

Real-Valued Logics for Typological Universals: Framework and Application

Tillmann Dönicke
Göttingen Centre for
Digital Humanities
University of Göttingen*
tillmann.doenicke@
uni-goettingen.de

Xiang Yu
Institute for Natural
Language Processing
University of Stuttgart
xiangyu@ims.
uni-stuttgart.de

Jonas Kuhn
Institute for Natural
Language Processing
University of Stuttgart
jonas.kuhn@ims.
uni-stuttgart.de

Abstract

This paper proposes a framework for the expression of typological statements which uses real-valued logics to capture the empirical truth value (truth degree) of a formula on a given data source, e.g. a collection of multilingual treebanks with comparable annotation. The formulae can be arbitrarily complex expressions of propositional logic. To illustrate the usefulness of such a framework, we present experiments on the Universal Dependencies treebanks for two use cases: (i) empirical (re-)evaluation of established formulae against the spectrum of available treebanks and (ii) evaluating new formulae (i.e. potential candidates for universals) generated by a search algorithm.

The availability of comparable treebanks – syntactically annotated corpora – for a growing number of typologically distinct languages (most prominently in the collaborative Universal Dependencies project (Nivre et al., 2016)) has led to a recent surge of interest in computational work aiming to detect systematic patterns in the grammatical systems of natural languages and/or to test hypotheses from theoretical work in language typology against empirical evidence. The treebank-based approach (Liu, 2010; Lochbihler, 2017; Gerdes et al., 2019; Bjerva et al., 2019c; Hahn et al., 2020) adds a more data-driven perspective to a strand of research in computational typology (Daumé and Campbell, 2007; Malaviya et al., 2017; Oncevay et al., 2019; Bjerva et al., 2019a; Bjerva et al., 2019b) that is based on carefully curated typological databases such as WALS¹ (Dryer and Haspelmath, 2013) or URIEL² (Littell et al., 2017).

The research strand in computational typology which relies on databases essentially builds on the language features that the long tradition of typological research has identified as most relevant for identifying language universals. Examples of such features are the relative order of verbs and their objects, and the order of nouns and their dependents such as adjectives, numerals and genitives. In the computational research relying on typological knowledge bases, the features are typically assumed to be Boolean and universals are formulated as propositional formulae. A major focus has been on (a) detecting universals that have the form of an implication between two typological variables, and (b) predicting the value of unknown features in typological databases based on systematic patterns in attested grammatical systems. Graphical models have been widely used to calculate the strength of an implication (Daumé and Campbell, 2007; Lu, 2013; Bjerva et al., 2019b; Bjerva et al., 2019a). While this approach is suitable if one wants to marginalize out the influence of confounding variables, it also constrains the investigated universals to have the form of an implication consisting of one implicand and usually one (but possibly multiple) implicant(s).

In principle, comparable treebanks can provide the basis for observing the empirical distribution of arbitrary grammatical patterns and thus explore a much larger space of potential candidates for universal typological properties – including combinations of more than two variables that cannot be reduced to logical implication. However, an integration of such an approach with linguistically grounded hypothesis

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

*Work performed while at University of Stuttgart.

¹<https://wals.info/>

²<http://www.cs.cmu.edu/~7Edmortens/uriel.html>

checking has to address two related challenges: (i) a theoretically guided way of navigating the enormous space of candidate propositions has to be developed, and (ii) a perspicuous framework is required for *expressing* multi-variable propositions and for *evaluating* them empirically against a full collection of comparable treebanks – while doing justice to the possibility of language-internal variation and tentative preferences by modeling features as real-valued.

This paper proposes an expressive framework that addresses the latter challenge. We specify a formalism and its semantics to evaluate typological formulae of arbitrary logical complexity. The core elements of our framework are customizable which gives prospective users the freedom to use their own implementations.

We also include a method for counteracting the bias in the sample of well-studied languages (which are much more likely to be included among the languages with a treebank), many of which are phylogenetically closely related, while other language families are only very sparsely represented.

To demonstrate its usefulness for empirical hypothesis testing, we run a number of experiments: 1) We re-evaluate established universals to test whether the evidence for universals provided by the framework reflects the broad consent. 2) Universals are always evaluated on a subset of all natural languages. We investigate how reliable it is to transfer the evaluation result obtained from a subset of languages or language families to unknown languages. 3) We use the framework to search new potential universals.

The framework and our experiments are available at <https://github.com/tidoe/typology-coling>.

1 Framework

1.1 Language–property matrix

Universals are modelled as logical formulae which consist of variables and logical connectives. The variables are organized in an $|L| \times |P|$ matrix V , where L is the set of investigated languages and P is the set of typological properties/features. The variable at v_p^ℓ represents the value for the property p of the language ℓ . In our framework, these values are truth values in the range $[0, 1]$. In the simplest case, they are binary truth values as in the example matrix below.

$$V = \begin{pmatrix} & \text{SOV} & \text{SVO} & \text{Postp.} & \text{Prep.} & \dots \\ \text{English} & 0 & 1 & 0 & 1 & \dots \\ \text{German} & 1 & 1 & 0 & 1 & \dots \\ \text{Japanese} & 1 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (1)$$

Formulae can be constructed from the language–property matrix by logical connectives, such as the conjunction in (2).³

$$\phi^\ell := v_{\text{SVO}}^\ell \wedge v_{\text{Prep.}}^\ell \quad (2)$$

1.2 Valuation function

The valuation function \mathcal{V} maps formulae to truth values. The function specifically defines the logical connectives for negation (\neg), conjunction (\wedge), disjunction (\vee) etc. For example, the valuation for conjunction in Boolean logic is defined as

$$\mathcal{V}(\phi_1^\ell \wedge \dots \wedge \phi_n^\ell) := \begin{cases} 1 & \mathcal{V}(\phi_1^\ell) = \dots = \mathcal{V}(\phi_n^\ell) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

³We use propositional logic and a language–property matrix in our framework to be more in line with traditional typological work, but equivalent formulae could be expressed using first-order logic. In doing so, the set of languages forms the universe and the properties correspond to one-place predicates. The formula in (2) would then correspond to $\text{SVO}(\ell) \wedge \text{Prep}(\ell)$. One could define a quantifier Q_w to perform the same weighting as described in Sec. 2.3, and equally evaluate $Q_w \ell \in L : \text{SVO}(\ell) \wedge \text{Prep}(\ell)$, a weaker variant than universal quantification with \forall .

With the language–property matrix in (1) and the valuation in (3), the formula in (2) evaluates as

$$\mathcal{V}(\phi^{\text{English}}) = \mathcal{V}(v_{\text{SVO}}^{\text{English}} \wedge v_{\text{Prep.}}^{\text{English}}) = 1 \quad (4)$$

since $\mathcal{V}(v_{\text{SVO}}^{\text{English}}) = \mathcal{V}(v_{\text{Prep.}}^{\text{English}}) = 1$.

