Linguistic Typology at the Crossroads 2-1 (2021): 94-120 ISSN 2785-0943

Verb-argument lability and its correlations with other typological parameters: a quantitative corpus-based study

NATALIA LEVSHINA[1], JOHN A. HAWKINS[2,3]

[1] MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS, NIJMEGEN,

[2]UNIVERSITY OF CALIFORNIA, DAVIS, [3]CAMBRIDGE UNIVERSITY

Submitted: 24/11/2021 Revised version: 28/02/2022

Accepted: 10/03/2022 Published: 14/04/2022

Abstract

We investigate the correlations between lability for verbal arguments with other typological parameters using large, syntactically annotated corpora of online news in 28 languages. We focus on A-lability, when the A argument alternates with S (e.g., *She is singing* vs. *She is singing a song*), and P-lability, when the P-argument alternates with S (e.g., *She opened the door* vs. *The door opened*). To estimate how much lability is observed in a language, we measure associations between Verbs or Verb + Noun combinations and the alternating constructions in which they occur. Our correlational analyses show that high P-lability scores correlate strongly with the following parameters: little or no case marking; weaker associations between lexemes and the grammatical roles A and P; rigid order of Subject and Object; and a high proportion of verb-medial clauses (SVO). Low P-lability correlates with the presence of case marking, stronger associations between nouns and grammatical roles, relatively flexible ordering of Subject and Object, and verb-final order. As for A-lability, it is not correlated with any other parameters. A possible reason is that A-lability is a result of more universal discourse processes, such as deprofiling of the object, and also exhibits numerous lexical and semantic idiosyncrasies. The fact that P-lability is strongly correlated with other parameters can be interpreted as evidence for a more general typology of languages, in which some have highly informative morphosyntactic and lexical cues, whereas others rely predominantly on contextual environment, which is enabled by fixed word order. We also find that P-lability is more strongly correlated with the other parameters than any of these parameters are with each other, which means that it can be a very useful typological variable.

Keywords: verb-argument lability; corpora; Universal Dependencies; word order; case marking; tight-fit and loose-fit languages.

1. Theoretical background

The usefulness of a typological parameter depends on how many other parameters it helps us to predict. Greenberg's (1963) word order correlations have been such a major achievement in linguistics because they connected many diverse and seemingly unrelated word order patterns. In this paper we demonstrate that the strength of attraction of verbs (as well as their arguments) to specific subcategorization frames, which can be defined in terms of verb-argument lability, can be a useful parameter, because it is strongly correlated with many others.

The attraction of verbs to specific subcategorization frames has been argued to be a part of the typology of tight-fit versus loose-fit languages. The terms were coined by Hawkins (1986: 121–127, 1995; see also Müller-Gotama 1994). Generally speaking, tight-fit languages have unique surface forms that map onto more constrained meanings, whereas loose-fit languages have vaguer forms with less constrained meanings and they rely on word-external and contextual features as defined in Hawkins (2019) for the assignment of meanings. For example, grammatical roles in tight-fit languages have a narrower semantic range than grammatical roles in loose-fit languages. The languages Jakaltek and Halkomelem strictly exclude inanimate subjects in transitive clauses (Aissen 2003), while English and Swedish merely strongly disprefer them (Dahl 2000). There are also more gradient distinctions. Both English and German allow for different kinds of subjects, but English is looser than German in the semantic roles that can fill the subject position (Hawkins 1986, 2019), and also than Russian and Korean.

The strength of the associations with grammatical roles is correlated with other linguistic parameters, including more explicit grammatical coding (e.g., formal case marking and use of complementizers and relativizers), avoidance of raisings and long distance WH-movements. Tight-fit languages have fewer instances of syntactic category ambiguity. For example, the English word *book* can be both a noun and a verb, while in German the corresponding noun and verb have different forms, *Buch* – *buchen*. Moreover, verb-final languages are often semantically tight (see also Levshina 2020).

If these parameters change, they often change together. English is a well-known case (Hawkins 1986, 2012). The loss of morphology correlated, in particular, with the emergence of SVO order, long distance movement and raising, greater category ambiguity and other features, including fewer restrictions on the semantics of

syntactic arguments. In contrast, German is more conservative. It preserves case marking, verb-final order (for all verbs in subordinate clauses and for non-finite verbs in main clauses), and it still has some variability in the order of Subject and Object. In addition, German has fewer instances of category ambiguity, tighter associations between semantics and roles, and very limited examples of raising. Generally speaking, English is more structurally ambiguous than German. For example, raising and control constructions are not distinguished formally in surface structure. Compare *Sue happened to win the lottery* (raising) and *Sue hoped to win the lottery* (control). In German, these are distinguished by formally different constructions. In Hawkins' terminology (2019), English relies more on word-external properties to derive meanings from ambiguous or vague surface forms, whereas German relies more on distinct grammatical and lexical patterns and on word-internal properties.

Importantly for our study, verbs in loose-fit languages have a broader set of subcategorization frames than in tight-fit languages. For example, the English verb *open* can be both transitive (e.g, *I opened the door*) and intransitive (*The door opened*), while German distinguishes formally between the transitive *öffnen* "open (tr.)" and the reflexive verb *sich öffnen* "open (intr.)".

The term 'lability' in the context of verb-argument structure is used in the typological literature to refer to alternations between the arguments of one and the same verb (as e.g., P or S with a verb like *open*) or between the arguments of alternating verbs in pairs such as *teach* and *learn* that describe a common event from different perspectives. The precise definitions proposed by different authors have varied in terms of how general or restrictive they are with respect to the alternations in question (compare, e.g., Nichols 1986; Nichols et al. 2004; Dixon 1994; Haspelmath 1993; Letuchiy 2009; Creissels 2014). In the present context we focus on a set of core patterns that most definitions subsume, which can be readily identified in our corpora and in which one and the same verb can be used transitively or intransitively without any formal change while the arguments alternate in their status as A or P or S. Examples from English are *break, open, eat* and *sing*. We are agnostic with regard to the question of which use (transitive or intransitive) is basic, and which is a result of valency derivation (but see Nichols et al. 2004 for insightful discussion of this issue and typological comparison). Syntactically annotated corpora enable us to track the arguments of individual verbs in many different languages and so permit us to quantify the degree of lability within and across languages. The resulting lability

scores can then be tested against other typological variables derived from the same corpora.[1]

More precisely, we consider two types of verb-argument lability, referred to as A-lability and P-lability. A-lability is defined in (1):

(1) A-lability occurs when the A argument of a verb in a transitive clause remains the same in a corresponding intransitive clause from which the P argument has been removed without any formal change to the verb.

In other words, with the same verb, the A-argument can turn into an S-argument: A=S (Dixon 1994).[2] Examples are the unspecified object alternation (2a), the understood body-part alternation (2b) and the characteristic property alternation (2c) (Levin 1993).

(2) a. Unspecified object alternation
Jack ate the cake. - Jack ate.
b. Understood body-part alternation
The Queen waved her hand at the crowd. - The Queen waved at the crowd.
c. Characteristic property alternation
The dog bites strangers. - The dog bites.

P-lability is defined in (3):

(3) P-lability occurs when the same argument can be used as the subject of a verb in an intransitive clause (S) and as the direct object (P) of a corresponding transitive clause without any formal change to the verb.

[1] Notice that the large-scale empirical study described here makes it straightforward to identify alternating arguments for one and the same verb, but does not lend itself readily to identifying alternating arguments among semantically related pairs of verbs such as *teach* and *learn* or *like* and *please*, in the absence of semantic features or tags that can identify the relevant pairs of verbs. Our definition is accordingly more restricted in this respect than that proposed in, e.g., Letuchiy's (2009) summary paper of the typology of lability, but less restricted in other respects (see fn.2).

