# Word order flexibility: a typometric study

Sylvain Kahane, Ziqian Peng, Kim Gerdes

# Word order flexibility: a typometric study

**Sylvain Kahane[1], Ziqian Peng[2], Kim Gerdes[2]**

[1] Université Paris Nanterre, Modyco (CNRS)
[2] Université Paris-Saclay, Lisn (CNRS)

sylvain@kahane.fr

{ziqian.peng,kim.gerdes}@universite-paris-saclay.fr

## Abstract

This paper introduces a typometric measure of flexibility, which quantifies the variability of head-dependent word order on the whole set of treebanks of a language or on specific constructions. The measure is based on the notion of head-initiality and we show that it can be computed for all of languages of the Universal Dependency treebank set, that it does not require ad-hoc thresholds to categorize languages or constructions, and that it can be applied with any granularity of constructions and languages. We compare our results with Bakker's (1998) categorical flexibility index. Typometric flexibility is shown to be a good measure for characterizing the language distribution with respect to word order for a given construction, and for estimating whether a construction predicts the global word order behavior of a language.

## 1 Introduction

For half a century, research in typology centers on the discussion of word order parameters, pioneered by Greenberg (1963, 1966), and elaborated by such authors as Hawkins (1983), Dryer (1992) and Nichols (1992). Bakker (1998) proposes a seminal study on *word order flexibility*, which we pursue here. First off, it is clear that languages differ in the flexibility of word order: Greek or Russian are more flexible than English or Chinese. Secondly, constructions differ in their flexibility across the diversity of languages: The relation between an adposition and its complement is less flexible than the relation between a verb and its direct object.

To give a first idea how this can be seen on a typometric scatter plot, consider Fig. 1 where each point corresponds to a language, with its x-value indicating the percentage of nominal dependents of adpositions on the right of the adposition and its y-value indicating the percentage of nominal object dependents of verbs to the right of the verb.

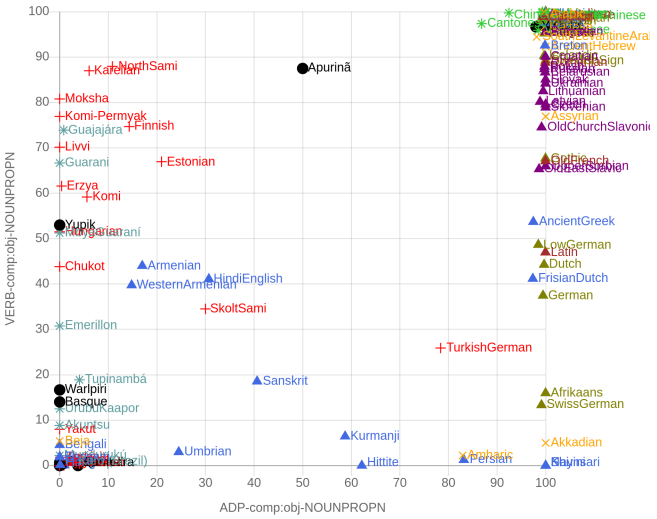


Figure 1: Two-dimensional scatter plots with ADP-comp:obj-NOUN/PROPN in the x-axis and VERB-comp:obj-NOUN/PROPN in the y-axis.

As a first observation, note that a large majority of languages have x=0 or x=100, leaving a rather empty space in the middle of the scatter plot. This shows that most languages have almost only prepositions or almost only postpositions; languages that mix prepositions and postpositions equally are rare.[1] However, many more languages accept both pre-verbal and post-verbal direct objects (Guzmán Naranjo & Becker 2018, Gerdes et al. 2021). Nonetheless, the postpositional languages on the left appear to be less strictly

[1]The scatter plot view understates the fact that many languages cluster around the bottom left (0,0) and the top right (100,100) of the plot: Many strictly postpositional languages also have their nominal objects on the left, and, inversely, many strictly prepositional languages also have their nominal objects to the right of their verbal governor. This is a well-established observation in typology since Greenberg (1963), and is at the base of our choice of SUD treebanks (see Section 2 for details).

postpositional than the prepositional languages on the right are prepositional. This motivates the definition of flexibility in Section 4. See Section 3 for more details on how to compute and understand these plots.

How can we measure the flexibility of languages and constructions? What are the properties of flexibility across languages and constructions? We will try to give answers to these questions in Section 6.

Most classical approaches to typology, including Bakker (1998) and previous works, are categorical in the sense that languages are grouped into categories based on their order constraints, and often only one basic word order is assumed per language from which other word orders are derived by movement, dislocation, or similar operations.

We propose a *typometric* approach (also called *token-based typology* by Levshina 2019): With the availability of a wide range of uniformly annotated treebanks in the Universal Dependencies (UD) project, it has become possible to approach these questions empirically. Syntactic typology outgrows the need for ad hoc categories and measures of distribution of languages across empirical observations become the center of interest (Futtrell et al. 2020, Levshina 2022). In Gerdes et al. (2021), quantitative universals describe empty or sparsely populated spaces in unidimensional or multidimensional spaces instead of qualitative universals that are claiming rare or impossible combinations of language features based on categories.

Tesnière (1959) proposed a classification of languages based on the dependency direction referring to Steinthal (1850) and Schmidt (1926). He opposes strict word order, when head-daughter relations mostly go in one direction, to *mitigated* when the head is amidst its dependents going out in both directions. Among languages with mitigated word order, there are languages with free order, as well as languages with mixed word order, where word order is quite strict in most constructions but inconsistent between constructions. This is what flexibility measures.

In this paper we propose measures of flexibility that can be applied to dependency treebanks and discuss the distributions of these measures compared to other observations on dependency treebanks. Similar measures have been first introduced by Futrell (2015) under the name of *word-order entropy* and have been studied by Levshina (2019).

In this paper, we try to characterize the distribution of all languages of our sample in terms of word order direction for each construction C: We compute for each language L, the number of head-initial realization of the construction C in L, what we call the head-initiality of language L under C (Section 3). We deduce from head-initiality a second measure we call flexibility and study the relation between head-initiality and flexibility for all languages in our sample, distinguishing flexible languages from mixed word order languages (Section 4). The typometric measure of flexibility we introduce is compared with Bakker's (1998) categorical measure of flexibility, as well as a more typometric measure à la Bakker (Section 5). We show that the distribution of head-initiality for every construction C can be characterized by the average head-initiality of C and the flexibility of C (Section 6). In Section 7, we explore the question of the predictability of word order distribution from one construction to another.