1.3 Averaging over languages

Let $\mathcal{V}(\phi^\ell)$ be the truth value of the formula ϕ for the language $\ell \in L$ under the valuation \mathcal{V} . Then $|L|$ truth values can be calculated for $\mathcal{V}(\phi^\ell)$, one for every language. A universal is a statement that is supposed to hold for “all” languages – in the statistical sense rather than the absolute sense – but a universal quantification (i.e. a conjunction of the values) would be misleading since it is not robust to outliers. The score for a formula is hence calculated as the (weighted) average

$$s(\phi) = \frac{\sum_{\ell \in L} w(\phi, \ell) \cdot \mathcal{V}(\phi^\ell)}{\sum_{\ell \in L} w(\phi, \ell)} \quad (5)$$

with w being the weight function. An unweighted average corresponds to a weighted average with all weights being 1.

2 Implementation

The framework allows implementing individual definitions for the variable matrix V , the valuation \mathcal{V} and the weight function w . We experimented with various implementations for all of those (Dönicke, 2020); this section explains the set-up used for the experiments in section 3.

The UD treebanks v2.5 (Zeman et al., 2019) consist of dependency treebanks for 90 languages from 20 families and 39 subfamilies. The treebanks have a uniform annotation which allows defining properties based on dependency constructions and extracting statistics for various languages of the world. This data can be used to investigate syntactic universals.

2.1 Variables from Universal Dependencies

We extracted a language–property matrix from the UD treebanks. The properties are specific constructions and the values are their relative frequencies. We extracted two types of properties:

- Single-link property: relative frequency of a construction involving one head and one dependent, e.g.

$$v_{\text{noun-amod:adj}}^\ell = \frac{\# \left[\begin{array}{c} \text{amod} \\ \swarrow \quad \searrow \\ \text{NOUN} \quad \text{ADJ} \end{array} \right] (\ell)}{\# \left[\begin{array}{c} \text{amod} \\ \swarrow \quad \searrow \\ \text{ADJ} \quad \text{NOUN} \end{array} \right] (\ell) + \# \left[\begin{array}{c} \text{amod} \\ \swarrow \quad \searrow \\ \text{NOUN} \quad \text{ADJ} \end{array} \right] (\ell)} \quad (6)$$

- Double-link property: relative frequency of a construction involving one head and two dependents, e.g.

$$v_{\text{verb-nsubj:noun-obj:noun}}^\ell = \frac{\# \left[\begin{array}{c} \text{obj} \\ \swarrow \quad \searrow \\ \text{nsubj} \\ \swarrow \quad \searrow \\ \text{VERB} \quad \text{NOUN} \quad \text{NOUN} \end{array} \right] (\ell)}{\# \left[\begin{array}{c} \text{obj} \\ \swarrow \quad \searrow \\ \text{nsubj} \\ \swarrow \quad \searrow \\ \text{VERB} \quad \text{NOUN} \quad \text{NOUN} \end{array} \right] (\ell) + \dots + \# \left[\begin{array}{c} \text{obj} \\ \swarrow \quad \searrow \\ \text{nsubj} \\ \swarrow \quad \searrow \\ \text{NOUN} \quad \text{NOUN} \quad \text{VERB} \end{array} \right] (\ell)} \quad (7)$$

six possibilities

Here, $\#[*](\ell)$ returns how often the construction $*$ appears for ℓ . The subscripts of v^ℓ represent the constructions: hyphens connect words, the head is represented by its part-of-speech tag, and the dependents are represented by their dependency relation and their part-of-speech tag, combined by a colon.

2.2 Fuzzy logic

The relative frequencies are in the interval $[0, 1]$, hence our valuation function has to define logical connectives for a real-valued logic. A common example for real-valued logic is fuzzy logic (Zadeh, 1965), which defines negation, conjunction and disjunction as follows:

$$\mathcal{V}(\neg\phi^\ell) := 1 - \mathcal{V}(\phi^\ell) \quad (8)$$

$$\mathcal{V}(\phi_1^\ell \wedge \dots \wedge \phi_n^\ell) := \min\{\mathcal{V}(\phi_1^\ell), \dots, \mathcal{V}(\phi_n^\ell)\} \quad (9)$$

$$\mathcal{V}(\phi_1^\ell \vee \dots \vee \phi_n^\ell) := \max\{\mathcal{V}(\phi_1^\ell), \dots, \mathcal{V}(\phi_n^\ell)\} \quad (10)$$

Implication and equivalence are shortcuts for combinations of the three previous connectives:

$$\mathcal{V}(\phi_1^\ell \Rightarrow \phi_2^\ell) := \mathcal{V}(\neg\phi_1^\ell \vee \phi_2^\ell) \quad (11)$$

$$\mathcal{V}(\phi_1^\ell \Leftrightarrow \phi_2^\ell) := \mathcal{V}((\phi_1^\ell \Rightarrow \phi_2^\ell) \wedge (\phi_2^\ell \Rightarrow \phi_1^\ell)) \quad (12)$$

All of these valuation functions are generalizations from Boolean logic. Other examples for many-valued logics are described in e.g. Smith (2012) and many other works. We chose fuzzy logic, in contrast to e.g. probabilistic logic (Mizraji, 1992), because the definitions of the fuzzy connectives do not make any assumption about the dependence of typological properties (cf. Dubois and Prade (1993)). Properties like “SVO” and “prepositions” could both be considered a subtype of head-initiality and therefore not be independent under a linguistic point of view. In probabilistic logic, the truth value of a conjunction is defined as the product of each conjunct’s truth value, which assumes the independence of the conjuncts.

2.3 Phylogenetic weighting

To confirm a typological universal, it is not sufficient to validate it on as many languages as possible, it is also necessary to validate it on languages from many language families. For example, if a candidate for a universal is tested on English, German, French, Italian, Spanish and Japanese and the universal holds for all of them but Japanese, then Japanese is not simply an outlier, it is also possible that the “universal” only holds within the Indo-European languages. Traditionally, typologists counteract the influence of overrepresented language families through different sampling methods (cf. Bickel (2011), Song (2018)), e.g. sampling only languages with different values (on the properties of interest) for each family (Dryer, 1989; Bickel, 2008). This “genealogical sampling”, however, requires binary/categorical values and a representative database of the world’s languages, and is not applicable in our experiments with the UD treebanks. Thus, we give each language family equal importance by setting the weight function to

$$w(\phi, \ell) = \begin{cases} \frac{w(\phi, \text{parent}(\ell))}{|\text{siblings}(\ell)| + 1} & \text{if } \ell \text{ has a parent}^4 \\ 1 & \text{otherwise} \end{cases} \quad (13)$$

The weights for the current example are shown in Figure 1. This approach does not undersample languages, instead all available data is used. To the best of our knowledge, this is a novel method that maximizes the usage of all available data while alleviating the sampling bias in the data.⁵ We believe its

⁴ ϕ could be dropped from this definition of w , as w only depends on ℓ . We still keep it because other implementations of w might also depend on ϕ (see Dönicke (2020) for examples).

