Discriminative Analysis of Linguistic Features for Typological Study

Hiroya Takamura, Ryo Nagata, Yoshifumi Kawasaki
Tokyo Institute of Technology, Konan University, Sophia University
takamura@pi.titech.ac.jp, nagata-acl@hyogo-u.ac.jp, kyossii@gmail.com

Abstract

We address the task of automatically estimating the missing values of linguistic features by making use of the fact that some linguistic features in typological databases are informative to each other. The questions to address in this work are (i) how much predictive power do features have on the value of another feature? (ii) to what extent can we attribute this predictive power to genealogical or areal factors, as opposed to being provided by tendencies or implicational universals? To address these questions, we conduct a discriminative or predictive analysis on the typological database. Specifically, we use a machine-learning classifier to estimate the value of each feature of each language using the values of the other features, under different choices of training data: all the other languages, or all the other languages except for the ones having the same origin or area with the target language.

Keywords: language typology, linguistic feature, WALS, machine learning

1. Introduction

There are numerous languages in the world. They are characterized from various viewpoints including the vocabulary, the syntactic rules, and the pronunciation system. In the language typology, the characteristics of languages are used to discuss the classification of languages and the similarity or dissimilarity between languages. The characteristics (or *features*[1] henceforth) are the backbone of the language typology. Part of the findings with regard to the linguistic features is aggregated as databases. One of the largest databases is the World Atlas of Language Structures (WALS) (Dryer and Haspelmath, 2014). WALS encompasses a wide range of linguistic features together with their values for various languages.

We would like to note that some linguistic features are sometimes informative to each other. There are believed to be tendencies or implicational universals between features (Comrie, 1981). For example, it is widely known that, if VSO is the dominant order of a language, then the language has prepositions (Universal 3 by Greenberg (1963)). In other words, the values of features provide clues to the value of another feature. This fact brings up the following two questions:

(i) how much predictive power do features have on the value of another feature?

(ii) to what extent can we attribute this predictive power to genealogical or areal factors, as opposed to being provided by tendencies or implicational universals?

To address these questions, we conduct a discriminative or predictive analysis on the typological database. Specifically, we use a machine-learning classifier to estimate the value of each feature of each language using the values of the other features in WALS, under different choices of training data: all the other languages, or all the other languages except for the ones having the same origin or area with the target language.

In addition to the scientific motivation above, we also have engineering motivations. It is widely known that WALS is sparse; the values of the majority of features are missing (e.g., (Daumé III and Campbell, 2007; Murawaki, 2015)). Evaluating the values of such features is a laborious task often requiring fieldwork. Our classifier can be used to estimate missing values in the database.[2] They may facilitate the statistical analysis on WALS (e.g., Albu (2006)).

We also take into account that some features in WALS are dependent on other features in a trivial manner as the order of V and O depends on the order of S, V and O. Such dependent features can obscure the findings pertaining to languages. We propose to remove dependent features from the attribute set for the classifier. We will distribute the resources of dependent features together with the estimation results so that other researchers can make use of them.

2. WALS and Related Work

2.1. WALS

As of June 2014, WALS contained 2,679 languages[3] and 192 features (Dryer and Haspelmath, 2014). Daumé III and Campbell (2007) reported that, of all the pairs of a language and a feature, only 16% are recorded. The remaining 84% are thus missing[4], suggesting that WALS is very sparse. In order to intuitively show its sparseness, we visualize the feature-language matrix in the left figure of Figure 1, where each line is associated with a feature, and each column is associated with a language. If the feature value of a language is recorded in WALS, the corresponding element is represented as a black dot, otherwise white. Since most part of this figure is white, it intuitively shows that WALS is very sparse.

[1] To avoid confusion with *feature* of machine learning as in *feature vector*, in this paper, we use the term *attribute* for machine learning, and *feature* for languages.

[2] We need to be careful in its use, because we can only obtain the estimated values that might be wrong.

