pattern without prior knowledge, whereas *PredLib* evaluates the consistency of a pattern once it is known. Mathematically, *PredCons3* values are always equal to or smaller than *PredLib3*, as reflected in the metric labels.

The difference between the two metrics is small when the valency class in the predictor language and its intersection with the target language are large. For example, the French verb *se distinguer* ("SBJ_de") and its Azerbaijani counterpart

[9] An additional rule stipulates that PredCons3 is set to 0 when the denominator equals zero, as in Table 5.

fərqlənmək (“NOM_ABL”) yield a *PredLib3* value of 0.47, calculated as 7/15 (see above) and only a slightly smaller *PredCons3* value of 0.43 (6/14), reflecting the number of *other* verbs showing the same correspondence. By contrast, for small classes, the difference can be dramatic. For example, the *PredCons3* values for all entries in Table 4 equal 0, as each correspondence is lexically unique and no other verbs can be used for prediction.

As with *PredLib3*, average *PredCons3* values were calculated within genealogically controlled language subsamples to yield *PredCons2*, reflecting the consistency of a language-specific lexical valency pattern with other languages. The formula for this two-argument function is given in (16).

(16) $\text{PredCons2}(\text{V}_\text{n}, \text{lang}_\text{m}) = \frac{\sum_{i=1}^{s} PredCons3(v_n, lang_i \rightarrow \text{lang}_m)}{s}$

Finally, one more averaging step was performed to assess the cross-linguistic predictability of a predicate’s valency patterns, as defined in (17).

(17) $\text{PredCons}(\text{v}_\text{n}) = \frac{\sum_{i=1}^{s} PredCons2(v_n, lang_i)}{s}$

As noted, *PredLib* and *PredCons* are largely similar. Nevertheless, I calculated both across the full dataset and found that each highlights distinct aspects of cross-linguistic valency pattern distributions, as discussed in the next section.

The quantitative analysis was conducted in R (R Core Team, 2025) using the following packages: colorspace (Zeileis et al. 2020), dplyr (Wickham et al. 2023), ggplot2 (Wickham 2016), ggpubr (Kassambara 2023), ggrepel (Slowikowski 2024), htmlwidgets (Vaidyanathan et al. 2023), infotheo (Meyer 2022), leaflet (Cheng et al. 2024), lingtypology (Moroz 2017), RColorBrewer (Neuwirth 2022), readxl (Wickham & Bryan 2025), writexl (Ooms 2025). The code is provided in the Supplementary materials.

5. Differences between predicates

5.1. Preliminary language-specific observations and comparison with intuition

The main empirical results of this study are presented in spreadsheets (available in the Supplementary materials) containing values of two predictability metrics—

PredLib2 and *PredCons2*—calculated for each language-specific verb–pattern combination in the BivalTyp database. For illustration, Table 6 shows a subset of 12 German entries, ordered by descending *PredCons2* values.

Table 6. Selected German entries and their *PredLib2* and *PredCons2* values

predicate label	German verb	pattern	PredLib2	PredCons2
hit	*schlagen*	TR	0.76	0.75
be short	*fehlen*	DAT_NOM	0.47	0.23
trust	*vertrauen*	NOM_DAT	0.31	0.19
believe	*glauben*	NOM_DAT	0.30	0.19
answer	*antworten*	NOM_DAT	0.28	0.17
help	*helfen*	NOM_DAT	0.24	0.16
flatter	*schmeicheln*	NOM_DAT	0.22	0.15
obey	*gehorchen*	NOM_DAT	0.18	0.11
resemble	*ähnlich sein*	NOM_DAT	0.39	0.11
follow	*folgen*	NOM_DAT	0.36	0.09
agree	*zustimmen*	NOM_DAT	0.21	0.08
wait	*warten*	NOM_aufACC	0.11	0.07

Several observations can be made from Table 6.

First, the data align with the intuitive observations made in Section 1. The use of the transitive pattern with *schlagen* 'hit', as in (1), is highly predictable—i.e., consistent with the typological behavior of its counterparts in other languages—reflected in high *PredLib2* and *PredCons2* values. In contrast, the "NOM_aufACC" pattern with *warten* 'wait', as in (2), yields much lower values. Intermediate cases also occur: for instance, *fehlen* 'be short', with the "DAT_NOM" pattern illustrated in (18), receives mid-range predictability scores, indicating that this pattern is less regular than that of *schlagen* but more typologically motivated than that of *warten*. This accords with the observation that the "DAT_NOM" pattern of *fehlen* aligns with semantically similar verbs such as *reichen* 'be enough'.

German (personal knowledge)
(18) *Mein-em Bruder fehl-t ein Euro*.
my-M.DAT.SG brother[DAT.SG] lack-PRS.3SG INDEF.M.NOM.SG euro[NOM.SG]

‘My brother is one Euro short.’

Second, the two predictability metrics are closely correlated. While individual values may diverge, the overall ranking of verb–pattern combinations in Table 6 remains largely consistent across *PredLib2* and *PredCons2*. This strong correlation is mathematically expected, given the similarity of their underlying formulas (see Section 4); for a visual confirmation of the correlation, see Figure S1 in the Supplementary materials. Unless stated otherwise, Sections 5.2–5.5 rely on the liberal estimate (*PredLib*), as it is mathematically more straightforward. Cases where the two metrics diverge in both values and interpretive implications will be discussed in Section 5.3.

Third, both metrics capture not only contrasts between patterns within a language (e.g., the higher values for the default transitive pattern versus the marginal “NOM_aufACC” pattern in German) but also differences among verbs sharing the same pattern. For instance, Table 6 includes all German “NOM_DAT” entries in BivalTyp. The variation among their scores reflects how typologically expected each verb’s use of this pattern is. The results are intuitively consistent: verbs such as *glauben* ‘believe’, *antworten* ‘answer’, and *helfen* ‘help’ show relatively high predictability for the German “NOM_DAT” pattern, reflecting the typological tendency of such verbs to align with each other and with other verbs whose second arguments are commonly believed to be recipients, addressees, and beneficiaries (often captured by the use of “dative” as a descriptive label). In contrast, *folgen* ‘follow’ and *zustimmen* ‘agree’ are less typologically straightforward, as many languages prefer structures that are literally modeled as “X follows after Y” or “X agrees with Y.”

While the range of values for the German “NOM_DAT” pattern is relatively narrow (0.21–0.31 for *PredLib2*), some patterns exhibit much greater variation. Table 7, which focuses on Finnish (fin; Uralic, Finnic) verbs with the “NOM_ILL” pattern, illustrates this with *PredLib2* values ranging from 0.21 to 0.78.

Table 7. Finnish entries with the “NOM_ILL” pattern

predicate label	Finnish verb	pattern	PredLib2	PredCons2
sink	*upota*	NOM_ILL	0.78	0.37
enter	*astua*	NOM_ILL	0.68	0.36
get stuck	*juuttua*	NOM_ILL	0.52	0.28

hit (target)	*osua*	NOM_ILL	0.41	0.32
match	*sopia*	NOM_ILL	0.37	0.24
mix	*sekoittua*	NOM_ILL	0.35	0.20
trust	*luottaa*	NOM_ILL	0.28	0.13
be content	*tyytyväinen* + COP	NOM_ILL	0.27	0.15
influence	*vaikuttaa*	NOM_ILL	0.26	0.16
fall in love	*rakastua*	NOM_ILL	0.25	0.15
touch	*koskea*	NOM_ILL	0.23	0.15
get to know	*tutustua*	NOM_ILL	0.21	0.13

The data in Table 7 likely reflect the layered structure of the Finnish "NOM_ILL" class. Verbs at the top form its typologically expected core:[10] they mainly denote concrete spatial motion, with the Y argument in the illative case marking an endpoint. Verbs lower in the hierarchy use this pattern in more idiosyncratic ways and mainly express abstract meanings related to emotion or cognition. Broader tendencies of this kind are discussed in Sections 5.3–5.4; see also the discussion in Section 2 of semantic roles as fuzzy entities whose centers can be identified quantitatively, including Bickel et al. (2014).

In summary, the language-specific predictability metrics developed here offer valuable insights for language-specific studies. Quantifying the degree of motivatedness in verb–pattern associations through typological data (*PredLib2* and *PredCons2*) can inform analyses of first- and second-language acquisition errors as well as language change. Exploring language-specific outcomes, however, lies beyond the scope of this paper. The remainder of Section 5 therefore focuses on cross-linguistic generalizations, primarily using the aggregate *PredLib* metric.

