Sergey Say*

Degrees of complexity in valency class systems: implications for efficiency

https://doi.org/10.1515/stuf-2025-2022

Abstract: The complexity of valency class systems remains an underexplored dimension of typological variation. Drawing on data from BivalTyp, a database cataloging valency patterns of 130 bivalent verbs across 120 languages, this paper proposes three metrics to quantify this complexity. Two of these metrics – the ratio of intransitive verbs among bivalent verbs and the entropy in the distribution of bivalent intransitive verbs across valency classes – are strongly positively correlated, reflecting different aspects of the overarching valency class complexity. The third metric, the entropy of all bivalent verbs across valency classes, captures both of these facets. Languages tend to gravitate toward moderate complexity values, avoiding overly simplistic or excessively complex valency class systems. However, significant variation exists, primarily driven by areal and genealogical factors. Valency class complexity, as measured and discussed in this paper, is empirically interwoven with numerous other parameters of morphosyntactic typology, suggesting that it is a fundamental factor in shaping linguistic diversity. Specifically, languages with large case inventories, a high prevalence of non-verbal predicates, and a preference for satellite-framed constructions in motion events are hypothesized to favor more complex valency class systems. Furthermore, complex argument-encoding devices tend to favor postverbal positions, aligning with the principles of incremental processing and the minimization of processing effort. Overall, this study contributes to understanding how efficiency pressures influence typological distributions.

Keywords: valency class; transitivity; complexity; entropy; word order

***Corresponding author: Sergey Say**, Institute of Slavic Studies, University of Potsdam, Potsdam, Germany, E-mail: say@uni-potsdam.de

1 Background and goals

1.1 Setting the stage

Languages differ in how they classify verbs into valency classes. This is evident in the German examples in (1) and (2) and their English translations.[1]

(1) German (Indo-European > Germanic; personal knowledge)
Ich *folge* *ihm.*
I.NOM follow.PRS.1SG he.DAT
'I follow him.'

(2) German (Indo-European > Germanic; personal knowledge)
Ich *sehe* *ihn.*
I.NOM see.PRS.1SG he.ACC
'I see him.'

The verbs in (1) and (2) are both *bivalent*, that is, have two arguments, but they differ in that *folgen* 'follow' (1) takes a Dative object, while *sehen* 'see' (2) takes an Accusative object. This encoding distinction is not observed in the English counterparts, as shown by the translations of (1) and (2). This is part of a larger picture: German has long been shown to exhibit finer argument-encoding distinctions than English (Hawkins 1986; Sauerland 1994). This difference can be understood in terms of complexity. Arguably, the German valency class system is more complex than the English system, as evidenced by both a larger case inventory and the extensive use of prepositions for argument encoding.

Despite occasional observations on differences between specific languages, a well-established quantitative method for evaluating valency class system complexity across languages is lacking. In this proof-of-concept paper, I present a technique for such purpose and apply it to a dataset comprising the valency patterns of 130 bivalent verbs in 120 languages. The main goal is to showcase the valency class system complexity as a relatively overlooked typological parameter. My further aim is to position this parameter within the broader scope of typological variation.

In the remainder of Section 1, I discuss some relevant analytic concepts (1.2), establish theoretical limits on valency class system complexity (1.3), and provide an overview of approaches to complexity (1.4). Section 2 introduces the dataset and two

1 This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317633480 – SFB 1287. I am profoundly grateful to the two anonymous reviewers, Peter Arkadiev, Kirill Kozhanov, Maria Ovsjannikova, and other members of the PoSla Typology Lab for their insightful comments on an earlier version of this paper. I thank George Moroz for his assistance in preparing the maps used in this paper. Any remaining errors are solely my responsibility.

valency class system complexity metrics. Main empirical findings are outlined in Section 3, with a focus on correlations between valency class system complexity and other typological parameters. Section 4 explores interactions with word order phenomena. In Section 5, I propose some hypotheses regarding complexity to be tested in the future. In the final Section 6, I summarize the findings and interpret them through efficiency-driven constraints.

1.2 Basic notions

The valency of a verb is "the list of its arguments with their coding properties" (Malchukov et al. 2015: 30). These properties include flagging, indexing, and word order (Malchukov et al. 2015: 31–33; Sinnemäki 2008). Flagging occurs when markers that indicate an argument's relationship to the verb are attached to or accompany the argument noun phrase (see Haspelmath 2019 for a detailed definition and a discussion of the relationship between flagging and dependent-marking). Flagging includes cases, adpositions, and combinations of these elements. For instance, in (3), the second argument of the German verb *warten* 'wait' is encoded by the combination of the preposition *auf* 'on, at' and the Accusative case.

(3) German (Indo-European > Germanic; personal knowledge)
Ich *warte* *auf* *ihn.*
I.NOM wait.PRS.1SG on he.ACC
'I am waiting for him.'

Indexing is also involved in the argument encoding in (1)–(3) as verb inflections cross-reference one of the arguments (cf. the 1SG forms in all these examples). However, the significance of indexing is relatively minor in German since it is entirely predictable from flagging (only Nominative subjects are indexed), while the reverse is not true. In other languages, indexing is more prominent, as demonstrated in (4), where noun phrases lack case marking, but their relationships to the verb are signaled by the shapes and positions of person, number, and gender indices within the verb form.

(4) Abaza (Northwest Caucasian; BivalTyp[2])
Fatíma *Murád* *jə-z-qá-l-ç̣-əj-ṭ.*
PN PN 3SG.M.IO-BEN-LOC-3SG.F.ERG-believe-PRS-DCL
'Fatima trusts Murad.'

2 Here and below, all examples taken from the BivalTyp database (Say (ed.) 2020–) are labeled accordingly; interlinear glosses follow the original database. For detailed references to language-specific contributions, consult the database itself.

Finally, in some languages, argument roles are primarily encoded by word order. In the English translation of (4), for example, word order is the primary cue allowing to identify the roles of the two referents.

Typological investigation into valency encounters challenges due to the language-specific nature of argument-encoding devices. Importantly, "the transfer of case labels from one language to another should not be understood as meaning that we are dealing with 'the same case'" (Haspelmath 2009: 510). For instance, the German Dative case in (1) cannot be directly equated with the Turkish Dative case, etc. To get around that problem, I rely on the concept of valency class. Two verbs belong to the same valency class if and only if their arguments are coded by the same combination of argument-encoding devices. Thus, the German verb *schauen* 'look', as used in (5), falls within the same valency class as the verb *warten* 'wait' in (3).

(5) German (Indo-European > Germanic; personal knowledge)

I	*schaue*	*auf*	*die*	*Wolken.*
I	look.PRS.1SG	on	ART.DEF.ACC.PL	cloud.PL

'I am looking at the clouds.'

The primary advantage of the valency class concept lies in its independence from the arbitrary descriptive labels applied to encoding devices. This allows for cross-linguistic comparisons of valency classes based on their lexical scope. For example, one can determine whether the equivalents of 'look' and 'wait' belong to the same valency in other languages, as they do in German.

1.3 The discriminatory function of case and other argument-encoding devices

Before delving into empirical data, it is important to consider the potential limits of complexity in valency class systems. The starting point is the common assumption that case and other argument-encoding devices primarily serve a discriminatory function, helping to syntagmatically differentiate discourse referents with respect to their semantic roles (Comrie 1989: 124–127; Dowty 1991; Levshina 2021: 4; Seržant 2019). For example, the use of the Nominative versus Accusative case in (2) allows the addressee to distinguish between 'the seer' and 'the person seen'. However, the vast majority of verbs in natural languages have no more than three arguments (Bickel et al. 2014; Dixon 2010). Therefore, in principle, a minimal system with just three distinct encoding devices could largely fulfill the discriminatory function. For example, a system where all monovalent verbs take a Nominative argument, all bivalent verbs use the Nominative-Accusative case frame, and all trivalent verbs use the Nominative-Accusative-Dative case frame would eliminate most confusion in terms of role-referent associations.

Such a hypothetical system could facilitate language production by reducing the cognitive load required to access verb-specific valency patterns from the lexicon. However, it would come with a drawback: individual argument-encoding devices in such a system would largely lack more specific semantic cues that aid speech perception. In this regard, an addressee can benefit from more complex systems (Givón 2009; Jaeger 2010). For instance, decoding the German sentence in (6), the addressee is likely to predict the sentence-final verb *abhängen* 'depend' before it is pronounced because there are very few verbs used with the preposition *von* 'of, from'. High levels of predictability of a syntactic element can in turn facilitate processing (Kuperberg and Jaeger 2016; Levy 2008; Stone et al. 2020).

(6) German (Indo-European > Germanic; personal knowledge; this standard formula is used, e.g., in SPD Government Program 2019–2024)

Teilhabe	*darf*	*nicht*	*vom*	*Alter*	*abhängen.*
participation	may.PRS.3SG	not	from.DEF.N.DAT.SG	age	depend.INF

'Participation must not depend on age.'

In an extremely complex hypothetical valency class system, each verb in the lexicon would have its own dedicated valency-encoding devices. Such a system would maximize cue reliability, so that successful decoding of any argument would be sufficient for identifying the entire scene. However, it would be excessively burdensome due to poor learnability and the need to process hundreds or thousands of different markers, some of which would inevitably be long.

In reality, all languages fall within the middle section of the spectrum between the two hypothetical extremes (see Section 3.1 for the empirical limits of valency class system complexity). Specifically, all languages seem to have a large transitive class forming the core of bivalent verbs, but also feature some intransitive bivalent verbs (Næss 2007; Tsunoda 1985). Thus, the existence and prevalence of intransitive bivalent verbs contribute to the overall valency class system complexity. This observation is relevant to the design of the dataset used in this study (see Section 2.1 for further details).

1.4 Approaches to complexity in linguistic typology

Linguistic complexity is a prominent subject in modern typology (Dahl 2004; Miestamo et al. (eds.) 2008; Sampson et al. (eds.) 2009; Shcherbakova et al. 2023; Stump 2017). Much of the discussion revolves around the equi-complexity hypothesis, which suggests that all languages tend to exhibit comparable levels of overall complexity

(Bentz et al. 2023). While there are compelling instances of partial complexity trade-offs, where complexity in one aspect of language is offset by simplicity in another, there is a growing consensus that the overarching equi-complexity hypothesis is either untestable or simply inaccurate (Fenk-Oczlon and Fenk 2014; Joseph and Newmeyer 2012; Shcherbakova et al. 2023; Shosted 2006; Sinnemäki 2014).

As interest in linguistic complexity grows, there arises a need for developing measurement techniques. One common approach involves enumerating the elements within specific domains, such as phonemes and tenses. This gist of this enumerative approach dates back to at least Sapir (1921), see also recent overviews in (Ackerman and Malouf 2013; Shcherbakova et al. 2023). However, this enumerative approach fails to capture many aspects of what is intuitively seen as hallmarks of complexity. This gives rise to approaches focusing on the complexity of rules in a system (McWhorter 2001) or the functional load of individual elements (Sinnemäki 2008). Many modern approaches to complexity draw insights from information theory and entropy (Ackerman and Malouf 2013; Juola 1998; Levshina 2019).

While complexity research encompasses various linguistic domains such as phonology, morphology, and syntax, valency classes have not received significant attention in this inquiry. The focus has primarily been on whether core arguments (A, S, and P) display encoding differences (Sinnemäki 2008, 2009; Shcherbakova et al. 2023). Apart from that, scholars have noted variations in "transitivity prominence", i.e., the lexical extent of the basic transitive scheme in the verbal lexicon (Creissels 2018; Haspelmath 2015). Lower transitivity prominence is associated, implicitly or explicitly, with higher complexity. The underlying logic is that lower transitivity prominence correlates with finer distinctions between semantic roles and verb classes (Drossard 1991: 435–436; Haspelmath 2015; Lazard 1994: 61–63). However, almost no attention has been paid to differences in the classification of intransitive polyvalent verbs, which also contributes to the valency class systems complexity. My paper aims to fill this gap from a quantitative typological perspective by proposing metrics to measure cross-linguistic differences in the complexity of valency class systems and by situating these differences within the broader context of typological variation.

2 Data and complexity metrics

2.1 The database

The actual arena for information transmission, where varying degrees of complexity can either facilitate or hinder communication, is *parole*, not *langue*. Consequently, assessing the complexity of argument-encoding systems from a token-based typological perspective emerges as the ultimate desideratum for future research. However, this approach faces challenges due to the current lack of broad cross-linguistic corpora with unified syntactic annotations (though see Schnell et al. 2021 for an overview of recent advances in this direction). As a compromise, this study will utilize a wordlist-based approach, whereby cross-linguistic comparison is based on a set of pre-selected verbal meanings. The primary data source will be BivalTyp, an online typological database of bivalent verbs and their encoding frames (Say (ed.) 2020–).

The BivalTyp data were contributed by language experts who elicited translations from native speakers using a questionnaire comprising 130 bivalent verbs presented in context. These sentences serve as discrete "probes" in the infinite semantic space. The focus on bivalent verbs stems from their proneness to display deviant valency behavior (Bickel et al. 2014). Moreover, the BivalTyp questionnaire primarily includes verbs that lack a true agent and/or a true patient, such as 'be afraid', 'follow', 'see', and 'touch' (Say 2014: 126). The focus on the verbs that deviate from the transitive prototype fosters observing cross-linguistic differences in valency class systems, as such verbs are known to typologically attract various non-canonical patterns (Bickel et al. 2014; Næss 2007; Tsunoda 1985). That said, limiting the scope of this study to bivalent verbs only could affect its results. Whether the current conclusions can be extrapolated to monovalent and trivalent verbs remains an issue for future exploration.

Each sentence in the questionnaire features two predefined arguments labeled "X" and "Y": for example, 'The boy (X) reached the bank (Y)' or 'The girl (X) hears the music (Y)'. This labeling is based on Dowty's "lexical entailments" (Dowty 1991), where "X" represents the argument that accumulates more agentive properties than the other. In this sense, the "X" and "Y" labels used in the BivalTyp database closely resemble the "A" and "P" labels in the "Bickelian" approach to alignment (Bickel 2011; Haspelmath 2011: 552–558).

Each entry in the database is annotated for its language-specific valency pattern, indicating the morphosyntactic devices used for encoding X and Y. For instance, in example (7), the valency pattern involves the X-argument in the Adlative case and the Y-argument in the Nominative case, which is also cross-referenced on the verb.

(7) Tsakhur (Nakh-Daghestanian > Lezgic; BivalTyp)

Murad-ı-sqha	*mık'ey-bı*	*uvaykı-mbı*
PN-OBL-ADLAT	key-PL.NOM	find.PFV-PL

'Murad found (his) keys.'

As a general rule, the valency classes in BivalTyp are defined based on all types of argument-encoding devices, namely flagging, indexing, and word order (see the brief overview in Section 1.2). In practice, however, the valency pattern can often be fully captured by reference to flagging alone. This is the case in example (7): here, the number indexing on the verb is predictable given the case frame, as only nominative arguments are indexed in Tsakhur. For this reason, the conventional label for the valency pattern in (7) is ADLAT_NOM, which directly corresponds to the use of flagging devices.

Overall, flagging is by far the most significant argument-encoding device for identifying valency classes in BivalTyp. This is due to two main reasons. First, typologically, word order and indexing are more crucial for distinguishing core arguments, while non-core arguments typically rely on flagging (Nichols 1986: 75). As a result, valency classes involving non-core arguments – central to the BivalTyp database – typically rely on flagging. Second, BivalTyp's language sample is biased towards Eurasia, a macro-area favoring dependent-marking. That said, some valency patterns in the BivalTyp dataset can only be identified when indexing and/or word order are taken into account. In a few languages in the dataset, including Abaza (mentioned in Section 1.2), indexing is the primary device for distinguishing valency classes. Additionally, many languages rely on indexing and/or word order to differentiate between unflagged arguments, often including, but not limited to, subjects and direct objects.

In BivalTyp, a language-specific verb is deemed transitive if its X and Y arguments are encoded by the same devices as the two arguments of the verb 'kill' in the same language (see Haspelmath 2015: 136 for a similar approach). Therefore, sentence (7) exemplifies one of the many intransitive bivalent patterns in Tsakhur, as X-arguments in Tsakhur transitive clauses are encoded by the ergative case, as in (8).

