Degrees of complexity in valency class systems: implications for efficiency

Abstract. The complexity of valency class systems remains an underexplored aspect of typological variation. Drawing on data from Bivaltyp, a database cataloging valency patterns of 130 verbs across 124 languages, I propose a method to quantify this complexity using various metrics. Two of these metrics – the ratio of intransitive verbs among bivalent verbs and the entropy in the distribution of intransitive verbs among valency classes – display a strong positive correlation, reflecting two facets of a broader fundamental complexity. The third overarching metric – the entropy in the distrubution of all bivalent verbs among valency classes – effectively captures both facets. Languages tend to gravitate toward moderate complexity values, avoiding overly simplistic or excessively complex valency class systems, which could impede efficient communication. However, significant variation persists, driven mainly by areal and genealogical factors, underscoring the relative diachronic stability of valency class system complexity. Entropy-based metrics exhibit positive correlations with structural properties such as the number of cases, prevalence of non-verbal predicates, and potentially a preference for satellite-framed constructions in motion events. Additionally, complex argument-encoding devices favor postverbal positions, aligning with the principle of incremental processing and processing effort minimization. Overall, this study contributes to understanding how efficiency pressures shape typological distributions.

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

1 Background and goals

1.1 Setting the stage

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

(1) *Ich folge ihm.*
'I follow him.'

(2) *Ich sehe ihn.*
'I see him'.

The German verb *folgen* (1) takes a Dative object, while *sehen* (2) takes an Accusative object. This reflects a semantic distinction that is not relevant for the English counterparts, as shown by the translations of (1) and (2). This is part of a larger picture: German has been long shown to exhibit finer argument-encoding distinctions than English (Hawkins 1986; Sauerland 1994). This difference can be captured in terms of complexity: arguably, the German valency class system is more complex than the English system.

Despite occasional observations on differences between specific languages, a well-established quantitative method for evaluating valency class system complexity across languages remains absent. In this proof-of-concept paper, I present a technique for such purpose and apply it to a large language sample. The main goal is to showcase the robustness of 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.

[1] ACKNOWLEDGEMENTS, INCLUDING REFERENCES TO FUNDING AUTHORITIES, REMOVED FOR PURPOSES OF ANONYMIZATION.

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 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, the most widely known argument-encoding technique, encompasses cases, adpositions, and their combinations. 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) *Ich warte auf ihn.*
'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. However, the significance of indexing is relatively minor in German since it is enitirely predictable from flagging (only Nominative subjects are flagged), while non-indexed arguments are further distinguished by flagging. 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)
Fatíma Murád jə-z-qá-l-ç-əj-ṭ
PN PN 3SG.M.IO-BEN-LOC-3SG.F.ERG-believe-PRS-DCL
'Fatima trusts Murad.'
(BivalTyp)[2]

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.

Typologically, word order and indexing are often crucial for distinguishing between core arguments, while non-core arguments typically rely on dependent marking (Nichols 1986: 75). Consequently, flagging is the key encoding type in the subsequent discussion.

Typological invetigation into valency encounters challenges due to the language-specific nature of argument-encoding devices. For instance, the German Dative case in (1) cannot be directly equated with any argument-encoding device in another language, even if they share the same descriptive label. 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).

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

(5) *I schaue auf die Wolken.*
'I am looking at the clouds.'

The primary benefit of the concept of valency class lies in its independence from arbitrary descriptive labels applied to the encoding devices themselves. The German verbs illustrated in (3) and (5) undeniably belong to the same valency class, regardless of how prepositions and cases in German are analyzed.

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. Our starting point will be the common assumption that case and other argument-encoding devices primarily serve a discriminatory function, helping to 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 vs. Accusative case in (2) allows the hearer to distinguish between 'the seer' and 'the person seen'. However, the vast majority of verbs in natural languages have no more than three semantic arguments. Therefore, in principle, a simplified system with just three distinct encoding devices could fully fulfill the discriminatory function. For example, a system where all monovalent verbs take a Nominative argument, all bivalents verbs use the Nominative-Accusative case frame, and all trivalent verbs use the Nominative-Accusative-Dative case frame would eliminate any confusion in terms of role-referent associations.

Such a hypothetical system could indeed enahnce speech production efficiency. However, it would come with a drawback: individual argument-encoding devices in such a system would largely lack semantic cues that aid speech perception. In this regard, a hearer or reader can benefit from more complex systems. For instance, in (6), the decoder of the signal in German can predict the sentence-final verb *abhängen* 'depend' early 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 (Stone et al. 2020).

(6) *Teilhabe darf nicht vom Alter abhängen.*
'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-encoded devices. Such a system would maximize cue reliability, so that successful decoding of any argument would be sufficient for identifying the whole scene. However, it would excessively burdensome for speakers due to poor learnability and the need to process long markers.

In reality, all languages fluctuate in the middle section of the space between the two hypothetical extremes. For instance, all languages seem to have a large transitive class forming the core of bivalent verbs, but also feature some intransitive bivalent verbs (Tsunoda 1985; Næss 2007). Consequently, observed valency class systems represent a compromise between uniformity, which facilitates production, and heterogeneity, which maximizes informativeness. In what follows, I will concentrate on the empirical limits of valency class system complexity.

1.4 Approaches to complexity
Linguistic complexity is a prominent subject in modern typology (Dahl 2004; Miestamo et al. (eds.) 2008; Sampson et al. (eds.) 2009; Shcherbakova et al. 2022). Much of the discussion

revolves around the equi-complexity hypothesis, which suggests all languages tend to exhibit comparable levels of overall complexity. While there are compelling instances of partial complexity trade-offs, where compexity 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; Shcherbakova et al. 2022; 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 (Shcherbakova et al. 2022). 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 2008; 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) are encoded differently (Sinnemäki 2008, 2009; Scherbakova 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 since the frequent use of non-default argument-encoding patterns presupposes 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 address this gap.