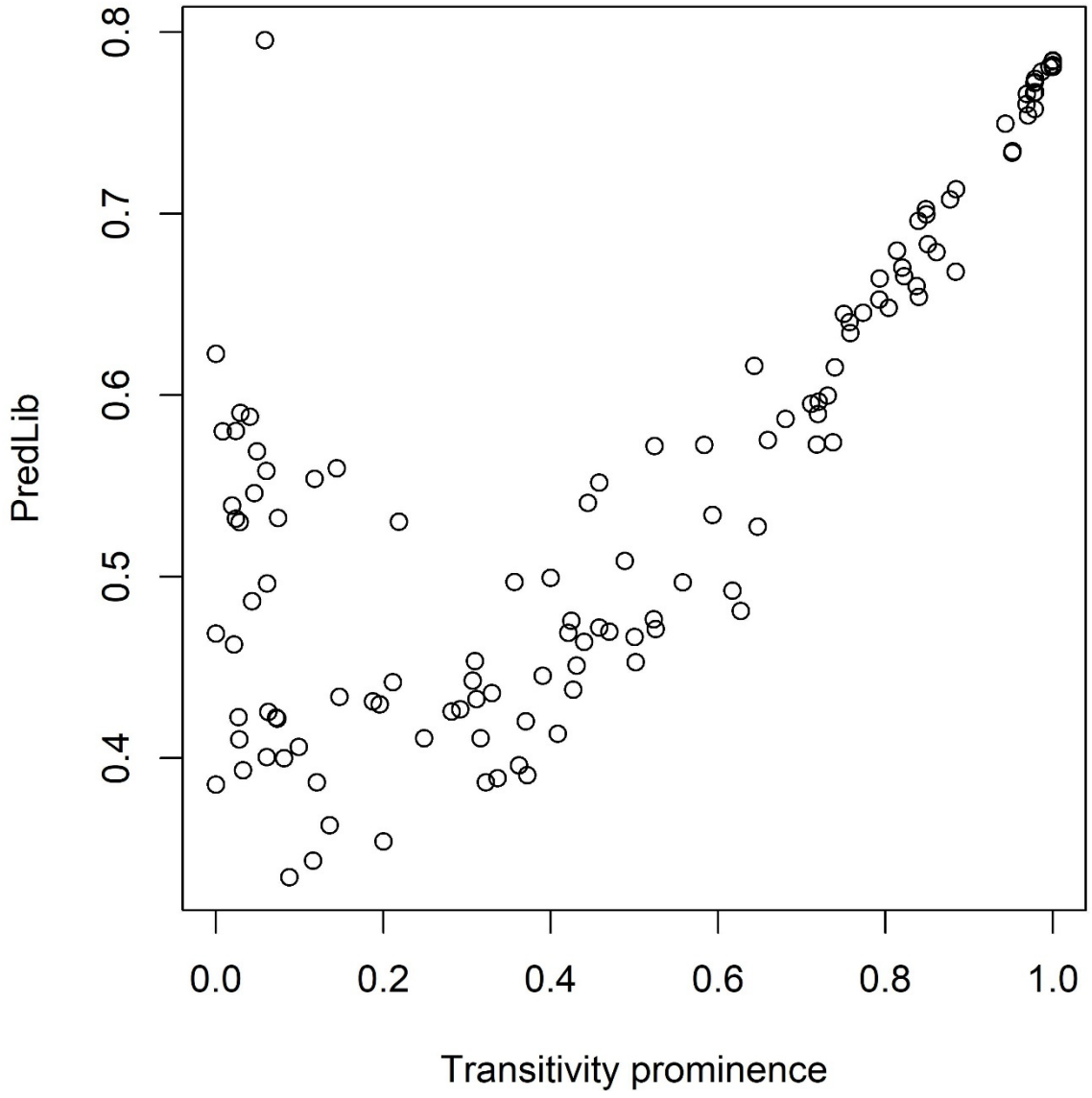
5.2. Predictability and high transitivity prominence

In this section, I discuss the main trends in the distribution of aggregate *PredLib* and *PredCons* values for the 130 predicates in the BivalTyp questionnaire. The most salient finding is a strong link between a predicate's predictability metrics and its **transitivity prominence**. Transitivity prominence measures the likelihood that a predicate is expressed by a transitive verb in a given language, calculated as the ratio

[10] Position at the top of this and similar scales could in principle be framed in terms of "prototypicality", the term I prefer to avoid mainly because it is used in too many different ways.

of transitive patterns to the total number of entries for that predicate (Haspelmath 2015).[11] The observed *PredLib*, *PredCons*, and transitivity prominence values for all predicates are provided in Table A1 (Appendix). Figure 1 show the raw observed values of transitivity prominence and *PredLib* as a scatterplot, with each dot representing one of the 130 predicates.

Figure 1. Transitivity prominence and *PredLib* and across 130 BivalTyp predicates



[11] To ensure data consistency and avoid unwanted genealogical biases, transitivity prominence values for individual predicates were calculated not as the mean across the entire sample, but as the average of mean values observed in the genealogically balanced subsamples described in Section 4.

The first observation is a very strong correlation between the metrics (Pearson's $r = 0.791$, $p << 0.001$; correlation with *PredCons* is even stronger), indicating that predicates with high transitivity prominence also display high predictability scores. This pattern reflects two linguistic facts: i) a substantial number of predicates in the BivalTyp sample have high transitivity prominence, meaning they are expressed by transitive verbs in most languages, and ii) the core of this class remains stable across languages (while languages may differ in the use of transitive patterns for predicates with lower prominence scores). These results align with previous evidence for a cross-linguistically stable set of semantic features associated with high transitivity—namely, the presence of a volitional agent acting physically on an affected patient (Hopper & Thompson 1980; Tsunoda 1985; Kittilä 2002; Næss 2007; Malchukov & Comrie 2015). In Figure 1, the top right-hand corner contains many predicates matching this semantic profile. For illustration, 23 predicates with transitivity prominence of 0.94 or higher are: 'make', 'drink', 'take', 'write', 'eat', 'sing', 'kill', 'wash', 'melt', 'throw', 'put on', 'break', 'read', 'fry', 'open', 'bend', 'take off', 'plough', 'pour', 'try to catch', 'drive', 'cover', and 'milk'. All have *PredLib* and *PredCons* scores ≥ 0.72. Predicates with transitivity prominence values ≥ 0.7 show an almost perfect correlation between transitivity prominence and predictability. Substantively, this means that their morphosyntactic behavior in individual languages is predictable precisely when they are assigned to the transitive class.

In summary, predicates expected to display robust semantic transitivity were indeed consistently realized as morphosyntactically transitive verbs, resulting in highly predictable valency patterns. While this finding is largely unsurprising, the interaction becomes less straightforward for predicates with lower transitivity prominence, visible on the left-hand side of Figure 1. In this region, some predicates still show typologically consistent behavior (top of the figure), whereas others do not (bottom). These two groups are discussed in Sections 5.3 and 5.4.

5.3. Predicates with low transitivity prominence but high predictability

5.3.1. Overview

This section focuses on predicates with low transitivity prominence but high *PredLib* values—predicates that display predictable behavior in terms of valency pattern selection across languages despite their low tendency to appear in transitive

constructions. These predicates occupy the upper-left area in Figure 1. Figure 2 zooms in on this area, with labels for individual predicates.

Figure 2. Predicates with low transitivity prominence and high predictability (*PredLib*)

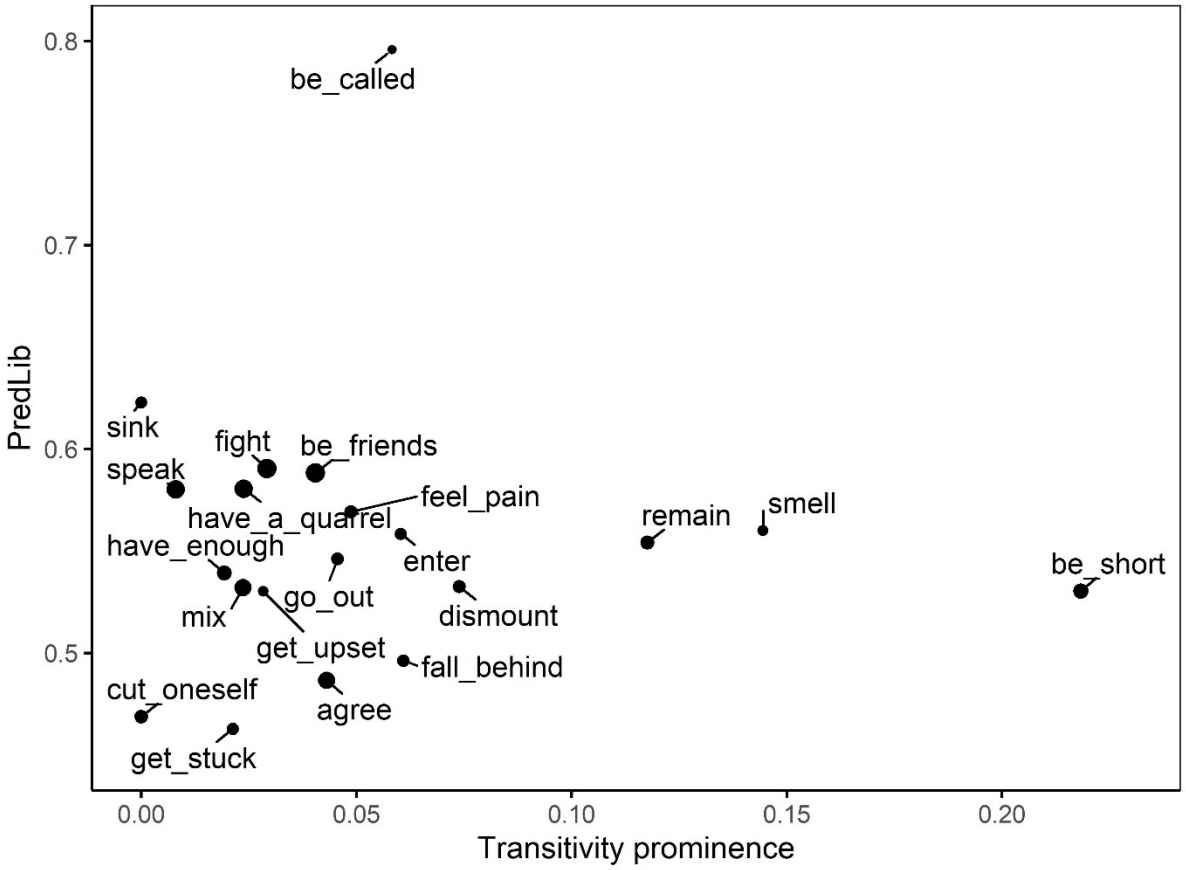


Figure 2 shows twenty predicates with transitivity prominence below 0.23 and *PredLib* above 0.45, with circle size reflecting *PredCons*. The very existence of this set is one of the key findings of this study. Unlike predicates with both high transitivity prominence and high predictability, this group is not homogeneous. It comprises i) symmetric predicates, ii) motion and action predicates, iii) predicates related to possession and iv) several idiosyncratic items.

5.3.2. Symmetric predicates

The first group includes 'fight', 'be friends', 'speak',[12] 'have a quarrel', 'mix', and 'agree'. Besides their high PredLib values (0.49–0.59) and low transitivity prominence

[12] The actual questionnaire sentence for this predicate is: "(I am looking for P. When I enter the room, I see that) P. is speaking with M."