(8) Tsakhur (Nakh-Daghestanian > Lezgic; BivalTyp)
Murad-ee *xoče* *givk'u*
PN-ERG snake(G3) G3.kill.PFV
'Murad killed the snake.'

BivalTyp's annotation of valency classes captures the encoding of arguments X and Y, regardless of whether the construction is headed by a simplex verb, non-verbal, or complex predicate. While this approach aids cross-linguistic comparison, the resulting classes may not always align with language-internal distinctions. For example, if a language expresses 'X answered Y' as literally 'X gave an answer to Y', the entry is classified as intransitive based on the encoding of X and Y, even if the simplex verb is considered transitive within the language. Non-verbal and complex predicates are noted in the database comments.

For this paper, I used the latest development version of BivalTyp available as of February 16, 2024. The initial dataset contained 124 languages, which was further reduced to 120 following the procedure discussed in Section 2.2. The BivalTyp sample is a convenience sample, heavily skewed both genealogically and geographically. The best-covered areas are Europe and the Caucasus, with a considerable number of languages from Asia (particularly Northern Asia) and some from Africa, while Australia, Papunesia and the Americas are largely unrepresented, see the Appendix for the list of languages and the maps in Section 3.3 for their geographic distribution. These biases undoubtedly impose limitations on the overall results obtained. However, they also enable the interpretation of differences observed within well-represented language families.

Some data are missing from the dataset; see the Appendix for details. The number of complete data points per language ranges from 90 to 130, with an average of 120.7, corresponding to 9.3 missing points per language. In addition to valency patterns in up to 130 constructions, BivalTyp languages are annotated for the total number of morphological cases and the basic word order pattern in the transitive construction (by default, these annotations are provided by contributors of BivalTyp's language-specific datasets).

The dataset employed in this study is available in the Supplementary Materials (https://github.com/serjozhka/Degrees_of_complexity_STUF). All calculations for this study were performed in R (R Core Team 2021). I utilized the following packages for data analysis and visualization: `colorspace` (Zeileis et al. 2020), `dplyr` (Wickham et al. 2023), `ggplot2` (Wickham 2016), `ggpubr` (Kassambara 2023), `ggrepel` (Slowikowski 2025), `infotheo` (Meyer 2022), `lmerTest` (Kuznetsova et al. 2017), `readxl` (Wickham and Bryan 2023), `tidyverse` (Wickham et al. 2019), and `writexl` (Ooms 2024). The R code used in this study is also available in the Supplementary Materials.

2.2 Two complexity metrics

I use two basic metric types designed to capture the level of complexity within valency class systems.[3]

The more straightforward metric is transitivity prominence (Haspelmath 2015, see also Drossard 1991 for an early insight). It is calculated as the ratio of transitive patterns to the total number of patterns obtained. For instance, there are 120 data points for Skolt Saami, with 67 representing the transitive pattern, resulting in a transitivity prominence of 0.56. Generally speaking, lower values are observed when

3 A more simplistic alternative to the techniques discussed here would be to calculate the number of distinct argument-encoding devices or valency classes in a language. In fact, the 'number of classes' column in the 'Languages' section of the online version of BivalTyp provides these numbers based on language-specific datasets. However, beyond the general limitations of this enumerative approach to complexity, BivalTyp's design is particularly ill-suited to relying on the number of distinct patterns observed, due to its small size, containing at most 130 entries, leaving many valency classes inevitably unrepresented.

the transitive class is more semantically restricted, indicating higher degree of semantic specialization and, ultimately, more complex valency class systems. The results obtained using this metric are discussed in Section 3.1.

However, transitivity prominence is "blind" to the number and nature of distinctions made between intransitive bivalent classes. Therefore, the primary metric used in this study is the entropy observed in the distribution of verb-specific patterns among language-specific bivalent valency patterns.[4] The formula for Shannon's entropy as used in this study is given in (9), where $p(x_i)$ corresponds to the probability that a certain observation displays the *i*-th of the *k* possible values of *x*. In this study, entropy is measured in *nats* with the base for the logarithm being *e*.

(9) $$H(x) = -\sum_{i=1}^{k} p(x_i) \cdot \log_e (p(x_i))$$

Possible values of the variable *x* correspond to distinct valency classes. For instance, in Finnish (Uralic > Finnic), *k* equals 15 reflecting the presence of 15 different valency classes in the Finnish dataset. Due to unavailability of actual probabilities of valency classes, the relative frequencies are used as an approximation, which is a standard step in entropy-based approaches to linguistic phenomena (Ackerman and Malouf 2013: 439; Levshina 2019). By way of illustration, there are 12 entries featuring the NOM_ILL valency pattern among a total of 129 data points for Finnish. Consequently, the p-value associated with the NOM_ILL class is 0.093 (= 12/129), contributing 0.22 (= $-0.093 \cdot \log_e(0.093) = -0.093 \cdot -2.37$) to the total observed entropy of bivalent verbs in Finnish, which equals 1.89.

In a nutshell, entropy of bivalent verbs (H) reflects the amount of information linked to a verb's valency class membership in a particular language. In a hypothetical language where all bivalent verbs are transitive, H would equal 0, as valency class membership would convey no information about verbs in such language. Conversely, high levels of H are observed in languages with many valency classes where the use of specific argument-encoding devices offers substantial cues for the sentence's overall meaning.

An initial analysis of the raw data revealed that H values are lower in languages with a higher incidence of missing data points. The occurrence of missing data, influenced by factors unrelated to the valency patterns themselves (e.g., the absence of speakers capable of expressing more complex or abstract meanings), may result in an underestimation of complexity levels for languages with fewer data points. In response to this problem, I took the following two steps. First, I disregarded the data

4 Entropy offers advantages over the enumerative approach to complexity, which relies on simply counting the number of elements in a system (e.g., the number of distinct valency classes). The primary issue with the enumerative approach is that it assigns excessive weight to very rare classes, making it highly susceptible to the vicissitudes of the data collection processes (Juola 1998: 207).

from the four languages with fewer than 90 entries (this arbitrary threshold excludes languages where more than approximately one-third of the maximum possible 130 entries are missing). Second, I employed a correction procedure for the remaining 120 languages. I generated 100 random subsamples, each consisting of 90 entries (without replacement), and computed the average entropy value observed within these subsamples. To illustrate, for Telugu (Dravidian), a language with 121 entries, the raw H value is 2.09, while the corrected value is 2.05. In all subsequent calculations, "H" refers to these corrected values; see the Appendix for the raw and corrected values of H. Theoretical limits on possible values of H with the procedure just discussed are 0 nats for the maximally simple system and the natural logarithm of 90 (≈ 4.5 nats) for the maximally complex system. In the next section, I discuss the empirically observed variation in the data.

In addition to transitivity prominence and the entropy of bivalent verbs (H), this study also uses the entropy of intransitive bivalent verbs (H_{intr}) as a third complexity metric. This metric is introduced in Section 3.2, following the discovery of a strong correlation between the first two metrics, which is also discussed in that section.

3 Results

3.1 General distributions of the two complexity metrics

The first complexity metric used in this study, transitivity prominence, was found to vary from 0.25 for Karata (Nakh-Daghestanian > Avar-Andic-Tsezic) and Estonian (Uralic > Finnic) to 0.79 for Joola-Fonyi (Niger-Congo > Jola). Substantively, this means that more than half of the verbs from the questionnaire display transitive patterns in Joola-Fonyi (aligning with 'kill') but intransitive bivalent patterns in, e.g., Karata.

The raw values may lack significance due to the arbitrary design of the questionnaire. However, the differences between values for individual languages are meaningful, indicating a key parameter of typological variation. These differences generally align with those reported by Haspelmath (2015) based on the ValPal database. The language samples used in BivalTyp and ValPal have a small overlap of 11 languages. Table 1 presents transitivity values reported for these languages in the two sources; a 12th row represents the data for Bezhta, the only Nakh-Daghestanian language in the ValPal sample, and Khwarshi, another Tsezic language available in BivalTyp, as its proxy. The rows are arranged in descending order of transitivity prominence as recorded in the ValPal database.

As shown in Table 1, the transitivity prominence values from ValPal and BivalTyp do not always match due to differences in verb samples. Notably, BivalTyp exhibits a broader range of values (with a standard deviation of 0.105 compared to

Table 1: Transitivity prominence values in ValPal (Haspelmath 2015) and BivalTyp.

Language	ValPal	BivalTyp
Mandarin Chinese (Sino-Tibetan > Chinese)	0.65	0.67
Italian (Indo-European > Romance)	0.62	0.6
Mandinka (Mande > Western Mande)	0.62	0.56
Japanese (Japanese)	0.61	0.54
Standard Arabic (Afro-Asiatic > Semitic)	0.60	0.57
Evenki (Altaic > Tungusic)	0.59	0.57
English (Indo-European > Germanic)	0.58	0.64
German (Indo-European > Germanic)	0.56	0.55
Eastern Armenian (Indo-European > Armenian)	0.54	0.5
Russian (Indo-European > Slavic)	0.50	0.42
Icelandic (Indo-European > Germanic)	0.47	0.36
Bezhta/Khwarshi (Nakh-Daghestanian > Avar-Andic-Tsezic)	0.40	0.28

0.077 for ValPal). This discrepancy arises because BivalTyp focuses on verbs that cross-linguistically can be either transitive or intransitive, while ValPal includes many verbs that are either uniformly intransitive (e.g., monovalent verbs) or uniformly transitive. Despite this difference, the datasets display a nearly perfect correlation (Pearson's $r = 0.94$, $p < 0.001$), indicating that the two sources mutually support each other's findings.

Figure 1 shows the overall distribution of transitivity prominence in the BivalTyp sample, with the actual values provided in the Appendix. Generally, this distribution resembles the bell curve of a normal distribution. The anomaly in the lower part of the distribution is due to the overrepresentation of the Nakh-Daghestanian family: of the 31 languages with transitivity prominence below 0.35, 23 are from this family, with only one Nakh-Daghestanian language, Udi (Lezgic), outside this range.

The second complexity metric, (corrected) entropy of bivalent verbs (H), also exhibits considerable variation across languages, ranging from 0.71 nats for Joola-Fonyi (Niger-Congo > Jola) to 2.71 nats for Khwarshi (Nakh-Daghestanian > Avar-Andic-Tsezic). These values highlight the simplicity of the valency class system in Joola-Fonyi and the complexity in Khwarshi. In Joola-Fonyi, the transitive class encompasses 91 entries, with the second most frequent class covering another 17 entries out of a total of 115 (with an exceptionally low total of only 6 classes). In Khwarshi, the transitive class comprises only 34 entries out of a total of 122, and there are 32 other classes. For illustration, Table 2 features 10 selected meanings represented by transitive constructions in Joola-Fonyi, corresponding to verbs from 10 distinct valency classes in Khwarshi (only one of which is transitive).

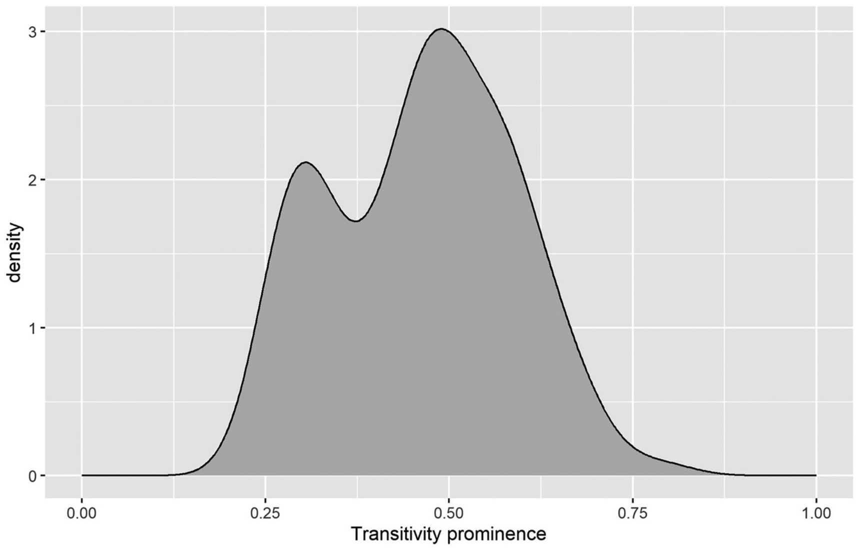

Figure 1: Distribution of transitivity prominence values in the sample.

Table 2: Selected verbs and valency patterns in Joola-Fonyi and Khwarshi.

	Joola-Fonyi		Khwarshi	
Meaning	**Verb**	**Pattern**	**Verb**	**Pattern**
'be afraid'	*kólí*	TR	*j/uƛ'a*	ABS_CONT
'avoid'	*ŋom*	TR	*j/iča*	ABS_CONT.EL
'wait'	*kob*	TR	*gic'a*	ABS_CONT.LAT
'attack'	*lóúm*	TR	*k'oλa*	ABS_SUPER
'win, beat'	*ŋoolen*	TR	*j/iža*	ABS_SUPER.EL
'see'	*juk*	TR	*j/akʷa*	DAT_ABS
'touch'	*gor*	TR	*j/etaχa*	ERG_CONT
'bite'	*rum*	TR	*hana*	ERG_GEN1
'be angry'	*leet*	TR	*semi mak'a*	GEN1_CONT.LAT
'eat'	*ri*	TR	*j/ac'a*	TR (ERG_ABS)

Table 2 provides a glimpse into the vast differences between valency class systems, highlighting the two extreme cases in the dataset. The overall distribution of H values in the sample is shown in Figure 2. See Section 3.3 for a map of the geographical distribution and Appendix for the actual H values in the sample.

The results for the entropy values (H) provide two key insights. First, the range of H values in the sample is substantial, with notable variation beyond the extreme

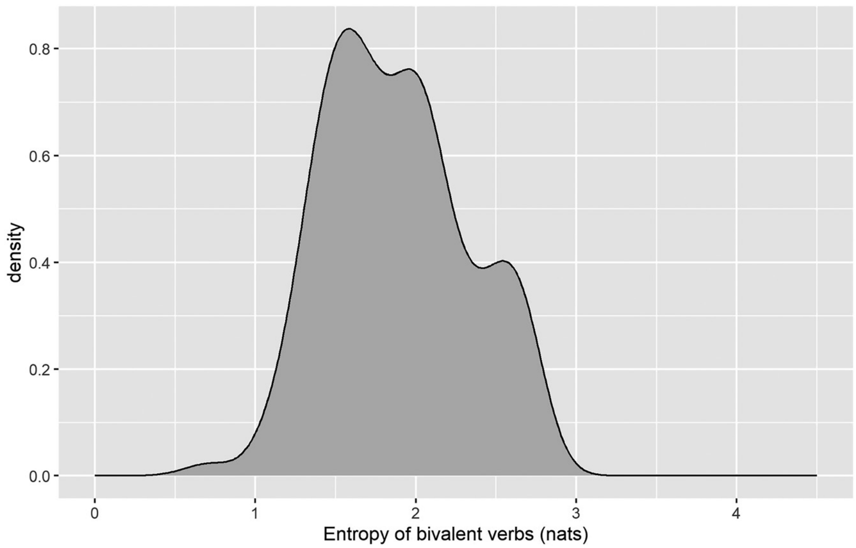

Figure 2: Distribution of entropy of bivalent verbs (H) in the sample.

cases of Joola-Fonyi and Khwarshi. Specifically, the bottom 20 % of languages exhibit H values below 1.48, while the top 20 % exceed 2.29. Second, this variation is constrained within empirical limits: although theoretically, H values could range from 0 to approximately 4.5, the observed values occupy less than half of this potential range. Sections 3.2–3.4 contextualize this variation in valency class system complexity within the broader context of linguistic diversity.