[2] Note that this type of lability is not considered as such by Letuchiy (2009), for whom lability necessarily involves a change in the semantic role of the subject. For him, verbs like *sing* and *drink* occurring with and without a direct object are 'pseudo-labile'. Letuchiy's definition of lability proper is accordingly more restrictive in this respect than ours (compare fn.1).

In other words, with the same verb S = P (Dixon 1994). Examples are the causative-inchoative alternation (4a), the middle alternation (4b) and the induced action alternation (4c) (Levin 1993).

(4) a. Causative-inchoative alternation
The boy broke the vase. - The vase broke.
b. Middle alternation
The publisher sells the book. - The book sells well.
c. Induced action alternation
She jumped the horse over the fence. - The horse jumped over the fence.

The above-mentioned contrast between English *open* (both transitive and intransitive) and German *öffnen* "open" (transitive only) with its reflexive verb counterpart in the corresponding intransitive *sich öffnen* "open (intr.)" suggests that English has more P-lability (causative-inchoative alternations, in particular) than German. However, this has not yet been examined in corpora and using quantitative measures.

In this paper we fill this gap, measuring A- and P-lability in languages with the help of large corpora, which are described in Section 2. We compute the Mutual Information between verbs, or combinations of verbs and nouns, and the alternating constructions in which they occur. The procedure and the scores are presented in Section 3. Then, we test the correlations between different measures of A- and P-lability and four other variables which have been used in the literature on tight-fit and loose-fit languages and more generally: word order rigidity; the position of the verb in the sentence; case marking; and the strength of associations between nouns and the grammatical roles of Subject and Object (Section 4). Finally, in Section 5 we discuss our findings and conclusions.

2. Data and method

We used the Leipzig Corpus Collection (Goldhahn et al. 2012)[3]. We first selected 30 online news corpora with 1M sentences in each of the following languages: Arabic, Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek (modern), Hindi, Hungarian, Indonesian, Italian, Japanese, Korean,

[3] http://wortschatz.uni leipzig.de/en/download/ (accessed 2022.02.27).

Latvian, Lithuanian, Persian, Portuguese, Romanian, Russian, Slovenian, Spanish, Swedish, Tamil, Turkish and Vietnamese. The corpora were annotated with the Universal Dependencies pipeline *udpipe* (Wijffels, Straka & Straková 2018), which allowed us to extract the subject, the direct object, the predicate, as well as their lemma, part of speech and morphological features, and other useful information. Due to our doubts about the quality and consistency of verb lemmatization in the data from Tamil and Turkish, these languages were later excluded. This is why we had 28 languages in the final sample.

In order to find patterns of A-lability, we extracted the frequencies of all verb lemmas with the same noun in subject position (represented by the Universal Dependency 'nsubj') with and without any kind of nominal or pronominal direct object (the Universal Dependency 'obj'). Consider the examples in Table 1.

A-lability Frequencies			
Verb	**Subject**	**Transitive**	**Intransitive**
be	idea	0	140
learn	student	21	35
play	team	55	47

Table 1: Examples of frequencies relevant for A-lability.

The table shows that the verb *be* with the noun *idea* as subject occurs 140 times (e.g., *the idea was…*), only in intransitive clauses. This is not surprising. The combination *student* + *learn* occurs 21 times with a direct object (e.g., *the students learn languages*) and 35 times without (e.g., *the students learn*). This is an example of A-lability. In this paper, we measure the degree of A-lability by estimating the degree of the skew of a Verb + Noun combination towards the Transitive construction, where the noun is the A-argument, or the Intransitive use, where the noun is the S-argument. The stronger the bias towards one or the other use across different Verb + Noun combinations in a corpus, the weaker the A-lability. If many combinations behave like *play* + *team*, the A-lability will be high. If more combinations behave like *be* + *idea*, the A-lability will be low.

In order to identify examples of P-lability, we extracted the frequencies of all verb lemmas (only predicates of main clauses) with the same noun occurring as direct object and as intransitive subject. Consider the examples in Table 2.

P-lability Frequencies			
Verb	**Noun**	**Intr. subject + Verb**	**Verb + Object**
die	people	64	0
open	door	36	149
begin	work	35	33

Table 2: Examples of frequencies relevant for P-lability.

The numbers should be read as follows. The verb *die* occurs with the noun *people* only as an intransitive subject (64 times), and never as an object. The verb *open* with the noun *door* as intransitive subject (*The door opened*) occurs 36 times, and as a direct object (*I opened the door*) 149 times. This is an example of P-lability. The stronger the skew of the Verb + Noun combinations towards one or the other use, the weaker the P-lability in a given corpus. If many rows contain the frequencies of *begin* + *work,* the P-lability will be high. If more Verb + Noun combinations behave like *die* + *people*, the P-lability will be low.

If we simply counted intransitive and transitive uses of verbs, it would be impossible to distinguish A-lability from P-lability. As will be shown below, making this distinction is crucial, and it is why it was necessary to control for the nouns as A, P or S.

Note that we only selected the verbs that served as predicates of main clauses. Particle verbs and verbs with separable prefixes were treated as one lemma (e.g., *break+out*, *um+leiten*). We also excluded verbs with reflexive, passive, antipassive, middle morphology or auxiliaries because of the substantial cross-linguistic differences in their semantics, formal properties and annotation. One consequence of this decision is that we are primarily measuring looseness vs. non-looseness (the formal marking of which can be quite variable across languages). We also excluded ditransitive clauses, in which one and the same verb had dependencies labelled as 'iobj' (indirect object) and 'obj' (direct object). The measures of lability presented below are based only on combinations of verbs and nouns that occur ten times or more in a corpus.

Lability measures were computed using two methods. According to the first, we controlled for both the verb and the noun, which means that our measures took into account not only the flexibility of the verb with regard to the alternation variants, but also the flexibility of the noun with regard to the roles of A or S (in cases of A-lability)

and S or P (in cases of P-lability). In the second method, we took into account the verbs only, adding up the frequencies of all nouns occurring as A and S, or as S and O with a given verb. The formulas and the scores are discussed in the next section.

3. Measures of lability

3.1. Mutual Information related to A-lability

Using the kinds of frequencies shown in Table 1, we computed Mutual Information (MI) related to A-lability for twenty-eight languages. For Verb + Noun combinations, the formula was as follows:

(5) $$I\,(V\&N;\ Cx) = \sum_{i,\,j} p\left(V\&N_i,\ Cx_j\right) log_2\ \frac{p\left(V\&N_i, Cx_j\right)}{p\left(V\&N_i\right) p\,(Cx_j)}$$

where *V&N* represents Verb + Noun combinations, and *Cx* stands for the constructional alternation, which includes the transitive construction ('nsubj' + Verb + some object) and the intransitive construction ('nsubj' + Verb). The higher MI, the stronger on average the associations between the Verb + Noun combinations and the particular constructions. Therefore, high MI scores indicate weak lability, characteristic of a tight-fit language, and low MI scores correspond to strong lability, characteristic of a loose-fit language.

For verbs only, the formula was as follows, where V stands for a verb:

(6) $$I\,(V;\ Cx) = \sum_{i,\,j} p\left(V_i,\ Cx_j\right) log_2\ \frac{p\left(V_i, Cx_j\right)}{p\left(V_i\right) p\,(Cx_j)}$$

Both types of scores are shown in Figure 1. The languages are ordered by their MI scores based on Verb + Noun combinations, but the two types of scores are strongly correlated: Spearman's rank-based correlation coefficient is 0.97, and the p-value $<$ 0.0001 (but see a more precise measure with genetic dependencies taken into account in Section 4). This means that the measures represent very similar information. The scores based on verbs only are lower in all languages, but the ordering is more or less the same, as the high correlation coefficient suggests. The highest scores are found in Portuguese, followed by Italian, Hindi, English and Slovene. The lowest score belongs to Lithuanian, followed by Vietnamese, Korean, Arabic and Persian. This ranking is

not predictable from any typological, genealogical or areal properties of the languages.