(0.01–0.04), they show unusually high PredCons scores (0.40–0.49), normally seen only in predicates with much higher transitivity prominence ($\geq$ 0.5). This pattern reflects their shared morphosyntactic stability across languages: verbs corresponding to these predicates typically belong to the same valency class, whose lexical composition is relatively constant cross-linguistically.

Semantically, these predicates share a feature of **reciprocity**: if *X fights with Y*, then *Y fights with X*. Most BivalTyp languages have a comitative pattern used uniformly for such symmetric predicates.[13] For instance, the South Levantine Arabic variety spoken in Nazareth uses a dedicated comitative construction, illustrated in (19), which appears with precisely these six predicates and not elsewhere.

Arabic (South Levantine, Nazareth) (ajp; Afro-Asiatic, Semitic; BivalTyp)
(19) *Shadi ʿam biḥke maʿ Reem*
PN PROG speak.PRS.3M.SG with PN
'Shadi is speaking with Reem.'

While South Levantine Arabic offers a particularly clear case, many other languages have a valency pattern likewise used for all—or nearly all—of these symmetric predicates and seldom elsewhere. Interestingly, although the lexical scope of the comitative pattern remains stable across languages, the markers (cases or adpositions) involved are not always diachronically stable. Preliminary evidence suggests that comitative markers often differ etymologically even among closely related languages (e.g., in Romance or Turkic), unlike other adpositions that are more conservative. This supports the idea that valency classes—sets of verbs sharing a valency pattern—can remain stable independently of the particular morphological markers involved (Say 2025).

5.3.3. Predicates related to possession

A second small group comprises 'have enough', 'be short', 'remain', and 'feel pain'. Their *PredCons* values (0.21–0.31) are lower than for symmetric predicates but still well within the mid-to-high predictability range shown in Fig. 2. Semantically, all

[13] The BivalTyp questionnaire includes two additional potentially symmetric predicates, 'encounter' and 'get to know' (= 'get acquainted'). Both show slightly lower predictability and higher transitivity prominence than the six predicates discussed in the main text, reflecting their cross-linguistic fluctuation between the transitive and comitative patterns.

these predicates are related to the idea of possession, evident not only in ‘have enough’, ‘be short’, and ‘remain’ (in the meaning used in the BivalTyp sentences, cf. 20) but also in ‘feel pain’ (21), with both examples coming from Tsugni Dargwa.

Tsugni Dargwa (dar; Nakh-Daghestanian, Dargwic; BivalTyp)
(20) *durħaˁ-la hanna wec'al qːuruš k-alg-un*
boy-GEN now ten money down-stay.PFV-PRET(3)
‘The boy has now ten roubles left.’

(21) *durħaˁ-la hanna bek' icː-ule = cabi*
boy-GEN now head hurt.IPFV-CNV.PROG = COP:N(3)
‘The boy now has a headache.’

Both examples involve the same “GEN_NOM” valency pattern, where the X-argument is coded in the genitive, the case typically used for adnominal possessors. Yet X-arguments in (20), (21) and further similar cases display some syntactic properties typical of clausal rather than adnominal dependents: for instance, they can be separated from the possessee (‘ten roubles’, ‘head’), as seen in (20) and (21). This supports Ovsjannikova & Say (2014), who show that genitive noun phrases can function as either clause- or phrase-level possessors, depending on the predicate and orther factors, as in Bashkir and other Turkic languages.

All the four predicates ‘have enough’, ‘be short’, ‘remain’, and ‘feel pain’ often involve patterns with an X-locus (Say 2014), where the X-argument is coded by an oblique coding device. As (20)–(21) show, clauses with the meaning like ‘X feels pain in Y’ can mirror possessive structures: while the assertion in (21) is the presence of a bodily sensation, it presupposes a part–whole relation between X and Y. Such parallels between verbs like ‘have enough’, ‘be short’, and ‘remain’, on the one hand, and ‘feel pain’, on the other hand, recur in many languages, explaining the relatively high predictability of their valency patterns—though this tendency is less stable than in symmetric predicates, yielding somewhat lower *PredLib* values (see Figure 2).

Given this, one may ask why the basic predicate ‘have’ does not appear in Figure 2. In fact, ‘have’ scores similarly high in *PredLib* (0.54) and even higher in *PredCons* (0.38), but its transitivity prominence is also higher (0.44). This reflects the typological variability of possessive encoding (Heine 1997; Stassen 2009), which often includes a transitive schema—especially common in Western Europe (Basque,

Romance, Germanic, Greek, Albanian, and to a degree Slavic and Baltic). In short, several predicates related to possession show consistent valency behavior cross-linguistically, though not always through the same structural pattern as used in the basic predicative possessive construction.

5.3.4. Motion and action predicates

The third discernible group includes 'sink', 'enter', 'go out', 'dismount', 'fall behind', and 'get stuck'. Like symmetric predicates, they have very low transitivity prominence (0–0.08) and moderate *PredCons* (0.15–0.25), somewhat lower than the possession-related predicates. Unlike the previous groups, their language-specific counterparts usually do not form a unified valency class but cluster around several motion and action schemas, such as motion toward ('sink', 'enter', 'get stuck') or from ('go out', 'dismount') a location.

The common feature of these predicates is that they denote concrete, observable events involving motion. In the vast majority of cases, their X-arguments align with canonical intransitive subjects (S), and their Y-arguments are often coded like spatial or instrumental adjuncts. Crucially, the same devices are sometimes extended to more abstract meanings through metaphorical extension.

For instance, in Brazilian Portuguese, the "SBJ_em" pattern involving the preposition *em* 'in', is observed in concrete spatial verbs such as *entrar* 'enter' (22), *afundar* 'sink', and *ficar grudado* 'get stuck'.

Portuguese (Brazilian) (por; Indo-European, Romance; BivalTyp)
(22) *Pedro entr-ou em casa*
PN(M) enter-3SG.PST in house(F)
'Pedro entered a house.'

Comparable uses of an adposition with the basic spatial meaning 'in' occur in equivalents of 'enter', 'sink', 'get stuck' in many other languages, including Swedish (*i*) and Modern Greek (*se*). However, the range of non-spatial uses differs: in Brazilian Portuguese, the same pattern appears with *acreditar* 'believe', *confiar* 'trust', and *pensar* 'think'; in Swedish, with *att förälska sig* 'fall in love' and *att ha ont* 'feel pain'; and in Modern Greek, with *léō* 'tell', *apantó* 'answer', and *tariázō* 'match'. These cross-linguistic mismatches show that concrete motion predicates have more consistent

coding across languages than abstract ones, explaining their relatively high predictability.

5.3.5. Residual predicates

Finally, three predicates in Figure 2—'be called', 'get upset', and 'smell'—do not fit into the previous groups. Their meanings are heterogeneous, but they share very low *PredCons* (0.03–0.12). The combination of high *PredLib* and low *PredCons* reflects the difference between the two predictability metrics as discussed in Section 4 and the tendency of the equivalents of these predicates to form unique valency classes not aligning with other verbs. In simple words, these verbs' specific valency behavior cannot be predicted by analogy with other verbs, but their uniqueness is itself predictable (hence high *PredLib* stemming from the low-base effect mentioned in Section 4). To illustrate, the equivalents of 'be called' often involve an unusual pattern in which both arguments appear in the default case (typically nominative or absolutive in descriptive grammars), but only X controls verb agreement, as in (23) from Polish.

Polish (pol; Indo-European, Slavic; BivalTyp)
(23) *Ten przedmiot nazywa-ł się klepsydr-a*
this.M.NOM.SG object(M)[NOM.SG] call:IPFV-PST[M.SG] compass(F)-NOM.SG
'This object was called a clepsydra.'

5.3.6. Interim summary

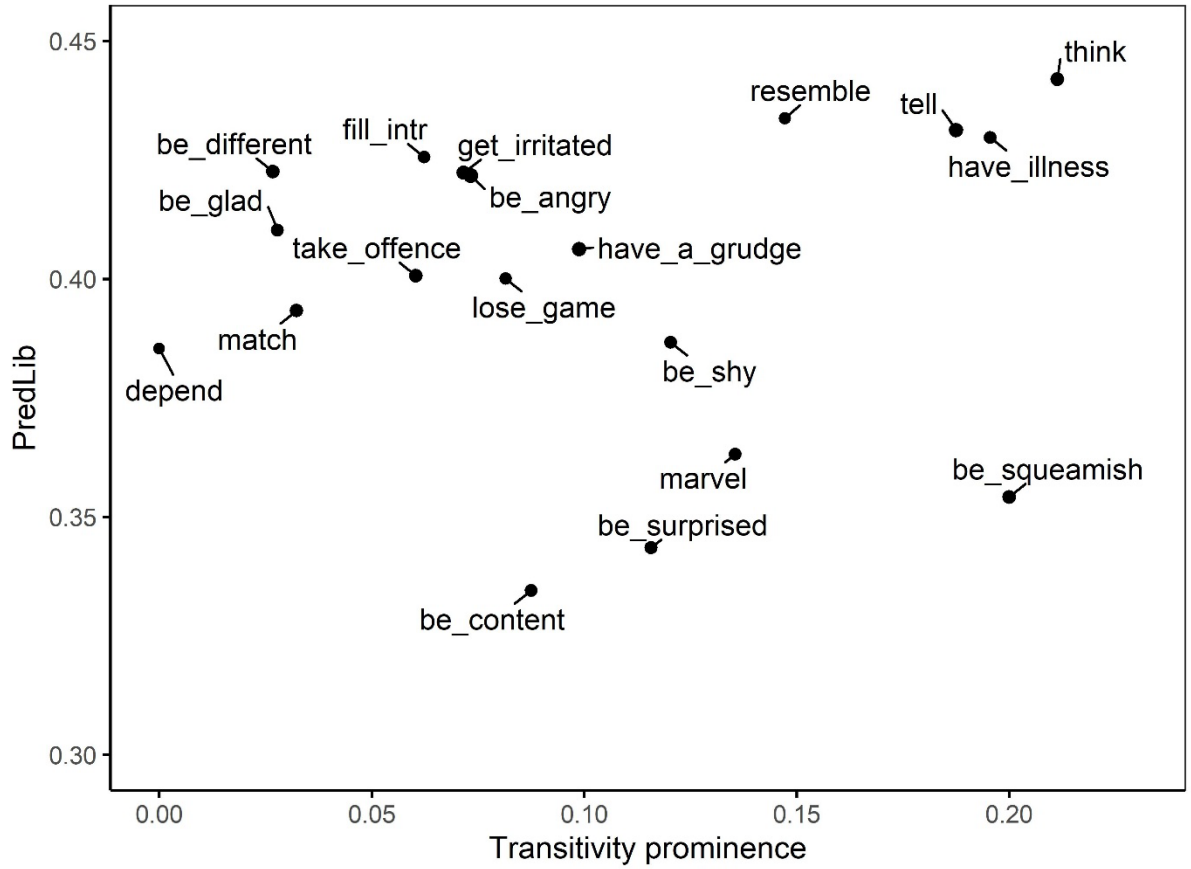
Predicates with low transitivity prominence but relatively consistent typological valency behavior are diverse, comprising several semantic clusters: symmetric, possession-related, and motion-related predicates.[14] Together they contrast with predicates displaying low valency pattern predictability, which are discussed in the next section.