3.2 Introducing the third complexity metric (H_{intr}) and dependencies between complexity metrics

The two complexity metrics introduced so far – transitivity prominence and entropy of bivalent verbs (H) – exhibit a negative correlation (Pearson's $r = 0.89$, $p < 0.001$). However, due to the genealogical bias in the sample, this observed correlation might be influenced by the overrepresentation of a few large and relatively homogeneous language families. To account for potential genealogical effects, a linear mixed-effects model was employed with linguistic family included as a random effect. In this model, entropy of bivalent verbs (H) is the predicted parameter, and transitivity prominence serves as the fixed effect. Table 3 provides a summary of the model results. The fixed effects reveal a significant negative relationship between transitivity prominence and entropy of bivalent verbs (Estimate = –2.87, $p < 0.001$). The

Table 3: Results of the linear mixed-effects model analyzing the relationship between transitivity prominence and entropy of bivalent verbs (H).

Effect	Estimate	Std. Error	df	t value	Pr(>|t |)
Fixed effects					
(Intercept)	3.173	0.099	61.13	32.08	< 0.001***
Transitivity prominence	−2.870	0.175	107.4	−16.38	< 0.001***
Random effects					
Variance (Intercept): Family	0.023				
Residual	0.025				

*Indicate significance levels: p < 0.001**, p < 0.001***.

random intercept for linguistic family captures a modest amount of variance (0.023), indicating some variability in entropy across language families, while the residual variance remains low (0.025).

However, the dependence reported here is partially due to the analytic relationship between the two parameters, as they are not entirely independent of each other. For instance, if the transitive class dominates a language's dataset and the transitivity prominence value is very high, the overall uncertainty decreases (since the membership of a verb in the transitive class will not be very informative), and H is inevitably low. Moreover, the transitive class empirically emerged as the largest class in all but one language in the sample,[5] confirming previous observations (Haspelmath 2015: 139; Lazard 1994: 131–158). In this context, languages with low transitivity prominence always feature numerous smaller classes, leading to higher entropy. Put differently, the overall result in the entropy formula presented in (9) is largely determined by the contribution of one summand, the one related to the transitive class, where $p(x_i)$ is precisely the value of transitivity prominence.

To investigate facets of complexity not determined by transitivity alone, I introduced another metric: the entropy of intransitive bivalent verbs, denoted as $\mathrm{H}_{\mathrm{intr}}$. Calculations followed the procedure outlined in Section 2.2, with the distinction that only intransitive bivalent verbs were taken into account. Similar to the overall entropy, I found a positive correlation between the total number of intransitive bivalent entries and their observed entropy. To address this, I again implemented a correction procedure: for each language, I generated 100 random subsamples (without replacement), each comprising 24 intransitive entries, and calculated the

5 The exception is Estonian (Uralic > Finnic), where the transitive class is smaller than the class of verbs that obligatorily require their object to be in the partitive case. In contrast, verbs like the Estonian equivalent of 'kill', which are annotated as transitive, allow for the partitive encoding of their object but do not require it. This situation highlights the general difficulty of delineating transitive and intransitive verbs in Estonian (and other Baltic Finnic languages).

corrected value of H_{intr} as the average entropy value observed within these subsamples. The subsample size of 24 was chosen because it is the lowest number of intransitive bivalent entries found in any language.[6]

Based on the calculations above, the theoretical limits for H_{intr} range from 0 to the natural logarithm of 24 (≈ 3.18). H_{intr} equals zero in a hypothetical language where all bivalent intransitive verbs belong to the same class, while the maximum value occurs when each bivalent intransitive verb forms its own class. The lowest empirical H_{intr} value, 0.98, is found in Joola-Fonyi, which also has the highest transitivity prominence and the lowest H value, as discussed in Section 3.1. The very low H_{intr} value reflects the homogeneity of Joola-Fonyi's small intransitive set (24 verbs), with 17 belonging to the second-largest class, where the second argument (Y) is marked by the preposition *di*. The highest H_{intr} value, 2.56, is observed in Alik Kryz (Nakh-Daghestanian > Lezgic), which also features an impressive 87 intransitive bivalent verbs and a low transitivity prominence score of 0.30. The bivalent intransitive verbs in Alik Kryz are highly diverse, spread across 30 valency classes, with the largest class containing only 12 verbs and all others having 7 or fewer. This diversity results in an H_{intr} value that reaches roughly four-fifths of the theoretical maximum.

Both raw and corrected H_{intr} values observed in the languages of the sample are provided in the Appendix. Figure 3 visualizes the overall distribution of the (corrected) H_{intr} values.

The distribution in Figure 3 is right-skewed, with a peak shifted towards higher values, showing a gradual ascent and abrupt descent. This suggests that many languages in the sample have bivalent intransitive class systems similar to, though less extreme than, that of Alik Kryz, while systems akin to Joola-Fonyi are relatively rare.

Building on these observations, the next step is to investigate whether there is an empirical relationship between H_{intr} and transitivity prominence, as suggested by the extreme values in Joola-Fonyi and some Nakh-Daghestanian languages. To address this question systematically, I fitted one more linear mixed-effects model with language family as a random effect. The results are summarized in Table 4.

The most important result is a significant negative relationship between transitivity prominence and the entropy of bivalent intransitives (Estimate = –0.927, $p \approx 0.001$). For illustrative purposes, Figure 4 presents the relevant scatterplot, where different colors correspond to language families.

6 In case where the initial set size was 24, the subsampling procedure resulted in identical subsets, as each 'random subsample' was simply a copy of the full initial sample. While this does not affect the integrity of the calculations, it should be noted that subsampling in these instances was not truly random.

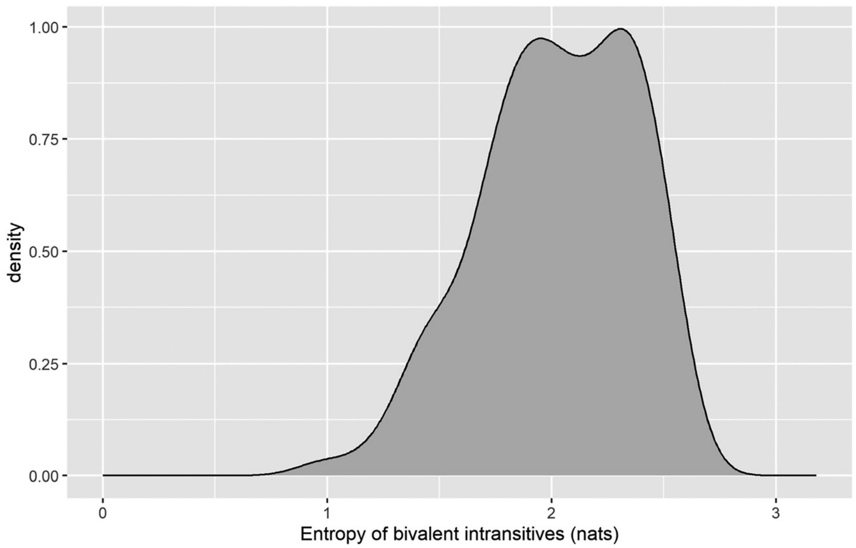

Figure 3: Distribution of entropy of bivalent intransitives (H_{intr}) in the sample.

Table 4: Results of the linear mixed-effects model analyzing the relationship between transitivity prominence and entropy of bivalent intransitives (H_{intr}).

Effect	Estimate	Std. Error	df	t value	Pr(>|t |)
Fixed effects					
(Intercept)	2.390	0.158	54.26	15.14	< 0.001***
Transitivity prominence	−0.927	0.276	108.20	−3.36	0.001**
Random effects					
Variance (Intercept): Family	0.065				
Residual	0.061				

*Indicate significance levels: p < 0.001**, p < 0.001***.

In essence, the observed correlation suggests that if one randomly selects 24 intransitive bivalent lexical entries in a given language, they are more likely to display a complex distribution among valency classes if the language has a lower transitivity prominence value. The link between these parameters is purely empirical; there is no inherent mathematical reason for the observed correlation. The substantial finding is that languages with a more restricted transitive class tend to make finer distinctions between intransitive bivalent classes. This generalization contributes to establishing the valency class system complexity as a fundamental parameter of typological variability, which manifests itself in at least two logically

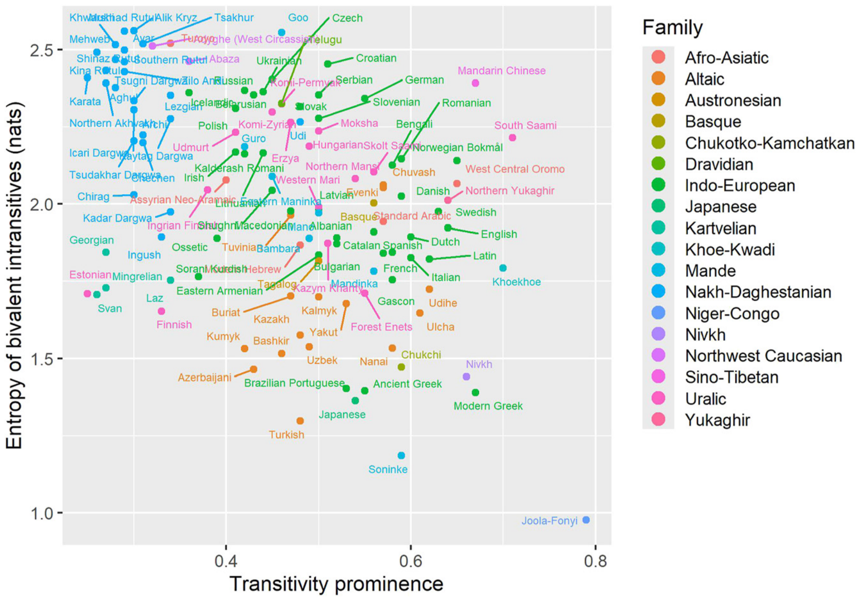

Figure 4: Transitivity prominence and entropy of bivalent intransitives (H_{intr}).

independent phenomena, transitivity prominence and entropy of bivalent intransitives.

3.3 Areal and genealogical distributions

All complexity metrics examined in this study exhibit notable geographic patterns. Figure 5 illustrates the distribution of transitivity prominence across the languages in the sample.

In areas well-represented by the BivalTyp sample, high transitivity prominence values are found in Western and Southern Europe as well as in the Far East, while the lowest values, indicating greater valency class system complexity, are concentrated in Eastern Europe and particularly in the Caucasus. These patterns align with earlier observations noted in the literature (Bossong 1998; Haspelmath 2015: 139–142; Lazard 1994: 63; Say 2014).

Figures 6 and 7 show areal distributions of entropy of bivalent verbs (H) and bivalent intransitives (H_{intr}), respectively. Figure 8 zooms in on the Caucasus, highlighting the distribution of H_{intr} in the area's languages.

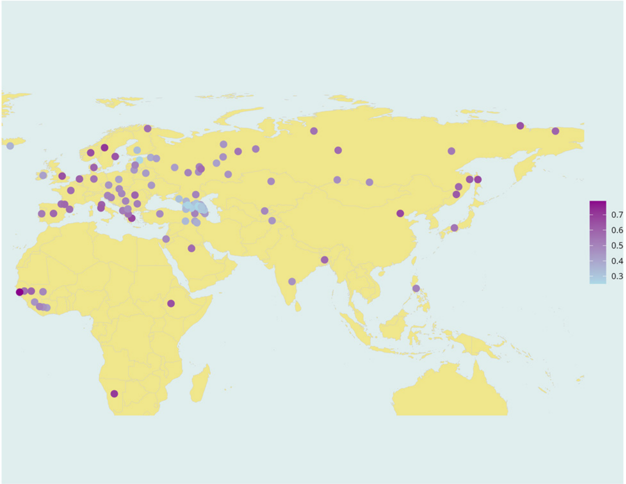

Figure 5: Transitivity prominence in the languages of the sample.

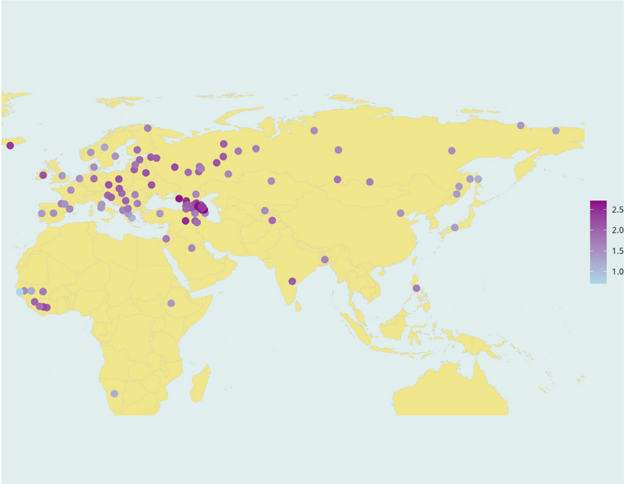

Figure 6: Entropy of bivalent verbs (H) in the languages of the sample.

Given the strong negative correlation between transitivity prominence and the entropy-based complexity metrics, H and H_{intr}, it's unsurprising that their distributions display an areal pattern largely inverse to that of transitivity prominence. The highest H values in Eurasia are found in the Caucasus, followed by a broad area

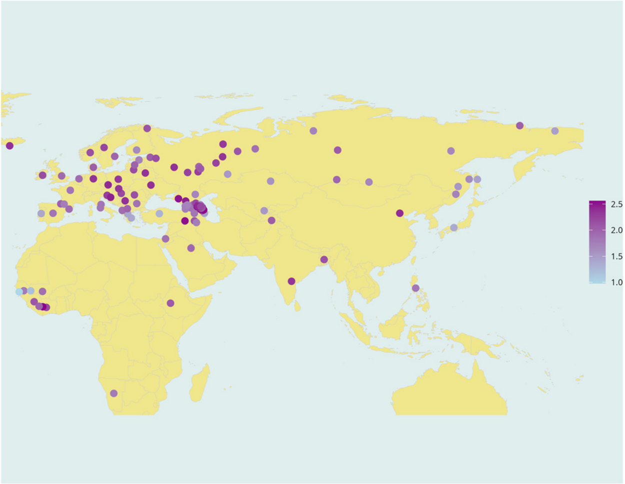

Figure 7: Entropy of bivalent intransitives (H_{intr}) in the languages of the sample.

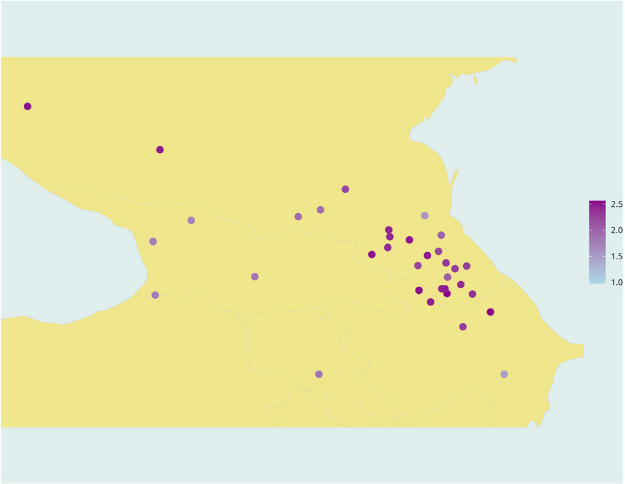

Figure 8: Entropy of bivalent intransitives (H_{intr}) in the languages of the Caucasus.

in Eastern Europe. This area is flanked by Western Europe and Siberia, where H values are low to moderate (Figure 6).

The final complexity metric, H_{intr}, displays a similar, but more nuanced distribution. For instance, Kartvelian languages in the West Caucasus display noticeably lower H_{intr} values than other languages of the Caucasus. German, with an H_{intr} of

2.34, stands well above the sample mean and thus aligns with Eastern Europe, which is dominated by Slavic languages.

The areal patterns in Figures 5–8 suggest that valency class system complexity may serve as a “typological marker”, as defined by Nichols (1992: 185). Nichols uses the term to describe structural features of languages that show “sufficient stability and persistence” (1992: 255), allowing their synchronic distribution to reflect large-scale historical processes. This term is intentionally analogous to genetic markers, which serve a similar purpose in studying biological populations. For instance, the large area in Eastern Europe where languages exhibit moderate to high entropy of bivalent verbs (H) is unlikely to have formed by chance; rather, these affinities between languages likely reflect historical processes in the region.