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 the running text, spoken or written. Consequently, assessing the complexity of argument-encoding systems from a corpus-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 sufficiently deep and unified 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, wherein 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 behaviour (Bickel et al. 2014). Moreover, the BivalTyp questionnaire primarily includes verbs that typologically deviate from the transitive prototype, such as ‘be afraid’, ‘follow’, ‘see’, and ‘touch’ (Say 2014: 126), which further fosters observing cross-linguistic differences in valency class systems.

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 (zero-marked) Nominative case, which is also cross-referenced on the verb.

(7) Tsakhur (Nakh-Daghestanian)

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

'Murad found (his) keys.'
(BivalTyp)

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 patterns in Tsakhur, as X-arguments in Tsakhur transitive clauses are encoded by the ergative case.

For this paper, I used the latest development version of BivalTyp available as of February 16, 2024. The raw dataset included 124 languages. It is important to note that 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 the New World is largely unrepresented. 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 in my dataset, with an average incidence of 11 out of 130 datapoints 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 BuvalTyp's language-specific datasets).

The dataset employed in this study is available at REMOVED FOR PURPOSES OF ANONYMIZATION. All calculations for this study were performed in R (R Core Team 2021). I utilized the following packages for data analysis and visualization: ggpubr (Kassambara 2023), infotheo (Meyer 2022), lingtypology (Moroz 2017), and lmerTest (Kuznetsova et al. 2017). The R code used in this study is available at REMOVED FOR PURPOSES OF ANONYMIZATION.

2.2 The metrics

I use two basic metrics designed to capture the level of complexity within valency class systems.

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 datapoints for Skolt Saami, with 67 representing the transitive pattern, resulting in a transitivity prominence of 0.56. While the may lack significant meaning due to the arbitrary design of the questionnaire, the differences between values obtained for individual languages are substantial and meaningful: the values range between 0.25 for Karata and 0.79 for Joola-Fonyi. Lower values are observed when the transitive class is more semantically restricted, indicating more complex valency class systems.

However, transitivity prominence is “blind” to the number and nature of distinctions made between intransitive classes. Therefore, the primary metric used in this study is the entropy observed in the distribution of verb-specific patterns among language-specific valency patterns.[3] The general formula for Shannon’s entropy is given in (8), where $p(x_i)$ corresponds to the probability that a certain observation displays the *i*-th of the *k* possible values of *x*.

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

In this study, entropy is measured in *nats* with the base for the logarithm being *e*. Possible values of a variable correspond to distinct valency classes. For instance, in Finnish, *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).

In a nutshell, the observed value of 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 entropy are observed in languages with many valency classes where the use of specific argument-encoding devices offers significant cues for the sentence’s overall meaning.

Upon conducting an initial analysis of the raw data and implementing bootstrapping simulations, it became evident that the raw H values are lower for languages with a higher incidence of missing datapoints. Given that the occurrence of missing data is influenced by factors unrelated to the study of valency patterns itself, relying on the raw H values would lead to an underestimation of the complexity levels for languages with a greater amount of missing data. In response to this problem, I took the following two steps. First, I disregarded the data from the four languages where fewer than 90 non-missing entries were obtained. Second, I employed a correction procedure for the remaining 120 languages. I generated 100 random subsamples, each consisting of 90 entries, and computed the average entropy value observed within these subsamples. To illustrate, for Turoyo, a language with 122 non-missing entries, the raw H value is 2.60, while the corrected value is 2.51. In all subsequent calculations, “H” refers to these corrected values. 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.

3 Results

3.1 Cross-linguistic variability

The corrected entropy values exhibit significant variation across languages, ranging from 0.71 nats for Joola-Fonyi to 2.71 nats for Khwarshi. 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

[3] 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).

17 entries out of a total of 115. In Khwarshi, the transitive class comprises only 34 transitive entries out of a total of 122, and there are 32 further classes. For illustration, Table 1 features 10 selected meanings represented by transitive constructions in Joola-Fonyi, corresponding to verbs from 10 distinct valency classes in Khwarshi.

Table 1. 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'oƛa*	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 1 provides a glimpse into the vast differences between valency class systems, highlighting the two extreme cases in the dataset. The range of possible variation among languages is the main finding so far. Sections 3.2–3.4 delve into this variation within the broader context of linguistic diversity.

3.2 Transitivity prominence and entropy of intransitives

A robust negative correlation is observed between transitivity prominence and H (corrected entropy), the primary complexity metrics employed in this study. This strong correlation persists in the linear mixed-effects model, where linguistic family is included as a random factor. The model reveals a regression of -2.86, accompanied by a very low *p*-value (<< .001).

Yet, this correlation is partially tautological. The transitive class emerges as the largest class in all languages of the sample, confirming previous observations (Haspelmath 2015: 139; Lazard 1994: 131–158). In this context, languages with low transitivity prominence are never dominated by a single valency class but instead feature numerous smaller classes, leading to higher entropy. Put simply, languages with low transitivity prominence automatically yield high entropy, regardless of the distribution of intransitive verbs among minor valency classes.

To investigate facets of complexity not determined by transitivity alone, I introduced another metric: the entropy of intransitive verbs, denoted as H_{intr}. Calculations followed the procedure outlined in Section 2, with the distinction that only intransitive verbs were taken into account. Similar to the overall entropy, I found a positive correlation between the total number of intransitive entries and their observed entropy. To address this, I again implemented a correction procedure: for each language, I generated 100 random subsamples, each comprising 24 intransitive entries, and calculated the corrected value of H_{intr} as the average entropy value observed within these subsamples. The subsample size of 24 stems from it being the lowest number of intransitive entries found in any language, ensuring subsampling for all languages.

Following these preparatory calculations, I asked whether there exists an empirical correlation between H_{intr} and transitivity prominence. To address this question, I fitted one more linear mixed-effects model with language family as a random factor. I observed a substantial regression coefficient of -0.89, coupled with a low *p*-value ($p < 0.002$). For illustration, I provide the relevant scatterplot in Figure 1 (disregarding genealogical information).