5.4. Predicates with low transitivity prominence and low predictability

[14] More precisely, what is typologically relatively consistent in the valency behavior of these verbs is the coding of the two arguments targeted in the frames used in the BivalTyp questionnaire: co-participants in action predicates (as in 'X fighting with Y'), possessor and possessee in possession predicates (as in 'X is short Y'), and figure and landmark in motion predicates (as in 'X dismounted from Y').

Figure 3 shows the 19 predicates with transitivity prominence below 0.23 and *PredLib* below 0.45, complementing the area in Figure 2 by covering its continuation toward lower *PredLib* values.

Figure 3. Predicates with low transitivity prominence and low predictability (*PredLib*)



The data in Figure 3 reveal that the lowest *PredLib* values are found in predicates denoting emotions and uncontrolled attitudes (‘be content’, ‘be surprised’, ‘be squeamish’, ‘be shy’) as well as in abstract predicates such as ‘depend’ and ‘match’. Among predicates with slightly higher transitivity prominence, the lowest PredLib values are likewise observed for abstract predicates (e.g. ‘influence’, PredLib = 0.39) and for predicates denoting emotions and attitudes (e.g. ‘fall in love’, 0.39; ‘enjoy’, 0.40).

In what follows, I set aside abstract predicates such as ‘depend’ and ‘influence’. These are infrequent in spontaneous discourse, often lack straightforward equivalents in many languages, and tend to yield periphrastic or missing entries in the BivalTyp

dataset. Instead, I focus on predicates of emotion, which dominate the area of both low predictability and low transitivity prominence.

The behavior of emotion predicates reflects their tendency to belong to multiple valency classes within a language, with their grouping and overlap with more concrete predicates varying unpredictably across languages. This is illustrated in Table 8 with data from the Kartvelian family. Kartvelian languages are particularly suitable for this comparison: (i) all four are represented in BivalTyp, and (ii) their morphosyntactic systems and case inventories are sufficiently similar for valency class labels to be directly comparable. Table 8 summarizes how nine emotion predicates are distributed across valency classes in the four languages.

Table 8. Selected predicates of emotion and their valency patterns in Kartvelian languages

predicate label	Georgian	Mingrelian	Laz	Svan
be content	NOM_INS	NOM_INS	NOM_ABL	DAT_NOM
be surprised	DAT_NOM	DAT_INS	NOM_DAT	DAT_NOM
be squeamish	DAT_NOM	DAT_NOM	n.a.	DAT_NOM
be shy	DAT_GEN	DAT_ALL	DAT_ABL	DAT_GEN
fall in love	DAT_NOM	DAT_ERG	NOM_DAT	DAT_NOM
enjoy	NOM_INS	ERG_INS	TR	DAT_NOM
take offence	DAT_GEN.ABL	ERG_DAT	DAT_ABL	DAT_ABL.EL
envy	DAT_GEN	DAT_GEN	NOM_DAT	DAT_BEN
trust	NOM_DAT	ERG_DAT	DAT_NOM	NOM_DAT

The data show that while small islands of stability can be identified (such as the use of the "DAT_NOM" pattern with 'be squeamish'), the overall distribution of emotion predicates across valency classes is highly variable. Each Kartvelian language assigns the equivalents of these nine predicates to between five and eight different classes—and crucially, each language does so differently. This supports the broader conclusion that verbs of emotion tend to show unpredictable valency behavior.

This observation aligns with previous studies, such as Malchukov (2005: 98–101), Bickel et al. (2014: 501–503), and Haspelmath (2001: 65), who notes cross-linguistic variation in the use of oblique marking and prepositions with experiencer predicates (e.g. *be amazed at*, *be interested in*, *be worried about* in English). Similarly, Ovsjannikova and Say (2020: 16–19) trace the emergence of idiosyncratic argument coding in

Russian reflexive emotion verbs to historical developments in which psychological meanings evolved from concrete uses formerly uniformly marked by the instrumental case.

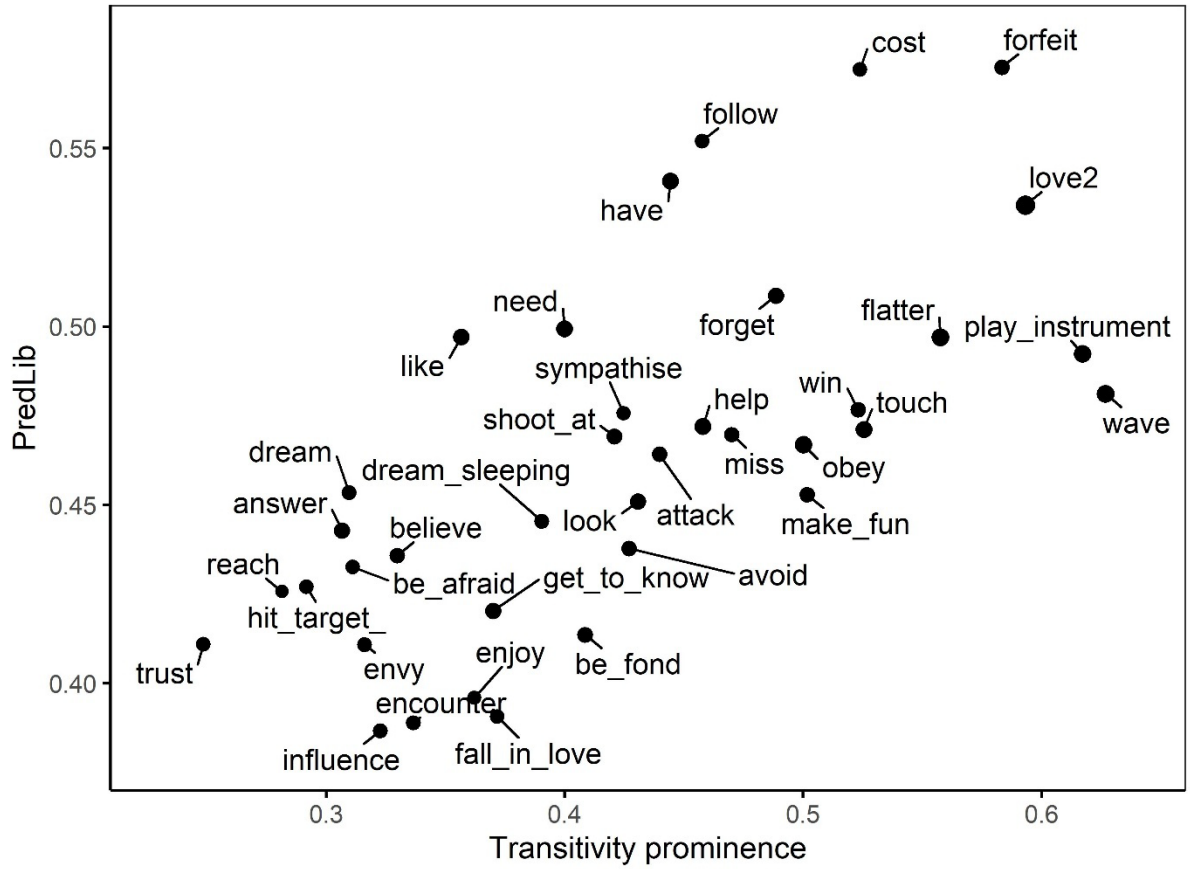
In sum, both language-specific and cross-linguistic evidence—including typological data from BivalTyp—indicate that predicates of emotion display highly variable valency behavior compared to predicates with more concrete meanings. This likely reflects a broader principle: the psychological meanings of emotion verbs rarely determine their valency pattern (Klein & Kutscher 2002).

5.5. Discussion

Sections 5.2–5.4 examined the distribution of transitivity prominence and predictability (mainly expressed through *PredLib* values) across the 130 BivalTyp predicates and related these measures to predicate semantics. For clarity, three major groups were identified: (i) predicates with high transitivity prominence, typically denoting a volitional action of X upon an affected Y (Section 5.2); (ii) predicates with low transitivity prominence but high predictability, often corresponding to symmetric, motion-related, or possessive meanings (Section 5.3); (iii) predicates with low transitivity prominence and low predictability, mainly covering emotional and abstract meanings (Section 5.4).

The preceding discussion focused on these extreme cases. However, a considerable number of predicates occupy an intermediate zone. Figure 4 displays the 37 predicates whose transitivity prominence values range between 0.23 and 0.64; all such predicates display *PredLib* values between 0.38 and 0.58.

Figure 4. Predicates with mid-range transitivity prominence



Although semantically heterogeneous, these predicates exhibit the same positive correlation between transitivity prominence and *PredLib* already noted for highly transitive predicates (Section 5.2). As a group, they typically involve an Y that is only weakly affected or not affected at all. In the typological literature, many of these predicates are described as predicates of contact, pursuit or interaction, including, for instance, 'touch', 'reach', 'follow', 'help', 'flatter', 'shoot at', 'answer', 'attack', 'make fun', 'avoid', and 'win' (Blume 1998; Malchukov 2005).