The key challenge with such distributions is determining the nature of the historical processes behind them. Two main hypotheses are typically considered: common ancestry (genealogy) and language contact (areality). Unfortunately, due to the sample’s size and nature, systematically separating these explanations is difficult. In the remainder of this section, I present anecdotal evidence suggesting that both factors contribute to shaping the observed patterns.

The major evidence for the areal factor comes from the high complexity zone in Eastern Europe. This area cross-cuts two large linguistic families, Uralic and Indo-European. As an illustration, 10 out of the 15 Uralic languages in the sample exhibit H values greater than 1.75. These languages are Udmurt, Komi-Permyak, Komi-Zyrian, Ingrian Finnish, Erzya, Estonian, Hungarian, Moksha, Finnish, and Western Mari. Most of these languages are spoken in Eastern Europe, except for Finnish and Estonian, and even these two languages are known to share many features with their Slavic and Baltic neighbors. By contrast, the remaining five Uralic languages with lower H values, South Saami, Skolt Saami, Northern Mansi, Kazym Khanty, and Forest Enets, are spoken in Northern Europe and Siberia. This geographical divide within the Uralic family does not correspond to genealogical relationships, as it does not align with the Uralic phylogenetic tree. Notably, the positions of the three Ugric languages differ, with Hungarian showing high complexity compared to Northern Mansi and Kazym Khanty.

A similar internal divide is observed within the Indo-European family, but here the more consolidated area is the one showing less complex valency class systems. The lowest H values are seen in Greek, Romance languages, some Germanic, and Albanian, which aligns with the “Standard Average European” Sprachbund (see Say 2014 for a similar pattern in transitivity prominence). Higher H values are found not only in Eastern Europe (Baltic and Slavic) but also in other parts of the European periphery, viz. in Irish and Icelandic (more data are needed for Indo-European languages in Asia). In summary, within both Uralic and Indo-European families, the patterning of complexity reveals an areal signal, highlighting the role of language contact.

It is challenging to prove that genealogical factors independently affect the distribution of complexity metrics. A major difficulty is that genealogically related languages are often spoken in neighboring regions. Therefore, to systematically demonstrate the role of genealogical factors, one would need to find examples where genealogically related languages maintain similar typological characteristics despite being spoken in geographically distant areas, a scenario that is hard to identify with the BivalTyp sample. Nonetheless, intuitively, the BivalTyp data seem to support the hypothesis that linguistic families can exhibit relative consistency in their valency class complexity, thus highlighting the genealogical factor.

Figure 9 presents a boxplot that summarizes the observed H_{intr} values for the four families in the sample with at least 10 languages: Nakh-Daghestanian (24 languages), Uralic (15), Indo-European (38), and Altaic (15).[7]

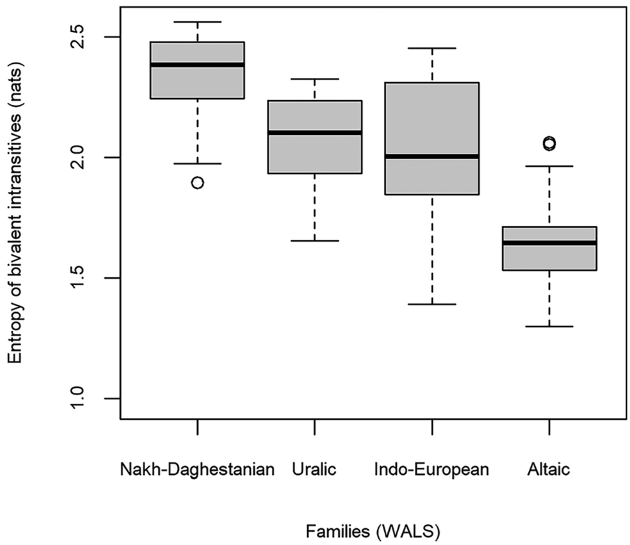

Figure 9: Range of entropy of bivalent intransitives (H_{intr}) across four linguistic families.

7 In this visualization, the boxes represent the interquartile range for each family, encompassing the middle 50 % of the data points. The whiskers extend to the minimum and maximum values within each family, excluding outliers, which are displayed as separate points. The median, shown by the mid bar within boxes, indicates the central tendency for each family.

The hierarchy of the four Eurasian families depicted in Figure 9 is as follows: Nakh-Daghestanian > Uralic (>) Indo-European > Altaic.[8] As previously mentioned, observed differences between these families cannot be conclusively attributed to the inheritance of complexity from a proto-state. It remains theoretically possible that within-family similarities and between-family differences might be explained by areal factors and/or common innovations rather than direct inheritance. Nevertheless, descriptively, some complexity-related patterns can be framed in genealogical rather than areal terms. For instance, Nakh-Daghestanian languages consistently exhibit high values of H_{intr} (> 1.89), surpassing most other languages in the Caucasus, including Kartvelian, Turkic, and Indo-European. Conversely, Altaic languages, encompassing Turkic, Mongolic, and Tungusic are characterized by moderate to low H_{intr} values, despite their extensive geographic range and multiple contacts with languages featuring highly complex valency class systems.

3.4 Entropy and morphological case

The literature suggests that "the degree of differentiation" in valency classes naturally depends "on the number of cases available" (Malchukov et al. 2015: 31). The creators of the ValPal project assume this syntax-to-morphology link, noting, for example, that the "finer coding distinctions among verb classes" in German, compared to English, are "due" to the availability of the dative and genitive cases (Malchukov et al. 2015: 31).

Intuitively, this hypothesis seems plausible: it's reasonable to expect that languages with richer morphological resources would use them to enhance valency-encoding patterns, as reflected in the complexity of their valency class systems (Malchukov 2015: 74). However, to my knowledge, this hypothesis has not yet been tested empirically. This is largely because, while the size of a language's morphological case inventory is a well-documented parameter of typological variation (Iggesen 2013), the complexity of valency class systems has not yet been established as a quantifiable and relevant typological parameter.

The BivalTyp dataset allows for an empirical examination of this hypothesis. Specifically, I analyzed the interaction between two parameters: the entropy of bivalent intransitives (H_{intr}), as introduced in Section 3.2, and the number of nominal morphological cases, as identified in BivalTyp's metadata (these data are available in the Appendix). Defining the number of morphological cases (as opposed to

8 The two-level genealogical classification provided by BivalTyp ("families" and "genera") is largely based on WALS (Dryer and Haspelmath (eds.) 2013), see Appendix for the details.

adpositions, for example) is a complex issue that cannot be fully addressed here due to space constraints. However, it is worth noting that BivalTyp's annotations follow the guidelines from the relevant WALS chapter (Iggesen 2013), with some minor adjustments.[9] Additionally, when exporting the BivalTyp data, I adjusted them for languages without nominal case: in the calculations and visualizations below, they were annotated as having "1" rather than "0" as the value of the parameter discussed (this reflects the number of forms, whereas the usual "0" denotes the number of contrasts in caseless languages, creating an artificial gap for the value of "1").

The interaction between the two parameters is visualized in Figure 10, where dots representing individual languages are colored according to their respective language families.

Figure 10 suggests a connection between the two parameters, as all languages with very large case inventories (>15) display high H_{intr} values. However, this trend is

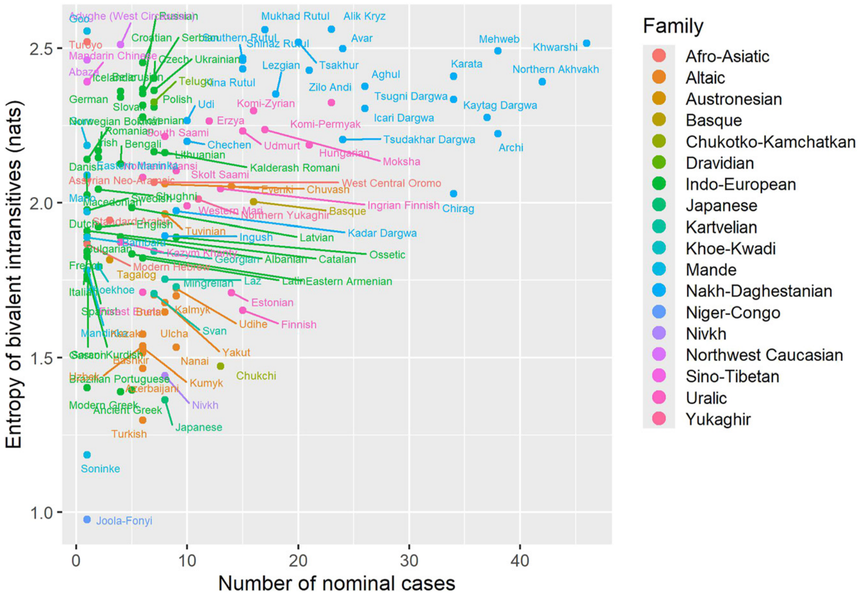

Figure 10: Number of nominal cases and entropy of bivalent intransitives (H_{intr}, in nats).

9 These adjustments include: (i) annotating languages not covered in the relevant WALS chapter, (ii) specifying the exact number of cases in instances where WALS groups possibilities together (e.g., "6–7 case categories"), and (iii) correcting occasional errors in the WALS database, as identified by language experts contributing to the BivalTyp project.

Table 5: Average entropy of bivalent verbs (H) and entropy of bivalent intransitives (H_{intr}) in languages of different nominal case inventory sizes.

	Number of languages	H (nats)	H_{intr} (nats)
No case	26	1.67	1.93
2–3 case categories	7	1.63	1.98
4–5 case categories	10	1.77	1.97
6–7 case categories	24	1.83	1.98
8–9 case categories	17	1.65	1.81
10–15 case categories	14	2.02	2.09
More than 15 case categories	22	2.39	2.34

largely driven by the Nakh-Daghestanian family, evident from the light blue dots in the upper-right corner. No clear relationship emerges for languages with case inventories below 10. Table 5 highlight this, summarizing average H and H_{intr} values across several bins of languages defined by their case inventory size.

Table 5 indicates that languages with large (10–15) or very large (> 15) case inventories may indeed favor complex valency class systems, whereas no clear pattern emerges for languages with smaller inventories. In other words, a rich case system might be a sufficient condition for a complex valency class system, but it is certainly not a necessary condition.

Further investigation with a more diverse sample and better control for confounding factors, including genealogy, is needed. However, it's clear that drawing a direct connection between case inventory size and valency class system complexity, as suggested by comparisons between English and German cited at the beginning of this section, can be overly simplistic. Languages with few or no cases can still develop complex valency class systems, often relying on adpositions or indexes, as seen in Northwest Caucasian.

3.5 Interim summary

The key findings from Sections 3.1–3.4 highlight that entropy-based metrics H and H_{intr} effectively capture significant differences in how languages organize their bivalent verbs into valency classes. Higher values of these metrics correspond to greater complexity, indicating a stronger functional load associated with argument-encoding devices. In contrast, the transitivity prominence indicates more complexity with lower values. Although entropy of bivalent intransitives (H_{intr}) and transitivity prominence are analytically independent, they demonstrate a robust negative correlation, emphasizing the relative complexity of valency class systems as a fundamental typological feature.

Overall entropy (H) serves as a comprehensive metric, capturing both the prevalence of intransitive bivalent verbs and the richness of valency-related distinctions between them. Valency class system complexity is not directly predictable from case inventories, a more familiar parameter in morphological typology. It remains relatively stable across linguistic families, though it can display large-scale areal patterns, indicating convergent processes over long periods.

4 Entropy and word order

4.1 Basic word order and entropy of valency classes

It is widely recognized that languages with a basic SVO order often lack morphological case and SOV languages favor the presence of a case system, a generalization dating back to at least Greenberg (1966: 96). While subsequent studies rejected the categorical universal, the idea still persists in recent quantitative studies (Maion 2018; Sinnemäki 2010). In fact, the data from BivalTyp also align with the observation that SVO languages favor the lack of morphological case.[10] Additionally, in languages where the presence of overt flagging of core arguments interacts with word order, verb-medial patterns allow for zero marking of arguments more freely than verb-edge patterns (Seržant et al. 2025). SVO languages have also been claimed to lack agreement morphology more often than other languages (Siewierska and Bakker 1996: 125). Overall, SVO languages are often seen as morphologically simpler than, e.g., SOV languages (Sinnemäki 2010).

The usual explanation for these generalizations is based on the idea that word order and overt morphological marking serve as cues for distinguishing arguments in speech production and perception. Consequently, the SVO pattern reduces the need for overt morphology, as the linear position of core arguments already facilitates their differentiation.

Given the widely accepted correlation between SVO word order and morphological simplicity, one might expect SVO languages to have less complex valency class systems – at least if one views morphological complexity as a driving factor for valency class system complexity; however, see Section 3.4 for a critical evaluation. In fact, Müller-Gotama (1994: 142–144) proposes a hypothesis suggesting that right-branching languages tend to have a wider range of subjects and objects, that is, have a higher ratio of clauses where arguments are encoded as canonical subjects and objects (see also a brief discussion in Haspelmath 2015: 132). However, Müller-

10 Among the 40 SVO languages in the sample, 15 (37 %) lack case marking, whereas among the 69 SOV languages, only 10 (14 %) are also caseless.

Table 6: Results of the linear mixed-effects model analyzing the relationship between word order and entropy of bivalent verbs (H).

Effect	Estimate	Std. Error	df	t value	Pr(>|t |)
Fixed effects					
(Intercept)	1.686	0.122	14.54	13.78	< 0.001***
Word order: SVO	0.081	0.084	102.40	0.96	0.34
Random effects					
Variance (Intercept): Family	0.197				
Residual	0.067				

*Indicate significance levels: p < 0.001**, p < 0.001***.

Gotama's findings were based on a small sample of 12 languages, and the observed contrasts were not particularly sharp. Nevertheless, his hypothesis predicts a higher transitivity prominence and, consequently, lower entropy of valency classes in SVO languages.

To test this hypothesis, I fitted two linear mixed-effects models, using word order (SOV vs. SVO; excluding the few languages with other annotations) as a fixed effect, and accounting for linguistic family as a random effect (the basic word order patterns in the languages of the sample, as identified in BivalTyp, are available in the Appendix). H and H_{intr} were used as dependent variables. Neither model showed a significant effect of word order. As an example, Table 6 presents the results from the model where H was the predicted variable.

Essentially, the results obtained here imply that there is no systematic difference between SVO and SOV languages in the number of valency-determined contrasts they make among bivalent intransitive verbs. Upon initial examination, these findings challenge the conventional assertion that SOV languages favor the presence of case distinctions (see references above). However, I hypothesize that there is, in fact, no contradiction. The balanced overall picture, with comparable entropy values in SOV and SVO languages, is shaped by two opposing factors. On the one hand, there is a greater need to morphologically distinguish A and P arguments in SOV languages, favoring the presence of at least some flagging contrasts. This aligns with available literature on the topic, briefly discussed earlier in this section, and also with the BivalTyp data, where SVO languages lack case altogether more often than SOV languages. On the other hand, there are indications suggesting that more specialized valency-encoding devices, and hence more complex contrasts, are favored by postverbal arguments compared to preverbal arguments. While I do not propose a systematic method to conclusively prove this last assertion, I observe some supporting evidence, which I discuss in the next section.

4.2 Postverbal positions accumulate more complexity in argument-flagging

In this section, I briefly outline four pieces of evidence supporting the overarching claim that postverbal positions exhibit more varied argument-flagging systems compared to preverbal positions. These observations are drawn from available literature and are not intended to introduce new findings. However, their combined impact lies in their potential to highlight the relative informativeness of pre- versus postverbal argument-coding devices.

The first piece of evidence is based on two trivial observations. i) Globally, subjects tend to occupy the clause-initial position, typically preceding verbs, and especially objects. This can be succinctly illustrated by the prevalence of SVO and SOV as the two most common types, followed by VSO languages (Dryer 2013; Hawkins 1983; Tomlin 1986). ii) Subjects exhibit the highest degree of role neutralization, while grammatical relations lower in the hierarchy exhibit more restrictions regarding semantic roles (Kibrik 1997; Van Valin and LaPolla 1997: 250–263).

