Supporting Information

Focal colors across languages are representative members of color categories

Joshua T. Abbott, Thomas L. Griffiths, and Terry Regier

1 Methods

1. (a) Data

Most of the color naming data we analyze are drawn from the World Color Survey (WCS), which collected color naming data from 110 unwritten languages worldwide [24]. We treated these data as follows. We considered any reference to a chip at lightness level A to correspond to A0 (English focal *white*), and any reference to a chip at lightness level J to correspond to J0 (English focal *black*). Therefore we interpreted any record indicating that a speaker had selected chips A1, A2, ..., A40 as a best example of some color term to indicate a single selection of A0, and we interpreted any record indicating that a speaker had selected chips J1, J2, ..., J40 as a best example of some color term to indicate a single selection of J0. In addition, because our analyses concern the relation between category extension (naming data) and best examples (focus data) on a per-speaker basis, we considered only those categories for which both naming and focus data were available for the same speaker. The WCS data are available at http://www.icsi.berkeley.edu/wcs/data.html.

We also conducted similar analyses for two languages outside the WCS: Dani and Berinmo. The data for Dani were derived from Heider [33] and the Berinmo data from Roberson et al. [6].

1. (b) Fitting Gaussian distributions to naming data

For each named color category used by each speaker in each language of the WCS, we modeled that category as a 3-dimensional Gaussian distribution in CIELAB space, and estimated the parameters of that distribution using a normal-inverse-Wishart prior, a standard estimation method for multivariate Gaussian distributions of unknown mean and unknown variance. Specifically, given a set of M chips $\mathbf{x}_i$ in color category t, where $\mathbf{x}_i$ holds the coordinates of that chip in CIELAB space, we obtain the estimates:

$$\mu_t = \frac{1}{M}\sum_i^M \mathbf{x}_i \ , \ \Sigma_t = \frac{SS_t + \lambda_0}{M + \nu_0} \tag{1}$$

where SS_t is the sum of squares for category t: $\sum_i^M (\mathbf{x}_i - \mu_t)(\mathbf{x}_i - \mu_t)^\mathsf{T}$, and λ_0 and ν_0 are the parameters of the prior. λ_0 was set by taking an empirical estimate of the variance in CIELAB coordinates over all chips in the stimulus array, and ν_0 was set to 1.

2 Effect of color space

We wished to know whether our results were robust to a change in choice of color space. To that end, we re-ran our analyses using CIELUV color space, in place of CIELAB as reported in the main text. All methods and analyses were the same as described using CIELAB, with the only difference being that stimulus chips were represented as points in CIELUV rather than CIELAB. The results are presented in Table 1 below, which re-displays the results provided in Table 1 of the main text, with the addition of model results assuming CIELUV color space. We find that under both color spaces, and under both evaluation measures, the representativeness model outperforms its competitors: it exhibits the best performance for the overall WCS distribution, and for a majority of the WCS languages. Since the results are similar, and for sake of brevity, we do not display the CIELUV results for further analyses within the SI.

Table 1: Quantitative evaluations of model predictions of empirical WCS focus choices (parentheses give number of languages for which this is the best performing model) under both CIELAB and CIELUV color space.

Model	QF(LAB)	QF(LUV)	RP(LAB)	RP(LUV)
Representativeness	**1.17 (76)**	**1.18 (65)**	**27.67 (104)**	**27.92 (102)**
Likelihood	1.74 (6)	1.62 (4)	42.64 (1)	42.48 (0)
Prototype	1.95 (3)	2.00 (1)	48.29 (0)	50.76 (0)
Exemplar	1.63 (24)	1.47 (37)	38.97 (5)	36.35 (8)
Chroma	2.13 (1)	1.91 (3)	78.50 (1)	75.41 (0)

3   Category unusualness

In this section we provide further treatment of the category unusualness analyses from the main text. We first present additional analyses that complement those in the main text, followed by figures of WCS categories at various levels of unusualness.

3. (a)   QF distance results

The main text presents results of a category unusualness analysis in which model performance was assessed by rank prediction; Figure 1 presents results of an analogous analysis in which model performance was instead assessed by QF distance.
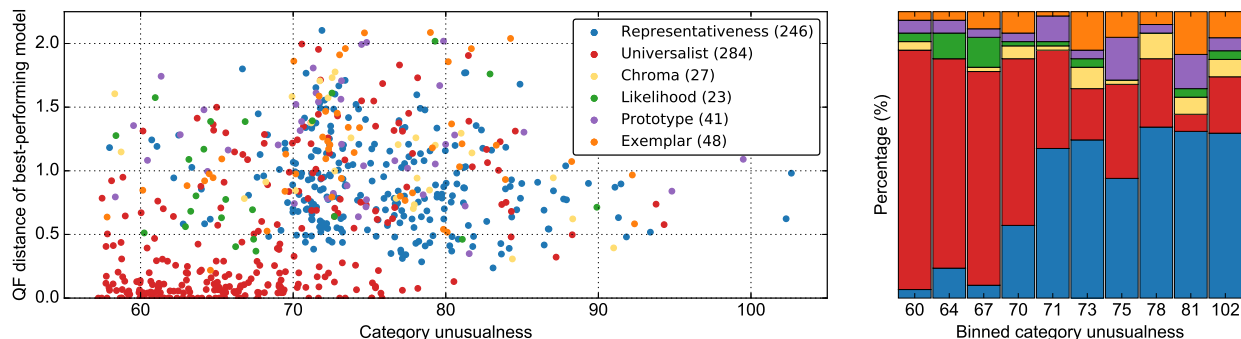
Figure 1: Effect of category unusualness. Left panel (scatterplot): Each dot represents a color category in the WCS, and the dot's color represents the best-performing model for that category. The horizontal axis represents category unusualness, and the vertical axis represents the model performance (QF distance) of the best-performing model for that category. Right panel (bar chart): The horizontal axis again represents category unusualness, this time partitioned into 10 bins with the same number of categories per bin. The stacked bars show, for each level of unusualness, the proportion of categories at that level of unusualness that were best predicted by each model.

The Universalist model performs well for the least unusual (most usual or common) categories; this is particularly apparent when using QF distance. These are typically categories corresponding roughly to English *black* and *white*, as can be seen in Figure 3 below displaying the 30 least unusual categories. As with the rank position analyses from the main text, the QF distance measure shows that the Representativeness model increases in performance, while the Universalist model decreases in performance, as a function of category unusualness.

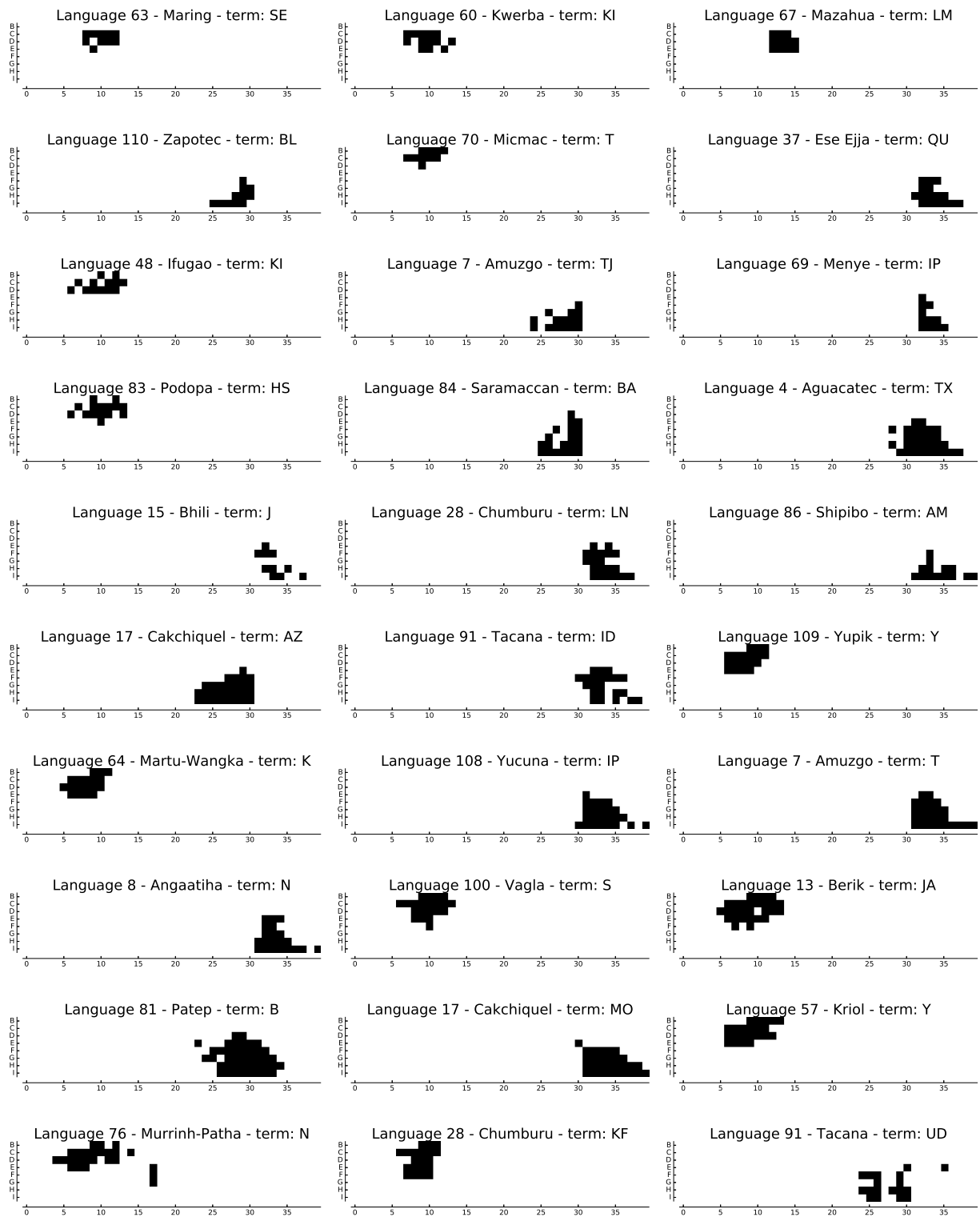## 3. (b)   WCS categories ranked by unusualness