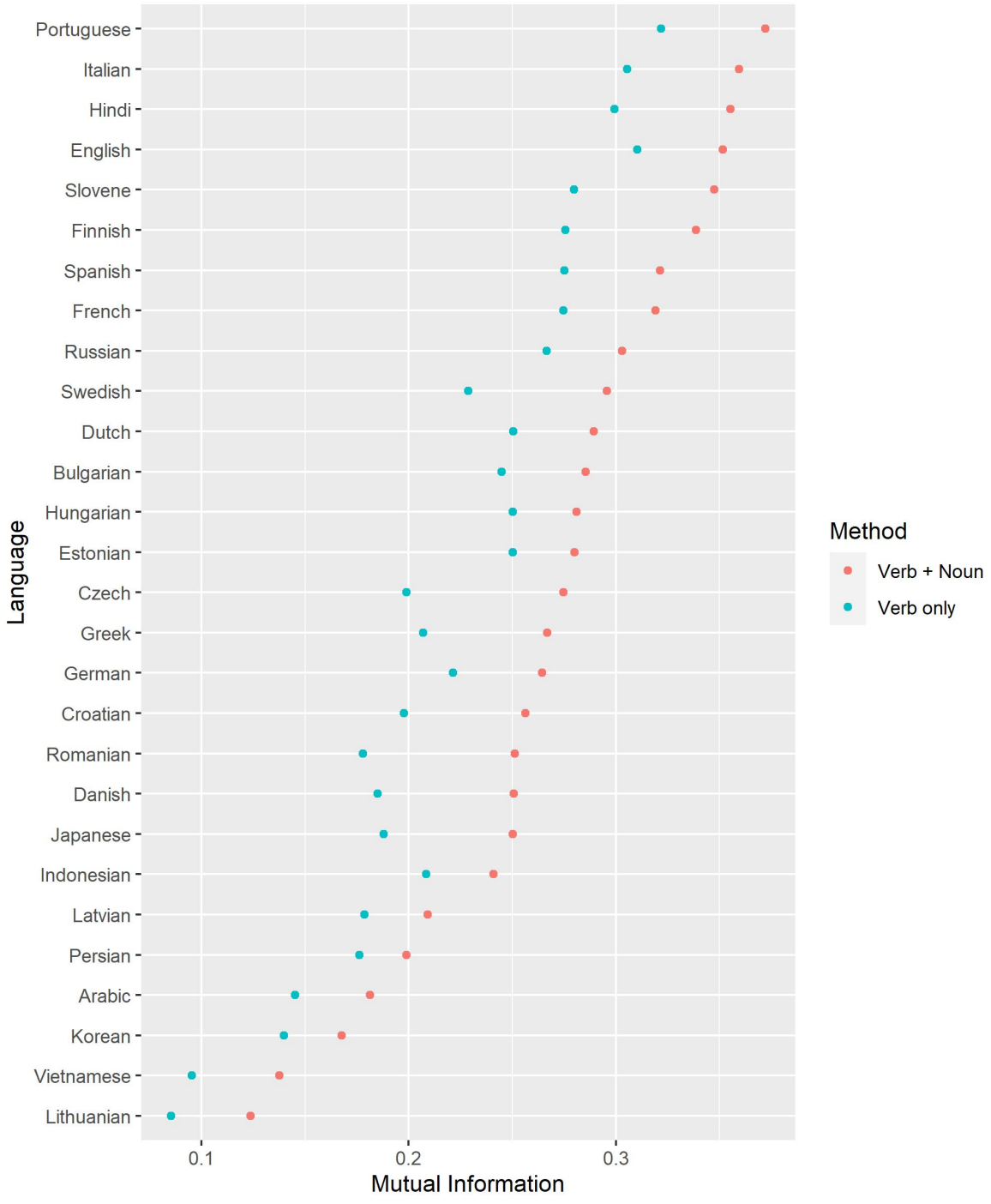

Figure 1: Distribution of MI scores representing A-lability. The higher the score, the weaker this type of lability in a language.

3.2. Mutual Information related to P-lability

To compute MI related to P-lability, we used the same approach as for A-lability, but took the frequencies of verbs and nouns in the construction 'nsubj' + Verb without object and the construction Verb + 'obj' (regardless of the presence or absence of any subject). The two methods, Verb + Noun (as 'nsubj' or 'obj') and Verb only, yield scores that are highly correlated: Spearman's correlation coefficient rho is 0.96, with the p-value < 0.001.

Figure 2 displays both types of MI scores. The top scores belong to Hungarian, Russian, Estonian, Latvian, Korean and Finnish. The high scores mean that these languages have strong associations between the Verb + Noun combinations and the constructions in which they appear as 'nsubj' or 'obj' respectively, characteristic of tight-fit languages. These languages also have formal case marking and relatively free word order of the core arguments. Many of the languages at the top are verb-final, or at least allow for the V-final order. The two languages at the bottom are Indonesian and Vietnamese, followed by English, French and Romanian. These have weaker associations between the Verb + Noun combinations and the constructions in which they appear as 'nsubj' or 'obj'. So they display stronger P-lability characteristic of loose-fit languages. They also have fairly rigid SVO order and no case morphology.

If we compare the range of values in Figure 1 and Figure 2, we see that the MI scores related to A-lability are on average lower than the MI scores related to P-lability. This impression is supported by paired Wilcoxon tests. The difference between the A- and P-lability scores is significant for both methods ($p = 0.028$ for verbs only, and $p < 0.001$ for Verb + Noun combinations). This means that languages are more tolerant with regard to A-lability in general. In addition, the spread of the P-lability scores is greater, which suggests more substantial cross-linguistic differences.

3.3. How variable is lability across individual verbs and arguments?

An important question is whether low MI scores, which characterize high-lability languages, are due to most verbs being weakly associated with a specific construction, or whether they are primarily influenced by a handful of very frequent idiosyncratic verbs with high verb-argument lability, e.g., English *break* or *learn.*[4]

[4] We thank Thomas Hörberg for pointing out this important distinction to us.