At this point, it is relevant to ask whether the groups defined by transitivity prominence and predictability can also be characterized in terms of **semantic roles**. Haspelmath (2014: 9–10) observes that "agents and locations are expressed very uniformly across languages," whereas "the salient differences concern the participants with theme, patient, experiencer and stimulus roles (and the like)."

The present findings broadly support this generalization. The coding of clear agents (X arguments) is highly consistent across languages, especially for predicates near the semantic core of transitivity. Beyond this group, most cross-linguistic variation concerns Y arguments (the Y-locus in the BivalTyp annotation; see also Say 2014). Deviations from canonical X-coding are largely confined to possessive and experiencer

predicates, where X does not fully meet standard criteria for agency. Haspelmath's observation regarding locations is likewise supported by the stability of motion predicates discussed in Section 5.3.

Such generalizations, however, typically assume that coding devices can be mapped onto discrete semantic roles defined a priori (e.g. 'agent', 'instrument', 'endpoint of motion'). While intuitive, this approach raises practical challenges, as it depends on a fixed inventory of meanings and on the use of diagnostic contexts as yardsticks, both of which are to some extent arbitrary and theory-dependent.

The present study instead adopts a gradable perspective, in which meanings are inferred from observed co-expression patterns. On this view, the association between coding devices and semantic domains can be assessed quantitatively, allowing for a data-driven characterization of their cores and peripheries. From this perspective, discrete semantic roles do not emerge as stable units governing valency behavior: although verb meaning is clearly relevant, the underlying semantic features operate on multiple levels that do not align neatly with traditional role distinctions.

Bickel et al. (2014) similarly approached semantic roles empirically, deriving them from co-expression patterns in a large cross-linguistic sample. Their fuzzy clustering identified both role-like groupings and degrees of membership, but notably failed to converge for less agentive arguments ($\approx$ Y), in contrast to more agentive ones ($\approx$ X). This suggests that such arguments do not form clearly delineated clusters. The present study builds on this insight but shifts the focus from identifying clusters to assessing degrees of membership.

A particularly revealing case concerns arguments whose semantic role is traditionally identified as *stimulus*. As shown in Section 5.4, stimuli of emotion predicates are encoded in highly variable ways, in line with Haspelmath's observation. By contrast, stimuli of perception predicates ('see', 'hear', 'look', 'listen') show moderate to high predictability across languages. Thus, even within this domain, "stimulus" does not constitute a uniform cross-linguistic category.

A similar conclusion emerges from symmetric predicates (Section 5.3). These form a relatively coherent group with high predictability, reflecting shared semantics. While this could be captured in terms of roles such as "protagonist" and "companion", the relevant commonality lies outside parameters such are commonly used to delineate semantic roles, such as volition or affectedness. Rather, it follows from reciprocity, as in predicates like 'fight', 'be friends', and 'mix'.

In sum, valency class assignment appears to operate on several partly independent levels—from default transitive alignment through semi-regular patterns (e.g. symmetric predicates) to more idiosyncratic cases. While each of these mechanisms has been noted in the literature (Section 2), the present results contribute to a more refined empirical understanding of how they interact across languages.

6. Differences between languages

This section examines cross-linguistic differences based on quantitative metrics reflecting the typological predictability of verb-specific valency patterns. While there are illuminating contrasts in specific verb–pattern combinations, the focus here is on aggregate metrics highlighting general cross-linguistic tendencies. To this end, I calculated mean values of *PredLib* and *PredCons* for each language-specific dataset; the results are available in the Supplementary materials.[15] The two metrics show an almost perfect correlation across languages. For simplicity, I will therefore focus on mean *PredCons* values (chosen because they display slightly higher standard deviation, i.e. sharper contrasts). All observations below equally apply to *PredLib*.

Conceptually, higher *PredCons* values are expected in languages with large valency classes whose lexical range is shaped by broad semantic generalizations, whereas lower values are typical of languages where valency class membership is largely idiosyncratic and less predictable from cross-linguistic patterns.

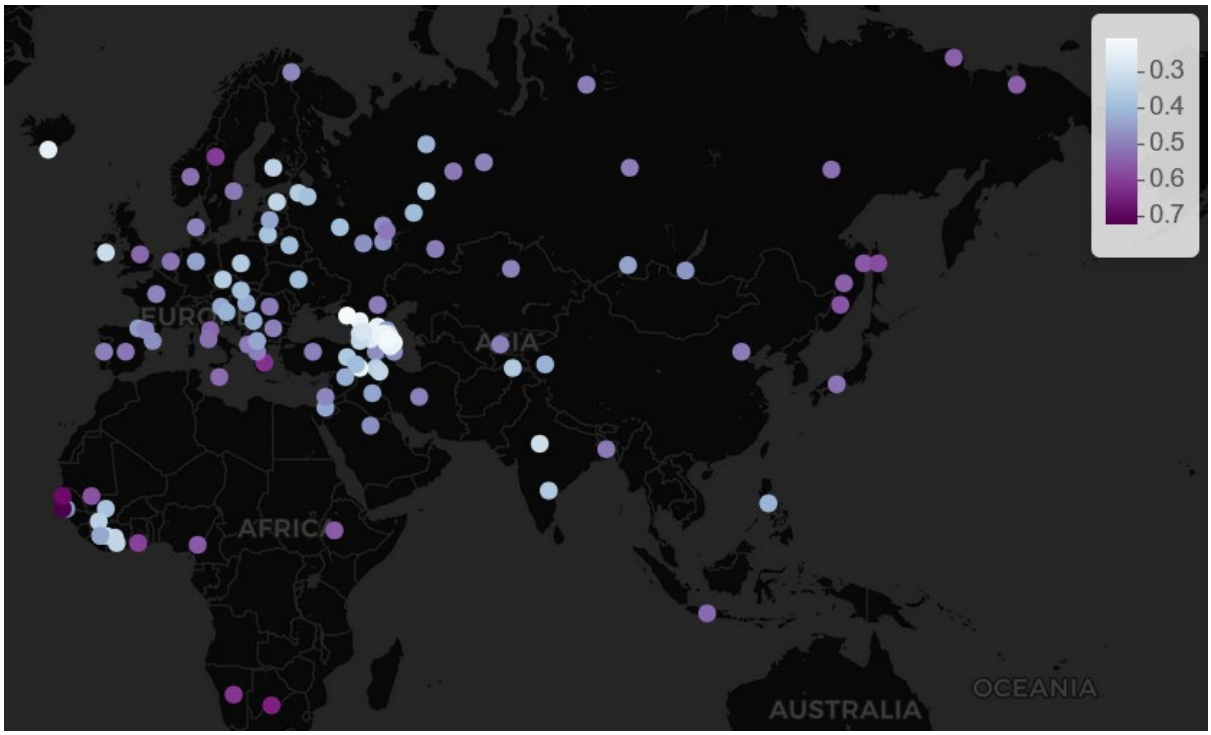
A note of caution is in order. As discussed in Section 4, *PredCons* and *PredLib* measure how closely a verb–pattern combination aligns with lexical patterns found in other languages. Consequently, a language's mean *PredLib* or *PredCons* value may reflect not only the internal logic of its valency system but also its similarity to systems represented in the BivalTyp sample. A language might thus receive a low *PredLib* value simply because its semantically motivated principles of valency classification are rare in the sample. To minimize genealogical bias, predictability values were calculated from subsamples containing only one language per genus, preventing inflation from closely related neighbors. However, broader areal or macro-genealogical biases were not controlled for. The following observations should

[15] At this stage, I excluded languages with fewer than 90 non-missing data points, reducing the sample from 149 to 145 languages. This step was intended to avoid potential biases, as languages with fewer data points often lack values for more "exotic" predicates whose valency behavior is less predictable.

therefore be read with this caveat in mind: when a language is said to show more predictable behavior, this means it aligns more closely with other languages in the BivalTyp sample. Nevertheless, as shown below, such sampling biases do not appear to have had a major effect.

Turning to the results, the average *PredCons* values vary considerably, from 0.21 in Adyghe (ady; Northwest Caucasian) to 0.72 in Joola-Fonyi (dyo; Niger-Congo, North-Central Atlantic). These extremes cover more than half of the theoretically possible range (0–1), reflecting profound differences among languages. The areal distribution of *PredCons* values is shown in Figure 5.