3

Figure 2: The 30 most unusual WCS categories (presented in descending order of unusualness).
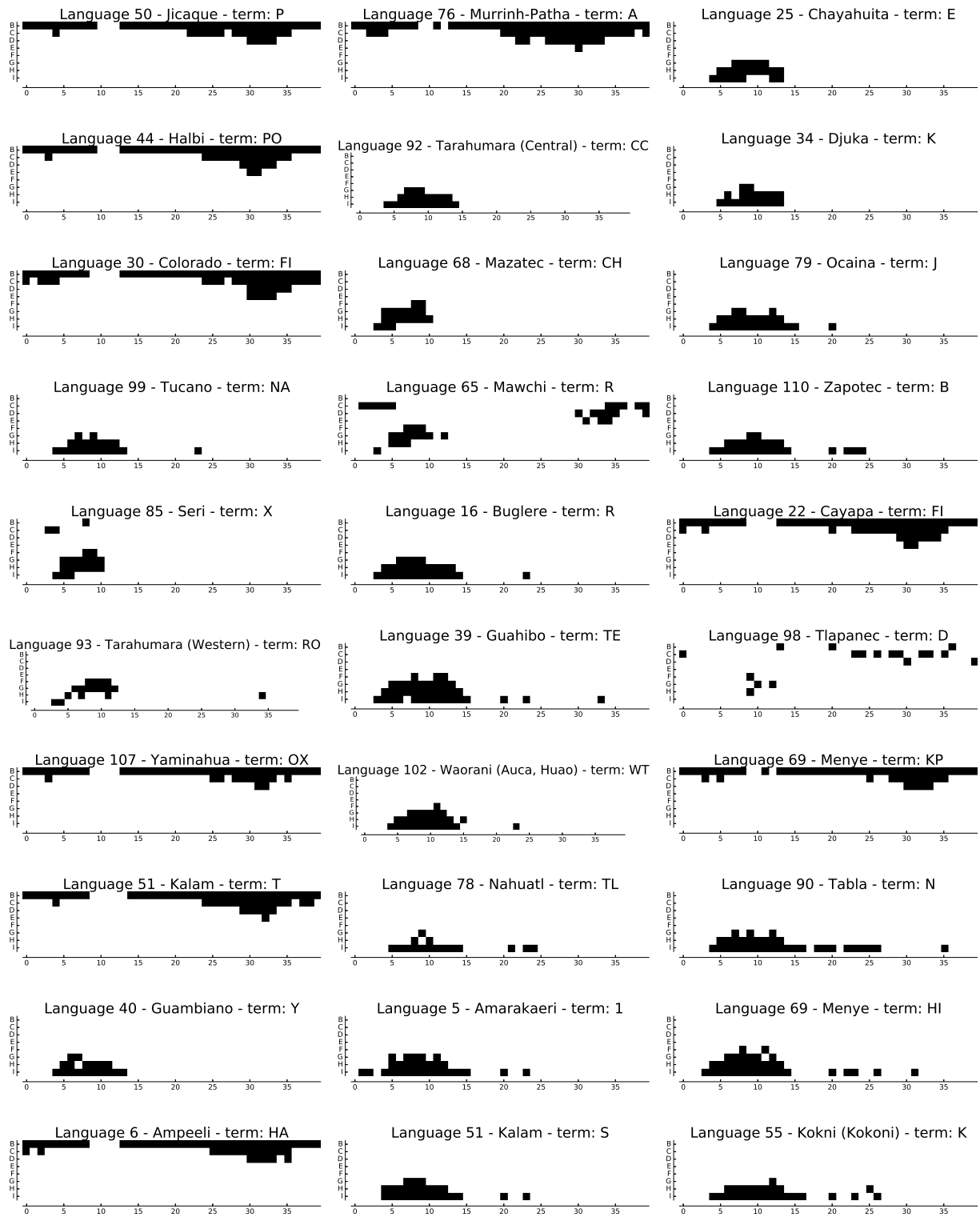
Figure 3: The 30 least unusual WCS categories (presented in ascending ranked order of unusualness).
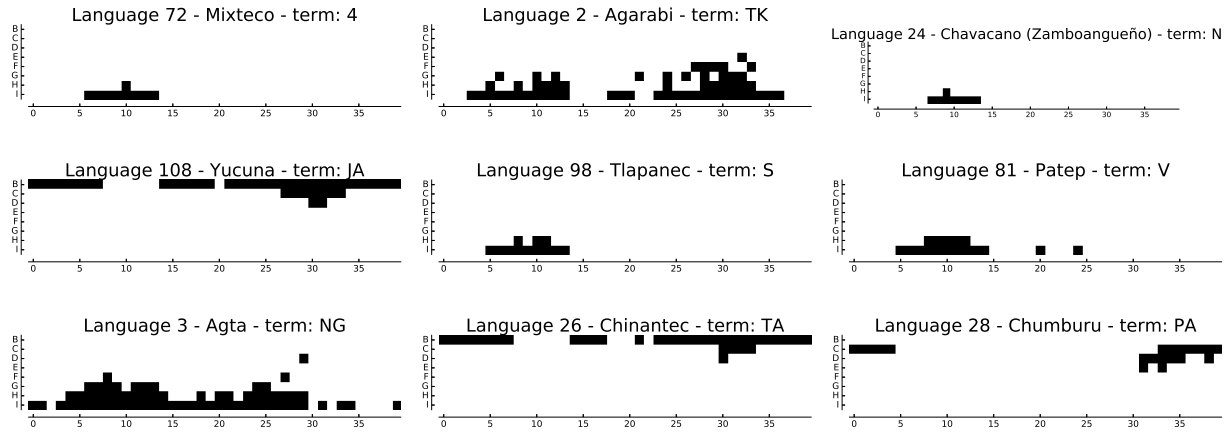
Figure 4: WCS categories in the 25th percentile of unusualness scores (presented in ascending ranked order of unusualness).
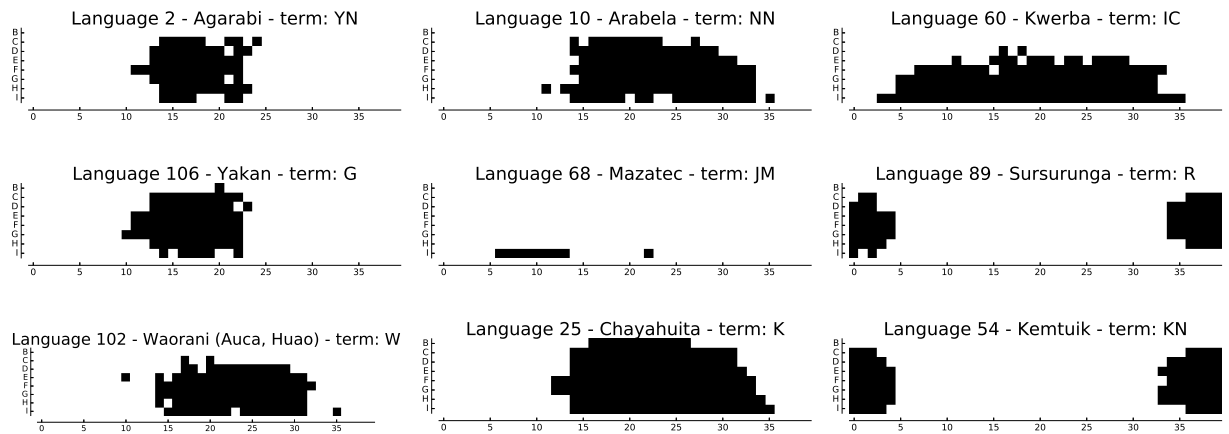

Figure 5: WCS categories in the 50th percentile of unusualness scores (presented in ascending ranked order of unusualness).
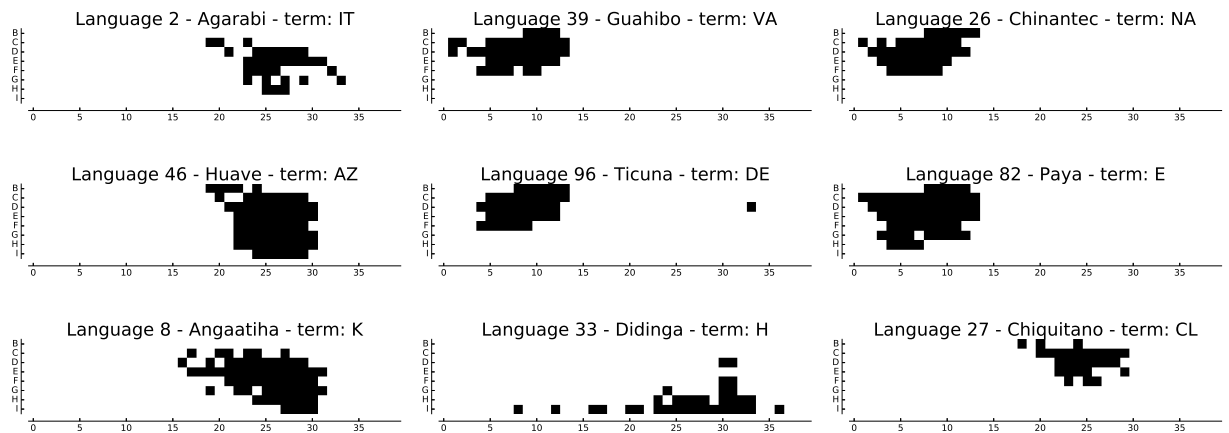

Figure 6: WCS categories in the 75th percentile of unusualness scores (presented in ascending ranked order of unusualness).

4   Language-level analyses

In this section, we provide language-level analyses of the data, and of model performance. We first present analyses for two languages not in the WCS—Dani [33] and Berinmo [6]—and then present analyses for WCS languages [24]. The format of the data to which we have access differs for these two classes of languages (Dani and Berinmo vs. WCS), and our analyses therefore necessarily differ accordingly. However we have attempted to keep the analyses as similar across languages as possible, given the differences in data format.

4. (a)   Dani

We considered Dani color naming data as reported by Heider [33]. Dani has been reported to use primarily a two-term color system, *mili* and *mola*, corresponding roughly to "cool" and "warm" colors, respectively, although Heider also found that roughly half the Dani participants also provided other terms for regions corresponding roughly to English *red*, *yellow*, and *blue*. Dani, with only two major color terms, has fewer major color terms than any of the languages of the WCS. Dani thus provides an opportunity to test our models against a system that is qualitatively different in an important respect from those of the WCS.

Our models require two sorts of data relative to the same set of stimuli: naming data and best example data. The experimental stimuli and procedure used by Heider [33] differed slightly from those used in the WCS and presented in the main text. A reduced set of 160 maximally saturated Munsell color chips were used, corresponding to every other column in the WCS chromatic grid, and data were not provided at the level of individual speakers. Instead, naming data for Dani were reported in the form of language-level responses, aggregated over speakers, distributed over the 160 chip chromatic grid. Such data were provided for all terms except for *mola*, which was described as the complement of *mili*. Focus data were also provided at the language level, i.e. again aggregated over speakers, in the form of a histogram of reported best examples for each term. These focus histograms were provided for *mili* and *mola*, but not for the less dominant terms mentioned above, consistent with the view of Dani as primarily a two-term language.

Preparation of the data. Since we do not have individual speaker data, we constructed a single naming map from the reported naming distributions to provide as input to our models. Naming data were not provided for *mola*, so we inferred data for that term by assuming that all 40 Dani participants provided a naming response for each chip in the stimulus array and that *mola* was the only missing term after summing the counts for other color terms at a given stimulus chip. This allowed us to create a mode map for Dani over the 160 chip chromatic grid, where each chip is assigned the color term used by a plurality of speakers (the modal term for that chip). Figure 7 below displays the resulting Dani mode map. Here, the extension of each color term is shown as a colored region, and the color assigned to that region is determined by taking the average RGB coordinates of the chips in the region. The number of focus hits for *mili* and *mola* per chip, aggregated over speakers in the language, is overlaid on top. Although Dani is generally regarded as a two-term color system, two color chips in the mode map were given names other than *mili* and *mola* by a plurality of speakers.

Analyses and results. Our analyses were conducted at the language level, because the data for this language were reported at the language level. We analyzed model predictions for only the terms *mili* and *mola* because these were the only terms for which focus data were reported. Each model returned a ranked list of all chips in the array. For each term, we recorded the number n of chips that received 1 or more focus choices. We then computed rank position by averaging together the rank positions of these n chips in the ranked list produced by the model. To compute QF distance, we compared the empirical histogram of focus choices for a given term to a model-predicted histogram in which each of the n top-ranked chips received a count of 1 and each other chip received a count of 0. The results of these analyses are provided in Table 2 below. We find that the Representativeness model outperforms its competitors by both metrics.
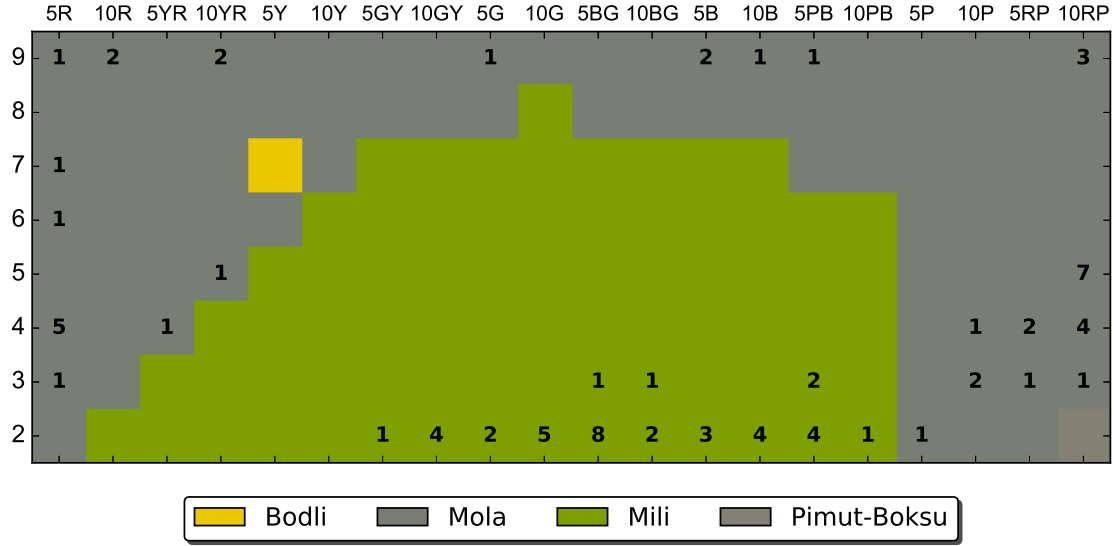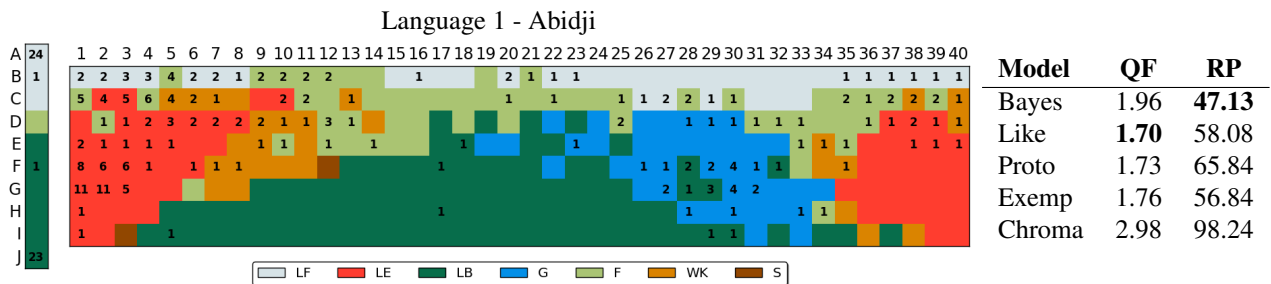
	5R	10R	5YR	10YR	5Y	10Y	5GY	10GY	5G	10G	5BG	10BG	5B	10B	5PB	10PB	5P	10P	5RP	10RP
9	1	2		2					1				2	1	1					3
8																				
7	1				1															
6	1																			
5					1															7
4	5			1														1	2	4
3	1									1	1				2			2	1	1
2						1	4	2	5	8	2	3	4	4	1	1				

Bodli Mola Mili Pimut-Boksu

Figure 7: Naming data for the Dani language, overlaid with the empirical focus distributions for *mili* and *mola*.

Table 2: Quantitative assessment of each model against Dani focus distribution.