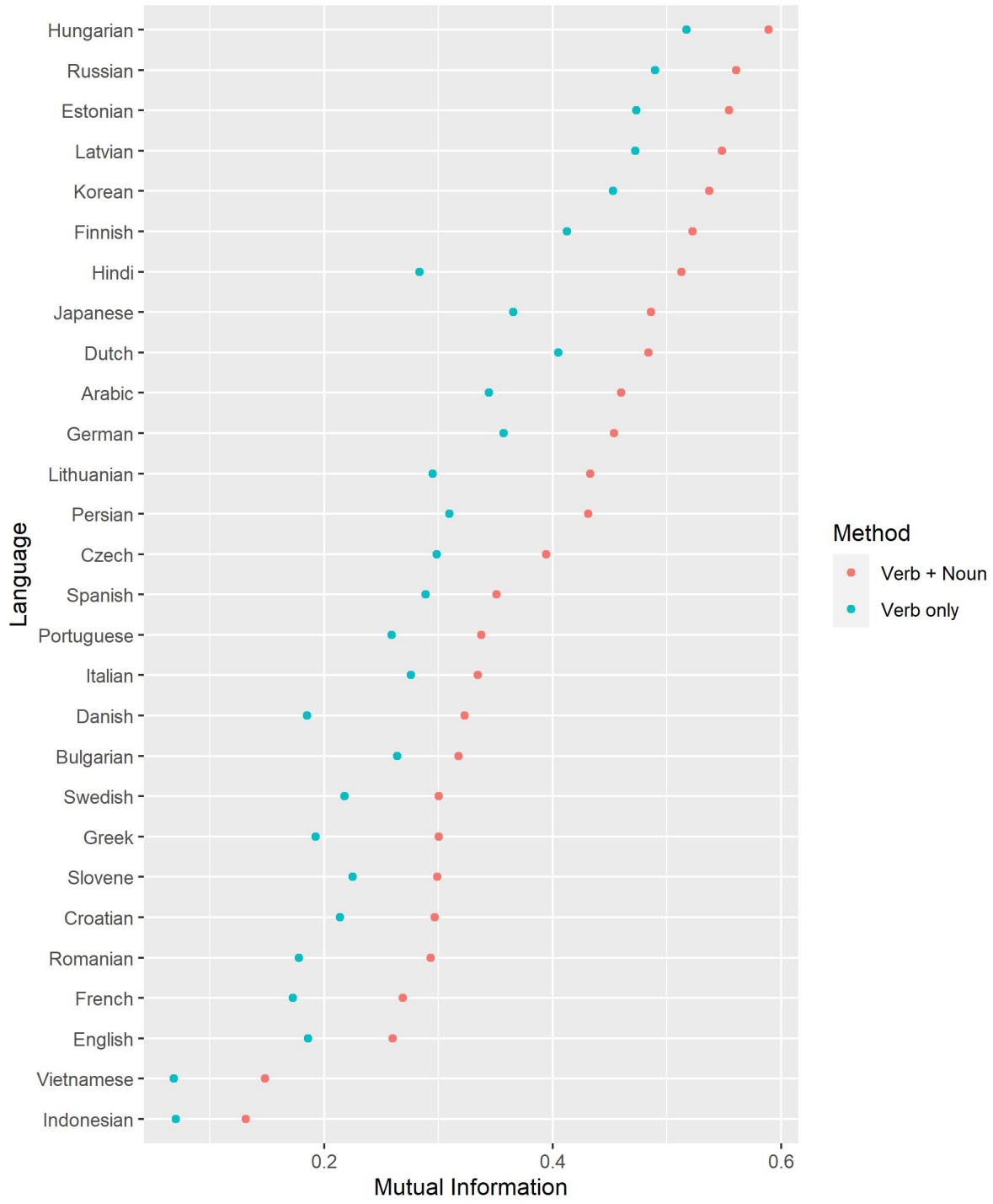

Figure 2: Distribution of MI scores representing P-lability. The higher the score, the weaker this type of lability in a language.

In order to answer this question, we computed how strongly the frequencies of a Verb + Noun combination (or of a Verb only) deviated from the expected proportion based on the total frequencies for the alternating constructions in the datasets. This measure is called Kullback—Leibler Divergence (D_{KL}) in information theory. It is also known as

relative entropy. The formula for a Verb + Noun combination *V&N* and Construction *Cx* is as follows:

$$(7) \quad D_{KL}\,(P||Q) = \sum_j P\left(V\&N,\ Cx_j\right) log_2\ \frac{P\left(V\&N, Cx_j\right)}{Q\ (V\&N, Cx_j)}$$

where P stands for the observed proportions of the specific Verb + Noun combination in each of the constructional variants, whereas Q represents the probability of this constructional variant for all Verb + Noun combinations, derived from the corpus frequencies. The higher the measure, the more this Verb + Noun combination contributes to MI. For verbs only, the procedure is identical, but instead of the observed and expected proportions of constructional variants in all Verb + Noun combinations we compute the proportions of constructional variants for Verbs only.

As a result, we obtained D_{KL} scores for individual combinations of Verb + Noun and for Verbs only. For example, the combinations *drive* + *people*, *open* + *room*, *begin* + *work*, *show* + *video* and *ask* + *woman*, had very low D_{KL} scores for P-lability because the Noun in these combinations occurred as S and P of the Verb with comparable frequencies. In contrast, the combinations *have* + *opportunity*, *do* + *job*, *play* + *role*, *score* + *point* and *make* + *sense* had very high D_{KL} scores for P-lability because the Noun was only used as P with these verbs.

The average D_{KL} scores were strongly correlated with the analogous MI scores presented in Sections 3.1 and 3.2. When computed on the A-lability data, Pearson's correlations were $r = 0.89$ (Verb + Noun combinations) and $r = 0.67$ (Verbs only), both $p < 0.001$. As for P-lability, the correlations were even stronger: $r = 0.98$ for Verb + Noun combinations and $r = 0.86$ for verbs, both $p < 0.001$. This means that both measures reflect similar information.

In order to estimate how strongly the individual Verbs or Verb + Noun combinations vary, we computed standard deviations of D_{KL} for every language. The higher the standard deviations, the more variability there is between the Verb + Noun combinations or individual verbs with regard to their association to one or the other construction.

Figures 3a and 3b display the MI scores and the standard deviations of the D_{KL} scores for A-lability. Figure 3a shows the scores for Verb + Noun combinations, and Figure 3b displays the scores for Verbs only. There are no obvious correlations between the measures in either plot.

Now let us turn to P-lability. Figure 4a shows the MI scores and the standard deviations of the D_{KL} scores for Verb + Noun combinations. There is a clear negative correlation: lower MI (and higher P-lability) scores correspond to higher D_{KL}, although there is also quite a lot of variability in the middle of the plot. Figure 4b displays the same data, but for Verbs only. In this case, the negative correlation is even more obvious.

How can we interpret these results? The looser languages on the left-hand side of the plot are loose due to certain individual Verb + Noun combinations which are highly labile, while other verbs are less or not labile. By contrast, if a language is generally tight, as are the languages on the right-hand side, the individual verbs in such a language are quite uniformly tight.

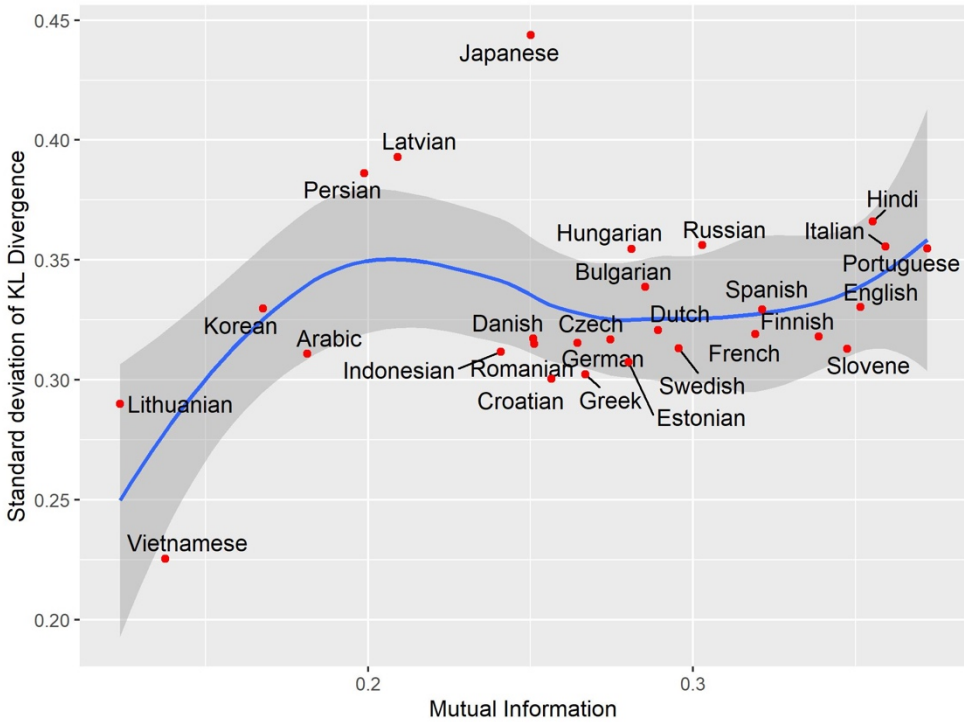

Figure 3a: MI scores representing A-lability (horizontal axis) and the standard deviations of the D_{KL} scores for individual Verb + Noun combinations (vertical axis). The curve is based on the LOESS smoothing method.