⁵There are other works also using phylogenetic information but in different ways (Dunn et al., 2011; Levinson and Gray, 2012; Jäger, 2018).

utility is worth further examination, however, in our preliminary experiments, the phylogenetic weighting demonstrates better agreement with the existing universals from the literature than the unweighted average.⁶

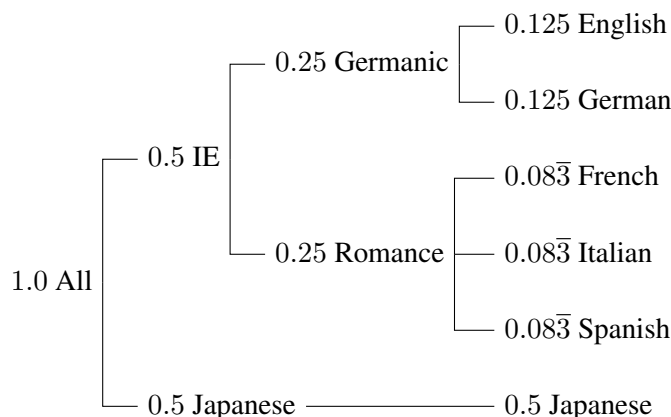


Figure 1: Example language family tree with weights.

As Bickel (2011) and Dryer (2018) point out, geography is also an important factor, i.e. neighboring languages are more likely to have common properties than distant languages. However, both works agree that measuring geographic distance of languages brings its own complications and therefore it is not taken up in this paper.

3 Experiments

The goal of our experiments is to demonstrate the usefulness of the proposed framework for empirical exploration of the language-typological space beyond simple implications and Boolean variables. The framework is to serve as a methodological tool and should ultimately be complemented with a theoretically motivated agenda for exploring systematic correlations among underexplored properties of grammatical systems. To establish how the framework can be used we proceed in three steps: First, we re-evaluate some famous universals of Greenberg (1963) and discuss the interpretations of varying truth values and standard deviations (Sec. 3.1).

In the second step, we demonstrate a scenario for testing unexplored candidates for typological universals (which can consist of arbitrary formulae from propositional logic) against the full spectrum of treebanks, taking into variability with respect to each property in the real-valued logic. We work with a simple procedure of generating candidate formulae: taking over the shape of potentially relevant logical combinations from established universals and replacing the properties they include with new ones. A framework is useful for empirical exploration if its diagnostics are robust: a statement capturing a systematic relationship should generalize from one sufficiently large sample of languages to another. We test this by splitting the set of treebanks in an estimation set and a held-out set (Sec. 3.2).

Finally, we enumerate implications of two variables from a selected set of single-link properties and discuss the top-scoring formulae, also in comparison to the findings from previous work on binary properties (Sec. 3.3).

3.1 Evaluation of Greenberg's universals

The framework can be used to evaluate universals proposed in typological literature. The most important example are the 45 universals by Greenberg (1963). 14 of them can be expressed with the UD variables

⁶In a comparison with the Universals Archive (UA) (Plank and Filimonova, 2000), we found that averaging with phylogenetic weighting achieved more similar results to the typological studies than an unweighted average. The similarity was measured by calculating the mean absolute error between the framework score and the UA score over a set of 51 universals in the UA. For each universal, the UA lists tested languages and counterexamples. The UA score was defined as the percentage of examined languages which are no counterexamples. The experiment is described in Dönicke (2020).

| s | σ | ϕ | Greenberg |
|------|----------|--|-----------|
| 1.00 | 0.01 | verb-nsubj:noun-obj:noun \Rightarrow case:adp-noun | #3 |
| 0.99 | 0.02 | verb-nsubj:noun-obj:noun \Rightarrow (verb-nsubj:noun-obj:noun + nsubj:noun-verb-obj:noun) | #6 |
| 0.99 | 0.02 | verb-nsubj:noun-obj:noun \Rightarrow noun-amod:adj | #17 |
| 0.99 | 0.02 | acl:verb-noun \Rightarrow (noun-case:adp \vee amod:adj-noun) | #24 |
| 0.96 | 0.12 | nsubj:noun-obj:noun-verb \Rightarrow noun-case:adp | #4 |
| 0.94 | 0.14 | (nsubj:noun-obj:noun-verb \wedge noun-nmod:noun) \Rightarrow noun-amod:adj | #5 |
| 0.91 | 0.13 | verb-obj:pron \Rightarrow verb-obj:noun | #25 |
| 0.91 | 0.14 | adj-advmod:adv \Rightarrow (noun-amod:adj \wedge verb-obj:noun) | #21 |
| 0.91 | 0.16 | nsubj:noun-obj:noun-verb \Rightarrow advmod:adv-verb | #7 |
| 0.91 | 0.12 | amod:adj-noun \Rightarrow (det:pron-noun \wedge nummod:num-noun) | #18 |
| 0.90 | 0.16 | verb-nsubj:noun-obj:noun + nsubj:noun-verb-obj:noun + nsubj:noun-obj:noun-verb | #1 |
| 0.82 | 0.30 | (verb-nsubj:noun-obj:noun \Rightarrow aux:aux-verb) \wedge (nsubj:noun-obj:noun-verb \Rightarrow verb-aux:aux) | #16 |
| 0.81 | 0.31 | case:adp-noun \Leftrightarrow noun-nmod:noun | #2 |
| 0.77 | 0.31 | obj:noun-verb \Rightarrow xcomp:verb-verb | #13 |

Table 1: Scores and standard deviations for Greenberg’s universals.⁷

from section 2.1. Most of these universals shown in Table 1 are implications, except for #1 (absolute) and #2 (equivalence). The most complex formula in the list is #16 which is composed of four variables.

The formulae are ranked by framework score, i.e. the weighted average truth value. 11 of the formulae score 0.90 or higher, i.e. they could be classified as “very true”. #2 is an equivalence and #16 is a quasi-equivalence which are generally harder to fulfil than one-way implications (in fuzzy logic, the truth value of an equivalence is defined as the minimum of the truth values of the two composed implications). The formula with the lowest score, #13, is an example for two conflicting universal tendencies, namely 1) the tendency to order subordinate clause and main verb the same as object and verb (this is predicted by different linguistic theories, e.g. the head-dependent theory), and 2) the tendency to put clauses after their single-word siblings (e.g. Dryer (2003)). One can reformulate #13 to take both tendencies into account by flipping the word orders on both sides of the implication or, equivalently, reversing the direction of the implication. This yields $s = 0.95$ ($\sigma = 0.09$).