Model	QF	RP
Representativeness	**2.09**	**35.29**
Likelihood	2.54	39.99
Prototype	2.39	42.70
Exemplar	2.88	36.44
Chroma	2.64	45.59

4. (b) Berinmo

The Berinmo data we consider [6] were originally collected in an attempt to replicate and extend earlier work based on Dani [5]. For this reason, the stimuli were the same as those of the earlier Dani work. Roberson et al. [6] reported results on Berinmo color memory that differed in important respects from those obtained from Dani, but the similarity in stimuli and procedure make the two studies directly comparable with respect to naming and focus data. As in the case of Dani, Berinmo naming data and focus data were both reported at the language level. The Berinmo naming data provided by Roberson et al. [6] were presented in the form of a mode map, which we illustrate in Figure 8 below. The focus data for each term were reported in the form of a histogram over the naming grid, here shown overlaid on top of the naming data.

Analyses and results. Because the data format for Berinmo was the same as that for Dani, we followed the same procedure as with Dani. The results of the Berinmo analyses are provided in Table 3 below. The Representativeness model outperforms its competitors by both metrics.

8
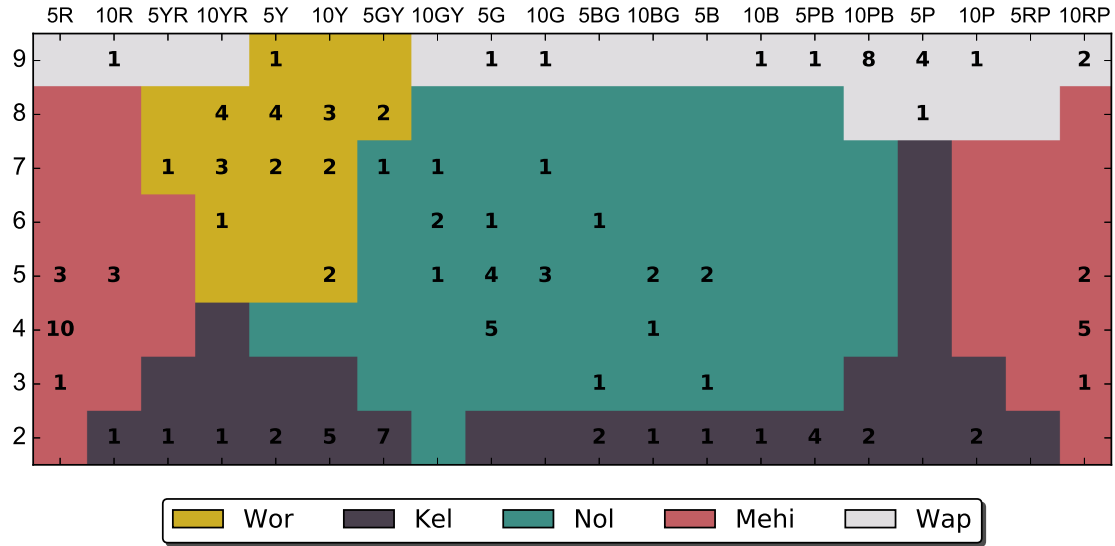
Figure 8: Naming data for the Berinmo language, overlaid with the empirical focus distribution.

Table 3: Quantitative assessment of each model against Berinmo focus distribution.

Model	QF	RP
Representativeness	**1.37**	**12.27**
Likelihood	1.69	14.31
Prototype	1.68	16.58
Exemplar	1.66	15.52
Chroma	2.64	24.75

4. (c) WCS languages

Here we present language-level analyses of the WCS data, along with model performance for each. We present the naming data as mode maps, displaying terms used by a plurality of the speakers. The number of focus hits per color chip, aggregated over speakers in the language, are overlaid on top, in line with our treatment of Dani and Berinmo above. Although the use of mode maps for visualization provides only a partial view of a language's color naming system, one can clearly see both similarities and differences in color naming patterns across languages.

Language 1 - Abidji

Model	QF	RP
Bayes	1.96	**47.13**
Like	**1.70**	58.08
Proto	1.73	65.84
Exemp	1.76	56.84
Chroma	2.98	98.24

Language 2 - Agarabi
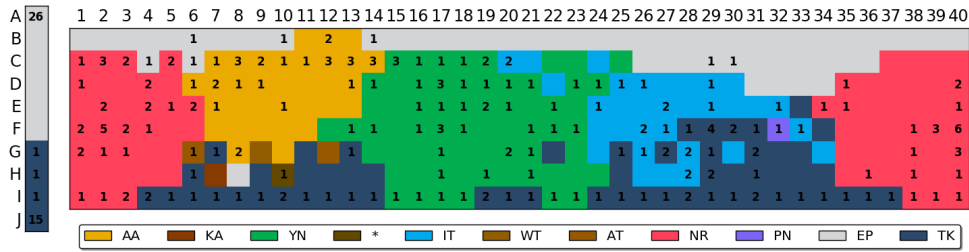
Model	QF	RP
Bayes	**1.10**	**31.89**
Like	1.37	44.18
Proto	1.35	43.90
Exemp	1.26	41.28
Chroma	2.32	81.43

Language 3 - Agta
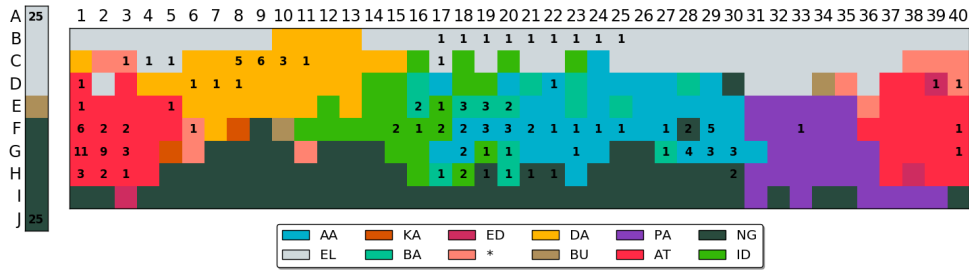
Model	QF	RP
Bayes	**1.35**	**36.10**
Like	1.65	43.42
Proto	1.83	47.54
Exemp	1.45	38.99
Chroma	1.98	78.71

Language 4 - Aguacatec
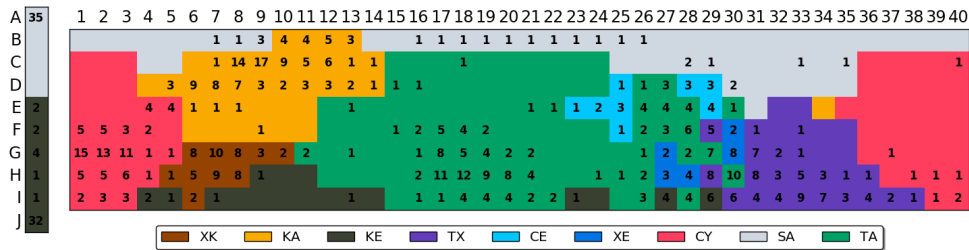
Model	QF	RP
Bayes	**1.01**	**21.80**
Like	1.88	34.16
Proto	1.83	40.63
Exemp	1.66	31.19
Chroma	2.23	81.47

Language 5 - Amarakaeri
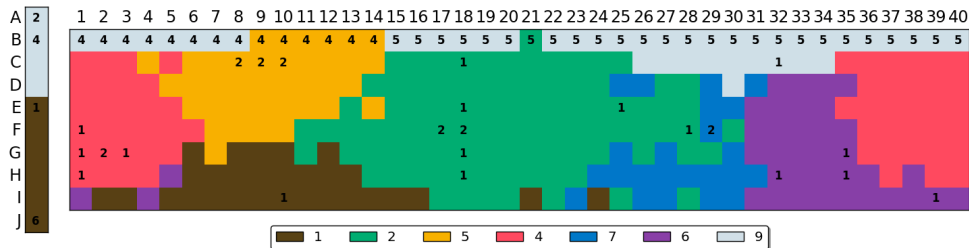
Model	QF	RP
Bayes	**1.08**	**21.14**
Like	1.25	36.02
Proto	1.21	43.92
Exemp	1.23	35.92
Chroma	3.13	57.58

Language 6 - Ampeeli
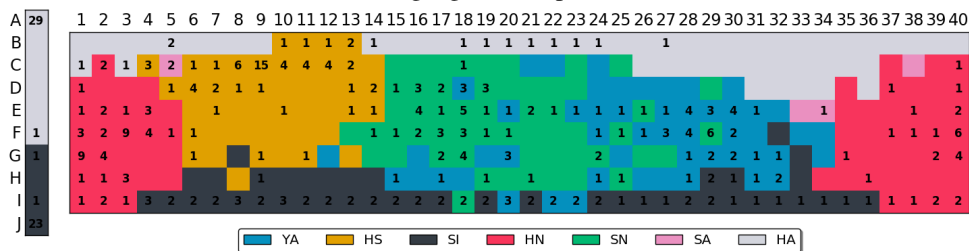
Model	QF	RP
Bayes	1.58	**33.98**
Like	1.68	46.04
Proto	1.80	51.85
Exemp	**1.57**	44.85
Chroma	2.22	81.05

Language 7 - Amuzgo
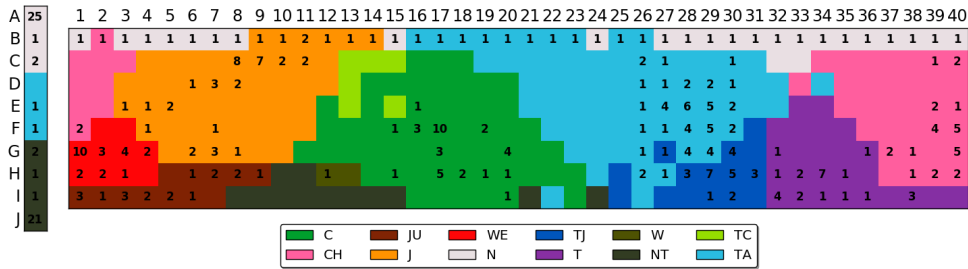
Model	QF	RP
Bayes	1.58	**19.69**
Like	1.56	28.04
Proto	**1.54**	27.06
Exemp	1.57	26.34
Chroma	1.75	28.90

Legend: C, CH, JU, J, WE, N, TJ, T, W, NT, TC, TA

Language 8 - Angaatiha
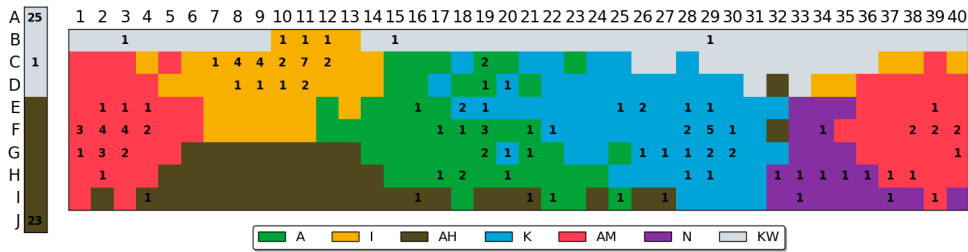
Model	QF	RP
Bayes	**1.02**	**26.52**
Like	1.48	40.03
Proto	1.68	45.69
Exemp	1.35	34.40
Chroma	1.58	77.99

Legend: A, I, AH, K, AM, N, KW

Language 9 - Apinayé
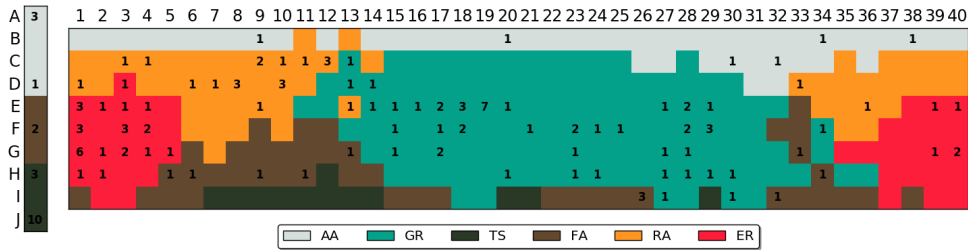
Model	QF	RP
Bayes	**1.52**	**46.08**
Like	1.75	50.38
Proto	1.73	75.86
Exemp	1.58	50.76
Chroma	2.21	93.68

Legend: AA, GR, TS, FA, RA, ER

Language 10 - Arabela
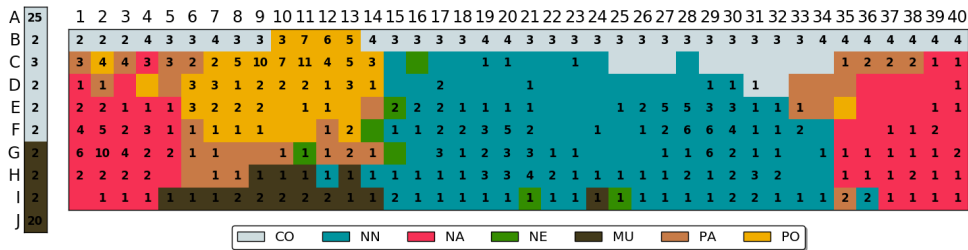
Model	QF	RP
Bayes	2.12	**42.12**
Like	2.27	57.78
Proto	2.42	78.39
Exemp	**2.08**	57.34
Chroma	3.32	97.55

Legend: CO, NN, NA, NE, MU, PA, PO

Language 11 - Bahinemo
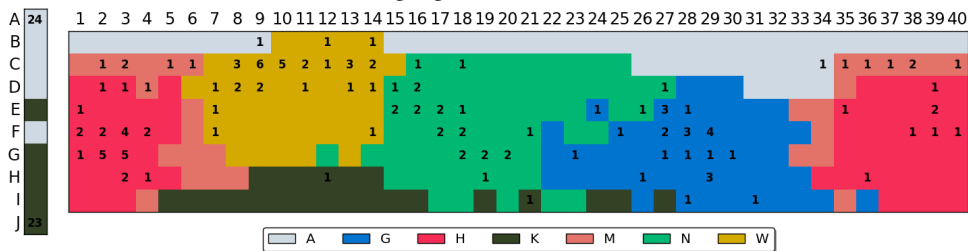
Model	QF	RP
Bayes	**0.94**	**36.71**
Like	1.36	53.03
Proto	1.55	58.03
Exemp	1.32	51.50
Chroma	1.59	81.31