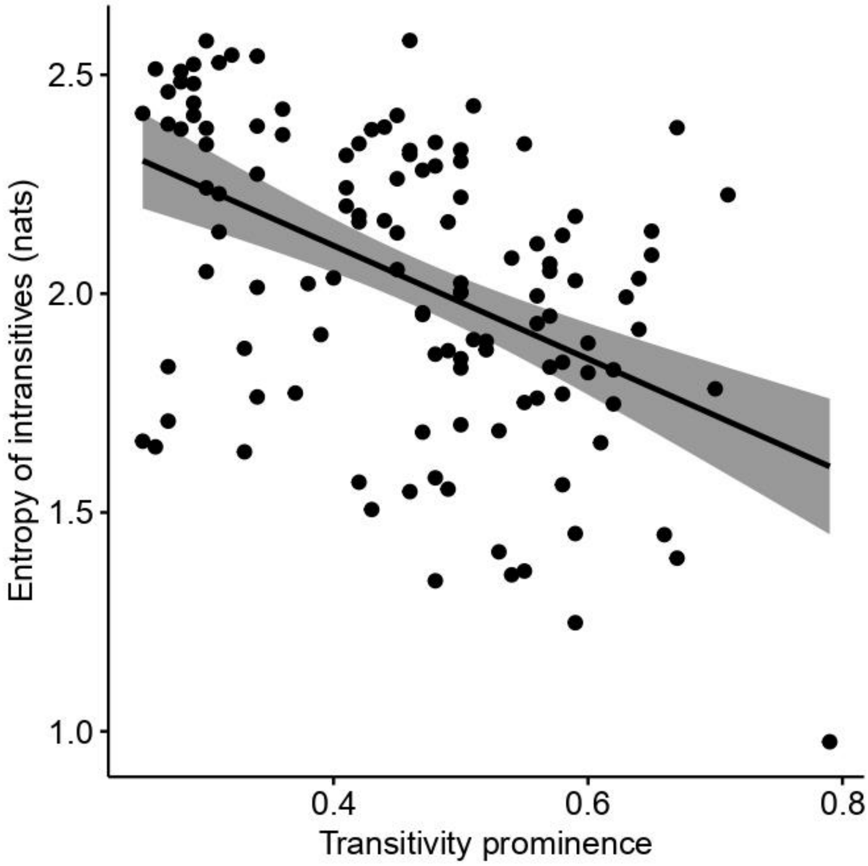

Figure 1: Transitivity prominence and H_{intr} (corrected entropy of intransitives).

In essence, the observed correlation suggests that if we randomly select 24 intransitive 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 classes. This generalization contributes to establishing the valency class system complexity as as a fundamental parameter of typological variability, which manifests itself in at least two logically independent phenomena.

3.3 Entropy and morphological case

An obvious hypothesis is that valency class system complexity is merely a syntactic manifestation of the size of case inventory, a well-known parameter of morphological variation (Iggesen 2013). I tested this hypothesis using data from BivalTyp.[4] Observationally, there is indeed a positive correlation between the number of nominal cases and H_{intr}, as shown in Figure 2.

[4] The BivalTyp's annotation was adjusted for languages without nominal case: for calculations, they were annotated as having 1 case instead of 0 cases.

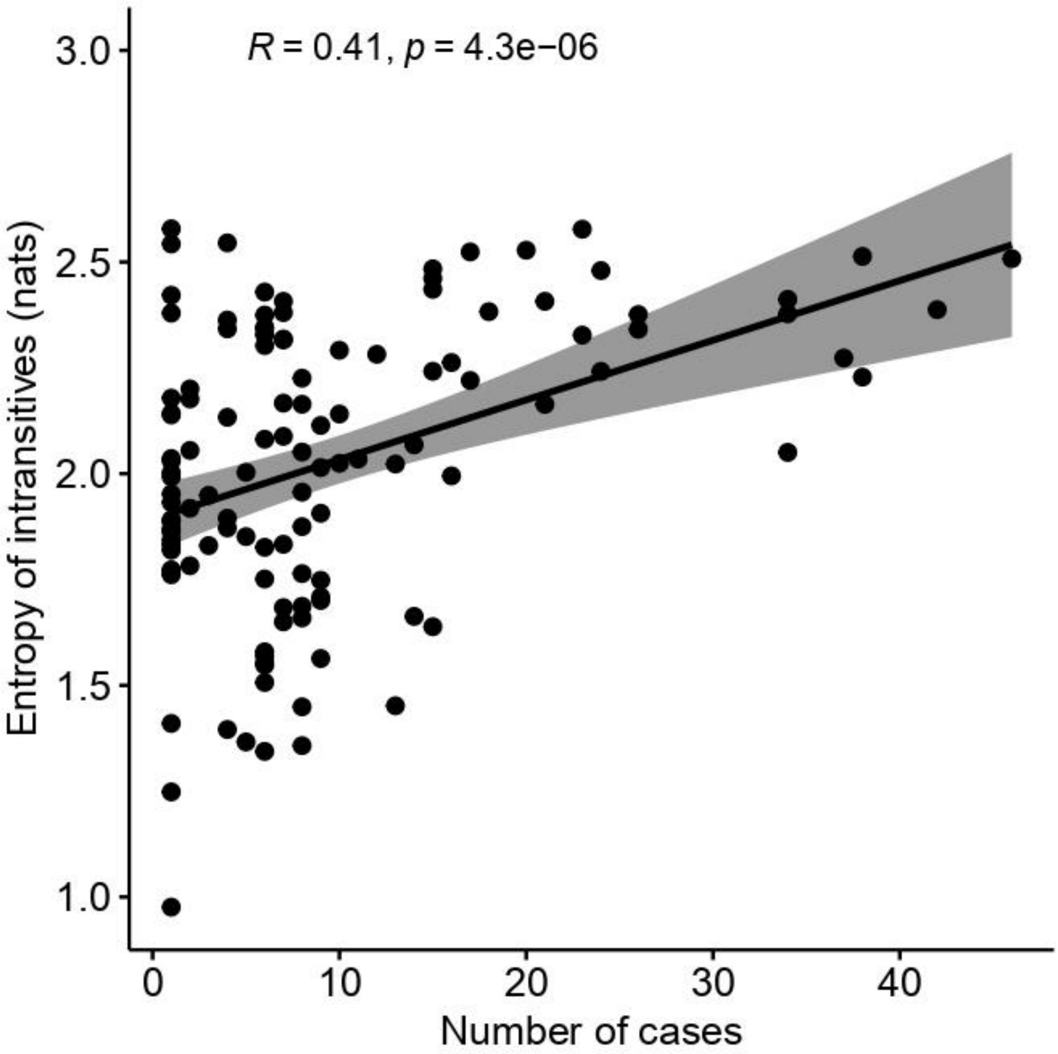

Figure 2: Number of nominal cases and H_{intr}.