σ denotes the (weighted) standard deviation which ranges between 0 and 0.5. A value of 0 means that the formula is equally true for all languages; a value of 0.5 means that the formula is absolutely true for half of the languages and absolutely false for the other half. Compare e.g. #3 (VSO \Rightarrow prepositions) and #4 (SOV \Rightarrow postpositions) – and the respective Figures 2 (a) and (b) – which have similar values for s but different values for σ . Most of the 90 languages in the treebanks show (near) 0% VSO order or 100% prepositions and therefore get a truth value of 1 for #3. Only some languages (at the bottom-left corner of Figure 2 (a)) get lower truth values for #3 which raises the standard deviation to 0.01. Regarding #4, there are more languages showing mid-range values for SOV order and postpositions, and therefore also mid-range truth values for #4. Since most languages still have a high truth value for #4, the score remains at 0.96 but the standard deviation of 0.12 signals a greater number of languages with truth values diverging from that score.

⁷Addition is not a logical connective but used here to calculate the value of properties which cannot be directly extracted from the UD treebanks. For example, the degree to which a language has S-O order is the percentage occupied by V-S-O, S-V-O and S-O-V constructions. Thus, the value $v_{\text{nsubj:noun-obj:noun}}^{\ell}$ can be calculated as $v_{\text{verb-nsubj:noun-obj:noun}}^{\ell} + v_{\text{nsubj:noun-verb-obj:noun}}^{\ell} + v_{\text{nsubj:noun-obj:noun-verb}}^{\ell}$. Addition is restricted to variables that are mutually exclusive, i.e. that are calculated with the same denominator (compare eq. (7)). That said, (out-of-logic) addition in our framework is comparable to addition of probabilities of disjoint events in probability theory; it is not as powerful and does not serve the same purpose as (in-logic) addition (“strong disjunction”) in certain real-valued logics, e.g. Łukasiewicz logic (Łukasiewicz and Tarski (1930); cf. Bergmann (2008, p. 179)), where it is possible to add arbitrary variables.

The result of an addition thus expresses the truth value for any (combination) of several alternative constructions being present in a language. Greenberg (1963), as others, uses the terms “dominant” and “alternative” to indicate two degrees of relative frequency of word orders. This binary distinction cannot be directly expressed with UD variables because they already express infinitely many degrees of relative frequency. We denote “ p is a dominant word order in ℓ ” as v_p^{ℓ} and “ p is an alternative word order of q in ℓ ” as $v_p^{\ell} + v_q^{\ell}$, since we think that these are the most appropriate real-valued equivalents.

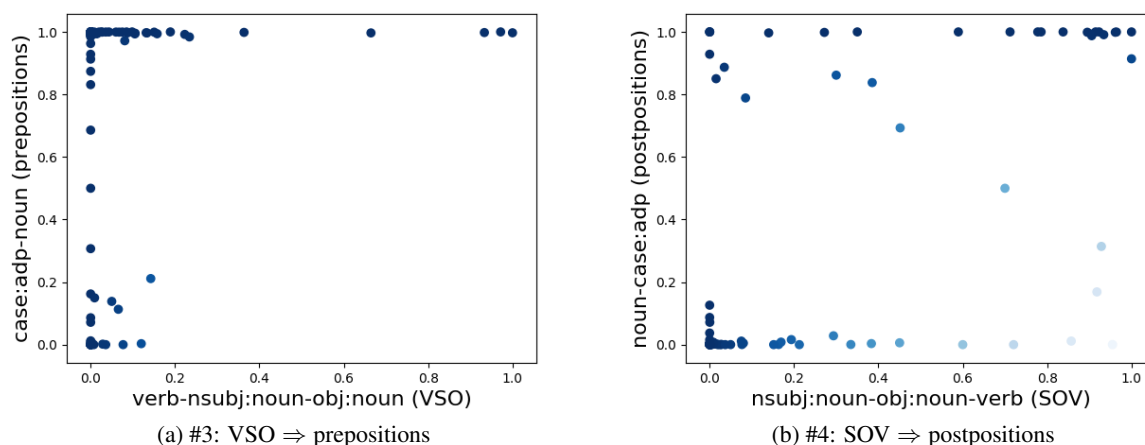


Figure 2: The properties of Greenberg’s universals #3 and #4 for all languages. Each point is a language and the color shading indicates the truth value of the universal (dark is true; light is false).

3.2 Universal prediction

For the second experiment, we generate over 200k formulae from the Greenbergian universals in Table 1 by keeping the structures and varying the variables. (The aim of keeping the structures is to reduce the search space and to generate only formulae that have the “usual shape” of traditional universals.) Logical tautologies are excluded by prohibiting the variables from appearing more than once within the same formula.

We evaluate these formulae in six runs. In each run, we randomly split the languages into two disjoint sets for every formula and evaluate it on both sets separately. The idea is to predict the score for the second set of languages from the score on the first set of languages. Ideally, the scores from both sets should be identical because this means that the initial score is reliable even for unknown data. After each run, we take the two values $s_1(\phi)$, $s_2(\phi)$ for every formula ϕ and calculate the overall root-mean-square error (RMSE):

$$RMSE = \sqrt{\frac{1}{|\Phi|} \cdot \sum_{\phi \in \Phi} |s_2(\phi) - s_1(\phi)|^2} \quad (14)$$

This measures how similar our predictions are to the actual values. Since we are especially interested in high-scoring formulae, we additionally calculate the RMSE on only those value pairs where $s_1(\phi) \geq 90\%$.

The six runs differ in the split method. 1) The languages are split either on the language, subfamily or family level. For the latter two, the languages of a (sub)family are either completely in the first or the second set. 2) The languages/subfamilies/families are either split 50%/50% or 20%/80%. The second percentages are more realistic since there is often data for a few languages/families from which one wants to predict how true a formula is in general.

The results in Table 2 illustrate two things. The less surprising finding is that evaluations on two balanced sets are more similar than evaluations on imbalanced sets. Arguably, because an evaluation on 20% of the data is much less representative than an evaluation on 80% of the data, whereas the two sets of a 50/50 split are about equally representative. Also, splitting on the language level yields more similar evaluation results than evaluating on two sets with completely different (sub)families. The second finding is that the RMSE is generally lower for formulae that are very true in at least one set of languages. This is also visible in Figure 3, the data points scatter in the shape of a lens around the identity graph, i.e. formulae that are found to be very true or very false on one set of languages are likely to achieve a similar score on a completely different set of languages. Formulae with mid-range truth values (0.2–0.8), on the