Legend: A, G, H, K, M, N, W

Language 12 - Bauzi
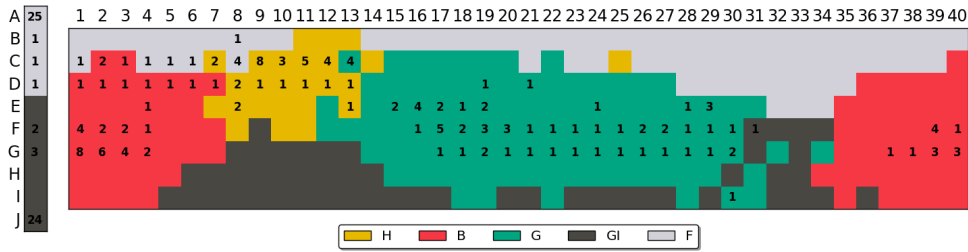
Model	QF	RP
Bayes	**1.09**	**15.65**
Like	1.31	40.01
Proto	1.36	42.21
Exemp	1.44	34.19
Chroma	2.07	92.83

Language 13 - Berik
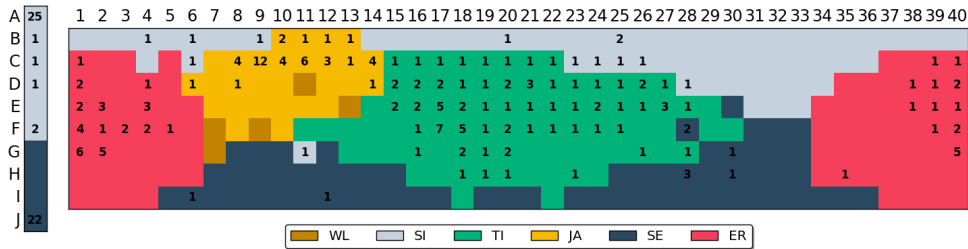
Model	QF	RP
Bayes	**1.23**	**30.17**
Like	1.66	47.64
Proto	1.87	62.48
Exemp	1.60	44.50
Chroma	2.44	85.83

Language 14 - Bété
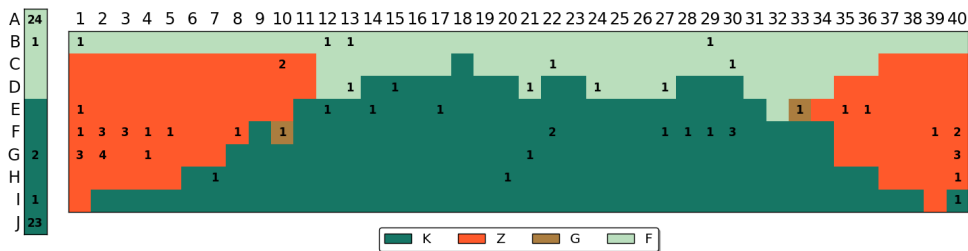
Model	QF	RP
Bayes	**1.04**	**28.20**
Like	1.46	51.64
Proto	1.61	52.11
Exemp	1.19	41.60
Chroma	2.14	115.71

Language 15 - Bhili
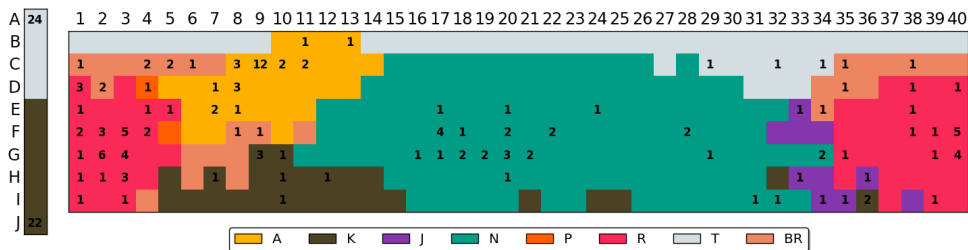
Model	QF	RP
Bayes	**0.95**	**22.26**
Like	1.34	35.51
Proto	1.34	38.52
Exemp	1.14	32.62
Chroma	1.69	87.41

Language 16 - Buglere
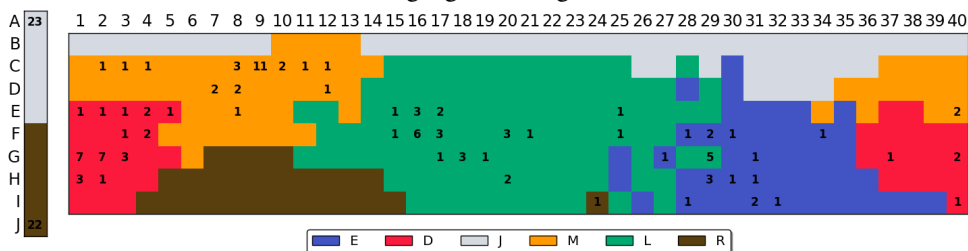
Model	QF	RP
Bayes	**1.00**	**18.13**
Like	1.69	46.92
Proto	1.77	62.91
Exemp	1.42	42.12
Chroma	1.36	82.95

Language 17 - Cakchiquel
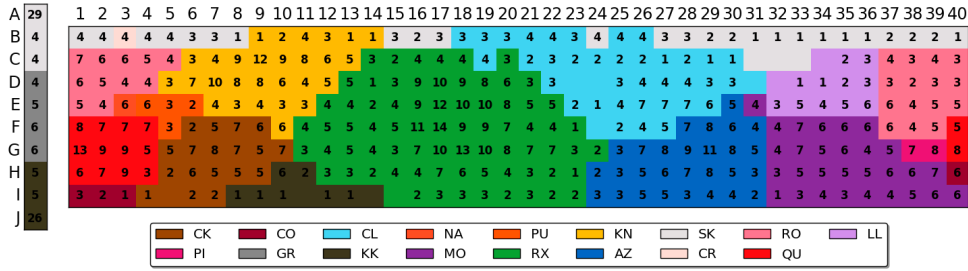
Model	QF	RP
Bayes	**1.04**	32.92
Like	1.39	30.71
Proto	1.34	30.39
Exemp	1.41	**28.94**
Chroma	2.72	56.93

Legend: CK, CO, CL, NA, PU, KN, SK, RO, LL, PI, GR, KK, MO, RX, AZ, CR, QU

Language 18 - Campa
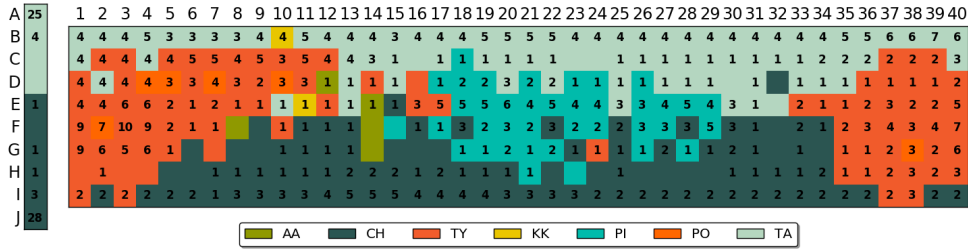
Model	QF	RP
Bayes	2.30	**56.91**
Like	1.99	72.67
Proto	2.10	74.47
Exemp	**1.82**	61.16
Chroma	4.36	119.79

Legend: AA, CH, TY, KK, PI, PO, TA

Language 19 - Camsa
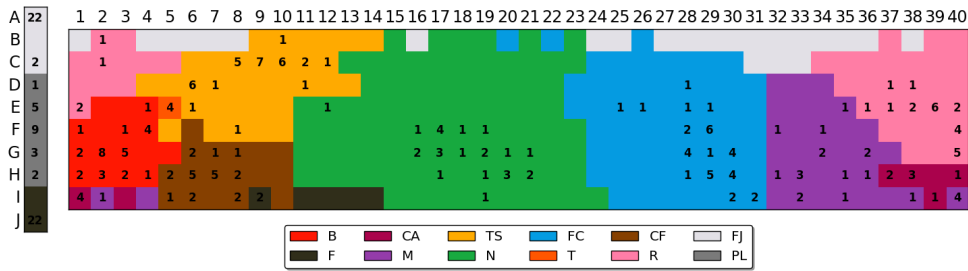
Model	QF	RP
Bayes	**0.91**	**15.37**
Like	1.14	23.55
Proto	1.19	24.81
Exemp	1.21	23.52
Chroma	1.17	30.92

Legend: B, CA, TS, FC, CF, FJ, F, M, N, T, R, PL

Language 20 - Candoshi
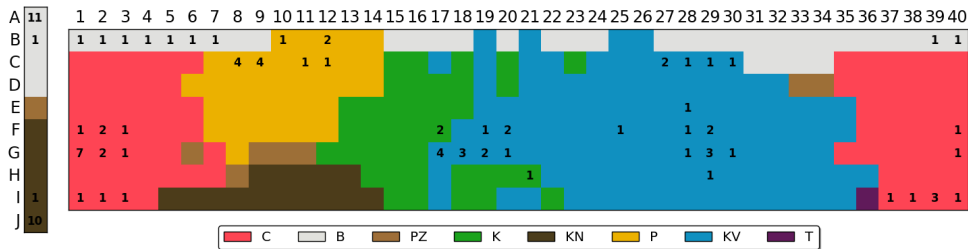
Model	QF	RP
Bayes	**1.21**	**22.26**
Like	1.43	37.13
Proto	1.36	27.59
Exemp	1.44	29.93
Chroma	2.02	58.86

Legend: C, B, PZ, K, KN, P, KV, T

Language 21 - Cavineña
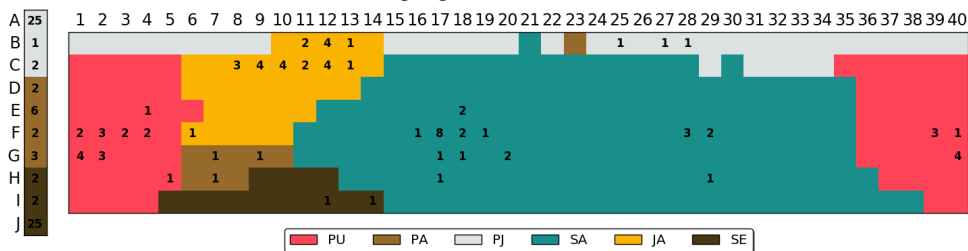
Model	QF	RP
Bayes	**0.98**	**17.08**
Like	1.10	37.07
Proto	1.15	48.14
Exemp	1.05	36.71
Chroma	1.55	70.20

Legend: PU, PA, PJ, SA, JA, SE

Language 22 - Cayapa
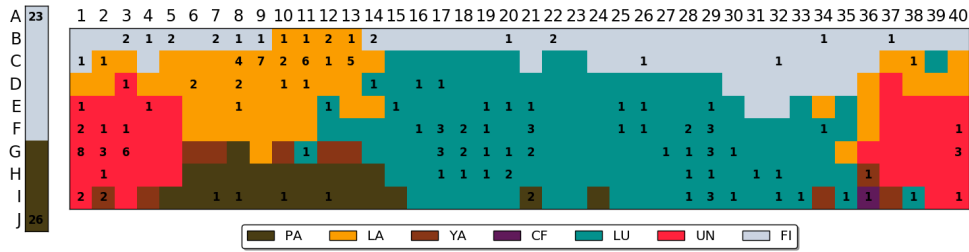
Model	QF	RP
Bayes	**1.18**	**21.64**
Like	1.61	42.00
Proto	1.75	61.55
Exemp	1.41	39.32
Chroma	2.01	90.03

Legend: PA LA YA CF LU UN FI

Language 23 - Chácobo
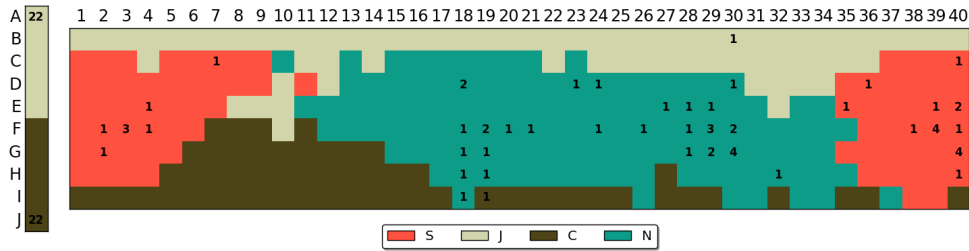
Model	QF	RP
Bayes	**0.98**	**25.83**
Like	1.52	47.23
Proto	1.63	63.38
Exemp	1.05	36.88
Chroma	2.10	156.58

Legend: S J C N

Language 24 - Chavacano (Zamboangueño)
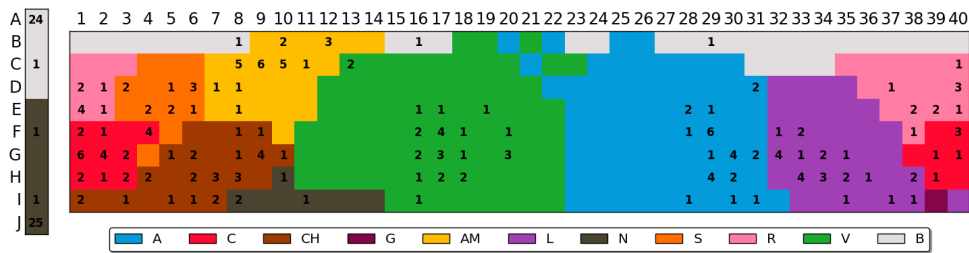
Model	QF	RP
Bayes	**0.79**	**12.54**
Like	0.95	22.29
Proto	0.96	22.52
Exemp	0.88	20.49
Chroma	1.41	37.08