However, this correlation might be epiphenomenal. For one thing, in a linear regression model predicting H_{intr} based on the number of nominal cases with linguistic family as a random factor, the observed coefficient's magnitude is moderate (0.001), though the *p*-value of 0.016 still indicates statistical significance. Importantly, the correlation becomes insignificant when transitvity ratio is also included in the model ($p \approx 0.07$). In fact, the observed correlation in the whole dataset is primarily due to one family, Nakh-Daghestanian. These languages, represented widely in the sample (24 languages), possess rich case systems (with 8 to 46 forms) highly complex valency class systems (H_{intr} ranging between 1.87 and 2.58, well above the sample mean). The peculiar position of these languages is discussed Section 3.4. For now, it would suffice to note that no correlation between H_{intr} and the number of nominal cases is observed in languages with no more than 10 distinct cases.

3.4 Areal and genealogical factors

All complexity metrics examined in this study, namely transitivity ratio and two versions of entropy (H and H_{intr}), show notable areal and genealogical patterns. For instance, Figure 3 illustrates the distribution of H across the languages in the sample. For enhanced readability, languages from the Caucasus are presented in a separate inset.

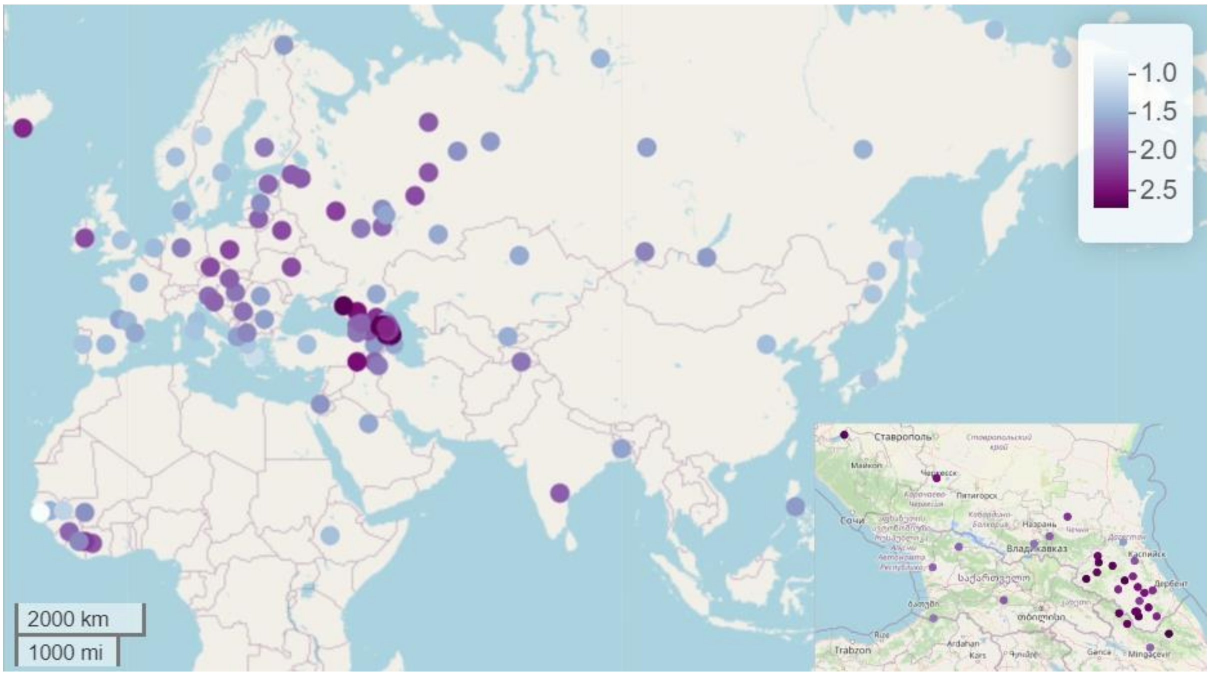

Figure 3: Entropy of valency classes (H) in the languages of the sample.

The data in Figure 3 allow for some broad areal generalizations. The Caucasus, particularly East Caucasus, exhibits the highest values of H in Eurasia. Eastern Europe also features relatively complex valency class systems, contrasting with Western Europe and Siberia. Similar patterns are observed for the entropy of intransitives (H_{intr}). As expected, the transitivity ratio displays reverse areal patterns, partly discussed in existing literature (Bossong 1998; Haspelmath 2015: 139–142; Lazard 1994: 63; Say 2014). The distinct areal patterns observed in suggest that valency class systems complexity can serve as as structural markers (in the sense of Nichols 1992: 185), offering insights into historical processes shaping linguistic diversity.

The usual question regarding such distributions is whether they are primarily shaped by genealogy or areality. Unfortunately, due to the sample's size and nature, systematically teasing apart these two explanations is challenging. However, there are indications that both genealogical and areal factors play a role. For example, the high complexity zone in Eastern Europe cross-cuts two large linguistic families (Uralic and Indo-European), with Eastern European languages exhibiting higher complexity levels than genealogically related languages in adjacent areas. This suggests language contact as a probable explanation for this pattern. Conversely, a comparison of the four families in the sample with at least 10 languages reveals systematic differences between them, as shown in Figure 4.