## 2 Dependency syntax and word order

Dependency syntax encodes constructions by relations between words representing combinations between larger units (Tesnière 1959, Hudson 1984). A dependency relation goes from one word to another, from governor to dependent. There is no a priori assumption on locality of a relation, and a long distance dependency, for example, does not need any special encoding in a dependency tree, which makes dependency treebanks the obvious choice when attempting to measure tendencies in word order across languages (Liu 2008).

A syntactic relation is a class of combinations of the same type, having similar properties. Dependency syntax makes the assumption that most constructions are asymmetric, with a head element controlling the distribution of the combination. In some languages, constructions are very rigid and combinations of a certain type tend to always have the same word order between the governor and the dependent. Examples of such rigid relations are the *subject* and the *object* relation in English.[2] Subject and Object are different

[2]Widely discussed exceptions to the rigidity of subject and object in English include the relative pronoun (*a person who I never met*) and marginal cases of dislocation such as *Chocolate I adore!* As in other typological studies, we restrict our object measures to nominal objects, thus excluding the

constructions and therefore are annotated as different relations.

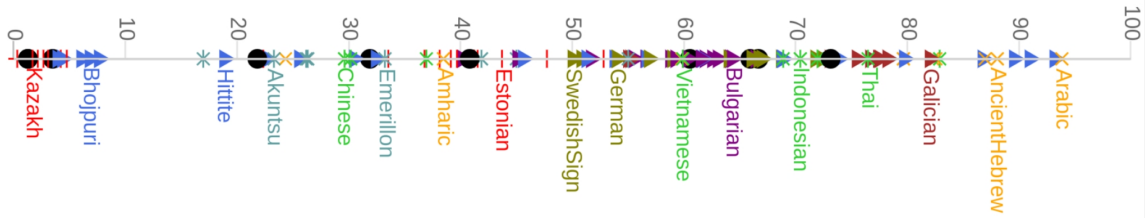


Figure 2: Head-initiality of core relations across SUD 2.11

Although the main criteria of distinguishing one relation from another is valency and morphology (for example an accusative case can be the main criterion for delimiting the direct object relation), in some cases the definition of a dependency relation relies on the word order itself and thus the relations have by definition a strict word order.[3]

Just as in the initial typometrics paper Gerdes et al. (2021), we rely for our measures on the Surface Syntactic Dependency (SUD) version of UD (Gerdes et al. 2018), in order to make our work comparable with previous work on word order typology, thus preserving "cross-category harmony" (Hawkings 1983) and avoiding complications in particular concerning adpositions and auxiliary verbs that are analyzed in an unusual manner in the original UD annotation scheme.[4]

Choosing SUD rather than UD has very little impact on the computation of the flexibility measures introduced in this paper.[5]

SUD's *comp* relation corresponds to UD's *aux*, *ccomp*, *iobj*, *obj*, *obl:arg*, *xcomp*, *cop*, *mark*, and *case*; *mod* corresponds to UD's *advcl*, *acl*, *advmod*, *amod*, *nmod*, *nummod*, and *obl:mod*. SUD's *subj* combines UD's *csubj* and *nsubj*. The relations *dislocated*, *det*, and *clf* remain unchanged between SUD and UD.

We base our work on the latest SUD version 2.11 which includes 243 treebanks in 138 languages in total. For our study, treebanks of the same language are combined and taken as one data point. 65 UD languages cover Indo-European languages. Afro-Asiatic, Uralic, and Tupian languages have 11 languages each. Turkic covers another 6 languages. Of the remaining languages only Basque, Chinese, Classical Chinese, Indonesian, Japanese, Korean, and Naija have more than 100k tokens. 21 of the UD languages are very small (less than 1000 tokens), which falls beneath our threshold for most of our measures.

## 3 Typometrics and scatter plots

A typometric analysis does not assume a basic word order or any threshold for categorizing languages or construction. Our basic observation is the measure of *head- initiality* defined for a language L and a construction C involving a unique dependency as follows:[6]

**head_initiality(L, C) =**
% of occurrences of C in L that are head-initial (governor < dependent)

In most cases the construction C limited to a dependency is defined as a *gov-rel-dep* triple (governor's POS, dependency relation, dependent's POS). In some cases the construction is defined as the sum of a series of gov-rel-*dep* triples. Note that any variable of the triplet (*gov*, *rel*, or *dep*) can be equal to *all*, denoting no restriction on this variable.

first case. Clearly, the measures we end up with will always depend on the annotation choices of each treebank.

[3]As an example, consider the annotation choices for Cantonese and Mandarin reported in Wong et al. (2017): Any element to the left of the verb is considered as "dislocated" even if it fills the verb's object slot.

[4]Guzmán Naranjo and Becker (2018), for example, find that UD's *case* relation stands out in their directional correlation measures.

[5]As SUD is obtained by a conversion of UD without any addition of information, the granularity remains similar, see Section 4. It only impacts locally some relations such as the subject, which, in SUD, is attached to the auxiliary rather than the content verb and whose direction can change in some cases (for instance, in German, where the subject can be between the auxiliary and the verb).

[6]Head-initiality is introduced in Gerdes et al. (2019, 2021), where it is called *direction*.

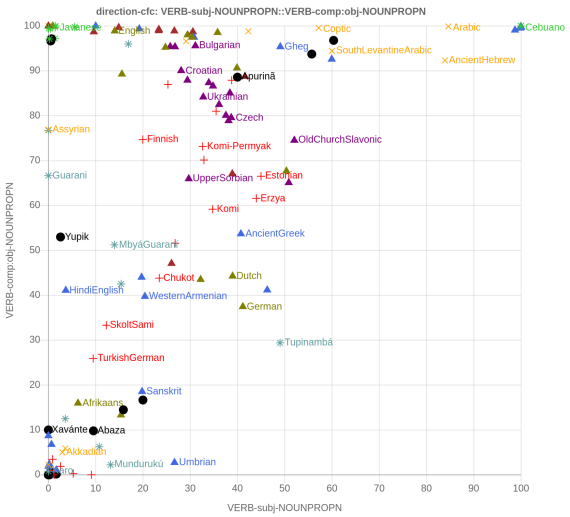


Figure 3: Two-dimensional scatter plots with verbal subjects and objects.