These two assertions oversimplify the issue: in reality, both are at the center of empirical and theoretical debates, starting with the enduring question of whether subjects and grammatical relations are universal. However, when taken together, they suggest a universal pressure for positioning the argument with the least transparent semantic role before the verb. While there is no definitive evidence for the preference of post-verbal position in the case of semantically specific arguments, the observation above provides significant support for less variability in coding preverbal arguments compared to postverbal arguments.

The second piece of evidence involves non-canonical A- and P-arguments (as discussed in Section 2.1, the labels X and Y used in the BivalTyp annotations largely correspond to what is referred to as A and P in some traditions). Typically, non-canonical arguments share some behavioral properties, often including word order, with their canonical counterparts in basic transitive constructions such as *the boy* (A) *broke a stick* (P), while differing in overt indexing or flagging (Cole et al. 1980; Haspelmath 2010). This suggests that non-canonical A's usually appear at the beginning of clauses, whereas non-canonical P's usually are not placed there but can occupy pre- and post-verbal positions with comparable ease across languages.

Both non-canonical A's and P's can display various flagging patterns, including case forms like datives, locatives, or benefactives. As a result, non-canonical A's add to the diversity of flagging choices in the preverbal domain. However, a noticeable imbalance emerges: non-canonical A's are typically much less varied than non-canonical P's (Bickel et al. 2014: 496–500; Say 2018: 565–566). This difference is evident in straightforward calculations based on the BivalTyp data: on average, languages in the dataset feature 4 distinct coding options for X-arguments (ranging between 1 and 12 options per language) and 13.5 options for Y-arguments (ranging between 4 and 22 options per language). In all languages of the sample, the number of coding options for Y is either equal to, as in Ingush (Nakh-Daghestanian > Nakh) and Soninke (Mande > Western Mande), or greater than that for X. This disparity indirectly supports the idea that varied flagging is more common in postverbal arguments than in preverbal ones.

The third piece of evidence relates to the linear position of non-core clause-level constituents, sometimes collectively represented as X, distinct from canonical subjects and direct objects (not to be confused with BivalTyp's use of X for the more agentive of the two arguments of the verb). In Section 4.1, I observed that there is no clear difference between VO and OV languages in terms of average entropy values (H and H_{intr}). Given that VO and OV languages are often interpreted as head-initial and head-final languages respectively, one might expect that the preferred linear position of non-core constituents would not consistently correlate with H or H_{intr} either. However, such an assumption might not hold true, as the preferred linear position of non-core constituents relative to the verb does not always align with that of O-arguments. Importantly, the two non-harmonic orderings are not equally common across languages: OVX patterns are significantly more widespread than XVO patterns (Dryer and Gensler 2013; Hawkins 2008: 170).

In this regard, the prevalence of languages exhibiting the OVX pattern supports the hypothesis that postverbal positions tend to accommodate a wider array of argument-encoding devices. This phenomenon is attested in several West African languages in the BivalTyp sample, as shown in (10a) and (10b).

(10) Bambara (Mande > Western Mande; BivalTyp)

a. *Sékù ye nàmasa` dún*
PN PFV.TR banana eat
'Seku ate a banana.'

b. *Sékù b'í túlomajɔ̀ àrajɔ` fɛ̀*
PN IPFV.REFL listen radio\ART by
'Seku is listening to the radio.'

In (10a), the pattern is transitive. Remarkably, the preverbal O argument lacks flagging, representing the only encoding option for preverbal NPs. In (10b), the postverbal argument is flagged by the postposition *fɛ̀* 'by', just one of several flagging choices for postverbal arguments. Such contrasts are characteristic of OVX languages. While the reverse XVO pattern is possible (seen in some Sinitic languages, for instance), it remains typologically uncommon.

All the evidence presented thus far concerned default word order patterns. The last piece of evidence emerges from dislocation phenomena. When an NP is dislocated, its flagging can change. Importantly, there is anecdotal evidence highlighting left-right asymmetries in such patterns. For example, Lambrecht (2001: 1069–1070) mentions a possible contrast between clause-fronted topics, lacking overt encoding of their thematic relation to the verb, and antitopics, which retain the usual flags, as shown in (11a) and (11b).

(11) Occitan (Indo-European > Romance; Lambrecht 2001: 1070)

a. *lo cinema, i vau sovent*
the cinema there I.go often
'The movies, I go there often.'

b. *i vau sovent, al cinema*
there I.go often to.the cinema
'I go there often, to the movies.'

As a side-effect of these patterns, expected coding contrasts are neutralized in fronted NPs. The prevalence of such phenomena typologically remains to be explored. Nevertheless, word order variations triggered by information structure could be relevant for the diachronic development of encoding differences between pre- and postverbal NPs. This can be exemplified by a constraint observed in northeastern Africa, referred to as the "no case before the verb" rule, or more appropriately, the "no case distinction before the verb" rule (König 2008: 240). This constraint manifests in various forms across the region, but it is most evident when a language permits alternative word order patterns, such that a certain argument is explicitly case marked when used after the verb but lacks case markers when used before the verb, as shown in (12a) and (12b).

(12) Päri (Eastern Sudanic > Western Nilotic; König 2008: 240, originally from Andersen 1988)

a. *joobi a-keel uburr-ì*
buffalo COMPL-shoot PN-ERG
'Ubur shot the buffalo.'

b. *ubur joobi a-keel-e*
PN buffalo COMPL-shoot-3SG.A
'Ubur shot the buffalo.'

As König hypothesizes (2008: 271–273), the absence of case markers in sentences like (12b) can be attributed to the grammaticalization of clefted constructions, initially associated with focusing the NP in the sentence-initial position; this dislocation scenario also accounts for the use of the A-index cliticized to the verb in examples like (12b). Regardless of their precise pathways of diachronic development, contrasts like those between (12a) and (12b) contribute to the greater encoding diversity in postverbal arguments as compared to preverbal arguments.

4.3 Interim summary and discussion

An exploration of how the complexity of valency encoding devices interacts with word order revealed that while there is no systematic disparity between OV and VO languages in terms of the overall complexity of their argument-encoding systems (4.1), postverbal arguments tend to attract more complex argument-encoding devices (4.2).

These two conclusions may initially appear contradictory. However, each is independently driven by processing-related functional pressures. On the one hand, (S)OV languages favor at least some overt flagging of core arguments compared to (S)VO languages, where the need for morphologically disambiguating subjects and objects is less urgent. Given that languages universally favor placing all non-subject dependents on the same side of the verb (Hawkins 2008: 170–171, 185), this tendency fosters more argument-encoding complexity in (S)OV languages. On the other hand, non-core dependents, commonly labelled "X" in word order studies, display greater coding diversity, which makes their processing more demanding before the specific verb lemma has been produced. Hawkins suggests this tendency as a tentative explanation for the universal preference to place O arguments before X (Hawkins 2008: 186). This generalization also supports the higher argument-encoding complexity in (S)VO languages: in these languages, O and X are almost always on the same side of the verb, increasing the need for differentiation through overt morphological means. In contrast, (S)OV languages can use (S)OVX patterns, reducing the need to distinguish X from O.

Thus, there are two counteracting tendencies that contribute to the observed equilibrium between (S)OV and (S)VO languages in terms of their argument encoding system complexity.

5 Further tentative correlations

In this section, I further explore the relationships between entropy in valency class systems and other typological parameters. Due to several factors, some of which are discussed in the following subsections, none of the connections identified here can be considered definitive or based on sufficient empirical evidence. Instead, these are working hypotheses and preliminary observations that require further testing in future research. Nonetheless, the primary aim of this section is to highlight the multiple potential links between valency class complexity, as measured by entropy-based metrics, and other typologically relevant parameters. The recognition of these links, even if they require further refinement, supports the broader goal of establishing valency class complexity as a fundamental and measurable parameter in typological research. All the data used in the following subsections, including additional annotations, are available in the Supplementary materials.

5.1 Non-verbal predicates

The BivalTyp database does not require that all language-specific predicative expressions consist of morphologically simplex verbs, cf. Haspelmath and Hartmann (2015: 58–61). Many entries feature non-verbal predicates, often accompanied by a copula, as illustrated in (13).

(13) Shinaz Rutul (Nakh-Daghestanian > Lezgic; BivalTyp)

Basir-ɨs	*k'vač'*	*ij*	*Karam*
PN-DAT	detestable	COP	PN

'Basir hates Karam.'

The distinction between verbal and non-verbal predicates is accessible for only 73 languages among the 120 languages used in this study. These languages can be referred to as "published" languages, meaning they feature interlinearized sentences, not just valency patterns, in the online version of BivalTyp. The remaining 47 languages ("unpublished" languages) so far lack these annotations. However, in both groups of languages, argument-encoding devices in non-verbal constructions are annotated using the same principles as in verbal constructions. For instance, the valency pattern of (13), schematically represented as DAT_NOM, is also found in various simplex verbal predicates in Shinaz Rutul, such as the equivalents of 'know', 'see' (14), and 'love'.

(14) Shinaz Rutul (Nakh-Daghestanian > Lezgic; BivalTyp)

basir-ɨs	*hagu-j*	*χal*
PN-DAT	[N]see.PF-AOR.N	house(N)

'Basir sees a house.'

The occurrence of non-verbal predicates varies widely across the published languages in the sample, ranging from just one entry in Mano (Mande > Eastern Mande) and Khoekhoe (Khoe-Kwadi) to over 30 entries in Irish (Indo-European > Celtic) and Alik Kryz (Nakh-Daghestanian > Lezgic). On average, there are 9.7 entries per language, with a median of 8.

More importantly, verbal and non-verbal predicates exhibit distinct distributions across valency classes. Non-verbal predicates are almost never transitive in the published part of the dataset, with only three entries identified as transitive. This low number aligns with expectations from the literature (see Dixon 2004: 5 for the general tendency against transitive non-verbal predicates and Lowe 2017 for a discussion of exceptions). Given this, it is natural to expect that languages with a prevalence of non-verbal predicates will exhibit greater valency class system complexity. For illustration, Figure 11 visualizes the relationship between the prevalence of non-verbal predicates and H_{intr}.

Observationally, the expected relationship between the two parameters is present: all languages with a prevalence of non-verbal predicates above 0.1 have H_{intr} values exceeding 1.75, indicating moderate to high valency class complexity.

However, proving these connections has been challenging. With only 73 published languages and a skewed genealogical distribution of non-verbal predicate prevalence, this parameter failed to show a significant effect on any of the three valency class complexity metrics once family was included as a random effect in the linear mixed models. While the intercept had the expected sign in all cases (positive for entropy-based measures, negative for transitivity prominence), none of the p-values were below 0.05. For now, the link between non-verbal predicates and valency class complexity remains a promising hypothesis, pending further testing with more genealogically balanced data.

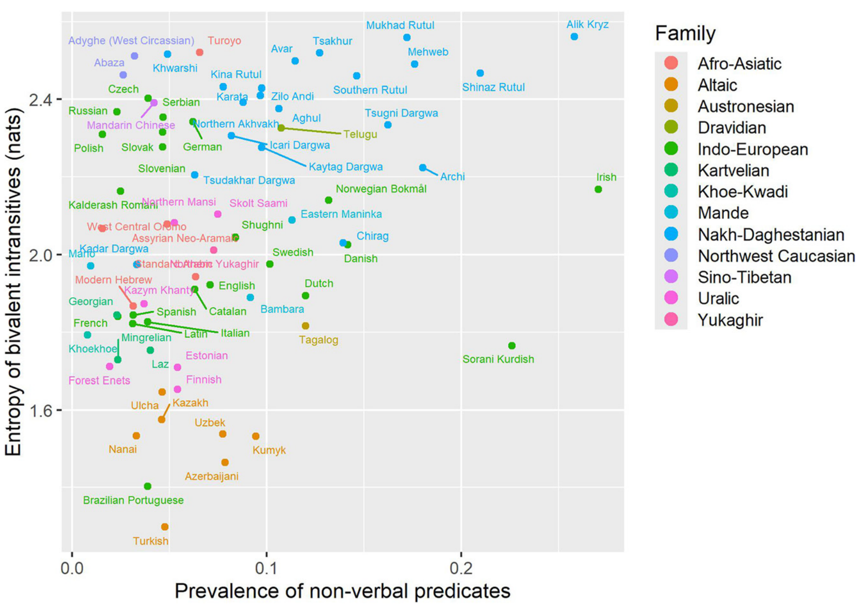

Figure 11: Prevalence of non-verbal predicates and entropy of bivalent intransitives (H_{intr}).

5.2 Preverbs

It is natural to anticipate some covariance between a verb's morphological structure and its valency pattern, as discussed extensively by Michaelis and Ruppenhofer (2001) in a German-focused context, see also Wood and Myler (2019) for a cross-linguistic overview from a formalist perspective. Although systematically investigating this covariance typologically poses significant challenges due to the versatility of verb structures and the lack of obvious *tertia comparationis*, the BivalTyp sample, comprising many languages of Europe and the Caucasus, suggests a positive correlation between the complexity of valency class systems and the prevalence of valency-related verbal prefixes.

Some evidence supporting this generalization comes from two Northwest Caucasian languages in the BivalTyp sample: Adyghe (West Circassian) and Abaza. Both languages exhibit high values of H_{intr}, ranking 7th and 11th in the 120-language sample. In both languages, the role of flagging is minimal, and the valency class of a verb is predominantly determined by the structure of "preverbs". Preverbs are prefixal elements associated with specific argument-indexing slots within the verb. Importantly, the system of preverbs encompasses a few dozens distinct elements, many of which feature in the BivalTyp dataset. In (15), for instance, the second argument is indexed in the slot introduced by the benefactive preverb *z*-. Thus, the preverb simultaneously affects the stem structure and the verb's valency pattern.[11]

(15) Abaza (Northwest Caucasian; BivalTyp)

Murad	*Fatima*	*d-lə-z-qʷə́c-əj-ṭ*
PN	PN	3SG.H.ABS-3SG.F.IO-BEN-think-PRS-DCL

'Murad is thinking about Fatima.'

Given the abundance of applicative preverbs in Northwest Caucasian languages (Arkadiev et al. 2024), it is unsurprising that these languages also display very complex valency class systems.

Further evidence arises from some differences within the Nakh-Daghestanian family. Some Nakh-Daghestanian languages showcase "correlation" between verbal prefixes and case markers, illustrated in (16).

11 In BivalTyp, the valency pattern of (15) is labeled as ABS_BEN, indicating that X, the thinking person, is encoded as a regular absolutive argument, whereas Y, the person being thought about, is indexed in a slot associated with the benefactive marker *z*-. Note that the marker *z*-, glossed as BEN, is best analyzed as part of the verb stem.

(16) Kryz (Nakh-Daghestanian > Lezgic; Authier 2009: 126)
yaba *xhin.a-k* *ke-b-cin*
fork.F hay-SUB SUB-F-stick.IMP
'Stick the fork in the hay!'

In (16), the preverb *ke-* "correlates" with the case marker attached to the second argument in the sense of sharing the same origin and displaying semantic similarity. Within Nakh-Daghestanian, such phenomena are observed in most Lezgic languages, except Udi and Archi, and in some Dargwic varieties (Gilles Authier, p.c.). Significantly, the Lezgic branch, with its 10 languages in the BivalTyp dataset, shows the highest H_{intr} values among the four Nakh-Daghestanian genera represented. Notably, Udi and Archi have the lowest H_{intr} values among these 10 languages.

The "correlation" between preverbs and case in Nakh-Daghestanian bears resemblance to the "oblique registration" patterns observed in certain Indo-European languages, notably in Slavic and Baltic (Zúñiga et al. 2024), as well as in some Germanic languages. In these languages, the formal correlation between prefixes and prepositions may be sometimes disrupted, but prefixes are abundant and contribute to the selection of valency patterns, especially in verbs related to motion. And again, these languages tend to display more complex valency class systems compared to branches where prefixation plays a relatively minor role in argument encoding, such as the Romance languages.

To sum up, a typological correlation emerges between the extensive use of (originally spatial) prefixes within verbs and the complexity of valency class systems. While systematic quantitative evidence is currently lacking, this hypothesis, if correct, can be naturally accounted for by diachronic considerations. Verbal prefixes and nominal flags can sometimes share a common origin, as seen in Balto-Slavic (Arkadiev and Kozhanov 2023: 13; Kuryłowicz 1964: 171–178) and Nakh-Daghestanian languages (Arkadiev and Maisak 2018: 130), and evolve non-spatial meanings through similar semantic processes. Hence, it is reasonable to expect richer flagging systems in languages that frequently employ valency-related preverbs.