[3] A general consensus is that currently there are approximately 7,000 living languages in the world (Lewis, 2009). It means that WALS contains less than half of all languages.

[4] We need to be aware that some features cannot be defined for some languages and this 84% part of the dataset contains both undefinable ones and actually missing ones.

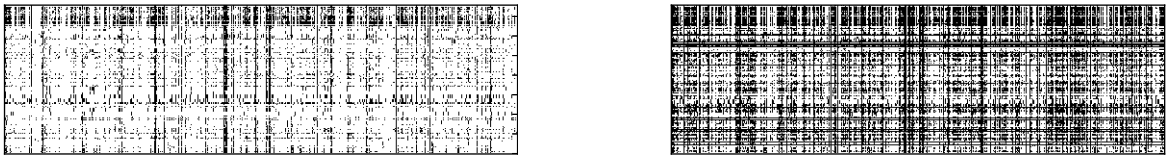

Figure 1: Feature-language matrix of WALS. Recorded features are represented as black points. Missing features are represented as white points. **Left**: original WALS. **Right**: original WALS and the features estimated with high confidence (the posterior probability is higher than 80%). If the output score of the classifier is positive, the estimation is regarded as confident.

The distribution of the number of non-empty features in WALS (more precisely, in the experimental dataset consisting of 2,370 languages used in the experiments in Section 4) is displayed as histogram in Figure 2. The figure exhibits a so-called long-tail style, showing that many languages have only a few non-empty features.

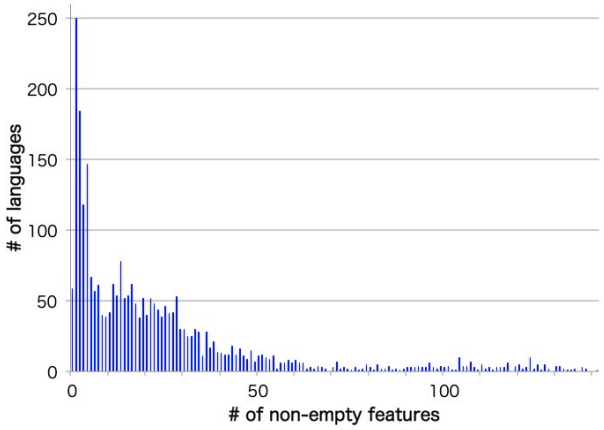

Figure 2: Histogram of the number of languages vs. the number of non-empty features

2.2. Computational analysis on or with WALS

Although there is a large amount of literature on the language typology, we name mathematical and computational work with WALS: Daumé III and Campbell (2007), Daumé III (2009), Lu (2013), Roy et al. (2014), Murawaki (2015). Daumé III and Campbell (2007) proposed a probabilistic model of the tendency or universal between linguistic features in WALS, where each feature is associated with a random variable, and the relations between features are captured by the statistical dependency between the random variables. Daumé III (2009) used a nonparametric bayesian model of linguistic features integrating both geographical and genealogical similarities, and calculated the measure indicating whether each feature value tends to be determined by a geographical reason or a genealogical reason.

Lu (2013) focused on the word order and extracted a directed asymmetrical graph structure with feature nodes in order to discover language universals. The feature pairs with a high dependency score are regarded as candidates of universals. Roy et al. (2014) focused on adpositions, and proposed an unsupervised method for determining whether each language uses prepositions or postpositions. Murawaki (2015) used linguistic features in WALS to represent languages with vectors. Only a small subset of the linguistic features was used due to the sparseness of WALS.

To the best of our knowledge, there have been no comprehensive efforts to estimate the feature values as is done in our work. The discriminative framework has not been exploited in the analysis of typological data.