A head-initiality of 0 for a given language and construction shows a strictly head-final construction, a head-initiality of 100 indicates a strictly head-initial construction. Measuring head-initiality across the UD languages for the combination of core dependency relations (to be defined in the next section) gives the unidimensional scatter plot of Figure 2, where, unsurprisingly, Japanese is the most head-final language, and Arabic the most head-initial language of our language sample set.[7]

A second two-dimensional scatter plot (Fig. 3) opposes the head-initiality of nominal subjects in the x-axis (VERB-subj- NOUN|PROPN) and nominal direct objects in the y-axis (VERB-comp:obj- NOUN|PROPN). We allow both the UD POS noun and proper noun as arguments. We observe a typical triangular shape of the resulting distribution indicating that nearly all languages have the tendency to have direct objects more to the right than subjects. Put differently, hardly any language has a higher head-initiality on subjects than on direct objects. See Gerdes et al. (2021) for a discussion of how this observation generalizes to the well-known absence of OVS languages.

[7]Colors and shapes of the language points follow the original typometrics paper with colored *triangles* for the different subgroups of Indo-European languages, *plus* signs for agglutinating languages, orange *x* signs for Afroasiatic and Semitic languages, and *circles* and *stars* for other groups. Data, scatterplots, and detailed captions are on https://typometrics.elizia.net/. Note that only some languages are labeled. This has no semantics and is done automatically to increase readability.

## 4 Flexibility of languages

For a language L, *flexibility* measures the distance of a construction C from a rigid construction. In this paper, we only consider constructions involving a governor G and a dependent D by a particular relation. The construction has a wider or narrower range depending on whether the relation between G and D or the categories of G and D are more or less constrained.

**flexibility(L,C)**
= 2 × min( head_initiality(L,C),
100–head_initiality(L,C) )
= twice the smallest distance of head_initiality(L,C) to 0 or to 100

The value of flexibility(L,C) ranges from 0 to 100 and measures the distance of C from a strictly head-initial or head-final construction. A very similar measure, *word order entropy*, has been proposed by Levshina (2019), inspired by Futrell et al. (2015).[8] She also considers the entropy for couples of dependencies, such as the relative position of subjects and objects.

For a given language L, we can compute the weighted average of flexibility(L,C) for a relevant set S of constructions C, which will be discussed below.

**head_initiality(L)** =
weighted average of head_initiality(L, C) on constructions C

**flexibility(L)** =
weighted average of flexibility(L, C) on constructions C.

A measure very similar to flexibility(L) has been introduced by Futrell et al. (2015), using conditional entropy. In information theory, the conditional entropy $H(Y|X)$ quantifies the amount of information needed to describe the outcome of the random variable Y given that the value of the random variable X is known. The more $H(Y|X)$ is close to 1, the more Y is independent from X, $H(Y|X)$ being equal to 0. In Futrell et al. (2015), X is used to select a set S of constructions, while Y describes the word

[8]Precisely, entropy(L,C) = $- p*\log_2(p) - (1-p)*\log_2(1-p)$, with p = head_initiality(L,C). This value also ranges from 0 to 100%, with value 0 for p=0 or 100% and 100% for p=50%. The only difference with our calculation is that entropy smoothes values for p around the 50% mark.

order on S. In other words, entropy, like flexibility, measures the extent to which word order choices depend on syntactic constructions.[9]

Let us discuss our choices of S for head-initiality and flexibility. The computation of head_initiality(L) and flexibility(L) is sensitive to the range D of data considered. Unlike head-initiality(L), the computation of flexibility(L) is sensitive to the granularity of the partition of D into a set S of constructions: the finer the partitioning S, the higher the yield of flexibility(L). In our case, we have adopted a rather fine granularity, as we consider any *gov-rel-dep* triplet as a different construction, where *gov* is the POS of the governor, *dep* is the POS of the dependent, and *rel* is the relationship between them. We could have used an even finer granularity, by taking into account certain features, for example by distinguishing relative pronouns (PronType= Rel) from personal pronouns (PronType=Prs) or by isolating demonstratives (PronType= Dem).[10] Moreover, when we have a direct complement of the verb, we will distinguish nominal complement (dep=NOUN) and pronominal complement (dep=PRON), but not when it is a prepositional complement (dep= ADP). Sometimes the granularity can be excessive, as when UD/SUD distinguishes proper nouns (PROPN) and common nouns (NOUN).[11] It must also be remarked that Levshina (2018) restricts her computation for verbal constructions to verbs that are roots, arguing that word order can be quite different between main and subordinate clauses in some languages (German and Wolof for instance). Our preference is to keep all occurrences, but it is certainly interesting to do a partition between main clauses and subordinate clauses.[12]

Unlike flexibility(L), the computation of head_initiality(L) is obviously very dependent on the choice made for the head of each construction. It is this question that motivated us not to work with UD, but to choose the SUD variant where adpositions, subordinating conjunctions, and auxiliaries are chosen as heads.[13] For auxiliaries, the question is delicate, because while for Indo-European languages, they are clearly heads, this is less obvious in languages where they are particles. On the other hand, the wh-words of Indo-European languages are treated as pronouns in both UD and SUD, even though they also have a head role, which explains in part their peculiar placement.

For head-initiality(L), we decide to consider the relations of type *comp*, *mod*, *udep*, *subj*, *dislocated*, that we call the *core* relations. We have included the *dislocated* relation, because the boundary between governed and dislocated elements is not always well defined.[14] We have eliminated the *det* relation because the direction of the determiner-noun relation is controversial (see the discussion around the DP-hypothesis since Hudson 1984 and Abney 1987), as well as *clf* (for classifiers), which is used inconsistently. For flexibility(L), we could keep *det* and *clf*, because the choice of the governor of a given relation does not play any role: Flexibility only measures the proportion of dependencies going in the same direction and remains the same when all dependencies are inverted. However the *det* and *clf* have only a small influence on the final result (cf. Table D) and, to be precise, we decided to use the *core* relations for the computation of flexibility(L) as well. Other SUD/UD relations are of little interest for our study as their direction is fixed in the UD annotation guidelines. This includes *conj*, *fixed*, *goeswith*, etc. It should also be noted that we have not considered the relations

[9]The entropy view of flexibility is very elegant, but, as mentioned by Levshina (2019), Futrell et al. (2015) gives "one aggregate score" for each language, rather than considering individual constructions before aggregating them.

[10]Levshina (2019) also considers constructions restricted to one dependent word form. This is only possible if the corpus contains enough occurrences of the word, which commonly implies for many languages to parse raw corpora that are bigger than the manually annotated corpora of UD.