Legend: A C CH G AM L N S R V B

Language 25 - Chayahuita
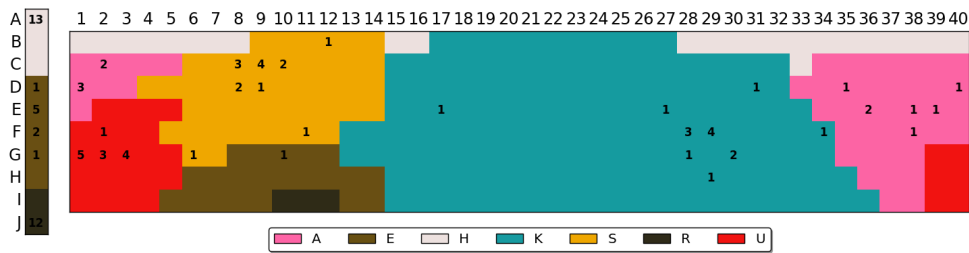
Model	QF	RP
Bayes	**1.00**	**19.57**
Like	1.06	30.31
Proto	1.08	38.98
Exemp	1.11	32.42
Chroma	1.23	61.36

Legend: A E H K S R U

Language 26 - Chinantec
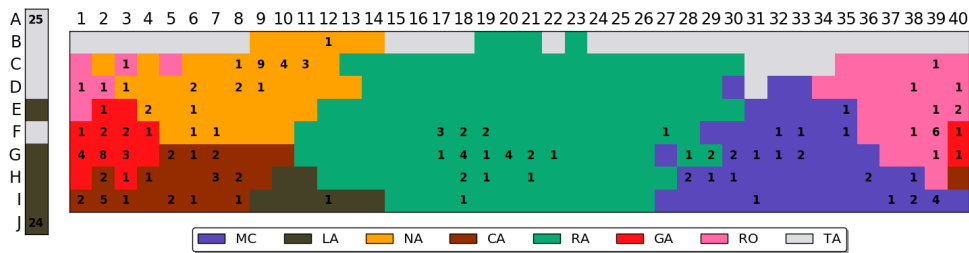
Model	QF	RP
Bayes	**0.99**	**17.71**
Like	1.34	31.81
Proto	1.38	34.09
Exemp	1.15	27.00
Chroma	1.48	60.02

Legend: MC LA NA CA RA GA RO TA

Language 27 - Chiquitano
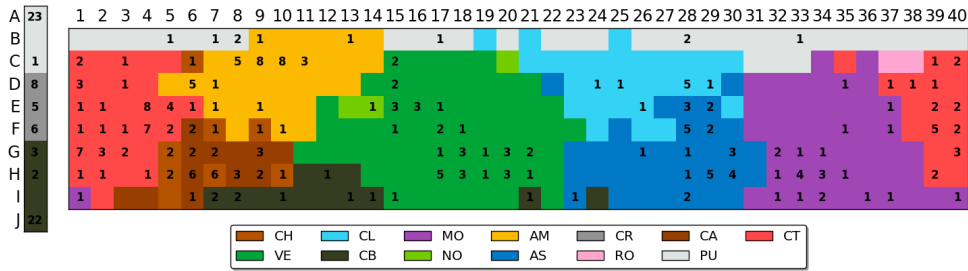
Model	QF	RP
Bayes	**0.92**	**22.56**
Like	1.16	28.68
Proto	1.26	30.24
Exemp	1.03	27.55
Chroma	1.37	46.45

Legend: CH, CL, MO, AM, CR, CA, CT, VE, CB, NO, AS, RO, PU

Language 28 - Chumburu
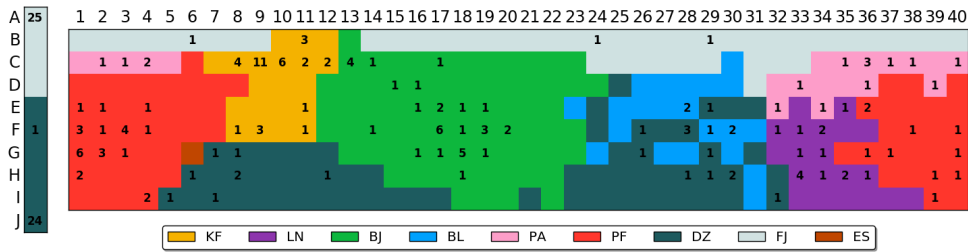
Model	QF	RP
Bayes	**1.03**	**17.77**
Like	1.33	31.86
Proto	1.37	34.61
Exemp	1.07	26.94
Chroma	1.56	69.44

Legend: KF, LN, BJ, BL, PA, PF, DZ, FJ, ES

Language 29 - Cófan
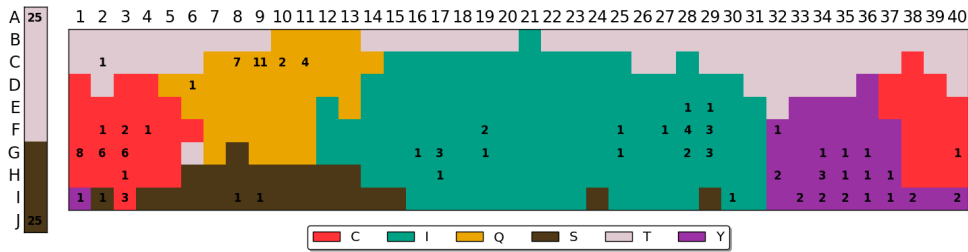
Model	QF	RP
Bayes	**1.02**	**18.44**
Like	1.16	30.54
Proto	1.20	37.58
Exemp	1.16	28.65
Chroma	1.51	77.24

Legend: C, I, Q, S, T, Y

Language 30 - Colorado
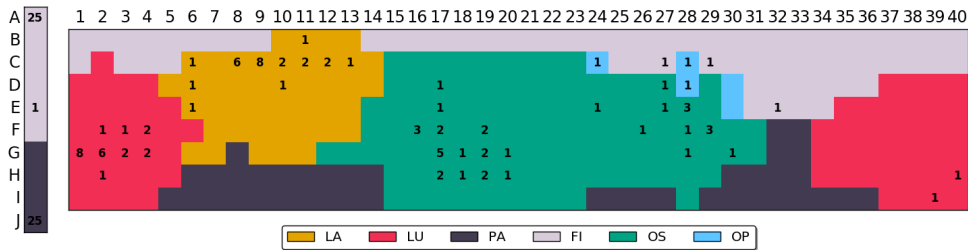
Model	QF	RP
Bayes	**1.26**	**16.73**
Like	1.38	37.98
Proto	1.35	39.84
Exemp	1.33	33.70
Chroma	1.57	72.92

Legend: LA, LU, PA, FI, OS, OP

Language 31 - Cree
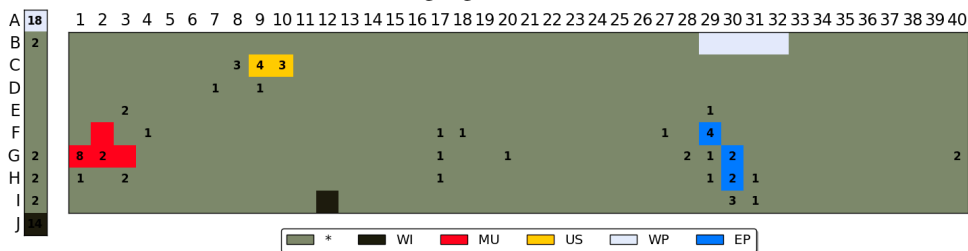
Model	QF	RP
Bayes	1.13	15.45
Like	1.13	16.06
Proto	1.13	20.21
Exemp	1.05	**14.99**
Chroma	**0.93**	21.95

Legend: *, WI, MU, US, WP, EP

Language 32 - Culina
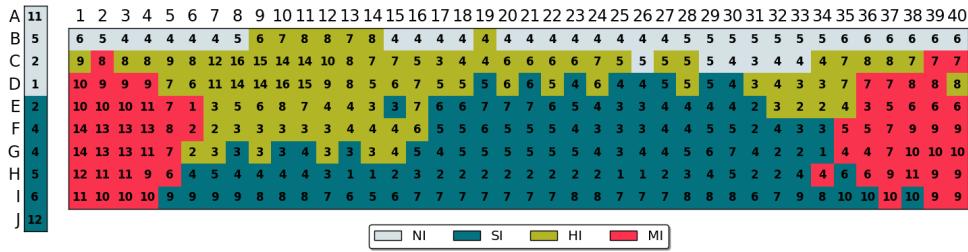
Model	QF	RP
Bayes	**1.32**	82.84
Like	1.75	81.94
Proto	2.04	89.68
Exemp	1.99	**81.20**
Chroma	3.67	113.37

Language 33 - Didinga
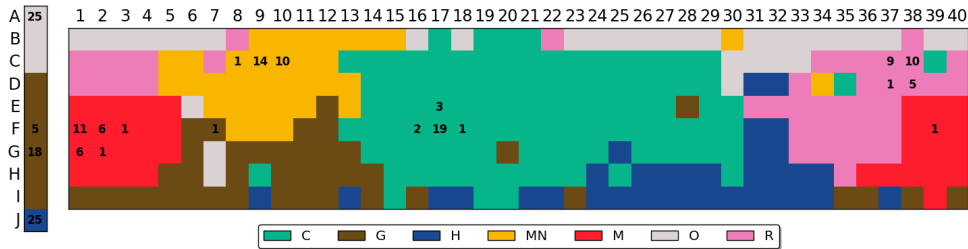
Model	QF	RP
Bayes	1.04	**15.42**
Like	0.95	31.70
Proto	0.96	37.87
Exemp	**0.92**	29.61
Chroma	1.78	88.03

Language 34 - Djuka
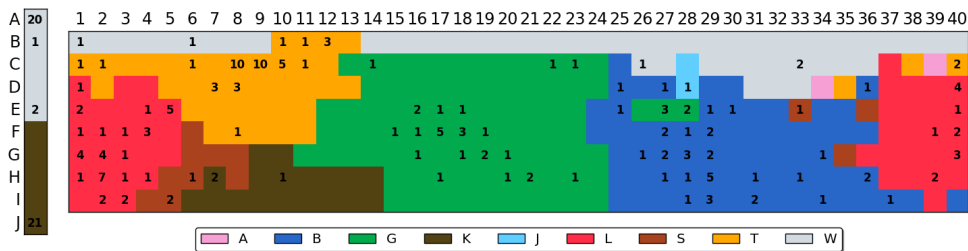
Model	QF	RP
Bayes	**0.92**	**24.54**
Like	1.46	35.87
Proto	1.48	46.66
Exemp	1.23	33.60
Chroma	1.52	65.37

Language 35 - Dyimini
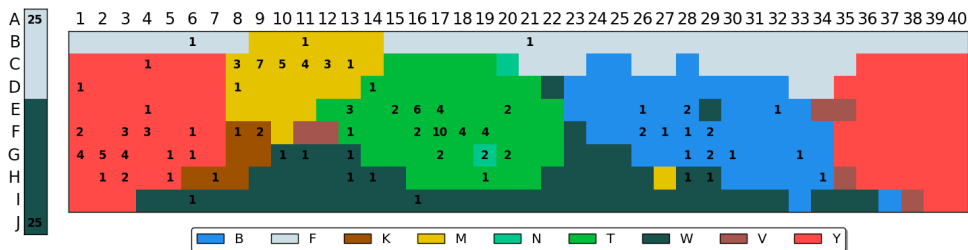
Model	QF	RP
Bayes	0.99	**23.78**
Like	1.09	38.32
Proto	1.10	41.97
Exemp	**0.94**	35.35
Chroma	1.40	69.14

Language 36 - Ejagam
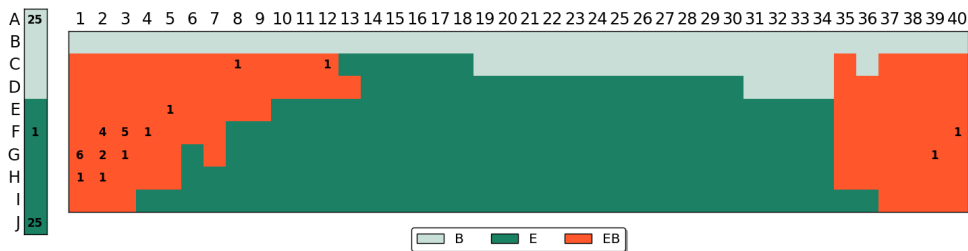
Model	QF	RP
Bayes	1.29	**29.88**
Like	1.13	55.21
Proto	1.20	38.06
Exemp	**1.09**	38.27
Chroma	1.89	109.83