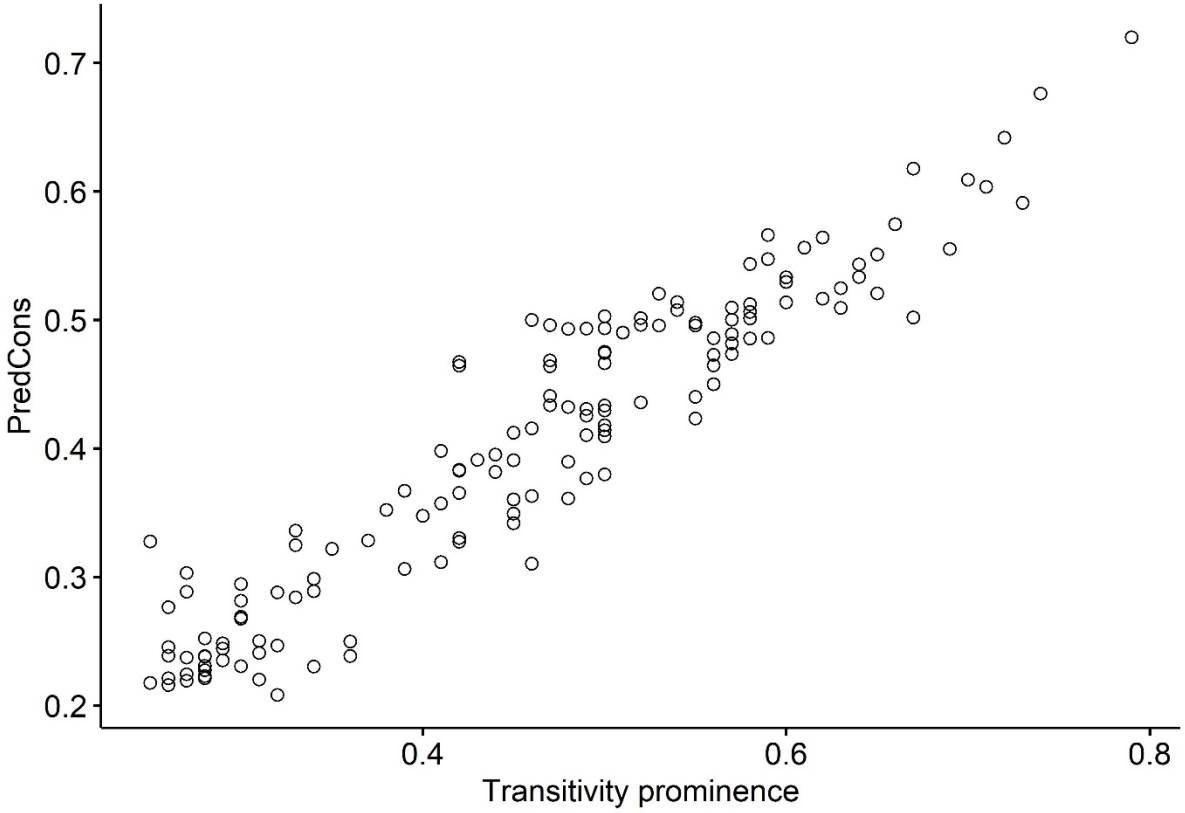
Figure 5. Mean *PredCons* values in BivalTyp languages (map)



As Figure 5 indicates, cross-linguistic differences in *PredCons* are not random but reveal clear areal patterns. The Caucasus stands out as an area with extremely low *PredCons* values; other regions with low values include Eastern Europe, parts of West Africa (notably among Mande languages) and probably South Asia (although the data are scarce). Higher *PredCons* values are observed in Western and Southern Europe, Mainland Scandinavia, Siberia, the Far East, and likely large parts of Africa—though the latter pattern remains tentative due to sparse sampling.

This areal distribution closely mirrors that of other complexity-related metrics, such as transitivity prominence and valency class system entropy.[16] The discussion below focuses on transitivity prominence, defined here as the ratio of transitive entries among all non-empty entries in a language dataset (see Haspelmath 2015; Say, forthc.; and Section 5.2 for similar calculations applied to predicates).[17] Predictability of valency class assignment (*PredCons*) correlates strongly with transitivity prominence across languages (Pearson's $r = 0.95$, $p << 0.001$), as shown in Figure 6.

Figure 6. Transitivity prominence and *PredCons* in BivalTyp languages (scatterplot)



The positive correlation between transitivity prominence and *PredCons* is expected both mathematically and conceptually. Mathematically, verbs belonging to smaller

[16] Say (forthc.) calculates valency class system entropy (*H*) from the same BivalTyp database, using the language-specific distribution of verbs across valency classes as probability estimates for different outcomes of the target variable. Both *PredLib* and *PredCons* show strong correlations with *H* (Pearson's $r = 0.96$, $p << 0.001$). Further details will be discussed elsewhere.

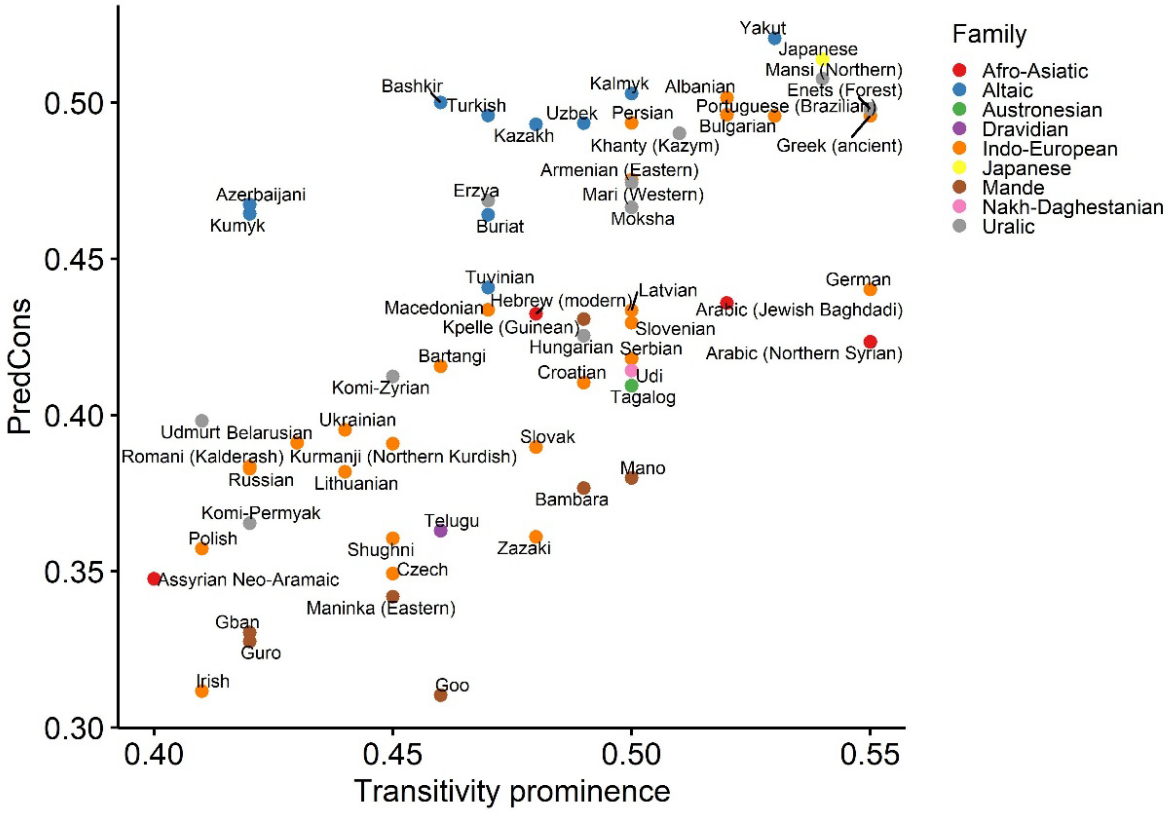
[17] For a map illustrating the areal distribution of transitivity prominence, see the online version of the BivalTyp database (Say 2020–), Section "Maps."

valency classes are less likely to achieve high *PredCons* values (see Section 4 for the relevant formulas). Conceptually, the transitive class is usually the largest and least semantically specialized bivalent class in any language; thus, a language with many transitive verbs tends to have a simpler, more regular system. From a functional perspective, such a system may also be easier to acquire in an L2 learning scenario than one with numerous non-default valency classes.

It is worth noting, however, that *PredCons* (like *PredLib*) is ultimately based on pairwise comparison of languages, whereas transitivity prominence is a system-internal parameter (see also footnote 11 on entropy). The fact that the two align so closely suggests that in a sufficiently large sample, pairwise cross-linguistic effects average out, and mean *PredCons* values come to reflect internal properties of valency systems—specifically, their semantic granularity—rather than mere resemblance to other languages in the dataset.

In sum, both transitivity prominence and mean *PredCons* values capture substantial cross-linguistic variation, distinguishing simpler valency systems with broad semantic generalizations from more intricate ones with fine-grained, often idiosyncratic, valency distinctions. Given their high overall correlation, it is instructive to look at cases where the two metrics diverge. Figure 7 zooms in on the mid-range of transitivity prominence, showing language names and color-coding language families.

Figure 7. Transitivity prominence and PredCons: zoom-in on mid-range transitivity prominence values



While the positive correlation remains visible in this section of the scatterplot, Figure 7 also reveals noticeable vertical fluctuations within this middle zone, demonstrating that *PredCons* captures something beyond transitivity prominence. These deviations align with genealogical distinctions, pointing to underlying grammatical differences. Specifically, within comparable ranges of transitivity prominence, Altaic languages[18] exhibit the highest *PredCons* values, reflecting more predictable valency systems, followed by Uralic, Indo-European, and finally Mande languages.[19] Among languages with lower transitivity prominence (not shown in Figure 7), an emerging hierarchy is Uralic > Kartvelian > Nakh-Daghestanian.

[18] There is ongoing debate over whether the so-called Altaic languages constitute a true genealogical group or an areal set comprising at least Turkic, Mongolic, and Tungusic languages. In any case, they share many typologically relevant features, including similar valency class system profiles. Tungusic languages are not shown in Fig. 7 due to their higher transitivity prominence values (in the range of 0.57–0.62), but they support the generalization that Altaic languages display high PredCons values given their transitivity prominence.

[19] The other families in Fig. 7 are represented too sparsely to draw firm conclusions. The position of the Mande languages in their areal context also remains to be explored once the sample includes more languages from Africa.

To conclude, systematic cross-linguistic differences in *PredCons* persist even when transitivity prominence is controlled for, and these differences display genealogical structuring, likely reflecting diachronically stable grammatical tendencies. A tentative hypothesis is that the observed variation reflects how extensively morphosyntactic devices originally expressing spatial relations are recruited for argument coding in non-spatial verbs. For example, Altaic languages—especially Turkic and Mongolic—tend to use their not too numerous morphological cases such as accusative, dative, and ablative to code verbal arguments, while postpositions with specific spatial meanings (e.g. 'under X', 'from near X') are rarely extended beyond literal use. In contrast, many Nakh-Daghestanian languages have very large case inventories, with numerous morphological cases with very specialized spatial meanings also employed for arguments of abstract verbs such as 'be afraid', 'wait', or 'think'. Similar developments are also seen in many prepositions in Indo-European languages such as German, Irish, and various Slavic and Baltic languages. These patterns suggest that the cross-linguistic differences sketched here are linked to the prominence of particular grammaticalization pathways—an issue that merits a more systematic investigation elsewhere.