5.3 Satellite-framed versus verb-framed languages

Languages exhibit significant variation in how they "pack" motion events into their lexical and syntactic units. Following Talmy's influential work (e.g., Talmy 2000), languages are often classified into two main types: satellite-framed and verb-framed. In satellite-framed languages, the verb conflates motion with the manner component, while paths are expressed by satellites (17). Conversely, in verb-framed languages, the verb conflates motion with the path component, with the manner of motion optionally specified by a prepositional phrase or adverbial clause (18).

(17) English (Indo-European > Germanic; Beavers et al. 2010: 333)
John limped into the house.

(18) French (Indo-European > Romance; Beavers et al. 2010: 333)

Je	*suis*	*entré*	*dans*	*la*	*maison*	*en*	*boitant.*
I	am	entered	in	the	house	in	limping

lit. 'I entered the house limping.'

Subsequent studies have shown that Talmy's original dichotomy is overly simplistic, particularly its categorization of entire languages rather than specific constructions (Beavers et al. 2010; Croft et al. 2010). Despite these nuances, I hypothesize that, all else being equal, satellite-framed languages tend to display more complex valency class systems as compared to verb-framed languages.

Presenting robust evidence to support this hypothesis is challenging, primarily due to the apparent scarcity of empirical typological studies systematically classifying a substantial number of languages into satellite- and verb-framed types. Nevertheless, in Table 7 below, I list several languages commonly considered representatives of the two types in the literature and provide their H_{intr} values, calculated using the BivalTyp data.[12]

While the data in Table 7 do not lend themselves to statistical testing, observationally, they do support the generalization that satellite-framed languages typically have more complex valency class systems compared to verb-framed languages: all values in the left-hand column exceed 1.87, while most languages in the right-hand column have lower values.

This generalization also receives support from the substantial characteristics of satellite- versus verb-framed constructions. In verb-framed constructions, the spatial relationships between the trajector and the landmark are by definition encoded within the verb. Therefore, encoding this information outside the verb may be redundant, and using less specialized flags with the noun phrase does not necessarily reduce the overall propositional content. In an extreme scenario, the clause may even display the basic transitive pattern, as in (19).

12 Talmy (2000: 222) provides a somewhat longer list but some of his groups, such as "Indo-European minus Romance" or "Semitic", are very broad. Consequently, gathering substantial language-specific evidence at this level of granularity is barely possible.

Table 7: Entropy of bivalent intransitives (H_{intr}) in presumed satellite- versus verb-framed languages. The literature on the topic often portrays cross-linguistic variation in broad strokes, mentioning entire genera, such as Romance or Germanic. Exceptions are sometimes made for 'major' languages, which are discussed individually. Reflecting this trend, the rows in Table 7 are also heterogeneous, with some presenting data for individual languages and others for entire genera.

Satellite-framed	H_{intr}	Verb-framed	H_{intr}
English	1.92	French	1.84
German	2.34	Spanish	1.84
Other Germanic	1.89–2.36	Other Romance	1.40–2.15
Russian	2.37	Turkish	1.34
Other Slavic	1.87–2.45	Japanese	1.36
Mandarin Chinese	2.39	Hebrew	1.86

(19) French (Indo-European > Romance; Barnabé 2016)

Elle	*a*	*traversé*	*la*	*rivière*	*à*	*la*	*nage.*
she	has	crossed	the	river	in	the	swim

'She swam across the river.'

In satellite-framed constructions, the path component should be specified in semantically specialized verb-external markers, including markers syntactically linked with the landmark noun phrase, such as the preposition *into* in (17). The variation of required dependent markers in satellite-framed constructions becomes especially apparent when multiple path components are combined as satellites to the same verb, as illustrated in (20). This contrasts with verb-framed languages, which "usually do not allow for accommodating multiple paths around a single verb (each path segment is typically expressed in a separate verb)" (Lewandowski 2018: 48).

(20) English (Indo-European > Germanic; Lewandowski, p.c.)
He ran out of the house through the yard down to the river.

Although languages are typically classified as satellite- or verb-framed based on lexicalization patterns, it is now clear that they also exhibit syntactic differences: satellite-framed languages tend to feature more intricate systems of spatial flags. To some extent, this fact can directly explain the differences in the H_{intr} values observed in Table 7, as the BivalTyp questionnaire covers several motion constructions (specifically, questionnaire entries #12 'enter', #14 'go out', #21 'reach', #65 'fall behind', #68 'cross', #74 'leave', and #90 'dismount'). However, I speculate that the same underlying differences transgress the limits of motion constructions and manifest in other constructions surveyed in the BivalTyp questionnaire, see (Stathi 2023) for a similar hypothesis on a slightly different material. This transgression can be attributed to processes of semantic extension, whereby spatial expressions get involved in argument encoding. Essentially, I hypothesize that nominal spatial expressions play a comparatively minor role in consistently verb-framed languages, contributing to their simpler valency class systems.

While this hypothesis cannot be statistically tested, there is anecdotal evidence supporting it, for example, in Turkic languages. These languages consistently exhibit lower entropy values compared to many other languages of Northern Eurasia (see Section 3.3). One possible explanation lies in how spatial markers are organized in Turkic languages. Turkic case forms typically convey minimal orientation contrasts such as between the Ablative for motion from or through a landmark, the Locative for static position relative to the landmark, and the Dative/Lative for motion towards the landmark. These generalized cases are sometimes used to encode arguments in non-spatial verbs; for example, Azerbaijani uses the Ablative case for verbs like 'be afraid', 'depend', and 'enjoy'.

In contrast, specific localizations in Turkic are usually expressed through secondary postpositions, as in Turkish examples such as *kadar* 'up to, until', *doğru* 'towards, in the direction of', *üstünde* 'above', *boyunca* 'along', or *yanında* 'next to, near'. Importantly, these postpositions almost never serve to encode arguments with non-spatial verbs, as shown by datasets from Turkish, Kazakh, Uzbek, and other Turkic languages in BivalTyp. The limited polysemy and weak grammaticalization of secondary spatial markers in Turkic languages can be attributed to the prevalence of verb-framed constructions. As is typical of verb-framed languages, Turkic languages compensate for the low frequency and generally minor role of specialized spatial markers by extensively employing spatial verbs, often combined in complex chains.

These limitations contrast with Slavic prepositions and Northwest Caucasian preverbs, which often have primary spatial meanings, sometimes very specific, but are also used to encode arguments for various non-spatial verbs. This is illustrated by examples (21) and (22) from Czech.

(21) Czech (Indo-European > Slavic; BivalTyp)

Petr	*vyhrá-l*	*nad*	*Michal-em*
PN(M)[NOM.SG]	win(PFV)-PST[M.SG]	above	PN(M)-INS.SG

'Petr beat Michal.'

(22) Czech (Indo-European > Slavic; BivalTyp)

Petr	*se*	*stýská*	*po*	*Michal-ovi*
PN(M)[NOM.SG]	REFL.ACC	miss(IPFV).PRS.3SG	along/after	PN(M)-LOC.SG

'Petr misses Michal.'

In summary, languages favoring satellite-framed constructions arguably tend to exhibit more intricate valency class systems. One possible explanation is that the prevalence of satellite-framed constructions stimulates the grammaticalization of spatial markers into semantically more abstract argument-encoding devices.

5.4 Colexification patterns

In discussions of cross-domain complexity trade-offs, researchers have examined the potential role of lexical homophony (Ke 2006). Fenk-Oczlon and Fenk (2008) regard homophony as a sign of semantic complexity, arguing that processing semantically ambiguous elements demands additional effort. They also provide evidence for a trade-off between morphological complexity and semantic complexity, as operationalized by homophony. However, they acknowledge alternative viewpoints regarding the contribution of lexical homophony to semantic complexity (see below).

Building on the overarching hypothesis that languages with lower morphosyntactic complexity tend to exhibit higher semantic complexity, such trade-offs might be expected in the realm of valency classes. To test this, both potential correlates must be empirically operationalized. For morphosyntactic complexity, I utilized the entropy of valency classes, as discussed above. However, assessing the degree of semantic complexity in individual languages poses challenges due to the lack of typological resources providing data on the prevalence of homophony across a large language sample. Therefore, I employed a section of the BivalTyp dataset as a proxy. Specifically, I analyzed datasets from 70 languages annotated for lexical predicative expressions (simplex verbs, complex verbs, and non-verbal predicates). This subset will be referred to as “lexically annotated languages”. For these languages, I identified instances where a single lexical item appeared in distinct entries, as illustrated in (23a–b).

(23) Eastern Maninka (Mande > Western Mande; BivalTyp)

a. *Sékù bára bɔ́ à lá só̀ lá*
PN PRF exit 3SG POSS village\ART at
‘Seku left his village.’

b. *Sékù bára à lá dìrikì bɔ́*
PN PRF 3SG POSS shirt\ART exit
‘Seku has taken off his shirt.’

The main verb in both sentences is the verb *bɔ́* ‘exit’. However, due to the use of different valency patterns, they convey clearly different predicative meanings. I labeled such instances as "colexifications" to sidestep the notoriously difficult differentiation between polysemy and homophony.[13]

Subsequently, I calculated the colexification index C by dividing the number of observed colexification instances by the total number of filled entries in a language. The initial hypothesis posited a negative correlation between C as a measure of semantic complexity and H as a measure of morphosyntactic complexity. However, in the linear mixed-effects model with family included as a random effect, no significant effect of C on H was found ($p = 0.169$). Notably, the intercept for C was positive (1.716), indicating a positive association, although not statistically significant.

Several factors may contribute to the lack of significant correlation, including the sample size of 70 lexically annotated languages and the inherently limited coverage of polysemy/homophony in the BivalTyp data. I hypothesize that with a larger sample, a significant positive correlation between C and H could emerge, particularly driven by languages predominantly characterized by simplex, monomorphemic verb stems. A few such languages are present among the lexically annotated languages, and they appear to support this expectation. Eastern Maninka (Mande > Western Mande) represents one extreme case, with a C value of 0.15, the highest in the sample, and an H value of 1.98, the highest among lexically annotated African languages. As an illustration, the verb *bɔ́* 'exit', exemplified in (23a–b), is also found in the database entries corresponding to meanings such as 'resemble', 'go out', 'depend', 'be different', and 'cover' and occurs in a total of four distinct valency patterns. Eastern Maninka stands in stark contrast to languages like Khoekhoe (Khoe-Kwadi), which lacks any colexification patterns in the dataset and has a very low H value of 1.15.

As mentioned earlier in this section, Fenk-Oczlon and Fenk (2008) viewed homophony as a case of semantic complexity and hypothesized its negative correlation with morphosyntactic complexity, framing it as a trade-off. My hypothesis suggests the opposite, viz. greater valency class complexity (a subtype of morphosyntactic complexity) in languages with a higher prevalence of colexifications (a phenomenon similar to homophony). Interestingly, this still aligns with the notion of a cross-domain trade-off. Both verbs and argument-encoding devices offer cues for clause interpretation. It seems plausible that languages with complex argument-encoding systems are more tolerant of underspecified verb lexemes compared to those with simpler argument-encoding systems.

13 However, I refrained from classifying pairs of dataset entries as colexifications when they shared both the verb and the valency pattern. This approach aimed to prevent the potential overestimation of "colexifications" stemming from instances where language experts were not able to provide non-identical translations for semantically nuanced entries such as 'hate' and 'despise'.

In short, my approach follows the general principle that "more distinctions indicate higher degrees of complexity" (Shcherbakova et al. 2023: 157; see also Sinnemäki 2009), thus avoiding the view that underspecification is a marker of complexity. Ultimately, this debate is terminological. The key substantive finding is the absence of a negative correlation between the prevalence of colexifications and valency class complexity. Instead, a potential positive correlation may be suggested, particularly in languages with reduced morphology, where argument-encoding devices and verbal roots appear to play a central role in expressing predicative meaning, though this trend has yet to be statistically confirmed.

5.5 Summary of hypotheses

Sections 5.1–5.4 contained a discussion of several typological features that may correlate with the complexity of valency class systems. These hypotheses are summarized in Table 8.

The four hypotheses summarized in Table 8 all require further quantitative research for validation. Moreover, even if they are confirmed, this study does not specify the direction of the causal relationships between the hypothesized correlates. Additionally, it is important to note that the four parameters summarized in Table 8 may be interrelated, potentially leading to spurious correlations with valency class system complexity. To detect this, the analysis should control for potential collinearity effects (e.g., by way of centering the variables), which is presently not feasible (due to the scarcity of data, the fact that verb-framing is a categorical rather than a numerical variable, etc.).

The possible correlations summarized in Table 8 can be subsumed under a broader denominator. If the correlations are valid, it appears that languages with simpler valency class systems tend to place a heavier communicative burden on their verbal roots. An ideal "verby" language would enjoy a wide repertoire of semantically specialized verbs, reducing the need for a elaborate system of non-verbal markers, including argument-encoding devices. Conversely, complex valency class

Table 8: Possible typological correlates of the complexity of valency class systems.

	High entropy	**Low entropy**
Prevalence of non-verbal predicates	High	Low
Prevalence of preverbs	High	Low
Verb-framing	Satellite-framed	Verb-framed
Prevalence of verb colexifications	High	Low

systems are likely associated with languages that prioritize nouns and other elements outside the verb itself. This differentiation can be linked to the overarching contrast between verb-based and noun-based languages; for further insights, see Polinsky (2012) for a quantitative-typological perspective and Qiu and Winsler (2017: 275–276) for an overview of psycholinguistic research. However, I defer this avenue of research for future investigation.

6 Discussion

This paper aims to introduce the concept of valency class system complexity and propose quantitative metrics for its assessment. While transitivity prominence has been previously discussed in the literature, the metrics of entropy of valency classes (H) and entropy of intransitive bivalent verb classes (H_{intr}) are novel. These three metrics were applied to a dataset from the online database BivalTyp, encompassing information on valency patterns of 130 bivalent predicates across 120 languages. Bivalent verbs were chosen to explore valency class system complexity because they often exhibit non-standard valency behavior and fall into multiple valency classes (Bickel et al. 2014). However, it remains to be seen whether these results extend to monovalent and trivalent verbs.

My main conclusion is that the metrics utilized in this study effectively capture a fundamental typological property that has been largely overlooked in existing research. Despite their logical independence, transitivity prominence and H_{intr} display a strong negative correlation. This suggests that these two metrics reflect different aspects of a broader underlying parameter – the valency class system complexity. Specifically, high values of H_{intr} indicate greater complexity, while high transitivity prominence values indicate greater simplicity. The entropy of the valency class system, denoted as H, serves as a cumulative metric that considers both facets. Ultimately, H quantifies the amount of information conveyed through contrastive valency patterns in a given language.

The calculations in this study presuppose theoretical limits on possible values of H, ranging between 0 and approximately 4.5 nats. However, the observed values in the sample fall within a narrower range, specifically 0.71 and 2.71. This observation leads to two main conclusions.

On the one hand, the variation in the complexity of valency class systems is empirically constrained: although theoretically conceivable, both maximally simple and maximally complex valency class systems are not attested in actual languages. Instead, languages gravitate towards mid-range degrees of complexity, covering less than half of the theoretically possible values. This tendency reflects a compromise between uniformity, which facilitates production, and heterogeneity, which maximizes cue reliability.

On the other hand, there is substantial variation in the degree of complexity of valency class systems, yet this variation is not random. For one thing, the observed patterns of variation reveal large-scale areal and genealogical signals, which again testifies to the fundamental character of this parameter. Even more importantly, the variation in the degrees of complexity in valency class systems interacts with several other structural parameters. Arguably, high valency class system complexity is common in languages with rich case systems and positively correlates with high prevalence of non-verbal predicates, preverbs, verbal colexifications, and a preference for satellite-framed patterns in the domain of motion. Given that complex valency class systems are largely determined by the extensive use of nominal flags rather than verbal indices, these correlations can be interpreted by appealing to the broad distinction between noun-based and verb-based languages, where more complex valency class systems are associated with noun-based languages.