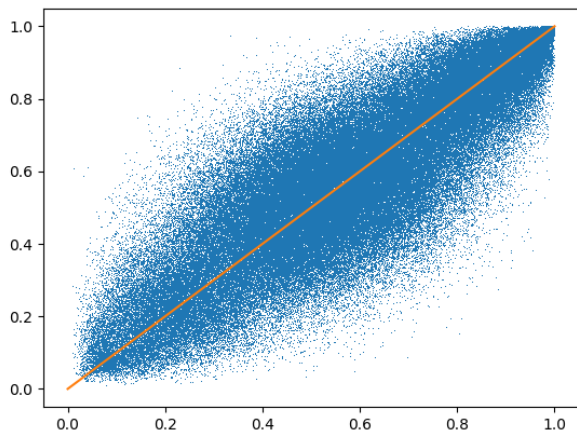


Figure 3: Truth values s_1 and s_2 from both sets (languages 50/50 condition) and the identity function.

| split level | split | | RMSE | |
|-------------|-------|-------|------|----------------|
| | % | abs. | all | $s_1 \geq 0.9$ |
| languages | 50/50 | 45/45 | 0.13 | 0.07 |
| | 20/80 | 18/72 | 0.15 | 0.12 |
| subfamilies | 50/50 | 19/20 | 0.14 | 0.08 |
| | 20/80 | 8/31 | 0.16 | 0.13 |
| families | 50/50 | 10/10 | 0.17 | 0.11 |
| | 20/80 | 4/16 | 0.22 | 0.23 |

Table 2: RMSEs for all split conditions, over all formulae and over formulae with a high score on the first set.

other hand, exhibit a much larger variance. This indicates that certain formulae have a higher potential of being universals than others. For subfamilies and families, the prediction error increases (the appendix contains scatter plots for all six runs).

3.3 Search for universals

Implications between two variables have been investigated in many previous studies. In the third experiment, we choose ten dependency relations⁸ and calculate pairwise implications between the corresponding single-link properties. Since each relation can be left- or right-directed, there are 20×20 possible implications (the appendix contains the table with all implications).⁹

Table 3 ranks the implications by average truth value. As a matter of fact, the subject precedes its head in the majority of UD languages, resulting in high scores for implications where “HEAD-nsubj” is the antecedent or “nsubj-HEAD” is the consequent (principle of explosion). Those implications are responsible for about half of the rows in Table 3. For example,

(#19) HEAD-advmod \Rightarrow nsubj-HEAD and

(#20) advmod-HEAD \Rightarrow nsubj-HEAD

are quite uninteresting implications since both achieve a score of 0.92 but have complementary antecedents. Thus, they only reveal that most languages exhibit subject-verb order, independently from the position of adverbs. However, there are also some formulae that were already discovered in different works and are also included in the Universals Archive, such as

(#22) a. amod-HEAD \Rightarrow nummod-HEAD

b. “IF the descriptive adjective precedes the noun, THEN, with overwhelmingly more than chance frequency, the demonstrative and the numeral do likewise.” (Universals Archive, #57),

which is part of the 18th universal of Greenberg, part of the 57th in the Universals Archive and the 7th in Daumé and Campbell (2007).

⁸acl, advmod, amod, aux, case, det, nmod, nsubj, nummod, obj

⁹We removed POS information in this experiment to lower the number of variables. The single-link properties without POS are very similar to those with the most frequent POS. For example, the mean absolute error of $\vec{v}_{\text{case-HEAD}}$ and $\vec{v}_{\text{case:adp-noun}}$ is 0.03 ($\sigma = 0.10$) and the mean absolute error over all single-link properties is 0.05 ($\sigma = 0.12$).

| # | <i>s</i> | σ | ϕ | UA | G. | D. | B. |
|----|----------|----------|--------------------------------------|-----|----|-------|----|
| 1 | 0.97 | 0.07 | acl-HEAD \Rightarrow nmod-HEAD | 176 | | | |
| 2 | 0.96 | 0.07 | acl-HEAD \Rightarrow nsubj-HEAD | | | | |
| 3 | 0.96 | 0.07 | HEAD-det \Rightarrow nsubj-HEAD | | | | |
| 4 | 0.95 | 0.11 | acl-HEAD \Rightarrow HEAD-case | 62 | 24 | 30 | |
| 5 | 0.95 | 0.10 | HEAD-amod \Rightarrow nsubj-HEAD | | | | |
| 6 | 0.95 | 0.07 | HEAD-aux \Rightarrow nsubj-HEAD | | | | |
| 7 | 0.95 | 0.07 | HEAD-case \Rightarrow nsubj-HEAD | | | | |
| 8 | 0.95 | 0.08 | HEAD-nsubj \Rightarrow nummod-HEAD | | | | |
| 9 | 0.94 | 0.09 | amod-HEAD \Rightarrow nsubj-HEAD | | | | |
| 10 | 0.94 | 0.09 | nmod-HEAD \Rightarrow nsubj-HEAD | | | | |
| 11 | 0.94 | 0.09 | HEAD-nsubj \Rightarrow HEAD-obj | | | | 3 |
| 12 | 0.93 | 0.11 | HEAD-acl \Rightarrow nsubj-HEAD | | | | |
| 13 | 0.93 | 0.12 | HEAD-aux \Rightarrow HEAD-case | | | | |
| 14 | 0.93 | 0.11 | HEAD-aux \Rightarrow nmod-HEAD | | | | |
| 15 | 0.93 | 0.10 | aux-HEAD \Rightarrow nsubj-HEAD | | | | |
| 16 | 0.93 | 0.18 | case-HEAD \Rightarrow nsubj-HEAD | | | | |
| 17 | 0.93 | 0.12 | case-HEAD \Rightarrow HEAD-obj | | | 2, 12 | 2 |
| 18 | 0.92 | 0.17 | acl-HEAD \Rightarrow obj-HEAD | | | 9 | |
| 19 | 0.92 | 0.15 | HEAD-advmod \Rightarrow nsubj-HEAD | | | | |
| 20 | 0.92 | 0.10 | advmod-HEAD \Rightarrow nsubj-HEAD | | | | |
| 21 | 0.92 | 0.12 | amod-HEAD \Rightarrow det-HEAD | 57 | 18 | 13 | 7 |
| 22 | 0.92 | 0.09 | amod-HEAD \Rightarrow nummod-HEAD | 57 | 18 | 7 | |
| 23 | 0.92 | 0.17 | HEAD-aux \Rightarrow det-HEAD | | | | |
| 24 | 0.92 | 0.12 | HEAD-aux \Rightarrow nummod-HEAD | | | | |
| 25 | 0.92 | 0.14 | HEAD-nmod \Rightarrow nsubj-HEAD | | | | |
| 26 | 0.92 | 0.11 | HEAD-nsubj \Rightarrow HEAD-nummod | | | | |
| 27 | 0.91 | 0.23 | HEAD-case \Rightarrow nmod-HEAD | 1 | 2 | 1 | 1 |
| 28 | 0.90 | 0.19 | acl-HEAD \Rightarrow advmod-HEAD | | | | |
| 29 | 0.90 | 0.12 | HEAD-advmod \Rightarrow aux-HEAD | | | | |
| 30 | 0.90 | 0.19 | det-HEAD \Rightarrow nsubj-HEAD | | | | |

Table 3: Top 30 implications with indices to similar formulae in previous works: Universals Archive, Greenberg (1963), Daumé and Campbell (2007), Bjerva et al. (2019b)

The implication with the highest score,

- (#1) a. acl-HEAD \Rightarrow nmod-HEAD
b. “IF the relative clause precedes the noun, THEN the genitive precedes the noun.” (Universals Archive, #176),

concerning the order of adjectival clauses and nominal noun modifiers and first described by Hawkins (1983), is listed in the UA but was not detected by the computer-assisted models of Daumé and Campbell (2007) and Bjerva et al. (2019b). This shows that corpus statistics can gain advantage over knowledge bases in some cases.