3. Estimation of Feature Values

We first evaluate the accuracy of estimation of feature values when the other features are used as attributes of the classifier; we are going to answer the first question in Introduction ((i) how much predictive power do features have on the value of another feature?). For this purpose, we employ leave-one-out within the languages, for which the value of the target feature is recorded in WALS. In other words, we (i) regard one such language as a test instance and the remaining languages as training instances, (ii) represent both the training and test instances with the features other than the target feature, (iii) see whether the value of the target feature in the test instance is correctly estimated or not, (iv) iterate this process for all those languages to calculate the estimation accuracy. Hence the whole process can be termed *leave-one-language-out*. Since features generally have multiple values, each of the features other than the target feature is binarized to make attribute vectors. Each attribute is 1 if the feature of the language is the value associated with the attribute, 0 otherwise.

3.1. Dependent features

WALS contains features with different granularities. Although there exist no equivalent features, relations between features are not systematically organized. For example, Feature 81A (Dryer, 2013g) indicates the order of S(ubject), V(erb), and O(bject) such as SOV or SVO, while 82A (Dryer, 2013f) indicates the order of only S and V. The difference between these two features is simply ascribed to their granularities, and if Feature 81A is SVO, then 82A must be SV. When a value of a feature restricts the possible values of another feature, we call the former feature *a dependent feature* of the latter. We should note that the dependency introduced here is meant to be a trivial dependency such as the one caused by the difference in granularity as in

the example above, and is different from linguistically interesting dependency such as the one between the order of O and V and the presence/absence of postposition.
Such dependent features can obscure the actual accuracy of the feature value estimation. The classification rules learned in the presence of dependent features are not important in terms of the nature of language. We will therefore evaluate the accuracy in two different situations; when the dependent features are used in the attribute set and when not. For this purpose, we manually created the list[5] of dependent features for each feature.

3.2. Languages in the same genetic or geographic groups

The similarity of languages is ascribed to the shared origin, the language contact, the language type, or the language universals (Moravcsik, 2013). Since the typological study concerns only the language type[6] and the language universals, the effect from the shared origin and the language contact needs to be eliminated. In other words, we are going to answer the second question in Introduction: (ii) to what extent can we attribute this predictive power to genealogical or areal factors, as opposed to being provided by tendencies or implicational universals?

3.2.1. The shared origin

The languages with the shared origin tend to have the same feature values. If the languages that share the origin with the target language are in the training data, the apparent estimation accuracy would be improved. In practice, however, we would like to estimate the feature values typically because the origin of the target language is unknown. It is also possible that the trained model fails to capture the linguistic universals and tendencies if the model simply learns the feature value distribution of the language family. We therefore evaluate the estimation accuracy under two settings; one is the setting where the languages with the shared origin are excluded from the training data, and the other is the setting with such languages. In the implementation of our experiments, if a language belongs to the same language family as the target language given by WALS, we regard it as sharing the same origin.
Note that such languages are excluded *from the training data*, while the features mentioned in Section 3.1. are excluded *from the attribute set*. The language family and the language genus are not used as attributes for classification, either.

3.2.2. The shared area (the language contact)

The other factor to be considered is the language contact. Two languages with significant mutual contact tend to become similar in many ways. The same argument as in Section 3.2.1. can, therefore, apply to the language contact. However, it is difficult to measure the degree of contact between languages. We assume that two languages that are less than 2,000km[7] distant from each other have had significant contact with each other, and examine the estimation performance without using the languages that have the same geographical area with the target language as training data.

[5]The list is available from `http://www.lr.pi.titech.ac.jp/~takamura/typology.html`.

[6]*Language type* in the context of language typology is not necessarily the same as genealogical type. For example, a head-final language would be similar to (i.e., in the same type with) another head-final language. However, it does not mean these two languages share an origin. The discussion on the definition of *type* was given by Paolo Ramat (1987).

4. Experiments

From each chapter of WALS, we choose the feature with A (e.g., 39A), removing the features with the other letters (e.g., 39B), since those features are highly relevant to the feature with A in the same chapter. Note that most chapters in WALS have only one feature, which is with A. Since Features 139A (Zeshan, 2013a) and 140A (Zeshan, 2013b) are defined for sign languages and should not be evaluated for the other languages, we removed these two features from the experimental dataset in all the experiments in this paper. As a result, we obtained 129 features for experiments.