Finally, the complexity of valency encoding devices interacts intriguingly with word order phenomena. While available data do not show any statistically significant difference in the average degree of valency class system complexity between languages with basic SVO and SOV orders, multiple pieces of evidence suggest that more complex valency-encoding devices tend to favor postverbal positions.

If this generalization holds true, one possible explanation for the observed asymmetry lies in processing mechanisms. The choice of a specialized valency-encoding flag often relies on idiosyncratic access to specific verbal lexemes. Consequently, producing preverbal arguments may require a form of "looking ahead" that violates the principle of radically incremental processing. This principle posits that speakers tend to proceed with whatever material is available first, see (Jaeger and Norcliffe 2009: 871–872) for an overview of psycholinguistic evidence supporting this principle.

Ultimately, this avoidance of "looking ahead" aligns with the "Maximize Online Processing" principle proposed by Hawkins (2014: 28–34). In this respect, my findings resonate with similar observations in the domain of subject indexing (Seržant and Moroz 2022), indicating a broader trend towards minimizing processing effort in speech production.

Supplementary Materials

Supplementary Materials for this paper are available in a GitHub repository at https://github.com/serjozhka/Degrees_of_complexity_STUF (accessed 3 February 2025). These materials include multiple spreadsheets with the data used as input for the analysis, the R code for the analysis, and a summary spreadsheet detailing various aspects of individual languages' valency systems as identified in my study.

Abbreviations

1	1st person
3	3rd person
A	A-argument
ABS	absolutive
ACC	accusative
AOR	aorist
ADLAT	adlative
ART	article
BEN	benefactive
COMPL	completive
CONT	localization 'on the vertical surface, in contact with'
COP	copula
DAT	dative
DCL	declarative
DEF	definite
EL	elative
ERG	ergative
F	feminine
GEN1	first genitive case
G3	third (animate) gender
H	human
IMP	imperative
INF	infinitive
INS	instrumental
IO	indirect object
IPFV	imperfective
LAT	lative
LOC	locative
M	masculine
N	neuter
NOM	nominative
OBL	oblique
PFV	perfective
PL	plural
PN	person name
POSS	possessive
PRF	perfect
PRS	present tense
PST	past tense
REFL	reflexive

SG	singular
SUB	localization ‘under’
SUPER	localization ‘on the horizontal surface or above’
TR	transitive

Appendix. Data summary for analyzed languages

Legend: C = number of nominal morphological cases, WO = basic word order, N = total number of entries, TP = transitivity prominence, H = entropy of bivalent verbs (raw and corrected, in nats), H_{intr} = entropy of intransitive bivalent verbs (raw and corrected, in nats).

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw		Corrected	
								H	H_{intr}	H	H_{intr}
Abaza	abaz1241	Northwest Caucasian	Northwest Caucasian	1	SOV	114	0.36	2.54	2.95	2.48	2.46
Adyghe	adyg1241	Northwest Caucasian	Northwest Caucasian	4	SOV	125	0.32	2.73	3.10	2.63	2.51
Aghul	aghu1253	Nakh-Daghestanian	Lezgic	26	SOV	113	0.28	2.61	2.81	2.56	2.38
Albanian	tosk1239	Indo-European	Albanian	4	SVO	128	0.52	1.67	2.06	1.64	1.89
Alik Kryz	kryt1240	Nakh-Daghestanian	Lezgic	23	SOV	124	0.3	2.77	3.08	2.70	2.56
Ancient Greek	anci1242	Indo-European	Greek	5	SVO	121	0.55	1.38	1.55	1.36	1.40
Archi	arch1244	Nakh-Daghestanian	Lezgic	38	SOV	122	0.31	2.39	2.58	2.35	2.22
Assyrian Neo-Aramaic	assy1241	Afro-Asiatic	Semitic	1	SOV	123	0.4	2.04	2.27	2.02	2.08
Avar	avar1256	Nakh-Daghestanian	Avar-Andic-Tsezic	24	SOV	122	0.29	2.70	2.95	2.66	2.50
Azerbaijani	nort2697	Altaic	Turkic	6	SOV	127	0.43	1.63	1.64	1.60	1.46
Bambara	bamb1269	Mande	Western Mande	1	SOV	109	0.49	1.75	2.06	1.73	1.89
Bashkir	bash1264	Altaic	Turkic	6	SOV	113	0.46	1.59	1.66	1.58	1.52
Basque	basq1248	Basque	Basque	16	SOV	116	0.56	1.65	2.20	1.62	2.00
Belarusian	bela1254	Indo-European	Slavic	6	NO	129	0.43	2.20	2.68	2.15	2.35
Bengali	beng1280	Indo-European	Indic	4	SOV	116	0.58	1.66	2.31	1.64	2.13

(continued)

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw H	Raw H_{intr}	Corrected H	Corrected H_{intr}
Brazilian Portuguese	port1283	Indo-European	Romance	1	SVO	129	0.53	1.40	1.52	1.38	1.40
Bulgarian	bulg1262	Indo-European	Slavic	1	SVO	129	0.52	1.68	2.06	1.65	1.87
Buriat	buri1258	Altaic	Mongolic	7	SOV	120	0.47	1.69	1.88	1.67	1.70
Catalan	stan1289	Indo-European	Romance	1	SVO	127	0.56	1.61	2.09	1.59	1.91
Chechen	chec1245	Nakh-Daghestanian	Nakh	10	SOV	115	0.31	2.33	2.49	2.30	2.20
Chirag	chir1284	Nakh-Daghestanian	Dargwic	34	SOV	122	0.3	2.30	2.40	2.25	2.03
Chukchi	chuk1273	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan	13	NO	121	0.59	1.32	1.56	1.29	1.47
Chuvash	chuv1255	Altaic	Turkic	8	SOV	128	0.57	1.67	2.29	1.64	2.06
Croatian	croa1245	Indo-European	Slavic	6	SVO	128	0.51	2.05	2.75	2.00	2.45
Czech	czec1258	Indo-European	Slavic	7	SVO	128	0.45	2.21	2.74	2.15	2.40
Danish	dani1285	Indo-European	Germanic	1	SVO	127	0.59	1.56	2.15	1.52	2.03
Dutch	dutc1256	Indo-European	Germanic	1	NO	125	0.6	1.49	2.04	1.47	1.89
Eastern Armenian	nucl1235	Indo-European	Armenian	5	NO	127	0.5	1.69	1.99	1.68	1.84
Eastern Maninka	east2426	Mande	Western Mande	1	SOV	115	0.45	2.02	2.42	1.98	2.09
English	stan1293	Indo-European	Germanic	2	SVO	127	0.64	1.41	2.07	1.38	1.92
Erzya	erzy1239	Uralic	Mordvin	12	SVO	110	0.47	2.01	2.51	1.99	2.26
Estonian	esto1258	Uralic	Finnic	14	SVO	129	0.25	2.00	1.92	1.97	1.71
Evenki	even1259	Altaic	Tungusic	14	SOV	118	0.57	1.65	2.24	1.62	2.05
Finnish	finn1318	Uralic	Finnic	15	SVO	129	0.33	1.89	1.86	1.85	1.65
Forest Enets	fore1265	Uralic	Samoyedic	6	SOV	104	0.55	1.53	1.85	1.52	1.71
French	stan1290	Indo-European	Romance	1	SVO	127	0.57	1.51	1.95	1.49	1.84
Gascon	occi1239	Indo-European	Romance	1	SVO	125	0.58	1.49	1.91	1.45	1.76
Georgian	nucl1302	Kartvelian	Kartvelian	7	SOV	130	0.27	2.11	2.10	2.08	1.84
German	stan1295	Indo-European	Germanic	4	NO	129	0.55	1.88	2.65	1.84	2.34

(continued)

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw		Corrected	
								H	H_{intr}	H	H_{intr}
Goo	gooa1234	Mande	Eeastern Mande	1	SOV	92	0.46	2.27	2.91	2.26	2.55
Guro	guro1248	Mande	Eastern Mande	1	SOV	92	0.42	2.08	2.42	2.07	2.19
Hungarian	hung1274	Uralic	Ugric	21	NO	126	0.49	1.95	2.47	1.90	2.19
Icari Dargwa	icar1234	Nakh-Daghestanian	Dargwic	26	SOV	122	0.3	2.62	2.81	2.48	2.31
Icelandic	icel1247	Indo-European	Germanic	4	SVO	126	0.36	2.43	2.77	2.38	2.36
Ingrian Finnish	finn1318	Uralic	Finnic	13	SVO	116	0.38	2.07	2.27	2.04	2.05
Ingush	ingu1240	Nakh-Daghestanian	Nakh	8	SOV	104	0.33	2.06	2.13	2.04	1.89
Irish	iris1253	Indo-European	Celtic	2	VSO	122	0.41	2.22	2.62	2.16	2.17
Italian	ital1282	Indo-European	Romance	1	SVO	129	0.6	1.44	1.94	1.40	1.83
Japanese	nucl1643	Japanese	Japanese	8	SOV	115	0.54	1.37	1.48	1.36	1.36
Joola-Fonyi	jola1263	Niger-Congo	Jola	1	SVO	115	0.79	0.72	0.98	0.71	0.98
Kadar Dargwa	kada1294	Nakh-Daghestanian	Dargwic	9	SOV	120	0.34	2.12	2.25	2.10	1.97
Kalderash Romani	vlax1238	Indo-European	Indic	8	SVO	121	0.42	2.07	2.41	2.03	2.16
Kalmyk	kalm1244	Altaic	Mongolic	9	SOV	121	0.5	1.62	1.87	1.60	1.70
Karata	kara1516	Nakh-Daghestanian	Avar-Andic-Tsezic	34	SOV	124	0.25	2.73	2.90	2.67	2.41
Kaytag Dargwa	kajt1238	Nakh-Daghestanian	Dargwic	37	SOV	123	0.34	2.37	2.62	2.33	2.28
Kazakh	kaza1248	Altaic	Turkic	6	SOV	130	0.48	1.59	1.72	1.56	1.58
Kazym Khanty	khan1273	Uralic	Ugric	4	SOV	108	0.51	1.68	2.02	1.67	1.87
Khoekhoe	nama1264	Khoe-Kwadi	Khoe-Kwadi	2	SOV	127	0.7	1.18	1.91	1.15	1.79
Khwarshi	khva1239	Nakh-Daghestanian	Avar-Andic-Tsezic	46	SOV	122	0.28	2.77	3.02	2.71	2.52
Kina Rutul	rutu1240	Nakh-Daghestanian	Lezgic	15	SOV	116	0.27	2.69	2.87	2.65	2.43
Komi-Permyak	komi1269	Uralic	Permic	23	SVO	120	0.46	2.13	2.66	2.08	2.32
Komi-Zyrian	komi1268	Uralic	Permic	16	SVO	119	0.45	2.10	2.54	2.05	2.30
Kumyk	kumy1244	Altaic	Turkic	6	SOV	127	0.42	1.69	1.73	1.66	1.53

(continued)

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw		Corrected	
								H	H_{intr}	H	H_{intr}
Latin	lati1261	Indo-European	Romance	6	NO	128	0.62	1.40	1.97	1.38	1.82
Latvian	stan1325	Indo-European	Baltic	5	SVO	105	0.5	1.77	2.13	1.76	1.98
Laz	lazz1240	Kartvelian	Kartvelian	8	SOV	124	0.34	1.92	1.94	1.90	1.75
Lezgian	lezg1247	Nakh-Daghestanian	Lezgic	18	SOV	116	0.34	2.43	2.73	2.38	2.35
Lithuanian	lith1251	Indo-European	Baltic	7	SVO	127	0.44	2.05	2.43	2.02	2.17
Macedonian	mace1250	Indo-European	Slavic	1	SVO	129	0.47	1.84	2.15	1.81	1.98
Mandarin Chinese	mand1415	Sino-Tibetan	Chinese	1	SVO	119	0.67	1.48	2.58	1.45	2.39
Mandinka	mand1436	Mande	Western Mande	1	SOV	115	0.56	1.56	1.96	1.52	1.78
Mano	mann1248	Mande	Eastern Mande	1	SOV	105	0.5	1.81	2.22	1.80	1.97
Mehweb	mege1234	Nakh-Daghestanian	Dargwic	38	SOV	125	0.26	2.76	2.97	2.70	2.49
Mingrelian	ming1252	Kartvelian	Kartvelian	9	SOV	128	0.27	2.02	1.98	1.99	1.73
Modern Greek	mode1248	Indo-European	Greek	4	NO	127	0.67	1.11	1.45	1.09	1.39
Modern Hebrew	hebr1245	Afro-Asiatic	Semitic	1	SVO	127	0.48	1.73	2.00	1.71	1.87
Moksha	moks1248	Uralic	Mordvin	17	SVO	115	0.5	1.89	2.42	1.87	2.24
Mukhad Rutul	rutu1240	Nakh-Daghestanian	Lezgic	17	SOV	122	0.29	2.77	3.05	2.70	2.56
Nanai	nana1257	Altaic	Tungusic	9	SOV	121	0.58	1.39	1.68	1.37	1.53
Nivkh	nivk1234	Nivkh	Nivkh	8	SOV	90	0.66	1.16	1.49	1.16	1.44
Northern Akhvakh	akhv1239	Nakh-Daghestanian	Avar-Andic-Tsezic	42	SOV	125	0.27	2.65	2.83	2.58	2.39
Northern Mansi	mans1258	Uralic	Ugric	6	SOV	95	0.54	1.74	2.28	1.74	2.08
Northern Yukaghir	nort2745	Yukaghir	Yukaghir	11	SOV	110	0.64	1.44	2.16	1.42	2.01
Norwegian Bokmål	norw1259	Indo-European	Germanic	1	SVO	129	0.65	1.44	2.28	1.41	2.14
Ossetic	osse1243	Indo-European	Iranian	9	SOV	116	0.39	1.94	2.07	1.92	1.89
Polish	poli1260	Indo-European	Slavic	7	SVO	129	0.41	2.21	2.60	2.17	2.31

(continued)

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw		Corrected	
								H	H_{intr}	H	H_{intr}
Romanian	roma1327	Indo-European	Romance	2	SVO	121	0.59	1.65	2.35	1.61	2.15
Russian	russ1263	Indo-European	Slavic	6	SVO	130	0.42	2.27	2.72	2.23	2.37
Serbian	serb1264	Indo-European	Slavic	6	SVO	128	0.5	2.01	2.63	1.97	2.35
Shinaz Rutul	shin1265	Nakh-Daghestanian	Lezgic	15	SOV	124	0.28	2.67	2.89	2.62	2.47
Shughni	shug1248	Indo-European	Iranian	2	SOV	119	0.45	1.94	2.28	1.90	2.04
Skolt Saami	skol1241	Uralic	Saami	9	SVO	120	0.56	1.71	2.33	1.68	2.10
Slovak	slov1269	Indo-European	Slavic	6	SVO	129	0.48	2.08	2.68	2.04	2.32
Slovenian	slov1268	Indo-European	Slavic	6	SVO	129	0.5	1.98	2.59	1.94	2.28
Soninke	soni1259	Mande	Western Mande	1	SOV	107	0.59	1.22	1.32	1.20	1.18
Sorani Kurdish	cent1972	Indo-European	Iranian	1	SOV	115	0.37	1.89	1.97	1.87	1.77
South Saami	sout2674	Uralic	Saami	8	SOV	108	0.71	1.27	2.33	1.24	2.21
Southern Rutul	borc1243	Nakh-Daghestanian	Lezgic	15	SOV	123	0.29	2.67	2.92	2.58	2.46
Spanish	stan1288	Indo-European	Romance	1	SVO	127	0.58	1.50	1.97	1.49	1.84
Standard Arabic	stan1318	Afro-Asiatic	Semitic	3	VSO	126	0.57	1.60	2.14	1.57	1.94
Svan	svan1243	Kartvelian	Kartvelian	7	SOV	129	0.26	2.05	2.00	2.01	1.71
Swedish	swed1254	Indo-European	Germanic	1	SVO	128	0.63	1.45	2.15	1.43	1.98
Tagalog	taga1270	Austronesian	Greater Central Phillipine	3	VSO	125	0.5	1.71	2.06	1.68	1.82
Telugu	telu1262	Dravidian	Dravidian	7	SOV	121	0.46	2.09	2.60	2.05	2.33
Tsakhur	tsak1249	Nakh-Daghestanian	Lezgic	20	SOV	118	0.31	2.70	2.99	2.63	2.52
Tsudakhar Dargwa	cuda1238	Nakh-Daghestanian	Dargwic	24	SOV	111	0.3	2.41	2.57	2.38	2.21
Tsugni Dargwa	sout3261	Nakh-Daghestanian	Dargwic	34	SOV	117	0.3	2.53	2.75	2.50	2.33
Turkish	nucl1301	Altaic	Turkic	6	SOV	126	0.48	1.45	1.45	1.43	1.30
Turoyo	turo1239	Afro-Asiatic	Semitic	1	SVO	122	0.34	2.60	2.98	2.53	2.52