Last but not least, universals that, as far as we know, have not been described so far are suggested, e.g. the following four and their composition in (1).

- (#13) HEAD-aux \Rightarrow HEAD-case
(#14) HEAD-aux \Rightarrow nmod-HEAD
(#23) HEAD-aux \Rightarrow det-HEAD
(#24) HEAD-aux \Rightarrow nummod-HEAD
- (1) a. HEAD-aux \Rightarrow (HEAD-case \wedge nmod-HEAD \wedge det-HEAD \wedge nummod-HEAD)
b. IF the auxiliary follows the verb, THEN the adposition follows the noun AND genitive, article and numeral all precede the noun.

Although the suggested universals are not guaranteed to be linguistically sound or meaningful, it could certainly provide inspiration to typologists for careful examination and interpretation.

4 Conclusion

This paper proposes a framework for the expression of typological statements which uses real-valued logics to capture the empirical truth value of a formula in a given collection of comparable treebanks. We demonstrate the application of the framework by evaluating established formulae as well as new formulae generated from a search algorithm. The components of the framework can be customized: users not working with the Universal Dependencies can simply exchange the data component; the weighting approach we provide can also be easily exchanged.

If a user wants to empirically check their own linguistic features, it is straightforward to obtain the relative frequencies from the treebank collection, complementing existing curated databases such as WALS. Very little manual work is needed. Both approaches have their pros and cons: typological knowledge bases could be subjective or incomplete, while data from the treebanks are subject to problems such as annotation inconsistencies, genre differences etc.

This paper discussed some experiments which have the purpose of providing an illustration of the expressiveness of the framework, focusing here on word-order variables. However, other variables could be expressed as well, and the framework also supports inclusion and combination of variables from completely different sources (e.g. phonological properties of a language), as long as the values range between 0 (false) and 1 (true). This especially allows using variables from corpora and from typological knowledge bases together.

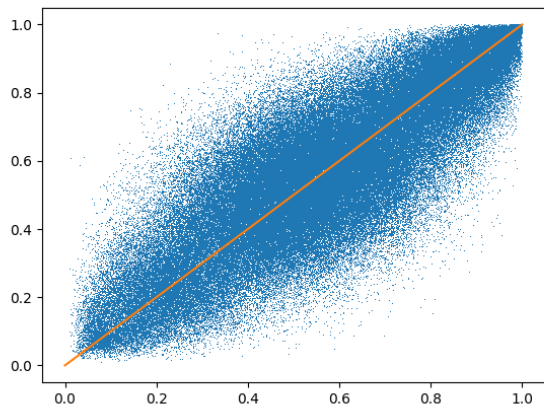
References

- Merrie Bergmann. 2008. *An introduction to many-valued and fuzzy logic: semantics, algebras, and derivation systems*. Cambridge University Press.
- Balthasar Bickel. 2008. A refined sampling procedure for genealogical control. *STUF - Language Typology and Universals*, 61:221–233.
- Balthasar Bickel. 2011. Absolute and statistical universals. In Patrick Colm Hogan, editor, *The Cambridge Encyclopedia of the Language Sciences*, pages 77–79. Cambridge University Press, Cambridge.
- Johannes Bjerva, Yova Kementchedjheva, Ryan Cotterell, and Isabelle Augenstein. 2019a. A probabilistic generative model of linguistic typology. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1529–1540, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- Johannes Bjerva, Yova Kementchedjheva, Ryan Cotterell, and Isabelle Augenstein. 2019b. Uncovering probabilistic implications in typological knowledge bases. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3924–3930, Florence, Italy, July. Association for Computational Linguistics.
- Johannes Bjerva, Robert Östling, Maria Han Veiga, Jörg Tiedemann, and Isabelle Augenstein. 2019c. What do language representations really represent? *Computational Linguistics*, 45(2):381–389.
- Hal Daumé, III and Lyle Campbell. 2007. A Bayesian model for discovering typological implications. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 65–72, Prague, Czech Republic, June. Association for Computational Linguistics.
- Tillmann Döncke. 2020. Evaluation of complex typological universals with language vectors and real-valued logics. Master's thesis, University of Stuttgart.
- Matthew S. Dryer and Martin Haspelmath, editors. 2013. *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Matthew S. Dryer. 1989. Large linguistic areas and language sampling. *Studies in Language*, 13:257–292.
- Matthew S. Dryer. 2003. Adverbial subordinators and word order asymmetries. In John A. Hawkins and Anna Siewierska, editors, *Performance principles of word order*, EUROTYPE Working Papers, pages 50–67. European Science Foundation.
- Matthew S. Dryer. 2018. On the order of demonstrative, numeral, adjective, and noun. *Language*, 94(4):798–833.