[11]On the other hand, UD does not usually distinguish prepositional dependents of the verb, which are all rel=*obl*, whether they are arguments or modifiers. This distinction is made only in a few treebanks, notably the native SUD treebanks (with the *comp* and *mod* labels). SUD uses the *udep* relation, for underspecified *obl* dependencies when the distinction in argument and modifier is not encoded.

[12]About the relations between constructions in main and subordinate clauses, see Schachter (1973).

[13]For instance, Futrell et al. (2015) based on computation on UD indicates that French and Italian are mostly head-final, while, based on SUD, they are head-initial at 76% and 77% respectively!

[14] For instance, in a pro-drop language such as Chinese, it is difficult to decide if preverbal objects are dislocated or not. See Note 3.

between co-dependents at all. Yet, some languages with a very strict head-final order, such as Japanese or Korean, can have much greater freedom in the placement of co-dependents, which is not taken into account in the present study.

Lastly, we chose to give each construction a weight equivalent to its frequency in the corpora, unlike Bakker (1998), who gives the same weight to each of the 10 constructions he considers (as well as Levshina (2019), who considers quantitative values for each construction but does an average with equal weight).

Figure 4 shows the head-initiality of SUD 2.11 languages in the x-axis and the flexibility in the y-axis. For treebanks with at least 1000 core relations, we observe that Ancient Greek, Tupinambá, Emerillon(Teko), Turkish-German (code switching corpus) and Old East Slavic are the most flexible languages (with flexibilities 59.3, 56.7, 55.2, 52.9, and 51.7 respectively), while Japanese, Hindi, Xibe, Kazakh and Telugu (flexibilities of around 0.5, 1.6, 2.1, 2.4, and 2.4 respectively) are the least flexible languages.

Languages with head-initiality equal to 0 or 100 have flexibility 0 and the closer they are to 50 the more likely they are to be flexible. But there are languages L with head_initiality(L) close to 50 and flexibility(L) close to 0, such as Bambara: these are *mixed order languages*. Languages with high flexibility(L), such as Ancient Greek, are *free order languages*.[15]

## 5 Comparison with Bakker's flexibility

Bakker (1998) proposes a computation of flexibility based on the same principles but does not take into account the greater or lesser flexibility of each construction for each language. In Bakker's computation, a language is either flexible or completely rigid.

**flexibility[Bakker](C,L) =**
0 if the construction C is completely rigid in L,
1 if it is flexible

**flexibility[Bakker](L)=**
(equal-weighted) average over 10 constructions C of flexibility[Bakker](C,L)

[15]Our measures are also dependent on the corpus chosen for the calculus and its genre. The flexibility measure of Ancient Greek is certainly increased by the fact that the corpus contains poetry and theater.

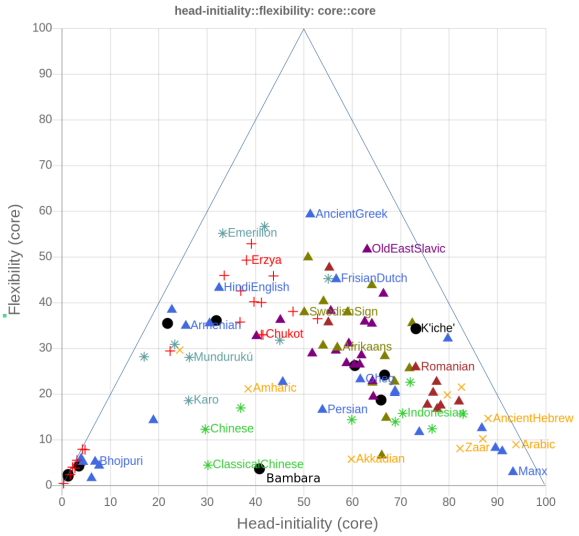

Figure 4: Head-initiality vs Flexibility (core)

Bakker gives the flexibility values for a sample of 86 European languages 47 of which are Indo-European. We propose to compare Bakker's values with a typometric index of the same type. Some of Bakker's constructions can be directly translated into SUD corpus queries, others can be approximated. For example, his "Adj/N" translates directly into the typometric measure NOUN-mod-ADJ. The Verb-Recipient relation (V/R) can only be approximated by VERB-comp:obl- ADP| NOUN (cf. Table A2 in Annex). The complete list of Bakker's constructions and their translation into typometric measures are provided in the construction flexibility Table A3 of the Annex.[16]

[16]Bakker (1998: 393ff) introduces another measure which he calls *consistency* and which is very dependent on the set of considered constructions, which are still the 10 same constructions (see Section 5):

**consistency[Bakker](L) =**

```
| #{ C / C is head-initial for L }
 - #{ C / C is head-final for L } |.
```

It seems to us that the consistency of a language L is well captured by our head_initiality(L), which is not dependent on the partitioning P into constructions undertaken by the linguist. Moreover, Bakker (1998: 401-2) notes that flexibility[Bakker] and consistency[Bakker] are correlated, but this is obvious as soon as there are languages whose head-initiality is close to 0 or 100. Likewise, our flexibility and head-initiality are related, since `flexibility(L) ≤ 2 × min ( head_initiality(L), 100 - head_initiality(L) )`, which we have visualized as a triangle in Figure 4.

As Bakker's flexibility measure is categorical per construction, we have to arbitrarily set a threshold at 5%, considering that languages with less than 5% variation for a given construction C are inflexible for C.

We can compare those 3 measures across the languages that we also find in UD: 1. Bakker's flexibility, 2. our recomputation of the flexibility à la Bakker, as a non-weighted average over Bakker's 10 constructions, with the 5% threshold as indicated above, and 3. our typometric flexibility.[17]

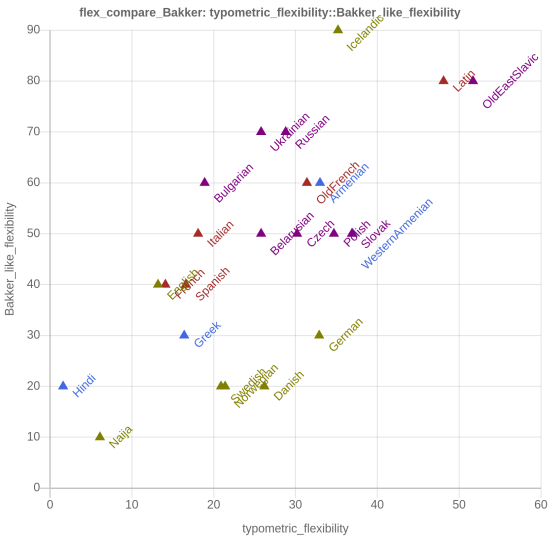


Figure 5: Typometric vs Bakker-like flexibility

The scatterplot of Figure 5 shows the strong correlation between the Bakker-like measure of flexibility and the typometric flexibility. The Bakker-like flexibility is also significantly correlated with Bakker's flexibility (Fig. C2), while there is only a weak correlation between the typometric flexibility and Bakker's flexibility (Fig. C1). See the complete results in Tables B1 and B2 in the Annex.