Out of the 2,679 languages contained in WALS (Section 2.1.), we removed 309 languages that have only one or none of the 129 features mentioned above and used the remaining 2,370 languages as the entire experimental dataset.

We will further remove some features and some languages respectively from the attribute set and the training dataset, depending on the target language and the experimental setting as explained in Section 3.

We use the logistic regression (LIBLINEAR)[8] as a classifier. We tune the regularization hyper-parameter C by selecting the optimal value out of 0.01, 0.1, 1, 10 and 100.

To calculate the distance between two languages from their latitude and longitude, we used a Perl module. [9]

4.1. Accuracy of feature value estimation

The results of the leave-one-language-out experiments are summarized in Table 1. The majority baseline in the table refers to the classifier that always outputs the majority class. The table shows that the trained classifier with the most practical setting (i.e., without dependent features as attributes, without languages with the shared origin or area as training data) achieves an accuracy of approximately 60% in macro and micro averages. It also shows that both dependent features and the languages with the shared origin or area always increase the accuracy.

On the right side of Figure 1, we visualize the feature-language matrix with empty elements being filled in. If the classifier outputs the positive score, we regard the feature as estimated with a high confidence, and fill in the corresponding element of the feature-language matrix. The comparison between the right and the left figures in Figure 1 intuitively shows how the sparseness of WALS is relieved.

Shared origin	Shared area	Accuracy (%)	
		Macro	Micro
		59.32	61.46
	✓	60.62	62.71
✓		60.65	62.69
✓	✓	64.36	66.24
Majority baseline		54.31	53.13

(a) Without dependent features

Shared origin	Shared area	Accuracy (%)	
		Macro	Micro
		66.27	72.15
	✓	67.32	73.04
✓		67.46	73.06
✓	✓	70.61	75.74
Majority baseline		54.31	53.13

(b) With dependent features

Table 1: Macro and micro averages of the estimation accuracy over different features (note that the datasets for different features can be of different sizes through leave-one-out, because the number of languages that have values for a feature can be different from that for another feature). The symbol ✓ in the *shared origin* column denotes that the languages in the same family are used as training data. The symbol ✓ in the *shared area* column denotes that the languages in the shared area are used as training data. (a) The features dependent on the target feature are not used as attributes, (b) All the features except the target feature are used as attributes.

WALS code	Language	Accuracy (%)
tha	Thai	78.9
hmo	Hmong Njua	77.9
kha	Khalkha	74.8
ndy	Ndyuka	74.0
khm	Khmer	74.0
knd	Kannada	70.9
lez	Lezgian	70.5
vie	Vietnamese	70.4
bag	Bagirmi	70.4
nht	Nahuatl (Tetelcingo)	70.0
...	...	...
wic	Wichita	52.9
mar	Maricopa	51.8
prh	Pirahã	51.3
goo	Gooniyandi	50.4
hix	Hixkaryana	50.4
ger	German	50.4
myi	Mangarrayi	50.0
brs	Barasano	50.0
grb	Grebo	46.2
pau	Paumarí	45.1

Table 2: 10 languages with the largest accuracy and 10 languages with the lowest accuracy. Only those that have more than 100 features recorded.

WALS code	Language	Increase (PT)
epe	Epena Pedee	11.2
khm	Khmer	7.7
nug	Nunggubuyu	7.6
kut	Kutenai	7.0
vie	Vietnamese	6.4
lez	Lezgian	6.1
lkt	Lakhota	5.1
asm	Asmat	4.8
hix	Hixkaryana	4.8
san	Sango	4.7
...	...	...
arp	Arapesh (Mountain)	-1.9
mao	Maori	-2.3
map	Mapudungun	-2.5
ain	Ainu	-2.6
mar	Maricopa	-2.6
ket	Ket	-2.8
klv	Kilivila	-2.9
khs	Khasi	-2.9
ram	Rama	-3.6
knd	Kannada	-4.7

Table 3: 10 languages with the largest increases in percentage points in accuracy that was caused by adding areal information and 10 languages with the lowest increases (i.e., the largest decreases).