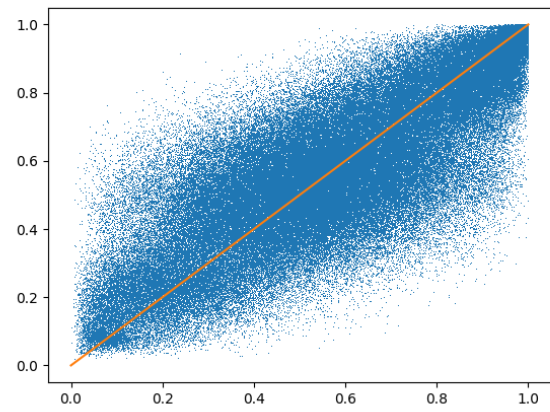
- Didier Dubois and Henri Prade. 1993. Fuzzy sets and probability: misunderstandings, bridges and gaps. In *[Proceedings 1993] Second IEEE International Conference on Fuzzy Systems*, volume 2, pages 1059–1068.
- Michael Dunn, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345):79–82.
- Kim Gerdes, Sylvain Kahane, and Xinying Chen. 2019. Rediscovering Greenberg’s word order universals in UD. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 124–131, Paris, France, 26 August. Association for Computational Linguistics.
- Joseph H. Greenberg. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2:73–113.
- Michael Hahn, Dan Jurafsky, and Richard Futrell. 2020. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5):2347–2353.
- John A. Hawkins. 1983. *Word Order Universals*. Academic Press, New York.
- Gerhard Jäger. 2018. Global-scale phylogenetic linguistic inference from lexical resources. *Scientific data*, 5(1):1–16.
- Stephen C. Levinson and Russell D. Gray. 2012. Tools from evolutionary biology shed new light on the diversification of languages. *Trends in cognitive sciences*, 16(3):167–173.
- Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 8–14, Valencia, Spain, April. Association for Computational Linguistics.
- Haitao Liu. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 120(6):1567–1578.
- Bethany C. Lochbihler. 2017. Typology across treebanks. Master’s thesis, University of Stuttgart.
- Xia Lu. 2013. Exploring word order universals: a probabilistic graphical model approach. In *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*, pages 150–157, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Jan Łukasiewicz and Alfred Tarski. 1930. Untersuchungen über den Aussagenkalkül. *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 23, cl. iii:39–50. English translation by J. H. Woodger, “Investigations into the Sentential Calculus.” In: Alfred Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 2nd ed., Indianapolis: Hackett Publishing Co., 1983, pages 38–59.
- Chaitanya Malaviya, Graham Neubig, and Patrick Littell. 2017. Learning language representations for typology prediction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2529–2535, Copenhagen, Denmark, September. Association for Computational Linguistics.
- Eduardo Mizraji. 1992. Vector logics: The matrix-vector representation of logical calculus. *Fuzzy Sets and Systems*, 50(2):179 – 185.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 1659–1666, Portorož, Slovenia, May. European Language Resources Association (ELRA).
- Arturo Oncevay, Barry Haddow, and Alexandra Birch. 2019. Towards a multi-view language representation: A shared space of discrete and continuous language features. *Phonology*, 27(28):84–51.
- Frans Plank and Elena Filimonova. 2000. The Universals Archive: A brief introduction for prospective users. *STUF-Language Typology and Universals*, 53(1):109–123.
- Nicholas J. J. Smith. 2012. Many-valued logics. *The Routledge companion to philosophy of language*, pages 636–651.
- Jae Jung Song. 2018. Language samples and sampling methods. In *Linguistic Typology*, chapter 5, pages 77–117. Oxford University Press.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Noëmi Aeppli, Željko Agić, Lars Ahrenberg, Gabrielè Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograinne Evelyn, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Olájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Kamil Kopacewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê H'ông, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Maria Liovina, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărânduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskiy, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisepp, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Luong Nguy~ên Thị, Huy ~ên Nguy~ên Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayo Olùòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvreliid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Răăbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roşca, Olga Rudina, Jack Rueter, Shoal Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Takaaki Tanaka, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uriá, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Manying Zhang, and Hanzhi Zhu. 2019. Universal dependencies 2.5. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

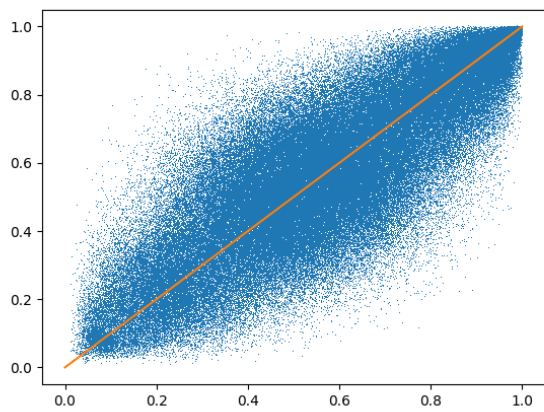
Appendix A. Scatter plots



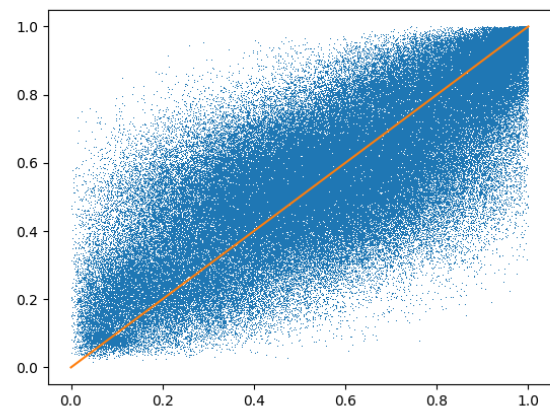
(a) languages 50/50 condition



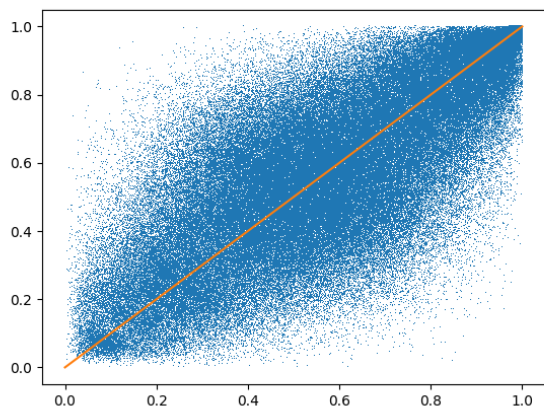
(b) languages 20/80 condition



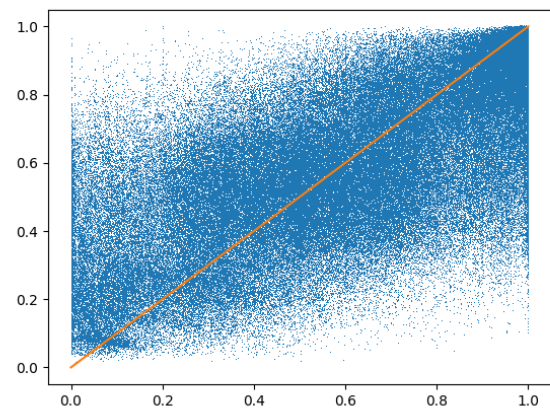
(c) subfamilies 50/50 condition



(d) subfamilies 20/80 condition



(e) families 50/50 condition



(f) families 20/80 condition

Figure 4: Truth values s_1 and s_2 from both sets and the identity function for six runs.