7. Conclusions

This paper contributes to the debate on how verbs' arguments are assigned to coding devices such as morphological cases, adpositions, and verbal indexes. As reviewed in Section 2, the literature distinguishes several mechanisms that differ in the generality of their application—from default case assignments with maximal generality, through mid-range mechanisms applying to semantically related verb classes, to fully idiosyncratic case assignments limited to individual lexical items.

The present study advances this discussion by proposing that broader, more general patterns should also be more consistent across languages, whereas lexical case assignments should be less consistent cross-linguistically. Based on this idea, I introduced statistical measures that estimate how predictable language-specific verb-pattern associations are from the behavior of verbs with the same meanings in other languages (Section 4).

The main empirical result is that predictability varies substantially both across predicates and across languages. This supports the view that valency patterns are

neither fully motivated (predictable) nor entirely idiosyncratic (unpredictable), but distributed along a continuum constrained by recurrent regularities. While this general theoretical claim is not new, the present study offers a proof of concept that such variation can be quantified empirically rather than merely described qualitatively on the basis of a few well-known languages.

Based on data from 130 bivalent predicates in 149 languages in the BivalTyp database (Section 3), I identified clear cross-predicate differences (Section 5). Predicates with high semantic transitivity (e.g., 'break') exhibit the most consistent cross-linguistic behavior, followed by a few other zones such as symmetric predicates ('X fights with Y'). More generally, verbs denoting concrete, observable events tend to be used in more predictable valency patterns than verbs with abstract meanings, especially emotion predicates. This challenges the empirical validity of some traditional semantic roles, particularly that of stimulus: instead of using dedicated forms to code the so-called stimuli of verbs of emotion, languages often employ diverse construals of emotions grounded in more concrete cognitive schemas, resulting in high intra- and cross-linguistic variability.

Significant cross-linguistic differences were also found (Section 6). While all languages combine motivated and idiosyncratic aspects in their valency systems, the relative weight of each varies. These differences appear to be fundamental, as the corresponding distributions show robust areal patterns: for example, the languages of the Caucasus display a substantially higher proportion of idiosyncratic valency patterns than Standard Average European languages. Moreover, the phenomenon seems to have a degree of diachronic stability, as suggested by noticeable family-level contrasts—Altaic languages, for instance, show more predictable valency class systems than Indo-European languages with comparable levels of transitivity prominence, such as Baltic and Slavic. A promising hypothesis for future research is that these quantitative differences may reflect the extent to which languages rely on spatial metaphors when construing non-observable events: the more diverse the underlying cognitive schemas, the less predictable the resulting valency behavior.

Acknowledgements

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317633480 – SFB 1287. I am greatly indebted to Peter Arkadiev, Maria Ovsjannikova, and two anonymous referees for their comments

on an earlier version of this paper, which I have gratefully incorporated into the revision. The usual disclaimers apply.

Abbreviations

1, 2, 3 first, second, third person
ABL ablative
ACC accusative
AUX auxiliary
BEN benefactive
CNV converb
COP copula
DAT dative
DCL declarative
DEF definite
ERG ergative
F feminine
GEN genitive
INDEF indefinite
IO indirect object
IPFV imperfective
LOC locative
M masculine
N neuter
NOM nominative
OBL oblique
PFV perfective
PL plural
PN person name
POSS possessive
PRET preterite
PROG progressive
PRS present tense
PTCP participle
PST past
REFL reflexive
SG singular
VER:SUP superessive versionizer

References

Apresjan, Jurij D. & Erna Páll. 1982. *Russkij glagol – vengerskij glagol. Upravlenie i sochetaemost'*. Vols. 1–2. Budapest: Tankyonvkiado.

Barðdal, Jóhanna. 2011. Lexical vs. structural case: a false dichotomy. *Morphology* 21. 619–654.

Bickel, Balthasar & Taras Zakharko & Lennart Bierkandt & Alena Witzlack-Makarevich. 2014. Semantic role clustering: an empirical assessment of semantic role types in non-default case assignment. *Studies in language* 38 (3). 485–511.

Blume, Kerstin. 1998. A contrastive analysis of interaction verbs with dative complements. *Linguistics* 36 (2). 253–280.

Bossong, Georg. 1998. Le marquage de l'expérient dans les langues d'Europe. In Jack Feuillet (ed.). *Actance et valence dans les langues de l'Europe*, 259–94. Berlin: Mouton de Gruyter.

Cheng, Joe & Barret Schloerke & Bhaskar Karambelkar & Yihui Xie. 2024. *leaflet: create interactive web maps with the JavaScript 'leaflet' library*. R package version 2.2.2. Available online at: https://CRAN.R-project.org/package=leaflet. Accessed on 2025.10.19.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Cysouw, Michael. 2014. Inducing semantic roles. In Silvia Luraghi & Heiko Narrog (eds.), *Perspectives on semantic roles*, 23–68. Amsterdam Philadelphia: John Benjamins.

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

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. WALS Online (v2020.4). Available online at: https://wals.info (Accessed on 2025.10.17.)

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

Fillmore, Charles. 1968. The case for case. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–88. New York: Holt, Rinehard and Winston.

Gaszewski, Jerzy. 2020. Does verb valency pattern areally in Central Europe? A first look. In Luka Szucsich & Agnes Kim & Uliana Yazhinova (eds.), *Areal convergence in Eastern Central European languages and beyond*, 13–53. Berlin: Peter Lang.

Guzmán Naranjo, Matías & Laura Becker. 2022. Statistical bias control in typology. *Linguistic typology* 26(3). 605–670.

Hartman, Iren & Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in language* 38(3). 463–484.

Haspelmath, Martin & Iren Hartmann. 2015. Comparing verbal valency across languages. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, Vol. 1, 41–71. Berlin, Boston: De Gruyter Mouton.

Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages In Alexandra Y. Aikhenvald & Robert M. W. Dixon & Masayuki Onishi (eds.), *Non-canonical marking of subjects and objects*, 53–83. Amsterdam: John Benjamins.

Haspelmath, Martin. 2009. Terminology of case. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 505–517. Oxford: Oxford University Press.

Haspelmath, Martin. 2014. Arguments and adjuncts as language-particular syntactic categories and as comparative concepts. *Linguistic Discovery* 12(2). 3–11.

Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, Vol. 1, 131–147. Berlin, Boston: De Gruyter Mouton.

Heine, Bernd. 1997. *Possession: cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.

Helbig, Gerhard & Wolfgang Schenkel. 1983. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Tübingen: Max Niemeyer. (first edition 1969).

Herbst, Thomas & David Heath & Ian F. Roe & Dieter Götz. 2004. *A valency dictionary of English*. Berlin, New York: Mouton de Gruyter.

Hopper, Paul J. & Sandra A. Thompson. 1980. *Transitivity in grammar and discourse*. *Language* 56(2). 251–350.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jackendoff, Ray S. 1990. *Semantic Structures*. Cambridge: The MIT Press.

Jolly, Julia A. 1993. Preposition assignment in English. In Robert D. Van Valin Jr. (ed.), *Advances in Role and Reference Grammar*, 275–310. Amsterdam: Benjamins.

Kassambara, Alboukadel. 2023. *ggpubr: 'ggplot2' based publication ready plots*. R package version 0.6.0. Available online at: https://CRAN.R-project.org/package=ggpubr. Accessed on 2025.10.19.

Kittilä, Seppo. 2002. *Transitivity: Towards a Comprehensive Typology*. Turku: Åbo Akademis Tryckeri.

Klein, Katarina & Silvia Kutscher. 2002. Psych-verbs and lexical economy. *Theorie des Lexikons (Arbeiten des Sonderforschungsbereichs 282)* 122. 1–41.

Klotz, Michael. 2007. Valency rules? The case of verbs with propositional complements. In Thomas Herbst & Katrin Götz-Votteler. *Valency: theoretical, descriptive and cognitive issues*, 117-128. Berlin, New York: De Gruyter Mouton.

Lazard, Gilbert. 1994. *L'actance*. Paris: Presses Universitaire de France.

Lazard, Gilbert. 2002. Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica* 36. 141–190.

Lehmann, Ch. 1991. Predicate classes and PARTICIPATION. In Hansjakob Seiler & Waldfried Premper (eds.), *Partizipation. Das sprachliche Erfassen von Sachverhalten*, 183–239. Tübingen: G. Narr.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge: Cambridge University Press.

Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation.* Chicago: University of Chicago Press.

Lyashevskaya, Olga & Egor Kashkin. 2015. Inducing verb classes from frames in Russian: morpho-syntax and semantic roles. In *Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference "Dialogue"* (Moscow, May 27–30, 2015), Vol. 1, 412–425. Moscow: RSUH Publishers.

Malchukov, Andrej L. 2005. Case pattern splits, verb types and construction competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and variation in natural languages: the case for case*. 73–117. Oxford: Elsevier.

Malchukov, Andrej and Leipzig Valency Classes Project team. 2015. Leipzig Questionnaire on valency classes. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, Vol. 1, 27–40. Berlin: Mouton de Gruyter.

Malchukov, Andrej L. & Bernard Comrie (eds.). 2015. *Valency classes in the world's languages*. Vols. 1–2. Berlin: Mouton de Gruyter.

Meyer, Patrick E. 2022. *infotheo: information-theoretic measures*. R package version 1.2.0.1. Available online at: https://CRAN.R-project.org/package=infotheo. Accessed on 2025.10.19.

Moroz, George. 2017. *lingtypology: easy mapping for linguistic typology*. Available online at: https://CRAN.R-project.org/package=lingtypology. Accessed on 2025.10.19.

Næss, Åshild. 2007. *Prototypical transitivity*. Amsterdam & Philadelphia: John Benjamins.

Nedjalkov, Vladimir P. 1969. Nekotorye verojatnostnye universalii v glagolnom slovoobrazovanii. In Igor' F. Vardul' (ed.), *Jazykovye universalii i lingvisticheskaja tipologija*, 106–114. Moscow: Nauka.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer palettes*. R package version 1.1-3. Available online at: https://CRAN.R-project.org/package=RColorBrewer. Accessed on 2025.10.19.