Language 37 - Ese Ejja
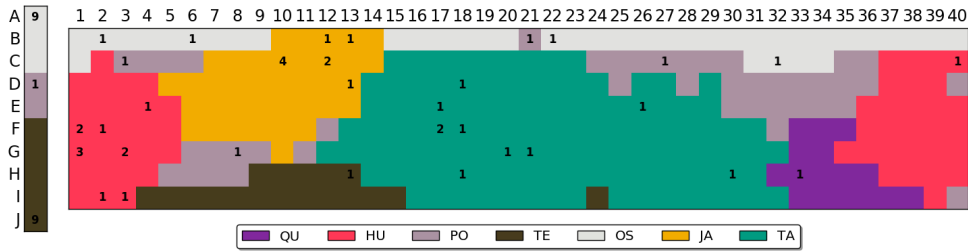
Model	QF	RP
Bayes	**0.85**	**22.27**
Like	1.17	27.64
Proto	1.17	33.69
Exemp	1.11	26.78
Chroma	1.73	77.91

Legend: QU, HU, PO, TE, OS, JA, TA

Language 38 - Garífuna (Black Carib)
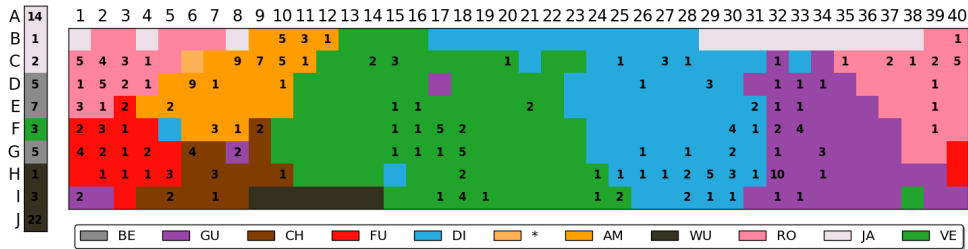
Model	QF	RP
Bayes	1.34	**37.51**
Like	1.63	50.88
Proto	1.61	65.41
Exemp	**1.26**	44.21
Chroma	2.29	119.70

Legend: BE, GU, CH, FU, DI, *, AM, WU, RO, JA, VE

Language 39 - Guahibo
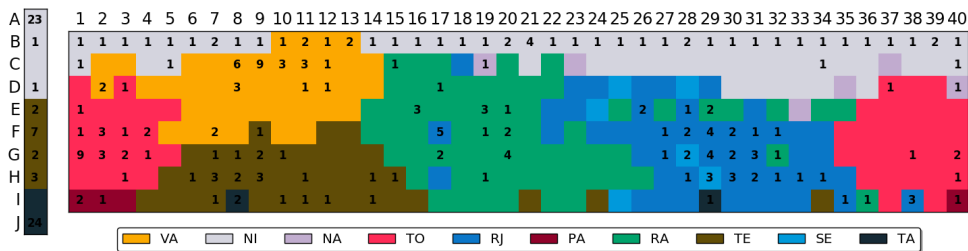
Model	QF	RP
Bayes	1.42	**29.84**
Like	1.40	40.66
Proto	1.51	54.59
Exemp	**1.34**	41.02
Chroma	2.24	84.67

Legend: VA, NI, NA, TO, RJ, PA, RA, TE, SE, TA

Language 40 - Guambiano
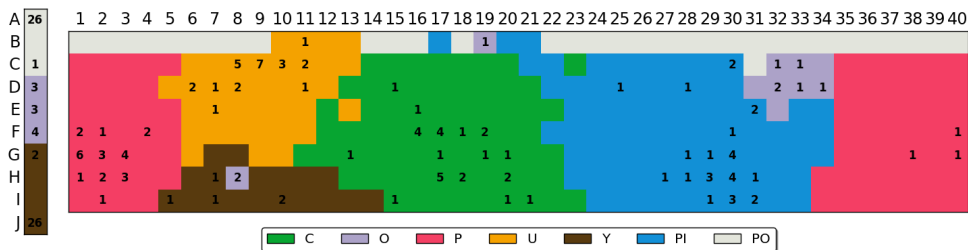
Model	QF	RP
Bayes	**1.09**	**24.06**
Like	1.30	38.76
Proto	1.25	41.61
Exemp	1.30	36.60
Chroma	1.40	61.16

Legend: C, O, P, U, Y, PI, PO

Language 41 - Guarijío
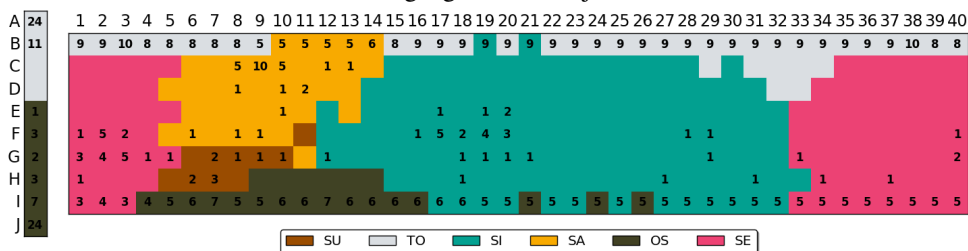
Model	QF	RP
Bayes	**1.22**	**20.27**
Like	1.47	39.28
Proto	1.51	44.67
Exemp	1.45	38.13
Chroma	4.27	69.73

Legend: SU, TO, SI, SA, OS, SE

Language 42 - Guaymí (Ngäbere)

Model	QF	RP
Bayes	**0.86**	**47.00**
Like	1.54	58.34
Proto	1.66	69.01
Exemp	1.26	56.17
Chroma	1.55	80.04

Language 43 - Gunu

Model	QF	RP
Bayes	**1.74**	**24.30**
Like	2.15	40.13
Proto	2.11	38.32
Exemp	2.35	34.58
Chroma	2.39	80.91

Language 44 - Halbi

Model	QF	RP
Bayes	**1.23**	**32.45**
Like	1.54	45.98
Proto	1.64	54.12
Exemp	1.39	41.69
Chroma	1.99	94.22

Language 45 - Huastec

Model	QF	RP
Bayes	**0.68**	**11.38**
Like	0.94	26.33
Proto	0.95	28.73
Exemp	0.98	25.47
Chroma	1.43	43.00

Language 46 - Huave

Model	QF	RP
Bayes	1.06	**16.51**
Like	1.01	26.16
Proto	1.05	26.03
Exemp	**0.99**	23.87
Chroma	1.43	41.13

Language 47 - Iduna

Model	QF	RP
Bayes	**1.01**	**20.49**
Like	1.20	44.25
Proto	1.28	50.20
Exemp	1.16	39.21
Chroma	1.57	73.41

Language 48 - Ifugao
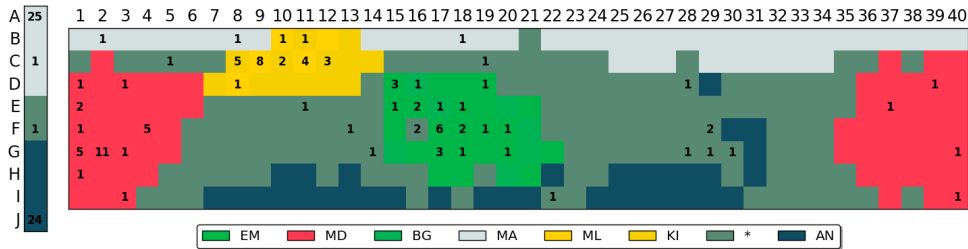
Model	QF	RP
Bayes	**1.13**	**25.93**
Like	1.31	33.67
Proto	1.44	35.83
Exemp	1.15	32.64
Chroma	1.53	69.81

Language 49 - Iwam
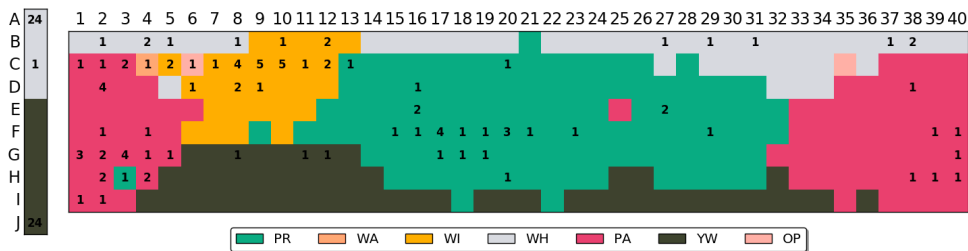
Model	QF	RP
Bayes	**0.73**	**41.90**
Like	1.68	75.47
Proto	1.75	83.44
Exemp	1.67	65.57
Chroma	1.88	115.54

Language 50 - Jicaque
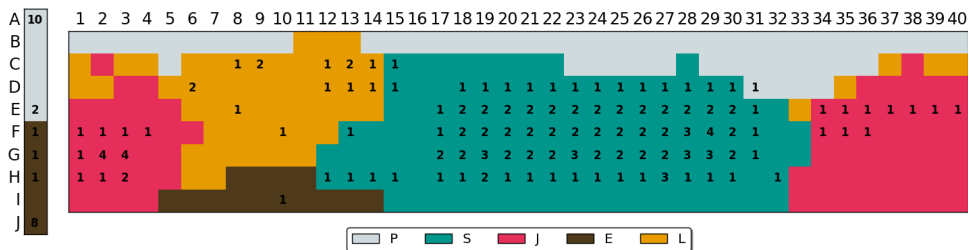
Model	QF	RP
Bayes	1.59	**22.09**
Like	**1.50**	46.00
Proto	2.08	56.76
Exemp	1.51	41.80
Chroma	2.90	72.19

Language 51 - Kalam
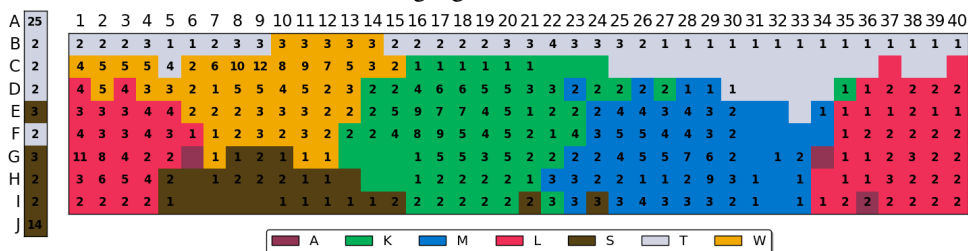
Model	QF	RP
Bayes	1.87	**31.38**
Like	1.68	43.43
Proto	1.78	52.42
Exemp	**1.53**	41.15
Chroma	3.19	89.89

Language 52 - Kamano-Kafe
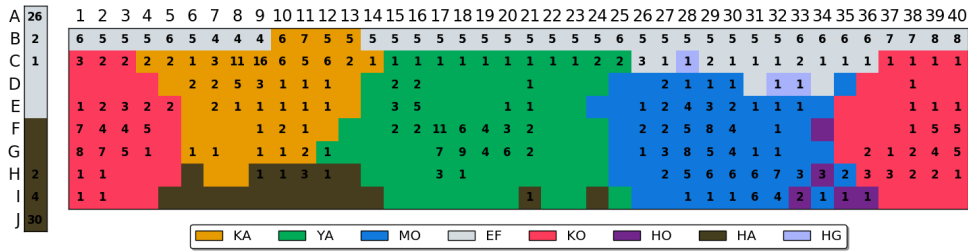
Model	QF	RP
Bayes	1.72	**21.82**
Like	1.64	35.36
Proto	1.70	35.71
Exemp	**1.55**	30.86
Chroma	3.67	63.81

Legend: KA, YA, MO, EF, KO, HO, HA, HG

Language 53 - Karajá
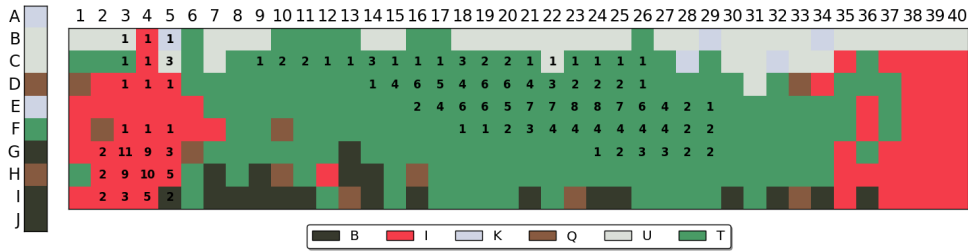
Model	QF	RP
Bayes	2.92	**38.14**
Like	2.71	55.93
Proto	3.04	81.09
Exemp	**2.21**	59.27
Chroma	4.38	118.70

Legend: B, I, K, Q, U, T

Language 54 - Kemtuik
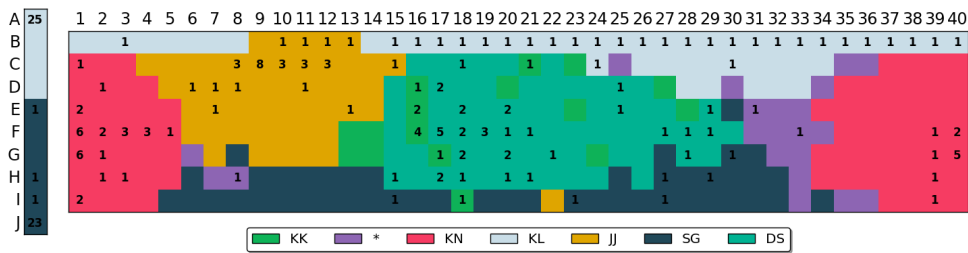
Model	QF	RP
Bayes	**0.83**	**23.09**
Like	1.34	43.17
Proto	1.40	46.31
Exemp	1.20	38.84
Chroma	2.02	80.87