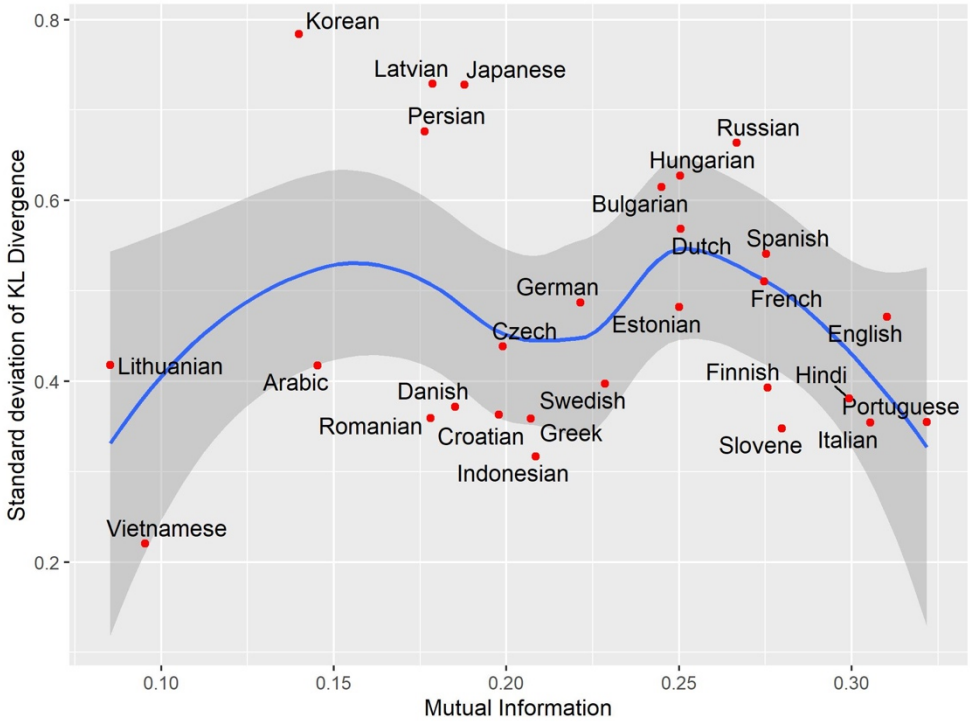

Figure 3b: MI scores representing A-lability (horizontal axis) and the standard deviations of the D_{KL} scores for individual Verbs (vertical axis). The curve is based on the LOESS smoothing method.

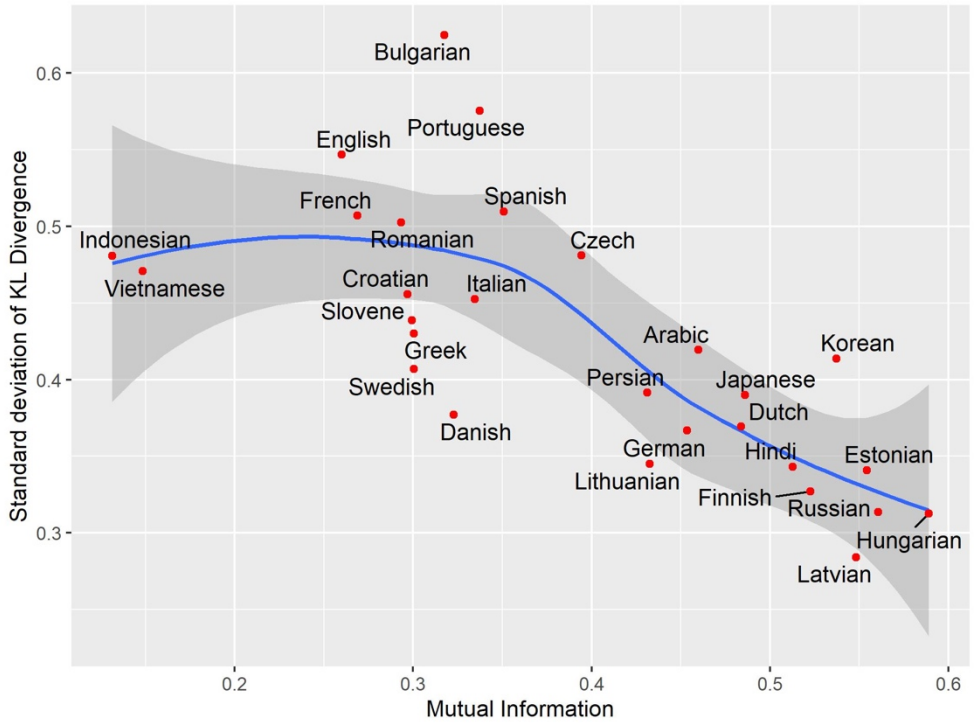

Figure 4a: MI scores representing P-lability (horizontal axis) and the standard deviations of the D_{KL} scores for individual Verb + Noun combinations (vertical axis). The curve is based on the LOESS smoothing method.

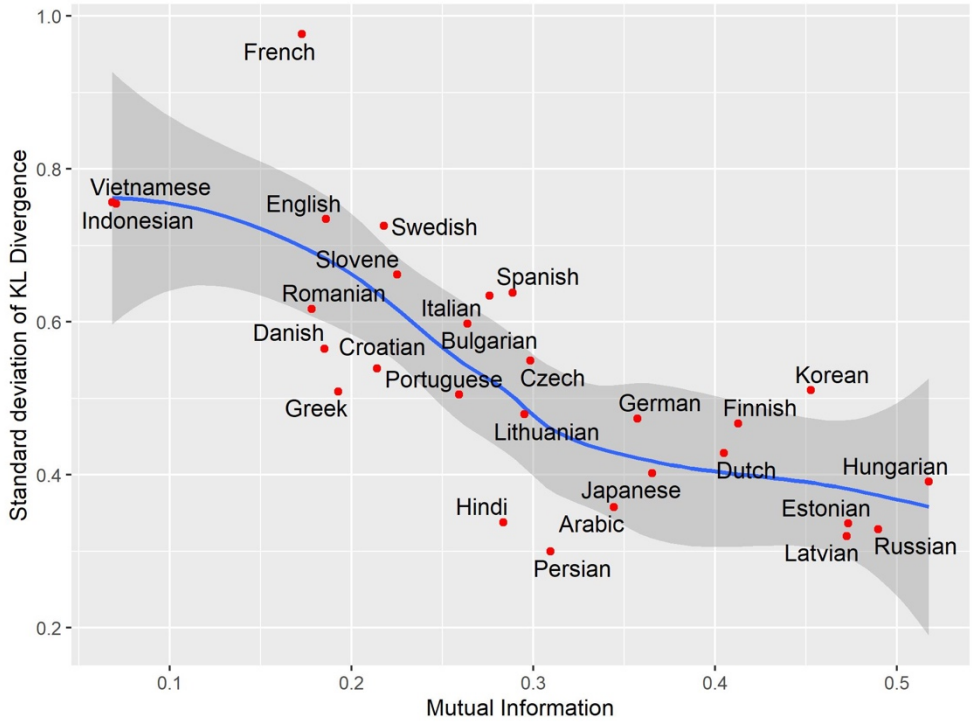

Figure 4b: MI scores representing P-lability (horizontal axis) and the standard deviations of the D_{KL} scores for individual Verbs (vertical axis). The curve is based on the LOESS smoothing method.

4. Correlations with other typological parameters

In this section we test for correlations between the A-lability and P-lability scores based on Mutual Information and the following parameters: rigidity of Subject and Object order; position of the lexical verb in the clause; case marking; and associations between lexemes and grammatical roles, which serves as a proxy for semantic tightness. We recycle the data from Levshina (2021), where the parameters were estimated by using the same online news corpora. More specifically, rigidity of Subject and Object order was computed as 1 minus entropy of SO and OS orders. To compute entropy, we used Shannon's (1948) famous formula, as shown below:

(8) $H(X) = -(P(SO)\,log_2\,P(SO) + P(OS)\,log_2\,P(OS))$

The proportions of SO and OS orders in transitive clauses were computed first based on the corpora, and then these entropy scores were computed (see Levshina 2019). If the proportions of SO and OS orders are equal (0.5), this leads to an entropy score of 1. If only one of the orders is used (either SO or OS), this leads to zero entropy. Since

entropy represents word order variability, we subtracted the entropy scores from 1 in order to obtain measures of word order rigidity. Lithuanian, Hungarian, Latvian, Czech and Estonian had the lowest scores and therefore the most variable orders, and Indonesian, French, English, Danish and Swedish had the highest scores and thus the most rigid orders. Note that in all languages, the SO order was the more frequent one. So we can speak about the rigidity of SO order. This variable was called 'Rigid Order (SO)'.