## 6 Flexibility of constructions

Having compared the overall flexibility of the languages, we can now see how the languages in a given sample S are distributed for each construction C and compare the constructions. Specifically, we are interested in how the head-initiality of the languages in our sample (the 138 languages in UD 2.11) is distributed for the different constructions C. Our hypothesis is that this distribution is reasonably well described by the following two values:

**average_head_initiality$_S$(C)** = average of head_initiality(L,C) over the Ls in S.

**average_flexibility$_S$(C) =** average of flexibility(L,C) over the Ls in S.

The less flexible C is on average, the more languages are attracted to 0 and 100. The average-head-initiality indicates whether 0 or 100 attracts more to one than the other. We propose two other values that will help us to better understand this attraction towards 0 and 100.

**head_initial_weightS(C) =**
average_head_initiality$_S$(C) / average_flexibility$_S$(C)

**head_final_weightS(C) =**
(100 – average_head_initiality$_S$(C)) / average_flexibility$_S$(C))

For a uniform distribution, flexibility = 50, head-initiality = 50, head-initial-weight = 1, and head-final-weight = 1. When head-final-weight > 1, the distribution is drawn towards 0 and when head-final-weight < 1, it is pushed away. The reverse holds for head-initial-weight. Our postulate is that the distribution of head-initiality is similar to a uniform distribution that has been distorted by stretching it from both sides.[18] Our head-initial and head-final weights give us an estimate of the strength of the force at each end.

We observe that for all the most frequent C constructions, both head_initial_weight$_S$(C) > 1 and head_final_weight$_S$(C) > 1 (see Table A3 in the Annex where all but one of the weights for the 10 Bakker constructions are greater than 1), which means that languages are attracted on both ends.

To give an idea of the different distributions we encounter, the three scatter plots below show three head-initiality distributions on the

[17]Some of Bakker's constructions, such as Dem/N or Pro/N, involve features that are not present in all UD treebanks (PronType=Dem and Poss=Yes in these two cases). Our computation is restrained to languages with all the required features.

[18]Levshina (2019), like us, uses the mean head initiality and the standard deviation to characterize the distribution of a head initiality for a given construction. The standard deviation is relevant for characterizing Gaussian distributions, but not for "stretched" distributions as here, particularly when elements tend to move away from the center and when these movements are asymmetric, with one end more attractive than the other.

treebanks: 1. Num/N (NOUN-any-NUM),[19] 2. aux-v (AUX-comp:aux-VERB), and 3. Adj/N (NOUN-mod-ADJ). The first (Num/N) distribution tends towards 0 and is pushed away from 100 (weights 4 and 0.6), while the two other distributions are attracted both by 0 and 100, with a bigger attraction to 0 for Adj/N (weights 4.2 and 2) and to 100 for Aux/V (weights 1.3 and 2.6).

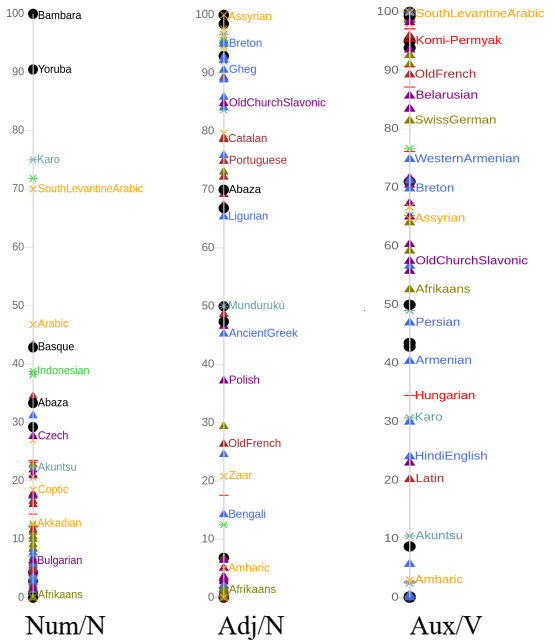


Figure 6: Head-initiality language distribution for three constructions Num/N, Adj/N, Aux/V.

Again, we can compare our flexibility results with two measures: the flexibility measure proposed by Bakker (1998) and a Bakker-like measure that we calculate from our sample.

**flexibility[Bakker](C)=**
% of languages in Bakker's sample that are flexible for C.

**flexibility[Bakker_like]$_S$(C) =**
% of Ls in S with flexibility(L,C) > 5.

Bakker restricts his study to a sample S of 86 European languages, 16 of them having enough data in UD to be compared.[20]
We find that V/O is the most flexible construction, followed by V/R and Adj/N. Two constructions do not behave at all as in Bakker's sample: Aux/V appears as the most flexible construction after V/O, while Rel/N appears as extremely inflexible.

Bakker also compares the flexibility of languages with head-initial and head-final basic order. Again we can introduce typometric Bakker-like measures. We consider that a language has head-initial basic order if more than 50% of core dependencies are head-initial.

**head_initial_flexibility[Bakker_like]$_S$(C) =** % of Ls in S with head_initiality(L,C)>50 that have flexibility(L,C)>5.

**head_final_flexibility[Bakker_like]$_S$(C) =** % of Ls in S with head_initiality(L,C)<50 that have flexibility(L,C)>5.

Bakker (1998: 392) "observed that head-modifier orders are, on the whole, more flexible than modifier-head orders." We have completely different results with our measures (see Table A3 in the Annex): Only for adpositions, languages with head-initial core order are more flexible than languages with head-final core order.

## 7 Predictivity

With these notions in place, it is now possible to measure which construction predicts best the overall core flexibility of a given language. For this, we measure the Spearman correlation between the distribution of flexibility(L,C) for various couples of construction C (Figure 8). We are particularly interested in the correlation with the *core* construction. Among the 10 Bakker constructions, the best predictors of the *core* flexibility is the *V/O* construction, unlike Adj/N. Note also some notable correlations: *Aux/V* and *V/O* that have a correlation of 0.59 as well as *subj* and *comp* that have a correlation of 0.53.