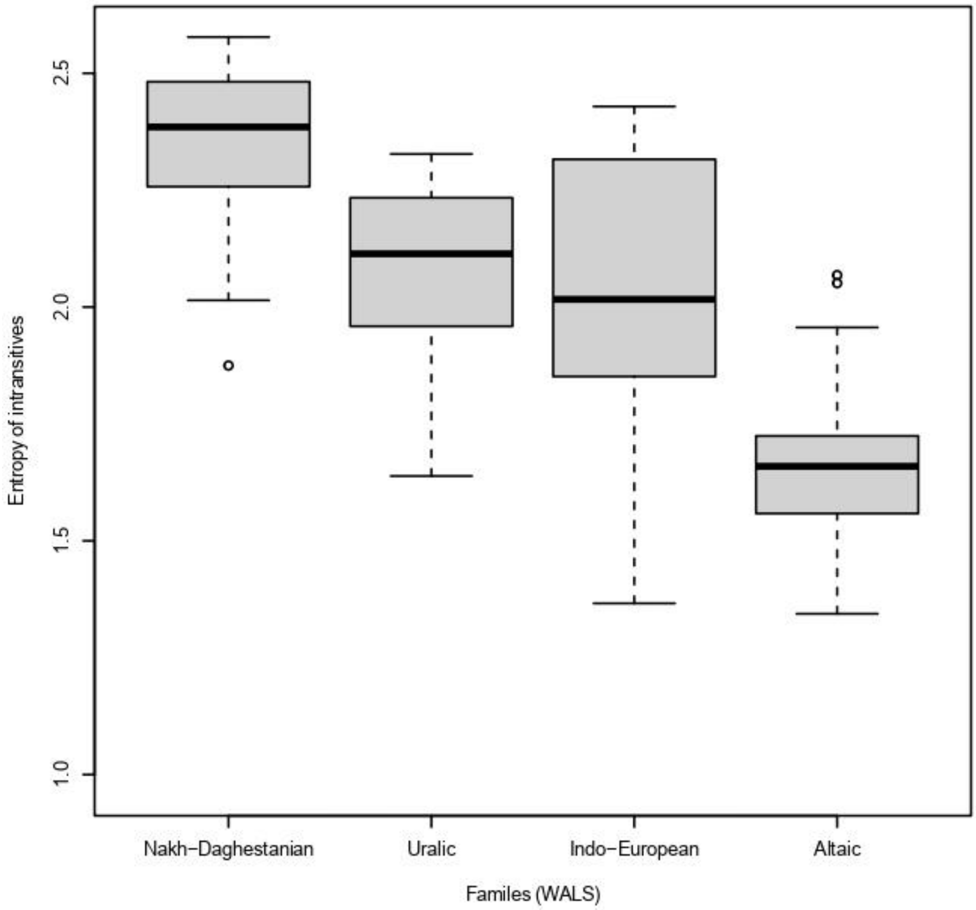

Figure 4: The range of entropy of intransitives across four linguistic families.

The hierarchy of the four Eurasian families that emerges from Figure 4 is Nakh-Daghestanian > Uralic (>) Indo-European > Altaic.[5] It is probable that the significant differences between families are, to some extent at least, shaped by the genealogical factor. Arguably, languages may retain certain features related to valency class systems complexity for millennia. A clear example of this phenomenon is evident in Nakh-Daghestanian languages.

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 corresponds to greater complexity, indicating a stronger functional load associated with argument-encoding devices. In contrast, the transitivity ratio indicates more complexity with lower values. Although entropy of intransitives (H_{intr}) and transitivity ratio are logically independent, they demonstrate a robust negative correlation, emphasizing the relative complexity of valency class system as a fundamental typological feature.

---

[5] The two-level genealogical classification provided by BivalTyp ("families" and "genera") is largely based on WALS (Dryer and Haspelmath (eds.) 2013). Despite using the term "Altaic", I maintain an agnostic stance regarding whether these languages constitute a genealogical taxon. Importantly, these languages are relatively homogeneous in terms of their H and H_{intr} values.

Overall entropy (H) serves as a comprehensive metric, capturing both the prevalence of intransitive 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 appears relatively stable diachronically but can exhibit large-scale areal patterns, suggesting convergent processes over extended periods of time.

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 rejected the categorical universal, the idea still persists in the quantitative era (Maion 2018; Sinnemäki 2010). In fact, the data from BivalTyp also align with the observation that SVO languages favor the lack of morphological case. 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 (Serzant et al., under review). SVO languages have also been claimed to lack agreement morphology more often than other languages (Siewierska and Backer 1996: 125). Overall, SVO languages are often seen as morhologically 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 as the basic word order and morphological simplicity, one would anticipate SVO languages to exhibit less complex valency class systems. 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 (see also a brief discussion in Haspelmath 2015: 132). However, Müller-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 ratio and, consequently, lower entropy of valency classes in SVO languages.

If tested straightforwardly, this prediction is, as it were, confirmed by the BivalTyp data: SOV languages in the sample display a higher average value of H (1.98) compared to SVO languages (1.77); T-test yields the *p*-value of 0.01. However, this result should not be taken at face value. Firstly, when accounting for genealogical family as a random factor in linear mixed-effects models, the influence of basic word order (specifically SOV vs. SVO) on H vanishes.[6] Secondly, none of the models used reveal any significant impact of word order on H_{intr}. Essentially, this implies that there is no systematic difference between SVO and SOV languages in the number of valency-determined contrasts they make among 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

[6] The notable difference highlighted by the plain T-test is likely driven by the Nakh-Daghestanian family, which is prominently featured in the sample. These languages consistently follow the SOV pattern and exhibit exceptionally high entropy values.

literature on the topic, briefly discussed earlier in this section, and also with the BivalTyp data, where SVO languages lack case altogether more often that SOV languages. On the other hand, there are indications suggesting that more specialised 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- vs. postverbal argument-coding devices.

The first 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 things: in reality, both are subject to empirical and theoretical debates, beginning 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 concerns non-canonical A- and O-arguments. 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* (O), 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 O'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 O'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 O'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). This disparity indirectly supports the idea that varied flagging is more common in postverbal arguments than in preverbal ones.

The third piece of evidence concerns the linear position of non-core clause-level constituents, collectively represented as "X", distinct from A, S, and O. 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 (X) 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 (9a) and (9b).

(9) Bambara (Mande)

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.'
(BivalTyp)

In (9a), the pattern is transitive. Remarkably, the preverbal O argument lacks flagging, representing the only encoding option for preverbal NPs. In (9b), the postverbal argument is flagged by the postposition *fɛ̀* 'by', just one of several flagging choices for postverbal arguments. Such contrast 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 hightlighting 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 (10a) and (10b).