Nichols, Johanna & David A. Peterson & Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8. 149–211.

Nichols, Johanna. 2008. Why are stative-active languages rare in Eurasia? Typological perspective on split subject marking. In Mark Donohue & Søren Wichmann (eds.), *The typology of semantic alignment systems*, 121-139. Oxford: Oxford University Press.

Ooms, Jeroen. 2025. *writexl: export data frames to excel 'xlsx' format*. R package version 1.5.4. Available online at: https://CRAN.R-project.org/package=writexl. Accessed on 2025.10.19.

Ovsjannikova, Maria & Sergey Say. 2014. Between predicative and attributive possession in Bashkir. In Pirkko Suihkonen & Lindsay Whaley (eds.), On diversity and complexity of languages spoken in Europe and North and Central Asia, 175–202. Amsterdam, Philadelphia: John Benjamins.

Ovsjannikova, Maria & Sergey Say. 2020. The instrumental case in the diachrony of Russian reflexive verbs of emotion: from cause to content. *Scando-Slavica* 66(1). 118–143.

R Core Team. 2025. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed 2025.08.12).

Say, Sergey (ed.). 2020–. *BivalTyp: typological database of bivalent verbs and their encoding frames*. Available online at: https://www.bivaltyp.info (Accessed on 2025.10.16).

Say, Sergey. 2014. Bivalent verb classes in the languages of Europe: a quantitative typological study. *Language dynamics and change* 4(1). 116–166.

Say, Sergey. 2018. Markirovanie aktantov dvuxmestnyx predikatov: predvaritel′nye itogi tipologičeskogo issledovanija. In Sergey Say (ed.), *Valentnostnye klassy dvuxmestnyx predikatov v raznostrukturnyx jazykax*, 557–616. St. Petersburg: ILI RAN.

Say, Sergey. 2025. Bivalent verb classes across Slavic: areal and genealogical patterns. *Russian Linguistics* 49: 13.

Say, Sergey. Forthc. Degrees of complexity in valency class systems: implications for efficiency. Accepted for publication in *STUF - Language Typology and Universals*, 2025, 4.

Slowikowski, Kamil. 2024. *ggrepel: automatically position non-overlapping text labels with 'ggplot2'*. R package version 0.9.6. Available online at: https://CRAN.R-project.org/package=ggrepel. Accessed on 2025.10.19.

Stassen, Leon. 2009. *Predicative possession*. Oxford: Oxford University Press.

Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396.

Vaidyanathan, Ramnath & Yihui Xie & J. J. Allaire & Jpe Cheng & Carson Sievert & Kenton Russell. 2023. *htmlwidgets: HTML widgets for R*. R package version 1.6.4. Available online at: https://CRAN.R-project.org/package=htmlwidgets. Accessed on 2025.10.19.

van den Eynde, Karel & Piet Mertens. 2006. Le dictionnaire de valence DICOVALENCE: manuel d'utilisation. Available online at: https://repository.ortolang.fr/api/content/dicovalence/1/documentation/DicovalenceManuel_v20_100625.pdf (Accessed on 2025.10.16)

Van Valin, Robert D., Jr. 1999. Generalized semantic roles and the syntax-semantics interface. In Francis Corblin & Carmen Dobrovie-Sorin & Jean-Marie Marandin (eds.), *Empirical issues in formal syntax and semantics 2*, 373–389. The Hague: Thesus.

Van Valin, Robert D., Jr. 2001. *An introduction to syntax*. Cambridge: Cambridge University Press.

Wickham, Hadley & Jennifer Bryan. 2025. *readxl: read excel files*. R package version 1.4.5. Available online at: https://CRAN.R-project.org/package=readxl. Accessed on 2025.10.19.

Wickham, Hadley & Romain François & Lionel Henry & Kirill Müller & Davis Vaughan. 2023. *_dplyr: A grammar of data manipulation* (R package version 1.1.4). Available online at: https://dplyr.tidyverse.org (Accessed 2025.10.16).

Wickham, Hadley. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer.

Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37 (1). 111–130.

Yip, Moira & Joan Maling & Ray Jackendoff. 1987. Case in tiers. *Language* 63(2). 217–250.

Zaenen, Annie & Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: the Icelandic passive. *Natural Language and Linguistic Theory* 3. 441–483.

Zeileis, Achim & Jason C. Fisher & Kurt Hornik & Ross Ihaka & Claire D. McWhite & Paul Murrell & Reto Stauffer & Claus O. Wilke. 2020. *colorspace: a toolbox for manipulating and assessing colors and palettes*. Journal of Statistical Software 96(1), 1–49.

Zolotova, Galina. 2006. *Sintaksicheskij slovar'. Repertuar sintaksicheskix edinic russkogo sintaksisa*. Moscow: Editorial URSS.

Appendix

Table A1. Aggregate predictability values (*PredLib* and *PredCons*) and transitivity prominence across 130 BivalTyp predicates, ordered by descending *PredLib* values

	PredLib	PredCons	Transitivity prominence
be called	0.80	0.03	0.06
make	0.78	0.78	1.00
drink	0.78	0.78	1.00
take	0.78	0.78	1.00
write	0.78	0.78	1.00
eat	0.78	0.78	1.00
sing	0.78	0.78	1.00
kill	0.78	0.78	1.00
wash	0.78	0.78	1.00
throw	0.78	0.78	1.00
melt	0.78	0.78	1.00
put on	0.78	0.77	0.99
break	0.77	0.77	0.98
fry	0.77	0.77	0.98
bend	0.77	0.77	0.98
read	0.77	0.76	0.98

take off	0.77	0.76	0.98
pour	0.77	0.76	0.97
try to catch	0.76	0.76	0.97
open	0.76	0.75	0.98
plough	0.75	0.74	0.97
milk	0.75	0.72	0.94
drive	0.73	0.72	0.95
cover	0.73	0.72	0.95
find	0.71	0.70	0.88
hold	0.71	0.66	0.88
hear	0.70	0.69	0.85
see	0.70	0.69	0.85
upset	0.70	0.61	0.84
drop	0.68	0.64	0.85
know	0.68	0.66	0.81
lose	0.68	0.64	0.86
look for	0.67	0.63	0.82
move (bodypart)	0.67	0.65	0.88
punish	0.67	0.63	0.82
understand	0.66	0.61	0.79
paint	0.66	0.63	0.84
give birth	0.65	0.61	0.84
hit	0.65	0.63	0.79
call	0.65	0.62	0.80
bite	0.65	0.62	0.77
cross	0.64	0.47	0.75
remember	0.64	0.58	0.76
love1	0.63	0.61	0.76
sink	0.62	0.15	0.00
surround	0.62	0.39	0.64
kiss	0.62	0.59	0.74
listen	0.60	0.56	0.73
despise	0.60	0.57	0.72
wait	0.60	0.56	0.71
fight	0.59	0.49	0.03

respect	0.59	0.51	0.72
be friends	0.59	0.49	0.04
hate	0.59	0.55	0.68
have a quarrel	0.58	0.45	0.02
speak	0.58	0.45	0.01
want	0.58	0.54	0.66
leave	0.57	0.49	0.74
catch up	0.57	0.52	0.72
forfeit	0.57	0.32	0.58
cost	0.57	0.29	0.52
feel pain	0.57	0.22	0.05
smell	0.56	0.12	0.14
enter	0.56	0.18	0.06
remain	0.55	0.26	0.12
follow	0.55	0.28	0.46
go out	0.55	0.21	0.05
have	0.54	0.38	0.44
have enough	0.54	0.30	0.02
love2	0.53	0.49	0.59
dismount	0.53	0.22	0.07
mix	0.53	0.40	0.02
be short	0.53	0.31	0.22
get upset	0.53	0.11	0.03
govern	0.53	0.44	0.65
forget	0.51	0.36	0.49
need	0.50	0.38	0.40
like	0.50	0.37	0.36
flatter	0.50	0.43	0.56
fall behind	0.50	0.19	0.06
play (instrument)	0.49	0.42	0.62
agree	0.49	0.40	0.04
wave	0.48	0.43	0.63
win	0.48	0.33	0.52
sympathise	0.48	0.27	0.42
help	0.47	0.39	0.46

touch	0.47	0.39	0.53
miss	0.47	0.32	0.47
shoot at	0.47	0.34	0.42
cut oneself	0.47	0.25	0.00
obey	0.47	0.41	0.50
attack	0.46	0.33	0.44
get stuck	0.46	0.18	0.02
dream	0.45	0.30	0.31
make fun	0.45	0.33	0.50
look	0.45	0.35	0.43
dream (sleeping)	0.45	0.30	0.39
answer	0.44	0.35	0.31
think	0.44	0.24	0.21
avoid	0.44	0.33	0.43
believe	0.44	0.33	0.33
resemble	0.43	0.18	0.15
be afraid	0.43	0.28	0.31
tell	0.43	0.27	0.19
have (illness)	0.43	0.20	0.20
hit (target)	0.43	0.29	0.29
reach	0.43	0.23	0.28
fill (intr)	0.43	0.19	0.06
be different	0.42	0.24	0.03
get irritated	0.42	0.26	0.07
be angry	0.42	0.29	0.07
get to know	0.42	0.36	0.37
be fond	0.41	0.35	0.41
trust	0.41	0.29	0.25
envy	0.41	0.30	0.32
be glad	0.41	0.20	0.03
have a grudge	0.41	0.28	0.10
take offence	0.40	0.24	0.06
lose game	0.40	0.19	0.08
enjoy	0.40	0.26	0.36
match	0.39	0.23	0.03

fall in love	0.39	0.29	0.37
encounter	0.39	0.31	0.34
influence	0.39	0.30	0.32
be shy	0.39	0.20	0.12
depend	0.39	0.15	0.00
marvel	0.36	0.20	0.14
be squeamish	0.35	0.24	0.20
be surprised	0.34	0.22	0.12
be content	0.33	0.20	0.09