Bakker (1998: 392) however states that "the best predictors of overall flexibility are the

[19]To be precise, SUD uses a special feature ExtPos, indicating the external POS of a word. Numerals, all categorized NUM in UD, are *nummod* or *det* when they are used as a quantifier (*my 7 cats*). In other uses, they work as a proper noun (*line 7, page 7, year 2023*) and receive the feature ExtPos=PROPN. It is this feature, rather than the POS, that is used in all our computations. It remains that the use of *nummod* is not consistent across all treebanks.

[20]When looking at a particular construction C, we only consider a language L if the treebanks of L have at least 50 occurrences of C. For Bakker-like measures to be calculated for L, the threshold of 20 must be reached for each of the 10 constructions considered.

flexibility of the recipient, genitive, numeral and relative clause. On the other hand, no prediction whatsoever can be drawn from the behavior of adpositions and articles." This is not backed up by our data with our weighted flexibility measure: The typometric genitive flexibility has a correlation of only 0.18 with the core flexibility, numerals have a correlation of 0.01, and relatives of 0.07.

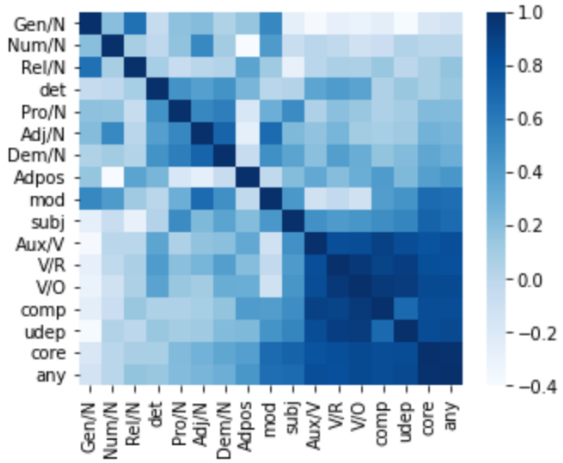


Figure 7: Heatmap of Spearman correlation between the distributions of flexibility(L,C) for various couples of construction C. See Table D of the Annex for detailed values.

Nonetheless, our data agrees with a common typologist view, most notably Dryer (1992), that sees V/O as a good predictor for word order constraints on other constructions.

Note also that the constructions with the biggest flexibility are the best predictors. This was expected because constructions with low flexibility tend to gather all languages around 0 and 100.

## 8 Conclusion

We have introduced a first measure, head-initiality, which measures the variable head-dependent word order on a language's treebank or on specific constructions. Based on this, we develop an operational notion of flexibility that renders the intuition that the average head-initiality can be far from 0 or 100 while the languages are strict per given construction.

We then show that our empirical notion of flexibility can be compared to previous definitions of flexibility of word order, notably to Bakker's work. Our notion of flexibility has the advantage that it can directly be computed from treebanks, that it does not require ad-hoc thresholds to categorize languages or constructions, and that it can be applied with any granularity of constructions.

Finally, we show which construction predicts overall word order flexibility of a language. For this, we rely on Spearman's rank correlation coefficient, which allows us to calculate a correlation between two distributions. We show that over UD's language sample, the highest correlation is obtained for nominal objects (V/O construction).

Since the Spearman correlation is a symmetric measure, we would like to continue our study by proposing an asymmetric measure that allows us to decide if one distribution can predict another. Our hypothesis, to be confirmed, is that constructions with the most uniform distribution, thus being flexible and well-balanced, provide better predictions of the behavior of other constructions. The V/O construction, which many authors take as a basic construction (see in particular the study of Dryer 1992) is thus an excellent candidate.

## Acknowledgements

We would like to thank the three anonymous reviewers of the Gurt-Syntaxfest 2023 for their careful and patient examination as well as for the many valuable comments they made.

This research was supported by the French National Research Project (ANR) Autogramm.

## References

Abney, Steven P. (1987). *The English noun phrase in its sentential aspect*. Doctoral dissertation, Cambridge: MIT.

Bakker, Dik (1998). Flexibility and consistency in word order patterns in the languages of Europe. In Siewierska A. (ed.) *Constituent order in the languages of Europe*, Berlin: Mouton de Gruyter, 383-419.

Chen, Xinying, and Kim Gerdes. (2017). Classifying Languages by Dependency Structure: Typologies of Delexicalized Universal Dependency Treebanks, Proceedings of the 4th Conference on Dependency Linguistics (Depling).

Dryer, Matthew S. (1992). The Greenbergian word order correlations, Language 68, 81-138.

Futrell, R., K. Mahowald, and E. Gibson (2015). Quantifying word order freedom in dependency corpora. In Proceedings of the third international conference on dependency linguistics (Depling), 91-100.

Futrell, R., R. P. Levy, and E. Gibson (2020). Dependency locality as an explanatory principle for word order. Language 96(2), 371-412.

Gerdes, Kim, and Sylvain Kahane. (2016). Dependency Annotation Choices: Assessing Theoretical and Practical Issues of Universal Dependencies. Proceedings of the 10th Linguistic Annotation Workshop (LAW X).

Gerdes, Kim, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. (2018). SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. Proceedings of the Universal Dependencies Workshop (UDW).

Gerdes, Kim, Sylvain Kahane, and Xinying Chen. (2021). "Typometrics: From implicational to quantitative universals in word order typology. Glossa: a journal of general linguistics 6:1.

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

Greenberg, Joseph H. (1966). *Language Universals*. The Hague: Mouton.

Guzmán Naranjo, Matías, and Laura Becker (2018). Quantitative word order typology with UD. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT),* 91-104.

Hawkins, John A. (1983). *Word Order Universals*. New York: Academic Press.

Hudson, Richard (1984). *Word Grammar*. Oxford: Basil Blackwell.

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

Levshina, Natalia (2022). Corpus-based typology: Applications, challenges and some solutions. *Linguistic Typology*, *26*(1), 129-160.

Liu, Haitao (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159-191.

Nichols, Joanna (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.

Schachter, Paul (1973). Focus and relativization. *Language* 49(1), 19-46.

Schmidt, P. W. (1926). *Die Sprachfamilien und Sprachenkreise der Erde: Atlas von 14 Karten*. Heidelberg: Winter.