Another measure was the proportion of main clauses with a lexical verb between the Subject and Object. As expected, it was near-zero in verb-final languages, such as Japanese, Korean, Persian and Hindi, and close to one in Indonesian, English, French, Vietnamese and Portuguese. This variable was labelled 'Verb between Subj and Obj'.

We also took into account how much case marking was present to help in identifying the Subject and the Object. In Levshina (2021), the scores represented Mutual Information between case and the corresponding grammatical roles (A and P). For languages with adpositional case marking, the data were extracted automatically. As an illustration, consider the frequencies for Spanish in Table 3.

Case	**Transitive Subject**	**Direct Object**
Zero marking	126,736	569,252
Preposition *a*	0	55,442

Table 3: Frequencies of zero case marking and the direct object marker *a* for Subject and Object in Spanish.

For languages with case morphology, random samples were drawn and analyzed manually in order to take account of case syncretism. Next, the counts were extrapolated to all occurrences of Subjects and Objects in transitive clauses in a corpus. Consider an illustration in Table 4, which contains frequencies for Russian. In languages with distinct forms for Subject and Object and also forms with case syncretism, as in Russian, these three situations were represented by separate rows.

Case	**Transitive Subject**	**Direct Object**
Nominative	47,521	0
Accusative	0	93,520
Nominative/Accusative (case syncretism)	42,884	246,361

Table 4: Frequencies of Nominative, Accusative and case syncretism forms in Russian

German was a special case, where all feminine, neuter and plural forms were treated as ambiguous, since their Nominative and Accusative forms are formally indistinguishable, whereas masculine nouns were analyzed as Nominative or Accusative only in the presence of determiners or adjectives, which normally carry the distinct marking in combination with the noun. See more details about the procedure in Levshina (2021).

Based on numbers like those displayed in Tables 3 and 4, we computed the Mutual Information between cases (*C*) and grammatical roles (*R*) for each language, using the formula in (9).

(9) $$I\,(C;\,R) = \sum_{i,\,j} p\left(C,\,R_j\right) log_2\,\frac{p\,(C_i, R_j)}{p\,(C)\,p\,(R_j)}$$

The higher the Mutual Information, the more strongly the case forms are associated with the grammatical roles in question. Languages with zero scores had no case marking on Subject and Object (Danish, Dutch, English, Indonesian, Swedish and Vietnamese). Languages with the highest scores were those with rich morphological case marking (Lithuanian, Hungarian, Latvian, Estonian and Japanese). Languages with some type of differential, lexically restricted or optional marking were in-between (the Slavic languages, Hindi, Korean, German, Persian and Turkish). The variable with these scores was called 'Case Marking'.

Finally, we took the Mutual Information between Nouns only and the grammatical role of Subject and Object as a proxy for semantic tightness. If the proportions with which a Noun is found as a transitive Subject and Object are similar to the baseline proportions of Subject and Object, this contributes to the semantic looseness of a language. If a Noun is strongly biased towards one of these roles, this increases its semantic tightness (see Levshina 2021 for more details). The higher the Mutual Information, the tighter the language. The languages with the highest scores were Hindi, Korean, Russian, Hungarian and Japanese. They are known as tight-fit languages in the literature with regard to the relationships between arguments and their semantics. Indonesian had the lowest score, followed by English and Spanish. These were the loosest languages in our sample. This tightness measure was labelled as 'MI Nouns'.

The correlation analyses were based on Spearman's rank-based correlations. In order to control for the genealogical dependencies in our data (i.e., the fact that many languages come from one and the same genus), we used a sampling procedure

whereby we created 1,000 samples. For every sample, we drew randomly only one language per genus and computed the correlation coefficient (ρ). For null hypothesis significance testing, the test statistic was first computed and logged for the original pairs of scores in every simulation. Next, we also ran 1,000 permutations, in which the original scores of the second variable were randomly reshuffled. The permutation scores helped us to capture the distribution of the test statistic under the null hypothesis. Next, we counted the number of cases out of 1,000 permutations where the permuted scores were equal to or more extreme than the original test. After we had these data for all samples, we averaged the coefficients and the *p*-values.

Figure 5 represents the correlation coefficients between the parameters, with the p-values being shown in Figure 6. One can see in Figure 5 that both types of P-lability scores (Verbs and Verb + Noun) are correlated with the other typological parameters.

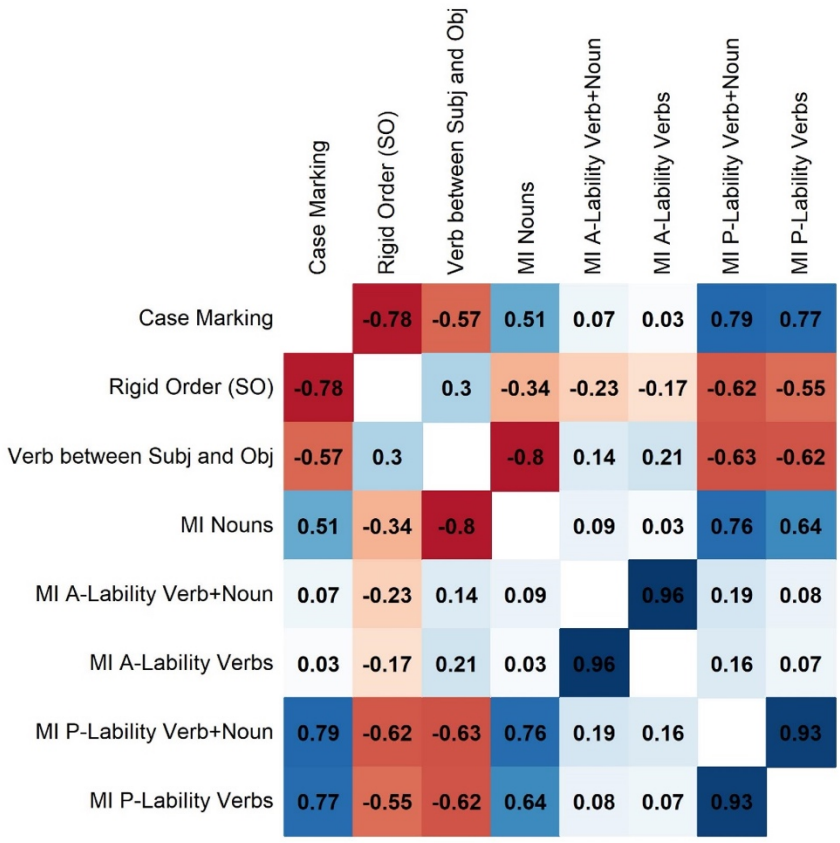

Figure 5: Correlations between the typological parameters and lability scores. The colour intensity represents the strength of the correlation. Blue cells stand for positive correlations. Red cells display negative correlations.