Appendix B. Implication matrix

| | HEAD-acl | acl-HEAD | HEAD-advmod | advmod-HEAD | HEAD-amod | amod-HEAD | HEAD-aux | aux-HEAD | HEAD-case | case-HEAD | HEAD-det | det-HEAD | HEAD-nmod | nmod-HEAD | HEAD-nsbj | nsbj-HEAD | HEAD-nummod | nummod-HEAD | HEAD-obj | obj-HEAD |
|-------------|----------|----------|-------------|-------------|-----------|-----------|----------|----------|-----------|-----------|----------|----------|-----------|-----------|-----------|-----------|-------------|-------------|----------|----------|
| HEAD-acl | 0.92 | 0.41 | 0.58 | 0.78 | 0.72 | 0.68 | 0.49 | 0.87 | 0.60 | 0.77 | 0.56 | 0.85 | 0.75 | 0.64 | 0.46 | 0.93 | 0.57 | 0.85 | 0.81 | 0.58 |
| acl-HEAD | 0.59 | 0.92 | 0.64 | 0.90 | 0.68 | 0.89 | 0.87 | 0.70 | 0.95 | 0.60 | 0.65 | 0.89 | 0.60 | 0.97 | 0.60 | 0.96 | 0.65 | 0.89 | 0.64 | 0.92 |
| HEAD-advmod | 0.90 | 0.78 | 0.77 | 0.71 | 0.86 | 0.81 | 0.75 | 0.90 | 0.81 | 0.87 | 0.74 | 0.83 | 0.82 | 0.82 | 0.74 | 0.92 | 0.76 | 0.85 | 0.85 | 0.78 |
| advmod-HEAD | 0.64 | 0.58 | 0.29 | 0.77 | 0.52 | 0.71 | 0.55 | 0.65 | 0.71 | 0.55 | 0.37 | 0.81 | 0.51 | 0.70 | 0.31 | 0.92 | 0.37 | 0.83 | 0.58 | 0.62 |
| HEAD-amod | 0.89 | 0.68 | 0.71 | 0.81 | 0.91 | 0.57 | 0.65 | 0.88 | 0.74 | 0.81 | 0.72 | 0.78 | 0.82 | 0.71 | 0.58 | 0.95 | 0.72 | 0.79 | 0.80 | 0.72 |
| amod-HEAD | 0.68 | 0.72 | 0.52 | 0.86 | 0.43 | 0.91 | 0.70 | 0.71 | 0.80 | 0.61 | 0.46 | 0.92 | 0.55 | 0.86 | 0.46 | 0.94 | 0.45 | 0.92 | 0.65 | 0.73 |
| HEAD-aux | 0.70 | 0.87 | 0.65 | 0.90 | 0.71 | 0.88 | 0.89 | 0.62 | 0.93 | 0.67 | 0.64 | 0.92 | 0.67 | 0.93 | 0.64 | 0.95 | 0.63 | 0.92 | 0.66 | 0.89 |
| aux-HEAD | 0.87 | 0.49 | 0.55 | 0.75 | 0.70 | 0.65 | 0.38 | 0.89 | 0.60 | 0.75 | 0.56 | 0.79 | 0.71 | 0.64 | 0.41 | 0.93 | 0.55 | 0.80 | 0.78 | 0.52 |
| HEAD-case | 0.60 | 0.77 | 0.55 | 0.87 | 0.61 | 0.81 | 0.75 | 0.67 | 0.95 | 0.45 | 0.56 | 0.89 | 0.52 | 0.91 | 0.48 | 0.95 | 0.56 | 0.88 | 0.56 | 0.87 |
| case-HEAD | 0.95 | 0.60 | 0.71 | 0.81 | 0.80 | 0.74 | 0.60 | 0.93 | 0.55 | 0.95 | 0.65 | 0.85 | 0.86 | 0.67 | 0.60 | 0.93 | 0.65 | 0.87 | 0.93 | 0.60 |
| HEAD-det | 0.89 | 0.85 | 0.81 | 0.83 | 0.92 | 0.78 | 0.79 | 0.92 | 0.85 | 0.89 | 0.84 | 0.76 | 0.87 | 0.84 | 0.77 | 0.96 | 0.83 | 0.83 | 0.87 | 0.85 |
| det-HEAD | 0.65 | 0.56 | 0.37 | 0.74 | 0.46 | 0.72 | 0.56 | 0.64 | 0.65 | 0.56 | 0.24 | 0.84 | 0.45 | 0.74 | 0.30 | 0.90 | 0.30 | 0.84 | 0.62 | 0.57 |
| HEAD-nmod | 0.97 | 0.64 | 0.70 | 0.82 | 0.86 | 0.71 | 0.64 | 0.93 | 0.67 | 0.91 | 0.74 | 0.84 | 0.92 | 0.59 | 0.63 | 0.92 | 0.74 | 0.86 | 0.88 | 0.67 |
| nmod-HEAD | 0.60 | 0.75 | 0.51 | 0.82 | 0.55 | 0.82 | 0.71 | 0.67 | 0.86 | 0.52 | 0.45 | 0.87 | 0.41 | 0.92 | 0.43 | 0.94 | 0.45 | 0.87 | 0.59 | 0.77 |
| HEAD-nsbj | 0.96 | 0.93 | 0.92 | 0.92 | 0.94 | 0.95 | 0.93 | 0.95 | 0.93 | 0.95 | 0.90 | 0.96 | 0.94 | 0.92 | 0.93 | 0.90 | 0.92 | 0.95 | 0.94 | 0.90 |
| nsbj-HEAD | 0.60 | 0.46 | 0.31 | 0.74 | 0.46 | 0.58 | 0.41 | 0.64 | 0.60 | 0.48 | 0.30 | 0.77 | 0.43 | 0.63 | 0.10 | 0.93 | 0.26 | 0.78 | 0.52 | 0.52 |
| HEAD-nummod | 0.89 | 0.85 | 0.83 | 0.85 | 0.92 | 0.79 | 0.80 | 0.92 | 0.87 | 0.88 | 0.84 | 0.83 | 0.87 | 0.86 | 0.78 | 0.95 | 0.87 | 0.77 | 0.85 | 0.86 |
| nummod-HEAD | 0.65 | 0.57 | 0.37 | 0.76 | 0.45 | 0.72 | 0.55 | 0.63 | 0.65 | 0.56 | 0.30 | 0.83 | 0.45 | 0.74 | 0.26 | 0.92 | 0.23 | 0.87 | 0.61 | 0.56 |
| HEAD-obj | 0.92 | 0.58 | 0.62 | 0.78 | 0.73 | 0.72 | 0.52 | 0.89 | 0.60 | 0.87 | 0.57 | 0.85 | 0.77 | 0.67 | 0.52 | 0.90 | 0.56 | 0.86 | 0.88 | 0.48 |
| obj-HEAD | 0.64 | 0.81 | 0.58 | 0.85 | 0.65 | 0.80 | 0.78 | 0.66 | 0.93 | 0.56 | 0.62 | 0.87 | 0.59 | 0.88 | 0.52 | 0.94 | 0.61 | 0.85 | 0.52 | 0.88 |

Table 4: Real truth values for pairwise implications (row implies column). Note that $\phi \Leftrightarrow \phi$ does not always evaluate to 1; in fuzzy logic, $\mathcal{V}_F(\phi \Leftrightarrow \phi) = \max\{1 - \mathcal{V}_F(\phi), \mathcal{V}_F(\phi)\}$.