4.2. Results from two perspectives

4.2.1. Language-wise summary

We next counted the number of correctly estimated features for each language, without using dependent features nor languages with the shared origin or area, and calculated the language-wise accuracy indicating how difficult it is to estimate the properties of the language. Due to space limitation, we show only the 10 languages with the largest accuracy values and the 10 languages with the lowest accuracy values, that have 100 or more recorded features in Table 2.

The language-wise accuracy can be further used to measure the sensitivity of a language to the areal effect. We calculate the increase in the language-wise accuracy that was caused by adding the languages with the shared area. We show the 10 languages with the largest increases and the 10 languages with the lowest increases (i.e., the largest decreases) in Table 3. For each of the languages shown in Table 3, there are other languages that are geographically close to and phylogenetically far from it. If this increase is a good indicator of sensitivity to areal effect, the languages with large increases should be affected by such other languages, while those with large decrease should be unaffected by such other languages, simply resulting in noise in training data. We can find some papers that support our result. For example, Enfield (Enfield, 2005) wrote "*Mainland South-*

[7]We follow the work by Hal Daumé III (2009), in which the effect of radius of a language is assumed to be 1,000km.

[8]We used LIBLINEAR available from `https://www.csie.ntu.edu.tw/~cjlin/liblinear/`.

[9]We used GIS::distance.

east Asia is one among many areas of the earth 's surface in which languages of different origins have come to share structural properties at multiple levels owing to historical social contact between speech communities", which supports the high ranks of Khmer and Vietnamese. For another example, Vajda (2010) wrote "*The prefixing verb structure of Ket differs strikingly from the surrounding Uralic, Turkic, Mongolic, and Tungusic languages of Inner Asia and Siberia*", which partially supports the low rank of Ket in the table.

4.2.2. Feature-wise summary

We first show the estimation accuracy for each feature both for the trained classifier and the majority baseline in Figure 3. The classifier was trained without dependent features, nor the languages with the shared origin or area. We can see that the trained classifier outperforms the baseline for most features.

To examine the above results more closely, we show the top 10 features with the largest differences in accuracy between the trained classifier and the majority baseline in Table 4. The trained classifier gained more than 30 points compared with the baseline for Features 85A, 83A and 95A. Most of the features in Table 4 pertain to the order of the head and the complement, suggesting that there is a certain tendency or universal with regard to the head-complement order. Since this is consistent with the findings in the typological study (Comrie, 1981), it suggests that our method works properly to estimate missing feature values, although our method is not the only one example that captures the implicational tendency with regard to the head-complement order.

4.3. Estimation of missing feature values

For each feature, we construct a classifier using all the languages, for which the value of the feature is recorded in WALS (i.e., without employing the leave-one-language-out approach), for the purpose of estimating the feature value that is actually missing. We will distribute the estimation result as a language resource together with the result of the leave-one-language-out experiment.[10]

We take Japanese as an example, and show the estimated values of features that are missing in WALS in Table 5.

Features 14A, 15A, 16A and 17A are defined for stress-accent languages. They are not defined for Japanese, which is a pitch-accent language (Tsujimura, 2002). Our method correctly estimated the grammatical gender to be absent in Japanese (Features 30A, 31A and 32A). Note that these features (30A, 31A and 32A) are dependent on each other and not used as training data of one another. As for Feature 141A, it is impossible to attain the correct value in the current setting, because there are only 6 training instances for this feature and all of them are *syllabic* or *alphasyllabic*.

We also show the estimated values of features of Italian and Spanish missing in WALS[11] in Tables 6 and 7. Since Italian have 71 missing values in our setting, we sample a small part of the entire set.

[10] The list is available from `http://www.lr.pi.titech.ac.jp/~takamura/typology.html`.