Steinthal, Heymann. *Die Classification der Sprachen dargestellt als die Entwicklung der Sprachidee*. Dümmler, 1850.

Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B., Cervone, C., ... & Carrier, A. (2021). Spatial agency bias and word order flexibility: A comparison of 14 European languages. *Applied Psycholinguistics* 42(3), 657-671.

Tesnière, Lucien (1959). *Eléments de syntaxe structurale*. Paris: Klincksieck. [Transl. by Osborne, T., Kahane, S. (2015) *Elements of structural syntax*. Benjamins].

Wong, T. S., K. Gerdes, H. Leung and J. Lee. (2017). Quantitative Comparative Syntax on the Cantonese-Mandarin Parallel Dependency Treebank. *Proceedings of the conference on Dependency Linguistics (Depling)*, 266–275.

# Annex

| Languages | Bakker-flexibility | Bakker_like_flexibility | typometric_flexibility |
|---|---|---|---|
| Armenian | 40 | 60 | 33 |
| Belarusian | - | 50 | 26 |
| Bulgarian | 60 | 60 | 19 |
| Czech | - | 50 | 30 |
| Danish | 30 | 20 | 26 |
| English | 40 | 40 | 13 |
| French | 10 | 40 | 14 |
| German | 40 | 30 | 33 |
| Greek | 60 | 40 | 16 |
| Hindi | - | 20 | 2 |
| Icelandic | 40 | **90** | 35 |
| Italian | **30** | 50 | 18 |
| Latin | **90** | 80 | **48** |
| Naija | _ | 10 | 6 |
| Norwegian | 40 | 20 | 21 |
| OldEastSlavic | - | **80** | **52** |
| OldFrench | - | 60 | 31 |
| Polish | 60 | 50 | 35 |
| Russian | 70 | 70 | 29 |
| Slovak | 50 | 50 | 37 |
| Spanish | 30 | 40 | 17 |
| Swedish | 40 | 20 | 21 |
| Ukrainian | - | 70 | 26 |
| WesternArmenian | - | 50 | 37 |

**Table A1. Various flexibility measures for languages where a treebank-based verification of Bakker's measures is available as described in footnote 20.**

| Bakker's 10 relations | Corresponding construction |
|---|---|
| V/O | VERB-comp:obj-NOUN/PROPN |
| Adj/N | NOUN-mod-ADJ |
| Pro/N | NOUN-any-[Poss=Yes] |
| V/R | VERB-comp:obl-ADP/NOUN |
| Gen/N | NOUN-mod[gen]-ADP/NOUN |
| Rel/N | NOUN-mod@relcl-VERB |
| Adpos | ADP-comp-NOUN |
| Num/N | NOUN-any-NUM |
| Dem/N | NOUN-any[PronType=Dem] |
| Aux/V | AUX-comp:aux-VERB |

**Table A2: the 10 Bakker's relation and their corresponding constructions**

| Measures | V/O | Adj/N | Pro/N | V/R | Gen/N | Rel/N | Adpos | Num/N | Dem/N | Aux/V |
|---|---|---|---|---|---|---|---|---|---|---|
| **freqSample** | 3.3 | 3.6 | 0.7 | 0.6 | 1.5 | 0.4 | 5 | 0.9 | 0.7 | 2.5 |
| **concerned_languages** | 113 | 96 | 51 | 59 | 68 | 42 | 103 | 85 | 64 | 87 |
| **typometric_flexibility%** | 26.2 | 16.2 | 15.9 | 31 | 20.2 | 2.5 | 5.5 | 22 | 10.3 | 25.7 |
| **Bakker-like_flexibility%** | 62.5 | 62.5 | 37.5 | 66.7 | 62.5 | 8.3 | 8.3 | 66.7 | 50 | 54.2 |
| **Bakker-like-flexibility(S)%** | 48.2 | 35.4 | 29.4 | 62.7 | 45.6 | 7.1 | 12.6 | 51.8 | 29.7 | 46 |
| **Bakker-like-head_initial(S)%** | 46.3 | 18.8 | 14.6 | 56.2 | 37 | 0 | 23.3 | 51.8 | 26.3 | 43.5 |
| **Bakker-like-head_final(S)%** | 49.3 | 68.8 | 90 | 65.1 | 51.2 | 7.9 | 8.2 | 50 | 57.1 | 46.9 |
| **head_initiality%** | 61 | 32 | 19 | 68 | 57 | 89 | 71 | 13 | 15 | 68 |
| **head_initial_weight** | 2.3 | 2 | 1.2 | 2.2 | 2.8 | 35.8 | 12.9 | 0.6 | 1.4 | 2.6 |
| **head_final_weight** | 1.5 | 4.2 | 5.1 | 1 | 2.1 | 4.3 | 5.3 | 4 | 8.3 | 1.3 |

**Table A3. Measures for the 10 constructions considered by Bakker (1998). Among them *Bakker-like flexibility* is normalized over the 24 languages in Table A1, Others are normalized with the amount of languages in the row 'concerned languages'**

| *Spearman* | Bakker-flexibility | Bakker_like flexibility | typometric flexibility |
|---|---|---|---|
| Bakker-flexibility | 1 | 0.458 | 0.479 |
| Bakker_like flexibility | 0.458 | 1 | 0.649 |
| typometric flexibility | 0.479 | 0.649 | 1 |

| *Pearson* | Bakker-flexibility | Bakker_like flexibility | typometric flexibility |
|---|---|---|---|
| Bakker-flexibility | 1 | 0.478 | 0.583 |
| Bakker_like flexibility | 0.478 | 1 | 0.692 |
| typometric flexibility | 0.583 | 0.692 | 1 |

Table B. Spearman correlation (left) and Pearson correlation (right) between Bakker's flexibility, Bakker-like flexibility and typometrics flexibility