Legend: KK, *, KN, KL, JJ, SG, DS

Language 55 - Kokni (Kokoni)
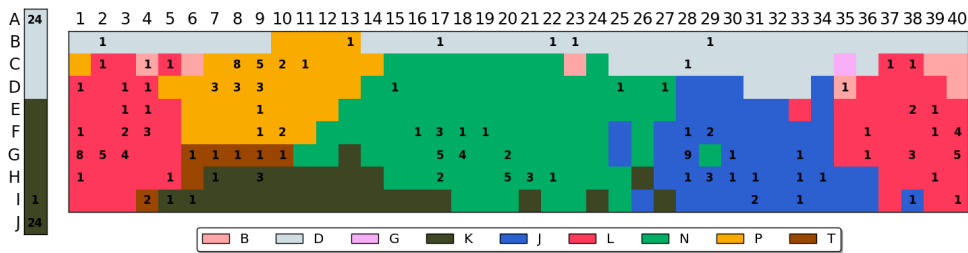
Model	QF	RP
Bayes	**0.82**	**31.73**
Like	1.61	47.70
Proto	1.69	56.99
Exemp	1.14	44.16
Chroma	1.61	85.39

Legend: B, D, G, K, J, L, N, P, T

Language 56 - Konkomba
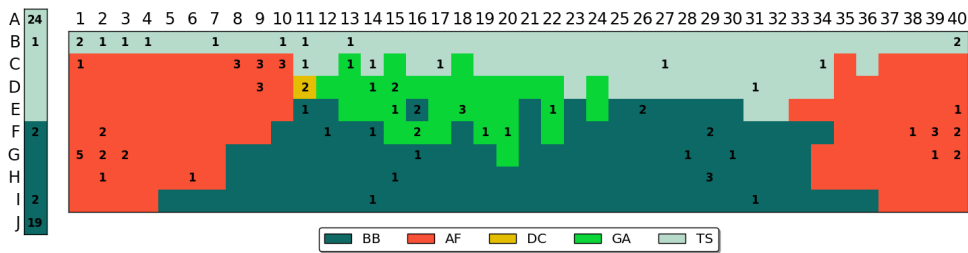
Model	QF	RP
Bayes	**1.18**	**29.14**
Like	1.24	52.62
Proto	1.27	56.38
Exemp	1.32	43.76
Chroma	1.85	97.32

Legend: BB, AF, DC, GA, TS

Language 57 - Kriol
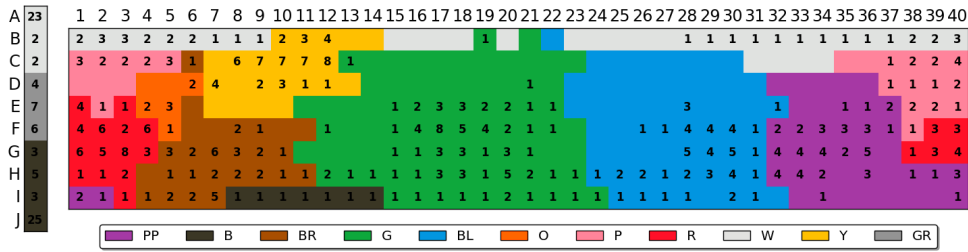
Model	QF	RP
Bayes	**1.02**	**22.89**
Like	1.30	31.25
Proto	1.33	30.06
Exemp	1.21	28.02
Chroma	2.20	52.26

Language 58 - Kuku-Yalanji
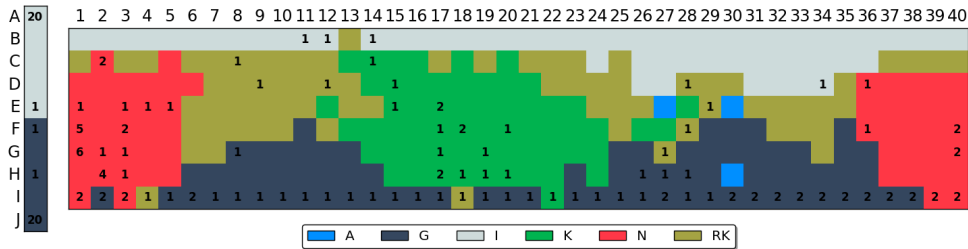
Model	QF	RP
Bayes	2.10	**51.77**
Like	1.99	64.31
Proto	2.12	59.12
Exemp	**1.92**	55.86
Chroma	2.91	101.62

Language 59 - Kuna
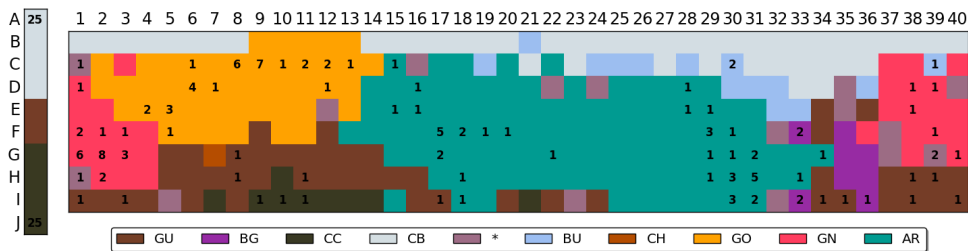
Model	QF	RP
Bayes	**1.16**	**32.00**
Like	1.52	43.58
Proto	1.63	55.13
Exemp	1.43	41.01
Chroma	1.33	65.30

Language 60 - Kwerba
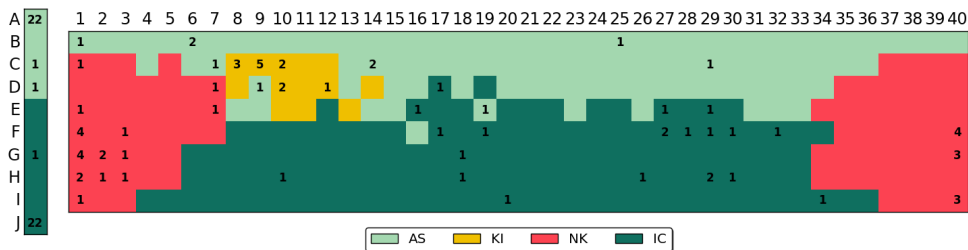
Model	QF	RP
Bayes	**0.95**	**32.16**
Like	1.33	60.03
Proto	1.44	56.42
Exemp	1.37	48.92
Chroma	1.80	107.25

Language 61 - Lele
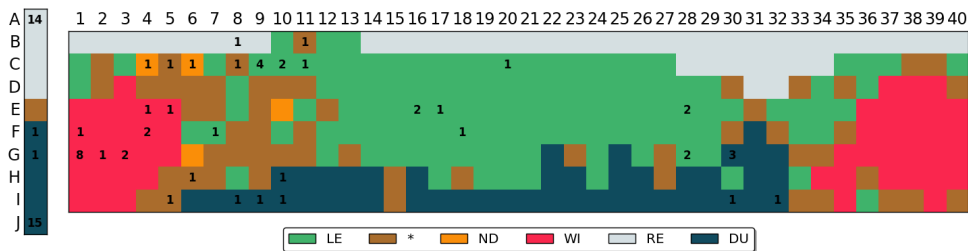
Model	QF	RP
Bayes	**0.81**	**27.96**
Like	1.53	60.55
Proto	1.35	78.90
Exemp	1.19	60.06
Chroma	1.65	88.81

Language 62 - Mampruli
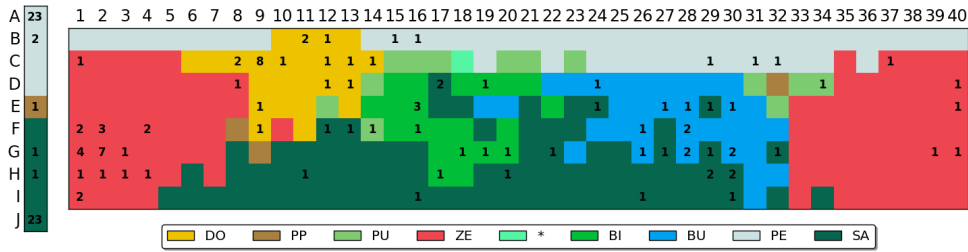
Model	QF	RP
Bayes	**0.91**	**28.14**
Like	1.38	42.37
Proto	1.61	47.21
Exemp	1.19	39.20
Chroma	1.69	91.27

Language 63 - Maring
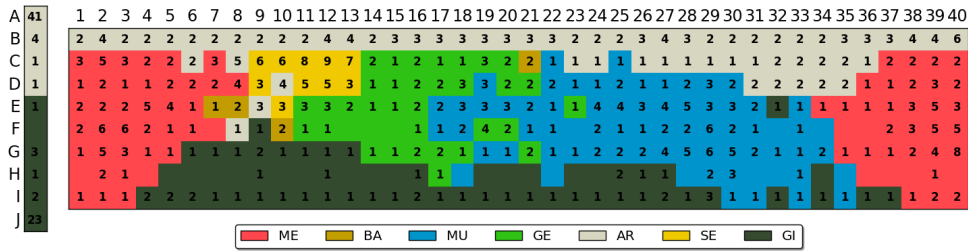
Model	QF	RP
Bayes	1.80	**55.27**
Like	1.88	68.78
Proto	2.08	68.91
Exemp	**1.52**	57.88
Chroma	3.51	107.36

Language 64 - Martu-Wangka
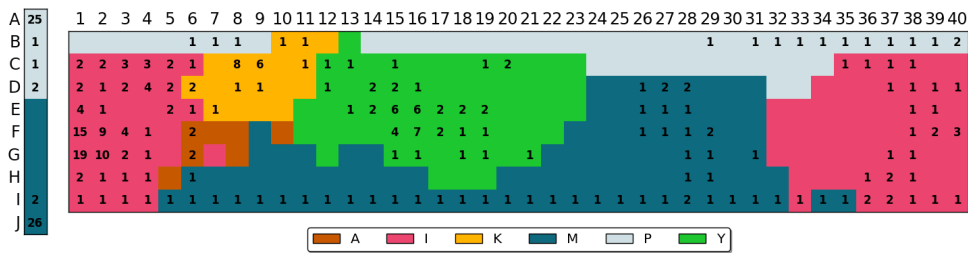
Model	QF	RP
Bayes	**1.66**	**31.32**
Like	1.89	50.19
Proto	1.93	47.66
Exemp	2.16	47.32
Chroma	2.23	80.73

Language 65 - Mawchi
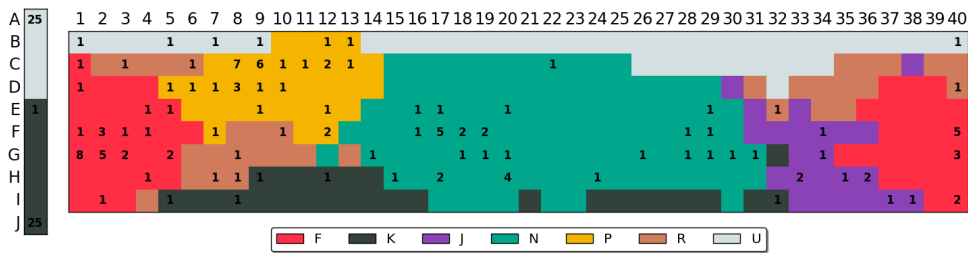
Model	QF	RP
Bayes	**0.77**	**28.00**
Like	1.42	43.75
Proto	1.51	56.13
Exemp	1.11	43.56
Chroma	1.65	77.26

Language 66 - Mayoruna
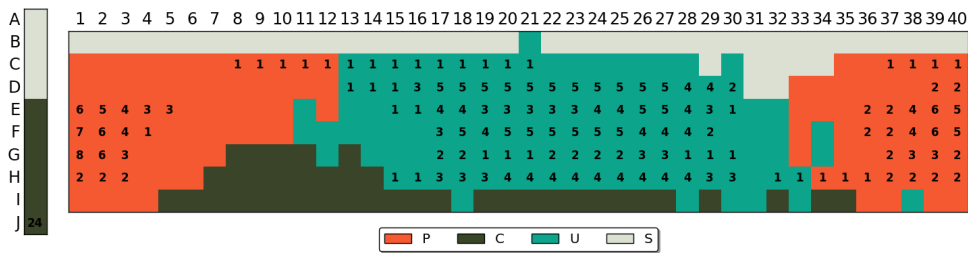
Model	QF	RP
Bayes	2.81	**31.69**
Like	2.14	38.25
Proto	2.14	46.27
Exemp	**1.91**	33.35
Chroma	3.96	111.51

Language 67 - Mazahua
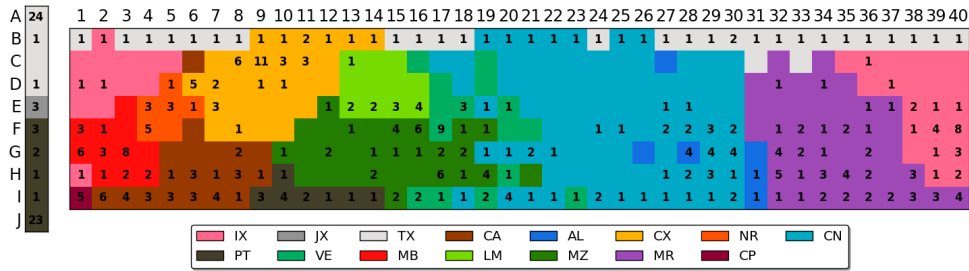
Model	QF	RP
Bayes	**1.22**	**17.03**
Like	1.60	24.83
Proto	1.55	26.67
Exemp	1.51	24.56
Chroma	1.96	38.59