	Case Marking	Rigid Order (SO)	Verb between Subj and Obj	MI Nouns	MI A-Lability Verb+Noun	MI A-Lability Verbs	MI P-Lability Verb+Noun	MI P-Lability Verbs
Case Marking		**0.001**	**0.025**	**0.052**	**0.352**	**0.351**	**0.002**	**0.004**
Rigid Order (SO)	**0.001**		**0.161**	**0.141**	**0.233**	**0.278**	**0.02**	**0.037**
Verb between Subj and Obj	**0.025**	**0.161**		**0.001**	**0.314**	**0.245**	**0.015**	**0.017**
MI Nouns	**0.052**	**0.141**	**0.001**		**0.336**	**0.34**	**0.002**	**0.01**
MI A-Lability Verb+Noun	**0.352**	**0.233**	**0.314**	**0.336**		**0**	**0.27**	**0.375**
MI A-Lability Verbs	**0.351**	**0.278**	**0.245**	**0.34**	**0**		**0.295**	**0.37**
MI P-Lability Verb+Noun	**0.002**	**0.02**	**0.015**	**0.002**	**0.27**	**0.295**		**0**
MI P-Lability Verbs	**0.004**	**0.037**	**0.017**	**0.01**	**0.375**	**0.37**	**0**	

Figure 6: Average p-values based on resampling and permutation

The correlation between both types of MI scores for P-lability and case marking is strong and positive. This means that languages with systematic case marking have low P-lability. There is also a positive correlation between MI related to P-lability and MI based on Nouns only. At the same time, P-lability scores are negatively correlated with verb-medialness and word order rigidity. This means, in turn, that languages with SVO and rigid SO order have more P-lability. Judging from the magnitude of the coefficients, we can also see that the P-lability scores based on verbs only are overall less strongly correlated with the other typological parameters than the P-lability scores based on Verb + Noun combinations.

We also observe significant negative correlations between rigid SO order and case marking, and between verb-medialness and case marking. In addition, there is a strong and significant negative correlation between Mutual Information based on associations between nouns and grammatical roles, and verb-medialness.

The A-lability scores do not participate in any significant correlations, however. They are only strongly correlated between themselves. This means that we do not find evidence that A-lability is correlated with any of the typological parameters.

5. Conclusions

Our quantitative analyses reveal that P-lability scores are systematically correlated with the other parameters related to tight and loose fit. Languages with low P-lability (and high MI scores) tend to have case marking, stronger associations between nouns and grammatical roles, relatively flexible order of Subject and Object, and verb-final order. These features are associated with tight-fit languages. In contrast, languages with high P-lability (and low MI scores) tend to have little or no case marking, quite rigid SVO order, and weaker associations between nouns and grammatical roles. These features are associated with loose-fit languages. Therefore, our data support Hawkins' (1986, 1995) observation that verbs in loose-fit languages are used in more diverse subcategorization frames, while the argument co-occurrences of verbs in tight-fit languages are more constrained. More generally, our analysis provides further evidence that the classification of languages into loose- and tight-fit is a theoretically useful one.

In order to explain why these and other grammatical features of this general typology correlate in the ways they do Hawkins (1995, 2014, 2019) appealed to language processing, and especially to the crucial role of the verb in online processing. At issue is: does the verb precede its co-occurring arguments (as in VSO and SVO languages), or do these latter precede the verb (i.e. SOV)? In the former case there is early activation in processing of the verb's co-occurrence possibilities (with temporary ambiguities and possible garden paths when parsing English-type languages as the intended frame is gradually selected). In SOV languages, co-occurrence frame activation and selection based on preceding material will be almost instantaneous.

Hawkins (1995, 2014, 2019) formulated predictions for correlating properties based on this difference. Languages with verb-final structures (like Japanese and Korean) should exhibit what he called greater 'predicate frame differentiation' and 'argument differentiation'. Predicate frame differentiation refers to the degree to which a verb is distinctive from others by virtue of its unique selectional restrictions or syntactic co-occurrence possibilities. A verb that is uniquely transitive is more differentiated, and less labile in the terminology of the present paper, than one that is ambiguously transitive or intransitive. A verb that selects restricted direct objects for 'putting on clothing' according to the body part and the type of clothing in question (hats on the head, a coat over the rest of the body, etc.) is more differentiated than one (like English *put on*) that is compatible with many different types of body parts

and clothing activities (Planck 1984, Hawkins 1986). Argument differentiation refers to the degree to which arguments are, e.g., differentially case-marked versus ambiguous as to case, and the degree to which they are assigned a narrow set of thematic roles like Agent and Patient rather than the broader set of English (which permits e.g., Locative and Instrumental subjects as in *This tent sleeps four people* and *The key opened the door*).

When the verb is the last constituent in the clause and the very next item to be parsed belongs in an altogether different clause, the parser must succeed instantly in selecting the correct frame and its arguments. The grammar and lexicon must, Hawkins hypothesized, help the parser of a verb-final language by ensuring that predicate frame selection is immediately successful. More differentiated predicate frames and arguments can accomplish this, and in a number of observable ways.

First, subcategorization restrictions can be made tighter in SOV languages. For example, it is desiderable to avoid transitive/intransitive ambiguities with the result that arguments can be paired with their predicates more uniquely and more easily and with less variability and lability. Second, additional selectional restrictions can be imposed so that certain verb-NP pairings are more constrained, frequently co-occurring and easily recognizable. Third, subjects and objects can be made less semantically diverse in SOV languages with the result that there are fewer co-occurrence possibilities to choose from and more constrained assignments of thematic roles to NPs (transitive subjects are agents, transitive objects are patients, etc.). We expect fewer assignments of thematic roles to a transitive subject such as Location and Instrument in these languages. Fourth, surface coding devices can be grammaticalized for arguments that permit immediate thematic role recognition. This is what case marking generally does. It constrains the thematic roles that can be assigned to surface NPs, making them less semantically diverse (and also making them less labile) compared with case-less languages. This has the dual advantage of making thematic role information available early on-line, prior to the verb, and of facilitating argument-predicate assignments once the verb is encountered. And indeed it has long been known that there is a strong correlation between SOV and case marking, as confirmed in the correlations of the present paper.

More generally we expect to see verb-final languages with a more constrained set of verb co-occurrence possibilities. We expect more predicate frame differentiation and more argument differentiation, but less 'argument trespassing', which Hawkins (1995, 2014, 2019) defined to mean less movement of NP arguments into clauses in

which they contract no semantic relations with their most immediate predicates. For verb-early (SVO, VSO, VOS) languages, however, no such constraints are predicted. These languages do not need to conventionalize devices that permit immediate and correct predicate frame recognition at the verb, because the parser still has the remainder of the clause in which to complete its predicate frame selection and argument recognition, and because aspects of the verb's processing (e.g. its precise interpretation and disambiguation) that depend on subsequent arguments cannot be identified at the verb anyway. As a consequence, the need for immediate and correct decision-making at the verb, and the resulting need for clear predicate-frame differentiation, argument differentiation, and for local argument-predicate matching will impose much weaker requirements on the grammars and lexicons of such languages. This is why verb-early languages are quite variable: Indonesian (SVO) reveals many similarities with English, for example, whereas Hebrew (SVO) has far fewer (Müller-Gotama 1994). Verb-final languages should be more constrained in these respects, therefore, whereas languages with earlier verbs in the clause are predicted to be more variable. Verb-final languages are accordingly generally tight-fit, whereas verb-early and verb-medial languages can be tighter (like Hebrew) or looser (like English), as we have seen in this paper.

What ultimately underlies the tight vs loose typology, as explained in Section 1, is the mapping between forms and meanings. Tight-fit languages have richer, more complex and more unique surface forms that map onto less ambiguous and more constrained meanings, i.e. there is more of a one-to-one correspondence between form and meaning. This simplifies the mapping between them, but at the expense of processing more complex forms, for example case-marked nouns. Loose fit languages involve simpler processing of their more minimal and semantically general forms like caseless nouns, but they require more complex contextual processing and disambiguation of meaning assignments through 'word-external properties', see Hawkins (2019). This trade-off can be measured and made more precise in terms of the different MI and lability scores that we have seen in this paper for different languages, in the assignments of different nouns to grammatical roles, and in the further correlations with case marking, SOV, verb-mediality and rigid order. For further exploration of the processing basis for these correlations in terms of competing efficiencies and the general theory of efficiency in processing (as laid out in Gibson et al. 2019; and Levshina 2022), and for suggestions for psycholinguistic experiments

that need to be conducted on different language types from this perspective, the reader is referred to Hawkins (2019).