[11] Although English should be a good example thanks to its familiarity to most researchers, there are hardly any missing features for English in WALS.

5. Conclusion

We used a machine learning classifier to estimate values of linguistic features. We proposed to remove dependent features from the attribute set, and the languages with the shared origin or area from training data. We calculated the approximate accuracy of estimation. To qualitatively evaluate the estimation result, we conducted a case study of examining estimated feature values of Japanese. We will distribute the list of dependent features and the estimation results for further study.

For future work, we would need more detailed evaluations including theoretical and empirical comparisons with other similar attempts. We should also examine the trained model; the features in the attribute set that are given large weights in the classifier can be good candidates for universals.

As suggested in Section 4.3., some features cannot be defined for some languages. Such information should also be summarized as a linguistic resource accompanying WALS. being supported by our computational method. Computational support for discriminating definable or not would be helpful.

Acknowledgement

Y. Kawasaki is supported by JSPS KAKENHI Grant Number 15J04335.

Albu, M. (2006). *Quantitative Analysis of Typological Data*. Ph.D. thesis, Fakultät für Mathematik und Informatik der Universität Leigzig, September.

Baerman, M. and Brown, D. (2013). Syncretism in verbal person/number marking. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Bhat, D. (2013). Third person pronouns and demonstratives. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Bickel, B. and Nichols, J. (2013a). Exponence of selected inflectional formatives. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Bickel, B. and Nichols, J. (2013b). Inflectional synthesis of the verb. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Brown, C. H. (2013). Finger and hand. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Comrie, B. (1981). *Language Universals and Linguistic Typology*. University of Chicago Press.

Comrie, B. (2013a). Alignment of case marking of pronouns. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

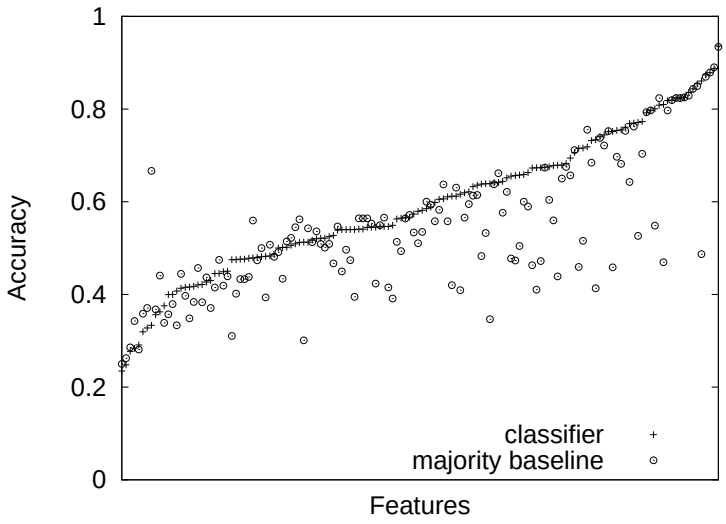

Figure 3: The estimation accuracy for each feature. The accuracy of the trained classifier is plotted with $+$, and that of the majority baseline is plotted with $\circ$. Features are ordered in the ascending order of the accuracy of the trained classifier, resulting in the monotonically increasing curve with $+$.

Feature ID	Feature name	Increase (PT)
85A	Order of Adposition and Noun Phrase	37.36
83A	Order of Object and Verb	33.25
95A	Relationship between the Order of Object and Verb and the Order of Adposition and Noun Phrase	31.87
88A	Order of Demonstrative and Noun	29.08
97A	Relationship between the Order of Object and Verb and the Order of Adjective and Noun	28.65
86A	Order of Genitive and Noun	25.70
81A	Order of Subject Object and Verb	25.41
99A	Alignment of Case Marking of Pronouns	24.42
89A	Order of Numeral and Noun	24.41
51A	Position of Case Affixes	23.82