(10) Occitan (Indo-European)

a. *lo cinema, i vau sovent*
the cinema there I.go often
'The movies, I go there often.'
(Lambrecht 2001: 1070

b. *i vau sovent, al cinema*
there I.go often to.the cinema
'I go there often, to the movies.'
(Lambrecht 2001: 1070)

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 (11a) and (11b).

(11) Päri (Nilotic)

a. *joobi* *a-keel* *uburr-ì*
buffalo COMPL-shoot Ubur-ERG
'Ubur shot the buffalo.'
(König 2008: 240)

b. *ubur* *joobi* *a-keel-e*
Ubur buffalo COMPL-shoot-3.SG.A
'Ubur shot the buffalo.'
(König 2008: 240)

As König hypothesizes (2008: 271–273), the absence of case markers in sentences like (11b) can be attributed to the grammaticalization of clefted constructions, initially associated with focusing the NP in the sentence-initial position. Ragradless of their precise pathways of diachronic development, constrasts like those between (11a) and (11b) contribute to the greater encoding diversity in postverbal arguments as compared to preverbal arguments.

4.3 Interim summary and discussion

Our 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), which also supports higher argument-encoding complexities in (S)VO languages. 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

5.1 Preliminary remarks

In this section, I further investigate the connections between entropy in valency class systems and other typological parameters. Howeverm here my focus shifts to parameters that are not fully annotated in the BivalTyp dataset or cannot be annotated at all. Consequently, the ideas discussed below are hypotheses derived from preliminary evidence and require systematic testing elsewhere.

5.2 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 (12).

(12) Shinaz Rutul (Nakh-Daghestanian)

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

'Basir hates Karam.'
(BivalTyp)

The distinction between verbal and non-verbal predicates is annotated only for 75 "published" languages in BivalTyp; the remaining 49 languages (so far) lack these annotations. However, for both "published" and "unpublished" languages, argument-encoding devices in non-verbal constructions are annotated using the same principles as in verbal constructions. For instance, the valency pattern of (12), schematically represented as "DAT_NOM", is also found in various simplex verbal predicates in Shinaz Rutul, such as the equivalents of 'know', 'see', and 'love'.

The prevalence of non-verbal predicates varies significantly across the languages of the sample, from just one entry in Mano and Khoekhoe to over 30 entries in Irish and Alik Kryz. More importantly, verbal and non-verbal predicates exhibit different distributions across valency classes. Non-verbal predicates are almost never be transitive (Dixon 2004: 5; although see Lowe 2017 for discussion of counterexamples). This observation is supported by the BivalTyp data: in the published pasrt of the dataset, only three non-verbal entries are identified as transitive patterns. Consequently, the prevalence of non-verbal predicates in a language-specific dataset shows a significant positive correlation with H (R = 0.41, p < 0.001; calculations were based on the 75 published languages). Less trivially, the prevalence of non-verbal predicates also correlates positively with H_{intr}, as illustrated in Figure 5.

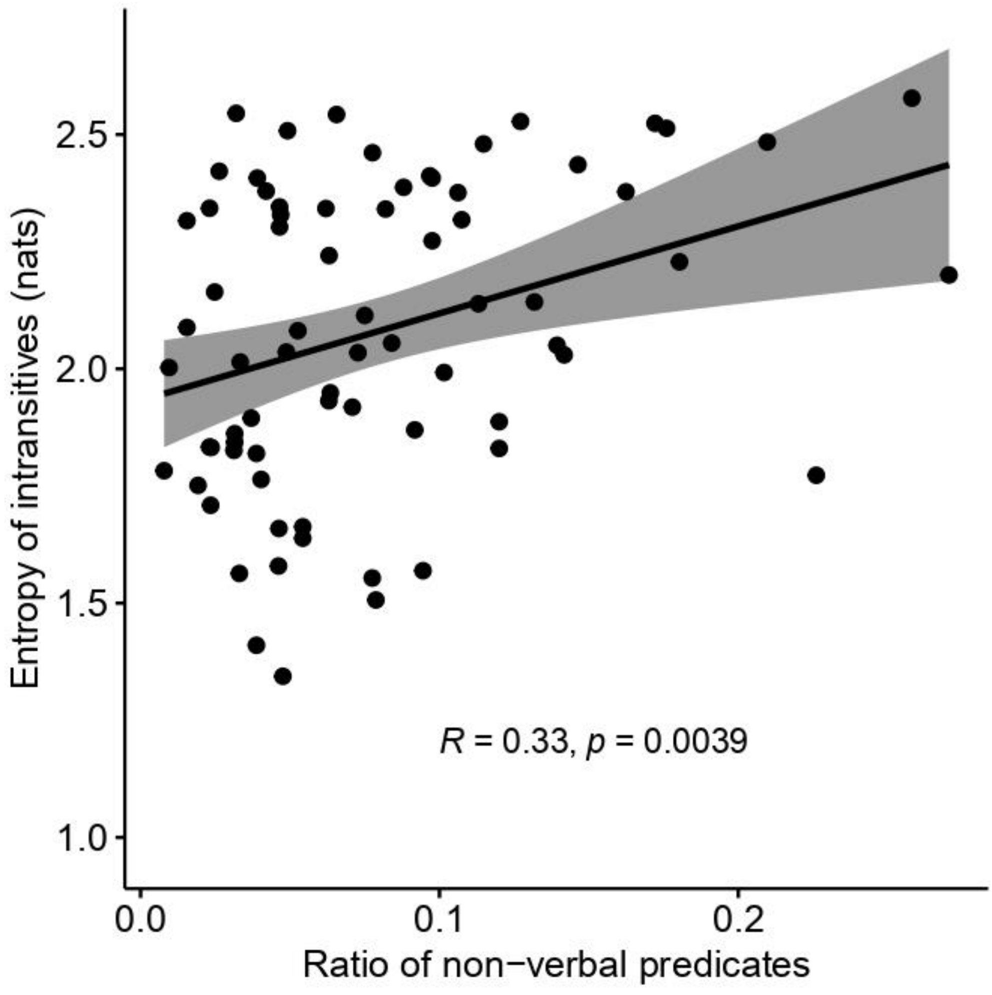

Figure 5: The prevalence of non-verbal predicates and the entropy of intransitives.