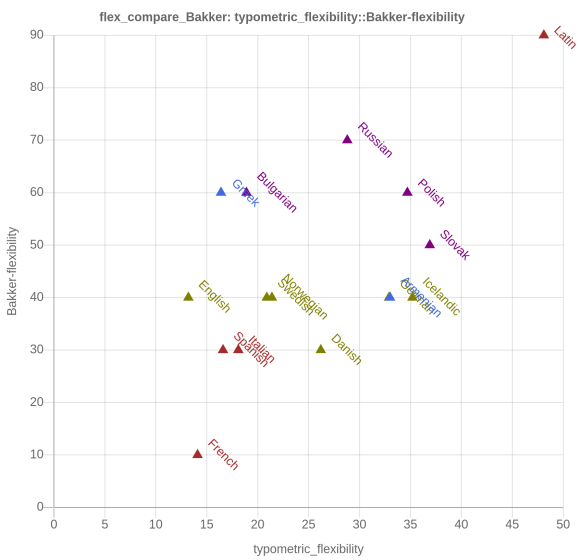


**Figure C1**: Typometric VS Bakker-flexibility

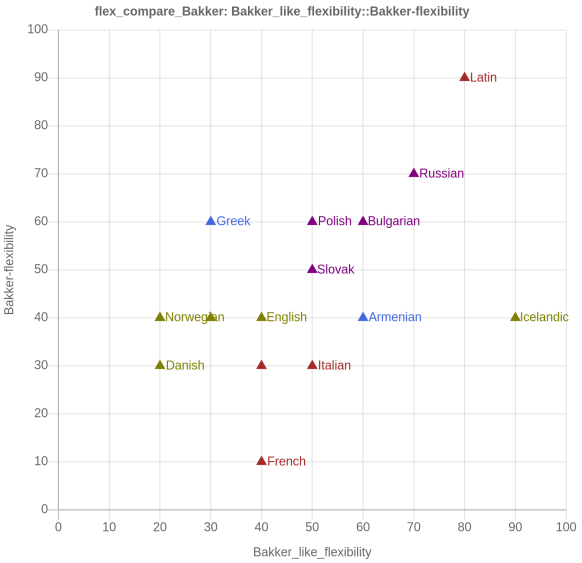


**Figure C2**: Bakker-like VS Bakker-flexibility

| | Gen/N | Num/N | Rel/N | det | Pro/N | Adj/N | Dem/N | Adpos | mod | subj | Aux/V | V/R | V/O | comp | udep | core | any |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gen/N** | 1 | 0.197 | 0.654 | -0.051 | 0.185 | 0.209 | 0.044 | 0.156 | 0.53 | -0.268 | -0.379 | -0.281 | -0.308 | -0.254 | -0.386 | -0.184 | -0.14 |
| **Num/N** | 0.197 | 1 | 0.086 | -0.02 | 0.169 | 0.526 | 0.104 | -0.4 | 0.423 | -0.068 | 0.003 | -0.023 | -0.121 | -0.087 | 0.04 | 0.008 | 0.01 |
| **Rel/N** | 0.654 | 0.086 | 1 | 0.09 | -0.062 | 0.004 | 0.037 | 0.362 | 0.121 | -0.292 | 0.003 | 0.069 | 0.062 | 0.145 | -0.006 | 0.074 | 0.167 |
| **det** | -0.051 | -0.02 | 0.09 | 1 | 0.469 | 0.397 | 0.469 | 0.251 | 0.009 | 0.032 | 0.359 | 0.418 | 0.375 | 0.056 | 0.146 | 0.079 | 0.144 |
| **Pro/N** | 0.185 | 0.169 | -0.062 | 0.469 | 1 | 0.532 | 0.583 | -0.177 | 0.296 | 0.503 | 0.056 | 0.195 | 0.142 | 0.057 | 0.103 | 0.224 | 0.218 |
| **Adj/N** | 0.209 | 0.526 | 0.004 | 0.397 | 0.532 | 1 | 0.716 | -0.267 | 0.68 | 0.229 | 0.171 | 0.257 | 0.107 | 0.087 | 0.124 | 0.273 | 0.256 |
| **Dem/N** | 0.044 | 0.104 | 0.037 | 0.469 | 0.583 | 0.716 | 1 | -0.064 | 0.482 | 0.365 | 0.193 | 0.399 | 0.307 | 0.165 | 0.219 | 0.352 | 0.296 |
| **Adpos** | 0.156 | -0.4 | 0.362 | 0.251 | -0.177 | -0.267 | -0.064 | 1 | -0.025 | 0.217 | 0.338 | 0.205 | 0.304 | 0.415 | 0.233 | 0.394 | 0.444 |
| **mod** | 0.53 | 0.423 | 0.121 | 0.009 | 0.296 | 0.68 | 0.482 | -0.025 | 1 | 0.442 | -0.119 | -0.03 | -0.123 | 0.407 | 0.462 | 0.684 | 0.664 |
| **subj** | -0.268 | -0.068 | -0.292 | 0.032 | 0.503 | 0.229 | 0.365 | 0.217 | 0.442 | 1 | 0.478 | 0.421 | 0.453 | 0.492 | 0.545 | 0.722 | 0.688 |
| **Aux/V** | -0.379 | 0.003 | 0.003 | 0.359 | 0.056 | 0.171 | 0.193 | 0.338 | -0.119 | 0.478 | 1 | 0.839 | 0.844 | 0.907 | 0.842 | 0.809 | 0.838 |
| **V/R** | -0.281 | -0.023 | 0.069 | 0.418 | 0.195 | 0.257 | 0.399 | 0.205 | -0.03 | 0.421 | 0.839 | 1 | 0.954 | 0.895 | 0.921 | **0.833** | 0.826 |
| **V/O** | -0.308 | -0.121 | 0.062 | 0.375 | 0.142 | 0.107 | 0.307 | 0.304 | -0.123 | 0.453 | 0.844 | 0.954 | 1 | 0.941 | 0.93 | **0.858** | 0.86 |
| **comp** | -0.254 | -0.087 | 0.145 | 0.056 | 0.057 | 0.087 | 0.165 | 0.415 | 0.407 | 0.492 | 0.907 | 0.895 | 0.941 | 1 | 0.689 | 0.843 | 0.843 |
| **udep** | -0.386 | 0.04 | -0.006 | 0.146 | 0.103 | 0.124 | 0.219 | 0.233 | 0.462 | 0.545 | 0.842 | 0.921 | 0.93 | 0.689 | 1 | 0.85 | 0.864 |
| **core** | -0.184 | 0.008 | 0.074 | 0.079 | 0.224 | 0.273 | 0.352 | 0.394 | 0.684 | 0.722 | 0.809 | 0.833 | 0.858 | 0.843 | 0.85 | 1 | 0.989 |
| **any** | -0.14 | 0.01 | 0.167 | 0.144 | 0.218 | 0.256 | 0.296 | 0.444 | 0.664 | 0.688 | 0.838 | 0.826 | 0.86 | 0.843 | 0.864 | 0.989 | 1 |

**Table D.** Spearman correlation between the distributions of flexibility(L,C) for various constructions