Table 4: Top 10 features with the largest differences in percentage points (PT) in accuracy between the trained classifier and the majority baseline. The dependent features are not used as attributes. The languages with the shared origin or area are not used as training data. (Dryer, 2013a; Dryer, 2013e; Dryer, 2013j; Dryer, 2013b; Dryer, 2013i; Dryer, 2013d; Dryer, 2013c; Dryer, 2013h; Dryer, 2013g; Comrie, 2013a)

Comrie, B. (2013b). Writing systems. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Corbett, G. G. (2013a). Number of genders. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Corbett, G. G. (2013b). Sex-based and non-sex-based gender systems. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Corbett, G. G. (2013c). Systems of gender assignment. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Cysouw, M. (2013a). Inclusive/exclusive distinction in independent pronouns. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Cysouw, M. (2013b). Inclusive/exclusive distinction in verbal inflection. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Daumé III, H. and Campbell, L. (2007). A Bayesian model for discovering typological implications. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 65–72.

Daumé III, H. (2009). Non-parametric Bayesian areal linguistics. In *Proceedings of the North American Chapter of the Association of Computational Linguistics (NAACL)*, pages 593–601.

Matthew S. Dryer et al., editors. (2014). *The World Atlas of Language Structures Online*. Leipzig: Max Planck In-

Feature ID	Feature name	Estimated value	Posterior (%)
14A	Fixed Stress Locations	1 No fixed stress ⋄	58.25
15A	Weight-Sensitive Stress	5 Unbounded: Stress can be anywhere ⋄	39.78
16A	Weight Factors in Weight-Sensitive Stress Systems	1 No weight ⋄	26.77
17A	Rhythm Types	5 Absent ⋄	32.06
30A	Number of Genders	1 None ✓	74.54
31A	Sex-based and Non-sex-based Gender Systems	1 No gender ✓	99.98
32A	Systems of Gender Assignment	1 No gender ✓	79.52
141A	Writing Systems	4 Syllabic *	82.39

Table 5: Estimated feature values of Japanese missing in WALS, with the optimal value of C. The score in the right column is the posterior probability of the estimated feature value given the other features. The symbol ✓denotes that the estimated value would be correct, while the symbol * denotes incorrect. The symbol ⋄ denotes that the feature is not defined for Japanese. In the training, the dependent features are not used. The languages with the shared area are not used. (Goedemans and van der Hulst, 2013a; Goedemans and van der Hulst, 2013b; Goedemans and van der Hulst, 2013c; Goedemans and van der Hulst, 2013d; Corbett, 2013a; Corbett, 2013b; Corbett, 2013c; Dryer, 2013g; Song, 2013; Brown, 2013; Comrie, 2013b)

Feature ID	Feature name	Estimated value	Posterior (%)
21A	Exponence of Selected Inflectional Formatives	5 No case ✓	28.79
22A	Inflectional Synthesis of the Verb	3 4-5 categories per word ✓	25.04
29A	Syncretism in Verbal Person/Number Marking	3 Not syncretic *	55.29
30A	Number of Genders	2 Two ✓	31.45
31A	Sex-based and Non-sex-based Gender Systems	2 Sex-based ✓	96.04
32A	Systems of Gender Assignment	3 Semantic and formal ✓	46.50
39A	Inclusive/Exclusive Distinction in Independent Pronouns	3 No inclusive/exclusive ✓	42.61
40A	Inclusive/Exclusive Distinction in Verbal Inflection	3 No inclusive/exclusive ✓	35.24
43A	Third Person Pronouns and Demonstratives	1 Unrelated to demonstratives ✓	26.29
55A	Numeral Classifiers	1 Absent ✓	66.12

Table 6: Estimated feature values of Italian missing in WALS, with the optimal value of C. The score in the right column is the posterior probability of the estimated feature value given the other features. The symbol ✓denotes that the estimated value would be correct, while the symbol * denotes incorrect. In training, the dependent features are not used. The languages with the shared origin or area are not used. (Bickel and Nichols, 2013a; Bickel and Nichols, 2013b; Baerman and Brown, 2013; Corbett, 2013a; Corbett, 2013b; Corbett, 2013c; Cysouw, 2013a; Cysouw, 2013b; Bhat, 2013; Gil, 2013c)

stitute for Evolutionary Anthropology. (Available online at http://wals.info, Accessed on 2014-07-03.).