Pinpointing one definitive reason for the observed correlation is challenging. One possibility is that non-verbal and verbal predicates favor slightly different types of intransitive patterns, leading languages with a substantial proportion of non-verbal predicates to display more mixed valency class systems compared to languages where most predicates are verbs. Another potential explanation could be related to the broader contrast between “nouny” and “verby” languages, as discussed in Section 5.6.

5.3 Verb structure

It is natural to anticipate some covariance between a verb’s internal structure and its valency pattern, as discussed extensively by Michaelis and Ruppenhofer (2001) in a language-specific context by Wood and Myler (2019) in 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 3rd and 14th 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”, which are prefixal elements associated with specific argument-indexing slots within the verb. In (13), for instance, the second argument is indexed in the slot introduced by the benefactive marker *z-*. Thus, the preverb simultaneously affects the stem structure and the verb’s valency pattern.

(13) Abaza (Northwest Caucasian)
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.’
(BivalTyp)

Given the abundance of applicative preverbs in Northwest Caucasian languages, 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 the so-called “correlation” between verbal prefixes and case markers, illustrated in (14).

(14) Kryz (Nakh-Daghestanian)
yaba *xhin.a-k* *ke-b-cin*
fork.F hay-SUB SUB-F-stick.IMP
‘Stick the fork in the hay!’
(Authier 2009: 126)

In (14), 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 some Dargwic varieties and in most Lezgic languages, except Udi and Archi (Gilles Authier, p.c.). Significantly, these branches exhibit the highest values of H_{intr} among Nakh-Daghestanian languages. Particularly noteworthy is the case of Udi and Archi, which display the lowest values of H_{intr} among the 10 Lezgic 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 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. Whule systematic quantitative evidence is currently lacking, this hypothesis, if correct, can be naturally accounted for by both diachronic and synchronic considerations. Diachronically, verbal prefixes and nominal flags often share a common origin and develop along similar lines. Hence, it is reasonable to expect richer flagging systems in languages that frequently employ valency-related preverbs. From a synchronic perspective, the correlated use of verbal prefixes and spatial flags enhances predictability regarding the choice of specific argument-encoding devices and provides additional cues for sentence processing. This may render hyper-complex valency class systems more manageable compared to languages lacking such cues.

5.4 Satellite-framed vs. verb-framed

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 (15). 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 (16).

(15) English
John limped into the house.
(Beavers et al. 2010: 333)

(16) French

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

Literally, 'I entered the house limping.'
(Beavers et al. 2010: 333)

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 emprical typological studies systematically classifying a substantial number of languages into sattelite- and verb-framed types. Nevertheless, in Table 2 below, I list several languages commonly considered typical representatives of the two types in the literature and provide their H_{intr} values, calculated using the BivalTyp data.[7]

Table 2. Entropy of intransitive verbs (H_{intr}) in presumed satellite- vs. verb-framed languages.

Satellite-framed	$\mathbf{H_{intr}}$	**Verb-framed**	$\mathbf{H_{intr}}$
English	1.92	French	1.83
German	2.34	Spanish	1.84
other Germanic	1.89-2.36	other Romance	1.41-2.18
Russian	2.34	Turkish	1.34
other Slavic	1.89-2.43	Japanese	1.36
Mandarin Chinese	2.38	Hebrew	1.86

While the data in Table 2 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 1.88, while most languages in the right-hand column have lower values.

This generalization also receives support from the substantial characteristics of satellite- vs. 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 (17).

(17) French

Elle traversa la rivière à la nage.

she 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 (14). 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 (17). 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).

(18) English

He ran out of the house through the yard down to the river.

(Lewandowski, p.c.)

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 2, as the BivalTyp questionnaire covers several motion constructions. However, I speculate that the same

[7] 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.

underlying differences transgress the limits of motion constructions and manifest in other constructions surveyed in the BivalTyp questionnaire. This transgression can be attributed to processes of semantic extention, 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 confirmed, 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 3.4). One possible explanation lies in the organization of their spatial markers. Turkic case forms typically convey minimal orientation contrasts, encompassing motion from or through the landmark (Ablative), static position relative to the landmark (Locative), and motion towards the landmark (Dative/Lative). Specific localizations such as ‘under’, ‘behind’, or ‘near’, are typically expressed using secondary postpositions. Crucially, these postpositions are almost never involved in argument encoding with non-spatial verbs, as evidenced by datasets from Turkish, Kazakh, Uzbek, and other Turkic languages in BivalTyp. This constraint sharply constrasts with, for instance, Slavic prepositions, which often have primary spatial meanings but can also encode arguments for a wide array of non-spatial verbs. The limited polysemy patterns observed in secondary spatial markers in Turkic can in turn be linked to the prevalence of verb-framed constructions in these languages.

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.

5.5 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 upon the overarching hypothesis that languages with lower morphosyntactic complexity tend to exhibit higher semantic complexity, we could expect such trade-offs in the realm of valency classes. To test this, we must empirically operationalize both potential correlates. 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) and identified instances where a single lexical item appeared in distinct entries, as illustrated in (19a-b).

(19) Eastern Maninka (Mande)

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 dìriki` bɔ́*
PN PRF 3SG POSS shirt\ART exit
'Seku has taken off his shirt.'
(BivalTyp)

Both sentences are headed by the verb *bɔ́* 'to 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.[8]

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, no significant correlation was observed. In fact, the observed coefficient was positive (R = 0.11), though statistically insignificant ($p = 0.35$).

Several factors may contribute to the lack of significant correlation, including the sample size of 70 properly annotated languages and the inherently limited coverage of polysemy/homophony in the BivalTyp data. I hypothesize that in a larger sample, a significant positive correlation between C and H might emerge, especially driven by languages with poor verb morphology. The few morphologically poor languages in the properly annotated portion of the BivalTyp dataset support this expectation. Eastern Maninka 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 properly annotated African languages. As an illustration, the verb *bɔ́* 'to exit', exemplified in (18a-b), also appears in the database entries corresponding to meanings such as 'to resemble', 'to go out', 'to depend', 'to be different', and 'to cover' and appears in the 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.