(continued)

Language	Glottocode	Family	Genus	C	WO	N	TP	Raw		Corrected	
								H	H_{intr}	H	H_{intr}
Tuvinian	tuvi1240	Altaic	Turkic	8	SOV	129	0.47	1.84	2.15	1.81	1.96
Udi	udii1243	Nakh-Daghestanian	Lezgic	10	SOV	123	0.48	2.05	2.61	2.00	2.27
Udihe	udih1248	Altaic	Tungusic	9	SOV	102	0.62	1.37	1.84	1.36	1.72
Udmurt	udmu1245	Uralic	Permic	15	SOV	126	0.41	2.19	2.58	2.14	2.23
Ukrainian	ukra1253	Indo-European	Slavic	7	SVO	129	0.44	2.17	2.66	2.14	2.36
Ulcha	ulch1241	Altaic	Tungusic	8	SOV	108	0.61	1.37	1.80	1.35	1.65
Uzbek	nort2690	Altaic	Turkic	6	SOV	129	0.49	1.57	1.71	1.55	1.54
West Central Oromo	west2721	Afro-Asiatic	Lowland East Cushitic	7	SOV	129	0.65	1.44	2.28	1.41	2.07
Western Mari	west2392	Uralic	Mari	10	SOV	122	0.5	1.78	2.17	1.76	1.99
Yakut	yaku1245	Altaic	Turkic	8	SVO	124	0.53	1.54	1.82	1.53	1.68
Zilo Andi	zilo1248	Nakh-Daghestanian	Avar-Andic-Tsezic	21	SOV	123	0.29	2.60	2.83	2.55	2.43

References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89(3). 429–464.

Andersen, Torben. 1988. Ergativity in Päri, a Nilotic OVS language. *Lingua* 75. 289–324.

Arkadiev, Peter & Kirill Kozhanov. 2023. Borrowing of morphology. With a case study of Baltic and Slavic verbal prefixes. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet & Antonio Fábregas (eds.), *The wiley blackwell companion to morphology*. Hoboken: Wiley & Sons.

Arkadiev, Peter, Yury Lander & Irina Bagirokova. 2024. Applicative constructions in the Northwest Caucasian languages. In Fernando Zúñiga & Denis Creissels (eds.), *Applicative constructions in the world's languages* [*Comparative handbooks in linguistics* 7], 869–912. Berlin, Boston: De Gruyter Mouton.

Arkadiev, Peter & Timur Maisak. 2018. Grammaticalization in the North Caucasian languages. In Heiko Narrog & Bernd Heine (eds.), *Grammaticalization from a typological perspective*, 116–145. Oxford: Oxford Academic. (online edition).

Authier, Gilles. 2009. *Grammaire kryz (Langue caucasique d'Azerbaïdjan, dialecte d'Alik)*. Leuven, Paris: Peeters.

Barnabé, Aurélie. 2016. Description verbale du mouvement dans le cadre de la typologie talmienne – une évaluation au croisement de deux paradigmes: le cognitivisme et l'enactivisme. *Anglophonia – French Journal of English Linguistics* 22. https://doi.org/10.4000/anglophonia.1064.

Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46. 331–377.
Bentz, Christian, Ximena Gutierrez-Vasques, Olga Sozinova & Tanja Samardžić. 2023. Complexity trade-offs and equi-complexity in natural languages: A meta-analysis. *Linguistics Vanguard* 9(s1). 9–25.
Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.
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.
Bossong, Georg. 1998. Le marquage de l'experient dans les langues d'Europe. In Jacques Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, 259–294. Berlin: Mouton de Gruyter.
Cole, Peter, Wayne Harbert, Gabriella Hermon & S. N. Sridhar. 1980. The acquisition of subjecthood. *Language* 56(4). 719–743.
Comrie, Bernard. 1989. *Language universals and linguistic typology*, 2nd edn. Chicago: Chicago University Press.
Creissels, Denis. 2018. Transitivity prominence in typological perspective: The case of Basque. *Anuario del Seminario de Filología Vasca Julio de Urquijo* 52(1–2). 175–187.
Croft, William, Jóhanna Barðdal, Willem Hollmann, Violeta Sotirova & Chiaki Taoka. 2010. Revising Talmy's typological classification of complex events. In Hans Boas (ed.), *Contrastive construction grammar*, 201–235. Amsterdam: John Benjamins.
Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
Dixon, R. M. W. 2004. Adjective classes in typological perspective. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Adjective classes: A cross-linguistic typology*, 1–49. Oxford: Oxford University Press.
Dixon, R. M. W. 2010. *Basic linguistic theory: Volume 2: Grammatical topics*. Oxford: Oxford University Press.
Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.
Drossard, Werner. 1991. Transitivität (vs. Intransitivität) und Transitivierung (vs. Intransitivierung) unter typologischem Aspekt. In Hansjakob Seiler & Waldfried Premper (eds.), *Partizipation: Das sprachliche Erfassen von Sachverhalten*, 408–445. Tübingen: Narr.
Dryer, Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. http://wals.info/chapter/81 (accessed 11 May 2024).
Dryer, Matthew S. & Orin D. Gensler. 2013. Order of object, oblique, and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. https://wals.info/chapter/84 (accessed 11 May 2024).
Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online* (v2020.3) [Data set]. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Fenk-Oczlon, Gertraud & August Fenk. 2008. Complexity trade-offs between the subsystems of language. In Matti Miestamo, Kaius Sinnemaki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 43–65. Amsterdam & Philadelphia: John Benjamins.
Fenk-Oczlon, Gertraud & August Fenk. 2014. Complexity trade-offs do not prove the equal complexity hypothesis. *Poznan Studies in Contemporary Linguistics* 50(2). 145–155.
Givón, Talmy. 2009. *The Genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam: John Benjamins.
Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 2nd edn., 73–113. Cambridge, MA: MIT Press.

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. 2010. The behaviour-before-coding principle in syntactic change. In Franck Floricic (ed.), *Essais de typologie et de linguistique générale: Mélanges offerts à Denis Creissels*, 541–554. Lyon: Presses Universitaires de l'École Normale Superieure.
Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567.
Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages, volume 1: Introducing the framework, and case studies from Africa and Eurasia*, 131–147. Berlin & Boston: De Gruyter Mouton.
Haspelmath, Martin. 2019. Indexing and flagging, and head and dependent marking. *Te Reo. The journal of the Linguistic Society of New Zealand* 62(1). 93–115.
Haspelmath, Martin & Iren Hartmann. 2015. Comparing verbal valency across languages. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages, volume 1: Introducing the framework, and case studies from Africa and Eurasia*, 41–71. Berlin & Boston: De Gruyter Mouton.
Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
Hawkins, John A. 1986. *A comparative typology of English and German: Unifying the contrasts*. Austin: University of Texas Press.
Hawkins, John A. 2008. An asymmetry between VO and OV languages: The ordering of obliques. In Greville G. Corbett & Michael Noonan (eds.), *Case and grammatical relations: Studies in honor of Bernard Comrie*, 167–190. Amsterdam: Benjamins.
Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
Iggesen, Oliver A. 2013. Number of cases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. https://wals.info/chapter/49 (accessed 11 May 2024).
Jaeger, Tim Florian. 2010. Redundancy and reduction: Speakers manage information density. *Cognitive Psychology* 61(1). 23–62.
Jaeger, Tim Florian & Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3/4. 866–887.
Joseph, John E. & Frederick J. Newmeyer. 2012. All languages are equally complex'. The rise and fall of a consensus. *Historiographia Linguistica* 39(3). 341–368.
Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5. 206–213.
Kassambara, Alboukadel. 2023. ggpubr: 'ggplot2' based publication ready plots (R package version 0.6.0). https://rpkgs.datanovia.com/ggpubr/ (accessed 11 May 2024).
Ke, Jinyun. 2006. A cross-linguistic quantitative study of homophony. *Journal of Quantitative Linguistics* 13. 129–159.
Kibrik, Aleksandr E. 1997. Beyond subject and object: Toward a comprehensive relational typology. *Linguistic Typology* 1(3). 279–346.
König, Christa. 2008. *Case in Africa*. Oxford: Oxford University Press.
Kuperberg, Gina R. & Tim Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1). 32–59.
Kuryłowicz, Jerzy. 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.
Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26.

Lambrecht, Knud. 2001. Dislocation. In Martin Haspelmath, Ekkehard Konig, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals*, 1050–1078. Berlin: Walter de Gruyter.

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

Levshina, Natalia. 2019. Token-based typology and word order entropy: A study based on Universal Dependencies. *Linguistic Typology* 23(3). 533–572.

Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 648200.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.

Lewandowski, Wojciech. 2018. A typological approach to the encoding of motion events. In María Ángeles Gómez González & Lachlam Mackenzie (eds.), *The construction of discourse as interaction*, 45–74. Amsterdam & Philadelphia: John Benjamins.

Lowe, John J. 2017. *Transitive nouns and adjectives. Evidence from Early Indo-Aryan*. Oxford: Oxford University Press.

Maion, Fabio. 2018. *Do certain word orders attract case marking? A typological survey on the dependency of syntax and morphology*. Tübingen: University of Tübingen BA thesis.

Malchukov, Andrej. 2015. Valency classes and alternations parameters of variation. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages, volume 1: Introducing the framework, and case studies from Africa and Eurasia*, 73–130. Berlin: Mouton de Gruyter.

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

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.

Meyer, Patrick E. 2022. infotheo: Information-theoretic measures. R package. version 1.2.0.1. https://CRAN.R-project.org/package=infotheo (accessed 11 May 2024).

Michaelis, Laura & Josef Ruppenhofer. 2001. *Beyond alternations. A constructional model of the German applicative pattern*. Stanford: CSLI.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam & Philadelphia: John Benjamins.

Müller-Gotama, Franz. 1994. *Grammatical relations: A cross-linguistic perspective on their syntax and semantics*. Berlin: Mouton de Gruyter.

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

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62(1). 56–119.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Ooms, Jeroen. 2024. writexl: Export data frames to Excel 'xlsx' format (R package version 1.5.0). Available at: https://CRAN.R-project.org/package=writexl.

Polinsky, Maria. 2012. Headedness, again. In Thomas Graf, Denis Paperno, Anna Szabolcsi & Jos Tellings (eds.), *Theories of everything. In honor of Ed Keenan*, 348–359. Los Angeles: UCLA Department of Linguistics.

Qiu, Chen & Adam Winsler. 2017. Language use in a 'one parent–one language' Mandarin–English bilingual family: Noun versus verb use and language mixing compared to maternal perception. *International Journal of Bilingual Education and Bilingualism* 20(3). 272–291.

R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/ (accessed 11 May 2024).

Sampson, Geoffrey, David Gil & Peter Trudgill (eds.). 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

Sauerland, Ulrich. 1994. German diathesis and verb morphology. In Douglas Jones (ed.), *Working papers and projects on verb class alternations in Bangla, German, English and Korean*, 37–92. Cambridge, MA: MIT AI Memo 1517. (non vidi).

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 S. 2018. Markirovanie aktantov dvuxmestnyx predikatov: Predvaritel'nye itogi tipologičeskogo issledovanija [Argument encoding of bivalent predicates: Preliminary results of a typological inquiry]. In Sergey S. Say (ed.), *Valentnostnye klassy dvuxmestnyx predikatov v raznostrukturnyx jazykax* [Valency classes of bivalent predicates in languages of various types], 557–616. St. Petersburg: ILI RAN.

Say, Sergey (ed.). 2020–. *BivalTyp: Typological database of bivalent verbs and their encoding frames*. https://www.bivaltyp.info (accessed 11 May 2024).

Schnell, Stefan, Geoffrey Haig & Frank Seifart. 2021. The role of language documentation in corpus-based typology. In Geoffrey Haig, Stefan Schnell & Frank Seifart (eds.), *Doing corpus-based typology with spoken language corpora: State of the art* (Language documentation & conservation special publication 25), 1–28. Honolulu: University of Hawai'i Press.

Seržant, Ilja A. 2019. Weak universal forces: The discriminatory function of case in differential object marking systems. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 149–178. Berlin: Language Science Press.

Seržant, Ilja A. & George Moroz. 2022. Universal attractors in language evolution provide evidence for the kinds of efficiency pressures involved. *Humanities and Social Sciences Communications* 9(58). https://www.nature.com/articles/s41599-022-01072-0 (accessed 11 May 2024).

Seržant, Ilja A., Sergey Say, Andreas Hölzl, Aigul Zakirova, Xinyi Gao & Andreas Pregla. 2025. A typology of positional differential argument marking. *Linguistic Typology*. https://doi.org/10.1515/lingty-2024-0021 (Epub ahead of print).

Shcherbakova, Olena, Volker Gast, Damián E. Blasi, Hedvig Skirgård, Russell D. Gray & Simon J. Greenhill. 2023. A quantitative global test of the complexity trade-off hypothesis: The case of nominal and verbal grammatical marking. *Linguistics Vanguard* 9(s1). 155–167.

Shosted, Ryan K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10(1). 1–40.

Siewierska, Anna & Dik Bakker. 1996. The distribution of subject and object agreement and word order type. *Studies in Language* 20(1). 115–161.

Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 126–140. Oxford: Oxford University Press.

Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 67–88. Amsterdam & Philadelphia: John Benjamins.

Sinnemäki, Kaius. 2010. Word order in zero-marking languages. *Studies in Language* 34(4). 869–912.

Sinnemäki, Kaius. 2014. Global optimization and complexity trade-offs. *Poznań Studies in Contemporary Linguistics* 50(2). 179–195.

Slowikowski, Kamil. 2025. ggrepel: Automatically position non-overlapping text labels with 'ggplot2' (R package version 0.9). https://cran.r-project.org/web/packages/ggrepel/index.html (accessed 3 February 2025).

Stathi, Katherina. 2023. *Granularity in the verbalization of events and objects: A cross-linguistic study*. Amsterdam, Philadelphia: John Benjamins.
Stone, Kate, Titus von der Malsburg & Shravan Vasishth. 2020. The effect of decay and lexical uncertainty on processing long-distance dependencies in reading. *PeerJ* 8. e10438.
Stump, Gregory. 2017. The nature and dimensions of complexity in morphology. *Annual Review of Linguistics* 3(1). 65–83.
Talmy, Leonard. 2000. *Toward a cognitive semantics vol. 2: Typology and process in concept structuring*. Cambridge, MA: MIT Press.
Tomlin, Russell S. 1986. *Basic word order: Functional principles*. London: Routledge.
Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385–396.
Van Valin, Robert D., Jr. & Randy LaPolla. 1997. *Syntax. Structure, meaning and function*. Cambridge: Cambrdige University Press.
Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer.
Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686.
Wickham, Hadley & Jennifer Bryan. 2023. Readxl: Read Excel files (R package version 1.4.3). Available at: https://CRAN.R-project.org/package=readxl.
Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller & Davis Vaughan. 2023. dplyr: A grammar of data manipulation (R package version 1.1.4). Available at: https://CRAN.R-project.org/package=dplyr.
Wood, Jim & Neil Myler. 2019. Argument structure and morphology. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.
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.
Zúñiga, Fernando, Peter Arkadiev & Veronika Hegedűs. 2024. Applicativizing preverbs in selected European languages. In Fernando Zuniga & Denis Creissels (eds.), *Applicative constructions in the world's languages*, 419–472. Berlin & Boston: De Gruyter Mouton.