Dryer, M. S. (2013a). Order of adposition and noun phrase. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013b). Order of demonstrative and noun. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013c). Order of genitive and noun. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013d). Order of numeral and noun. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013e). Order of object and verb. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013f). Order of subject and verb. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013g). Order of subject, object and verb. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013h). Position of case affixes. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013i). Relationship between the order of object and verb and the order of adjective and noun. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013j). Relationship between the order of object and verb and the order of adposition and noun phrase. In Matthew S. Dryer et al., editors, *The World At-*

Feature ID	Feature name	Estimated value	Posterior (%)
52A	Comitatives and Instrumentals	1 Identity ✓	47.87
55A	Numeral Classifiers	1 Absent ✓	98.34
56A	Conjunctions and Universal Quantifiers	1 Formally Different ✓	94.34
60A	Genitives, Adjectives and Relative Clauses	6 Highly Differentiated ✓	94.34
141A	Writing Systems	3 Alphasyllabic *	95.20

Table 7: Estimated feature values of Spanish missing in WALS, with the optimal value of C. The score in the right column is the posterior probability of the estimated feature value given the other features. The symbol ✓denotes that the estimated value would be correct, while the symbol * denotes incorrect. In training, the dependent features are not used. The languages with the shared origin or area are not used. (Stolz et al., 2013; Gil, 2013c; Gil, 2013a; Gil, 2013b; Comrie, 2013b)

las of Language Structures Online, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Enfield, N. (2005). Areal linguistics and mainland southeast asia. *Annual Review of Anthropology*, 34:181–206.

Gil, D. (2013a). Conjunctions and universal quantifiers. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Gil, D. (2013b). Genitives, adjectives and relative clauses. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Gil, D. (2013c). Numeral classifiers. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Goedemans, R. and van der Hulst, H. (2013a). Fixed stress locations. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Goedemans, R. and van der Hulst, H. (2013b). Rhythm types. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Goedemans, R. and van der Hulst, H. (2013c). Weight factors in weight-sensitive stress systems. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Goedemans, R. and van der Hulst, H. (2013d). Weight-sensitive stress. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg, editor, *Universals of Language*, pages 73–113. MIT Press, Cambridge, MA.

Paul M. Lewis, editor. (2009). *Ethnologue: Languages of the World, 16th edition*. SIL International.

Lu, X. (2013). Exploring word order universals: a probabilistic graphical model approach. In *Proceedings of the ACL Student Research Workshop*, pages 150–157.

Moravcsik, E. A. (2013). *Introducing Language Typology*. Cambridge University Press.

Murawaki, Y. (2015). Continuous space representations of linguistic typology and their application to phylogenetic inference. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics and Human Language Technologies (NAACL-HLT2015*, pages 324–334.

Ramat, P. (1987). *Linguistic Typology*. Walter de Gruyter.

Roy, R. S., Katare, R., Ganguly, N., and Choudhury, M. (2014). Automatic discovery of adposition typology. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1037–1046.

Song, J. J. (2013). Periphrastic causative constructions. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Stolz, T., Stroh, C., and Urdze, A. (2013). Comitatives and instrumentals. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Natsuko Tsujimura, editor. (2002). *The Handbook of Japanese Linguistics (Blackwell Handbooks in Linguistics)*. John Wiley & Sons.

Vajda, E. (2010). A siberian link with the na-dene. *Anthropological Papers of the University of Alaska*, 5:31–99.

Zeshan, U. (2013a). Irregular negatives in sign languages. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Zeshan, U. (2013b). Question particles in sign languages. In Matthew S. Dryer et al., editors, *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.