The hypothesis I propose here contradicts the initial hypothesis drawn from observations in Fenk-Oczlon and Fenk (2008). However, it can still be viewed as an instance of a cross-domain trade-off. Both the verb and the argument-encoding devices provide cues for interpreting the clause. It is reasonable to assume that underspecified verb lexemes are less tolerated in languages with poor argument-encoding systems, especially those that don't compensate for this deficiency with extensive verb morphology.

From the perspective presented here, upholding Fenk-Oczlon and Fenk's characterization of homophony as a hallmark of semantic complexity becomes problematic. Instead, considering verb lexemes as linguistic tools for conveying information, homophonous verbs inherently possess lower discriminatory potential compared to non-homophonous ones. From this standpoint, high incidence of colexifications resembles, for instance, a limited case inventory, a trait typically associated with simplicity. Ultimately, it seems reasonable to adhere to a general principle whereby "more distinctions indicate higher degrees of complexity" (Shcherbakova et al. 2022: 157; see also Sinnemäki 2009).

However, this controversy primarily revolves around terminology. The key substantive finding is the lack of a negative correlation between the prevalence of colexifications and the complexity of valency class systems. Instead, we might propose a potential positive correlation,

[8] 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'.

especially in languages with poor morphology, where argument-encoding devices and verbal roots are the primary tools for conveying predicative meaning.

5.6 Generalization

Sections 5.2–5.5 contained a discussion of several typological features that may correlate with the complexity of valency class systems. These hypotheses are summarized in Table 3.

Table 3. 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

The four hypotheses summarized in Table 3 all require further quantitative research for validation. Furthermore, even if they are confirmed, this study does not imply a causal direction between the hypothesized correlates. Additionally, it is important to note that the four parameters summarized in Table 3 may be interrelated, potentially leading to spurious correlations with valency class system complexity. To detect this, the analysis should control for potential collinearity effects, which is presently not feasible.

The possible correlations summarized in Table 3 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 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 Summary and 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 verb classes (H_{intr}) are novel. These thre metrics were applied to a dataset from the online database BivalTyp, encompassing information on valency patterns of 130 predicates across 124 languages.

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 prominince 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, gravitate towards mid-range degrees of complexity, covering less than half of the theoretically possible values. This tendency reflects a compromise between speakers and hearers: overly complex systems would be too cognitively demanding for speakers in terms of learnability and production efforts, while overly simple systems would lack important semantic cues, requiring excessive contextual guesswork from hearers.

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. This suggests that valency class system complexity is relatively stable diachronically, which again testifies to the fundamental character of this parameter. Even more importantly, the variation is the degrees of complexity in valency class systems interacts with several other structural parameters. Arguably, high valency class system complexity positively correlates with rich case systems, 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 appealling 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 interact 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.

Abbreviations

3 3rd person
A A-argument
ABS absolutive
ADLAT adlative
ART article
BEN benefactive
COMPL completive
COP copula
DAT dative

DCL declarative
ERG ergative
F feminine
H human
IMP imperative
IO indirect object
IPFV imperfective
LOC locative
M masculine
NOM nominative
OBL oblique
PFV perfective
PL plural
PN person name
POSS possessive
PRF perfect
PRS present
REFL reflexive
SG singular
SUB location 'under'
TR transitive

References

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

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

Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of motion expressions revisited. *Journal of Linguistics* 46. 331–377.

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. In Seppo Kittilä & Fernando Zúñiga (eds.), *Advances in research in semantic roles* [*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.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Drossard, Werner. 1991. Transitivitat (vs. Intransitivitat) 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. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online* (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (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).

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.

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, *Universals of language* (2nd edn), 73–113. Cambridge, MA: MIT 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. Vol. 1. Introducing the framework, and case studies from Africa and Eurasia*, 131–147. Berlin & Boston: De Gruyter Mouton.

Haspelmath, Martin & Iren Hartmann. 2015. Comparing verbal valency across languages. In In Andrej Malchukov & Bernard Comrie, (eds.). *Valency classes in the world's languages. Vol. 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, T. Florian & Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3/4. 866–887.
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. *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.
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. doi: 10.3389/fpsyg.2021.648200.
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 and Leipzig Valency Classes Project team. 2015. Leipzig Questionnaire on valency classes. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages. Vol. I. 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 Sinnemaki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam & Philadelphia: John Benjamins
Moroz, George. 2017. lingtypology: linguistic typology and mapping. R package. Version 1.1.17. https://CRAN.R-project.org/package=lingtypology (accessed 11 May 2024).
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.

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. URL 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.

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. MIT AI Memo 1517. (non vidi).

Say, Sergey. 2014. Bivalent verb classes in the languages of Europe: A quantitative typological study. In *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., Sergey Say, Andreas Hölzl, Aigul Zakirova, Xinyi Gao, Andreas Pregla & Nina Adam. A typology of positional differential argument marking. Under review.

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. Article 58. https://www.nature.com/articles/s41599-022-01072-0 (accessed 11 May 2024).

Shcherbakova, Olena, Volker Gast, Damián E. Blasi, Hedvig Skirgård, Russell D. Gray & Simon J. Greenhill. 2022. A quantitative global test of the complexity trade-off hypothesis: the case of nominal and verbal grammatical marking. *Linguistics Vanguard*, 9(s1), 155–167. doi:10.1515/lingvan-2021-0011 (accessed 11 May 2024).

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

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

Sinnemäki, Kaius. 2014. Global optimization and complexity trade-offs. *Poznan studies in contemporary linguistics* 50(2). 179–195.

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. https://doi.org/10.7717/peerj.10438 (accessed 11 May 2024).

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.

Wood, Jim & Neil Myler. 2019. Argument structure and morphology. In Mark Aronoff (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. http://doi.org/10.1093/acrefore/9780199384655.013.605 (accessed 11 May 2024).

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.