We have found further that P-lability scores based on verbs only are overall less strongly correlated with the other typological parameters than the P-lability scores based on Verb + Noun combinations. This is not surprising, because these latter scores also include the attraction of nouns to different grammatical roles. At the same time, both of these scores are more strongly correlated with word order rigidity and case marking than semantic tightness scores based on nouns only, but the latter has a stronger correlation with verb-final order (since the languages in the sample, except for Arabic, are either verb-medial, or verb-final). This may have to do with the fact that the attraction of nouns to one or the other role helps to avoid costly reanalysis when the verb comes last. Whether or not the verb has special marking depending on the roles of its arguments is less important for that purpose.

It is remarkable that our P-lability scores are more strongly correlated with the other typological parameters than the latter are among themselves. This is an unexpected finding, but it can be explained by the fact that P-lability scores convey information not only about the verbs, but also (explicitly or implicitly) about the nouns in different roles. These scores can thus be a useful indicator of the word-external or word-internal orientation of the language in question (Hawkins 2019).

We should also mention that high P-lability for a language in general seems to be due to some highly labile verbs, which behave 'promiscuously', appearing in both constructional frames. At the same time, there are also lexemes with low promiscuity. In contrast, the verbs or combinations of verbs and nouns in tight languages with low P-lability tend to be uniformly faithful to one or the other constructional frame. These conclusions are based on our analysis of lexical variability based on Kullback—Leibler Divergence scores (also known as relative entropy).

In contrast to P-lability, the A-lability scores are not correlated with any of the typological properties we have examined here. A-lability is also found more frequently in our corpora than P-lability, as we see from the lower MI values in the former. A possible explanation for this is that A-lability is often driven by general pragmatic factors. For example, the object can be omitted due to its high accessibility, e.g., *And Italy wins [the final]!* Many objects are omitted due to specific conventionalized inferences, *e.g., He drinks again [liquor].* Object omission is also possible if the focus is on the action, while the object has low discourse prominence, e.g., *She chopped and chopped [e.g., meat]* (Goldberg 2005). Other reasons are cultural.

For example, the object can be omitted when it is taboo, e.g., *Pat sneezed [mucus] onto the computer screen*, or for feelings of tact, *I contributed [$1,000] to UNICEF* (Goldberg 2005). In addition, many rules allowing for object omission are also lexically and semantically specific (Fillmore 1986). All these pragmatic factors and lexical idiosyncrasies explain the lack of systematic correlations between A-lability and the other typological properties of the languages in question. It would be very interesting to investigate if there exist cross-linguistic regularities at the level of individual verbs and verb classes, and if so which semantic roles alternate (cf. Letuchiy 2009). As mentioned above, it would not be realistic to annotate manually all verbs for their semantics in the twenty-eight languages, so we have to leave this question for future research.

We also hope to have demonstrated that information-theoretic measures, such as Mutual Information, entropy and Kullback-Leibler Divergence, can be fruitfully used for language comparison. As initiatives such as the Universal Dependencies progress and typologically diverse large corpora are made available it becomes increasingly possible to infer typological variables directly from texts and to build on, and refine, earlier typological patterns derived largely from grammars, as in the present study. We hope that new corpora and tools will allow for further testing of the findings of this study using a larger and more diverse sample of languages and genres. A further question is then whether there are causal relationships between the parameters themselves, and what they would look like exactly. A causal analysis in Levshina (2021) showed that case marking is more likely to be affected by other typological parameters (word order and associations between lexemes and syntactic roles) than the other way round. We need a larger sample of languages in order to answer this question and to test all possible causal links. It would also be interesting to add the other parameters of Hawkins (1995, 2014, 2019), such as the frequency of long-distance syntactic dependencies or of categorial ambiguity, and to test their relationships with the parameters examined here.

Data availability statement

The dataset with the Mutual Information and other values for every language in the sample is available as supplementary materials on the journal website.

Acknowledgements

The first co-author's work on this paper was funded by the Netherlands Organisation for Scientific Research (NWO) under Gravitation grant "Language in Interaction", grant number 024.001.006. We thank the reviewers for their insightful and constructive feedback, which has helped us improve the paper. All remaining errors are only ours.

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21. 435-483.

Creissels, Denis. 2014. P-lability and radical P-alignment. *Linguistics* 52(4). 911-944.

Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of Language* 7(1). 37-77.

Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Fillmore, Charles J. 1986. Pragmatically Controlled Zero Anaphora. *Proceedings of the Berkeley Linguistics Society* 12. 95-107.

Gibson, Edward, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, & Richard Levy. 2019. How efficiency shapes human language. *Trends In Cognitive Sciences* 23(5). 389-407.

Goldberg, Adele E. 2005. Argument realization: the role of constructions, lexical semantics and discourse factors. In Jan-Ola Östman & Miriam Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*, 17-44. Amsterdam: John Benjamins.

Goldhahn, Dirk, Thomas Eckart & Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck et al. (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 759-765. Istanbul: ELRA.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of human language*, 73–113. Cambridge, Mass: MIT Press.

Haspelmath, Martin. 1993. More on the typology of inchoative/causative alternations. In Bernard Comrie & Maria Polinsky (eds.), *Causatives and transitivity*, 87–120. Amsterdam: John Benjamins.

Hawkins, John A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London: Croom Helm.

Hawkins, John A. 1995. Argument-predicate structure in grammar and performance: A comparison of English and German. In Irmengard Rauch & Gerald F. Carr (eds.), *Insights in Germanic Linguistics, Vol. 1 Methodology in Transition*, 127-44. Berlin: Mouton de Gruyter.

Hawkins, John A. 2012. The drift of English towards invariable word order from a typological and Germanic perspective. In Terttu Nevalainen & Elizabeth C. Traugott (eds.), *The Oxford Handbook of the History of English*, 622-632. Oxford: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Hawkins, John A. 2019. Word external properties in a typology of Modern English: A comparison with German. *English Language and Linguistics* 23(3). 701-723.

Letuchiy, Alexander. 2009. Towards a typology of labile verbs: Lability vs. derivation. In Patience Epps & Alexandre Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 247-268. Berlin / New York: De Gruyter Mouton.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Levshina, Natalia. 2019. Token-based typology and word order entropy. *Linguistic Typology* 23(3). 533-572.

Levshina, Natalia. 2020. How tight is your language? A semantic typology based on Mutual Information. In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, 70-78. Düsseldorf: ACL.

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.

Levshina, Natalia. 2022. *Communicative Efficiency: Language Structure and Use*. Cambridge: Cambridge University Press.

Müller-Gotama, Franz. 1994. *Grammatical Relations: A Cross-Linguistic Perspective on Their Syntax and Semantics*. Berlin: Mouton de Gruyter.

Nichols, Johanna. 1986. On form and content in typology. In Winfred P. Lehmann (ed.), *Language Typology 1985: Papers from the Linguistic Typology Symposium, Moscow, 9-13 Dec. 1985*, 141 - 162. Amsterdam: John Benjamins.

Nichols, Joanna, David A. Peterson, & Jonathon Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8. 149-211.

Plank, Frans. 1984. Verbs and objects in semantic agreement: Minor differences between English and German that might suggest a major one. *Journal of Semantics* 3(4). 305-360.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27. 379-423, 623-656.

Wijffels, Jan, Milon Straka & Jana Straková. 2018. udpipe: Tokenization, parts of speech tagging, lemmatization and dependency parsing with the UDPipe NLP Toolkit. R package version 0.7.

Corpora

Leipzig Corpora Collection

https://wortschatz.uni-leipzig.de/en/download

CONTACT

natalia.levshina@mpi.nl