Language 68 - Mazatec
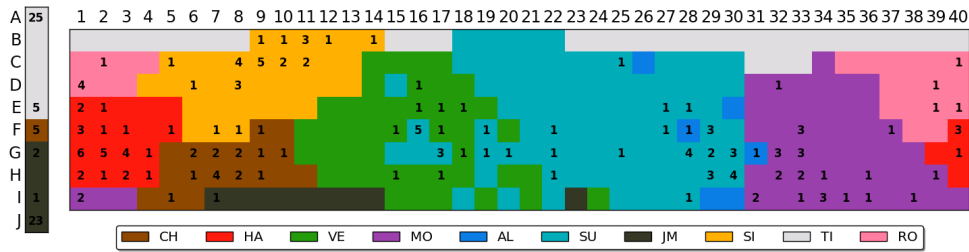
Model	QF	RP
Bayes	**0.85**	**20.38**
Like	1.18	28.92
Proto	1.30	32.47
Exemp	1.23	28.77
Chroma	1.42	52.61

Language 69 - Menye
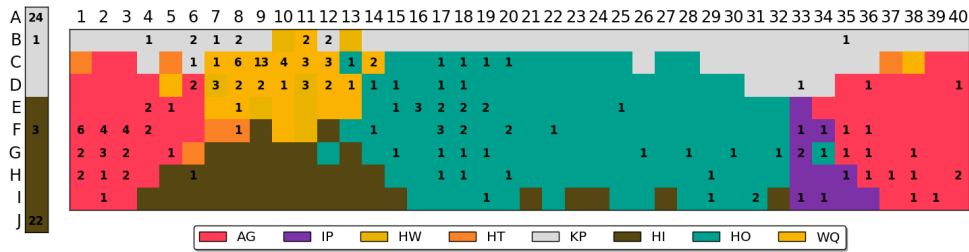
Model	QF	RP
Bayes	**0.87**	**29.62**
Like	1.62	48.40
Proto	1.90	69.59
Exemp	1.29	44.54
Chroma	1.58	90.57

Language 70 - Micmac
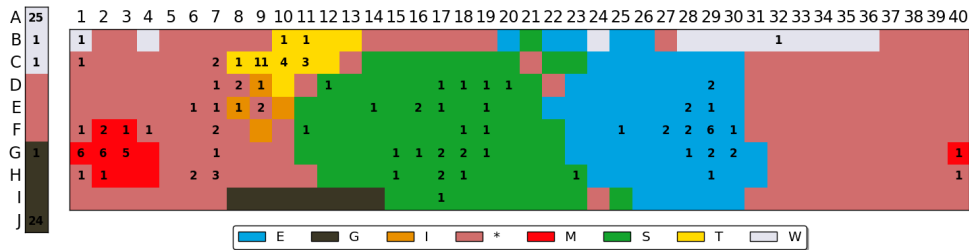
Model	QF	RP
Bayes	1.15	**17.76**
Like	0.97	21.71
Proto	**0.88**	22.98
Exemp	1.03	19.56
Chroma	1.32	42.31

Language 71 - Mikasuki
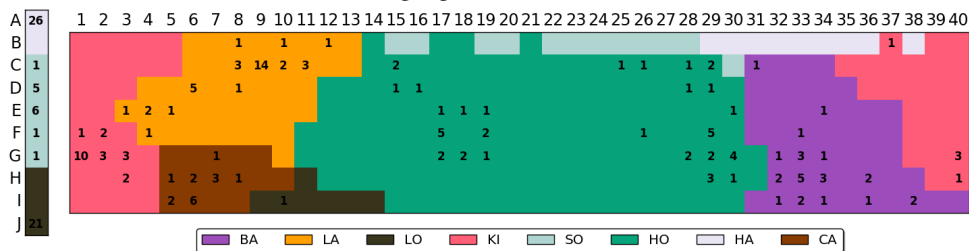
Model	QF	RP
Bayes	**0.87**	**14.79**
Like	1.51	32.31
Proto	1.51	42.61
Exemp	1.38	33.34
Chroma	1.28	54.15

Language 72 - Mixtec
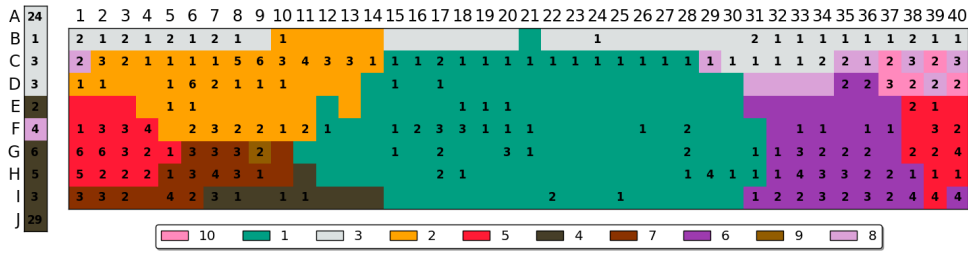
Model	QF	RP
Bayes	**1.30**	**25.51**
Like	1.50	35.29
Proto	1.49	37.52
Exemp	1.48	33.93
Chroma	2.12	57.58

Legend: 10, 1, 3, 2, 5, 4, 7, 6, 9, 8

Language 73 - Mundu
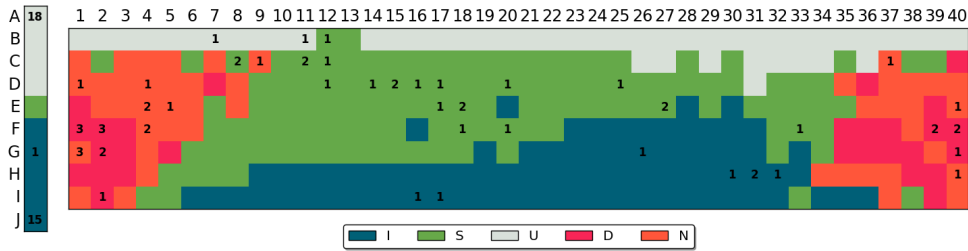
Model	QF	RP
Bayes	**0.92**	**45.63**
Like	1.59	74.48
Proto	1.52	83.72
Exemp	1.26	67.17
Chroma	1.77	106.59

Legend: I, S, U, D, N

Language 74 - Múra-Pirahá
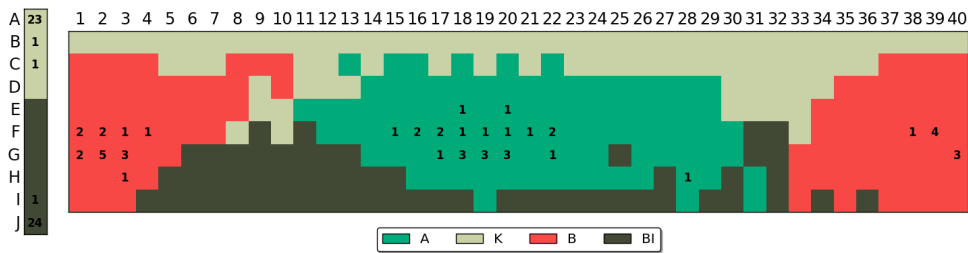
Model	QF	RP
Bayes	1.03	**12.07**
Like	1.09	43.60
Proto	1.05	34.77
Exemp	**0.90**	29.74
Chroma	2.02	115.75

Legend: A, K, B, BI

Language 75 - Murle
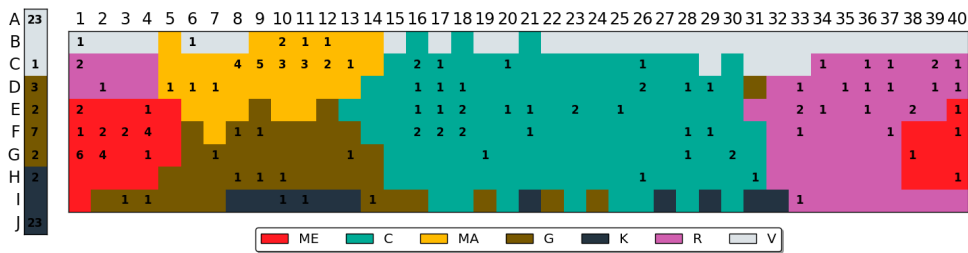
Model	QF	RP
Bayes	0.99	**24.18**
Like	**0.99**	34.13
Proto	1.06	36.84
Exemp	1.05	33.54
Chroma	1.67	74.55

Legend: ME, C, MA, G, K, R, V

Language 76 - Murrinh-Patha
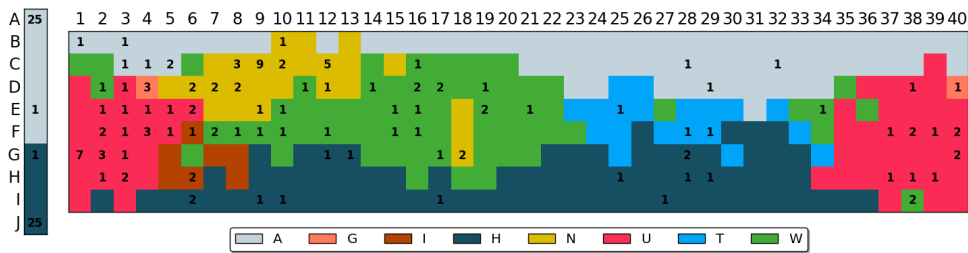
Model	QF	RP
Bayes	**0.85**	**37.60**
Like	1.39	57.32
Proto	1.58	68.33
Exemp	1.28	55.53
Chroma	1.60	78.68

Legend: A, G, I, H, N, U, T, W

Language 77 - Nafaanra
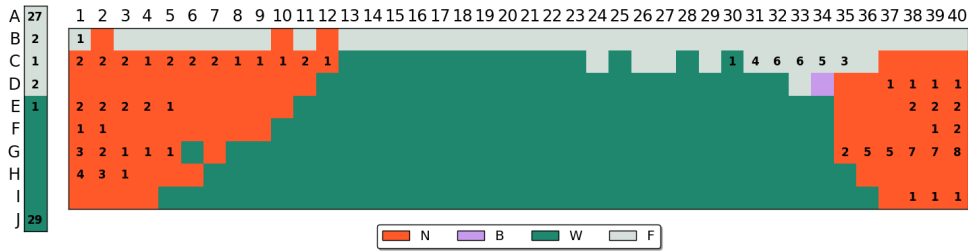
Model	QF	RP
Bayes	1.90	40.27
Like	1.49	51.09
Proto	1.70	47.47
Exemp	**1.31**	**37.00**
Chroma	2.53	132.76

Legend: N B W F

Language 78 - Nahuatl
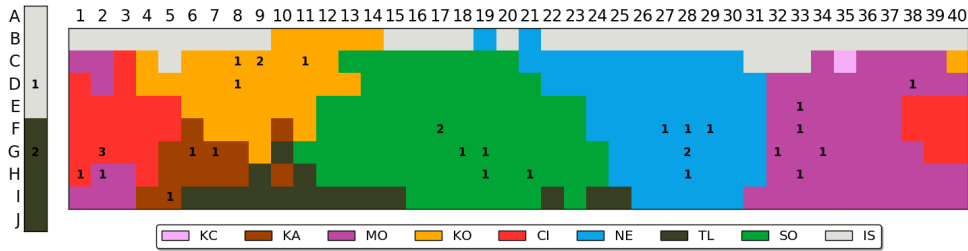
Model	QF	RP
Bayes	1.13	**25.23**
Like	1.14	35.64
Proto	1.16	35.63
Exemp	**1.03**	30.13
Chroma	1.40	38.17

Legend: KC KA MO KO CI NE TL SO IS

Language 79 - Ocaina
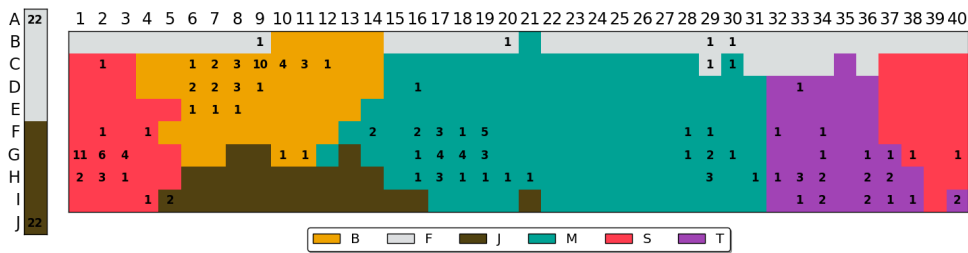
Model	QF	RP
Bayes	**1.09**	**20.44**
Like	1.37	40.63
Proto	1.44	51.11
Exemp	1.34	38.85
Chroma	1.74	65.92

Legend: B F J M S T

Language 80 - Papago (O'odham)
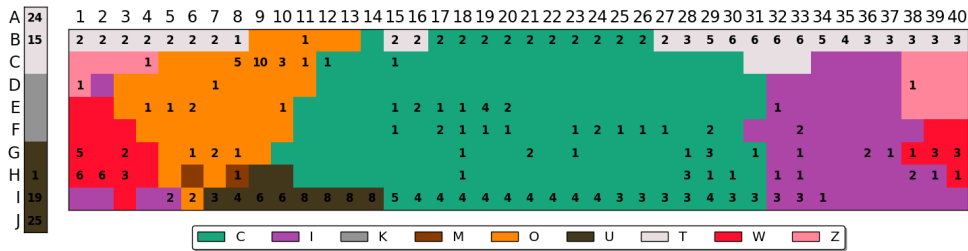
Model	QF	RP
Bayes	1.26	**19.22**
Like	**1.01**	29.40
Proto	1.13	40.19
Exemp	1.10	30.38
Chroma	2.24	60.58

Legend: C I K M O U T W Z

Language 81 - Patep
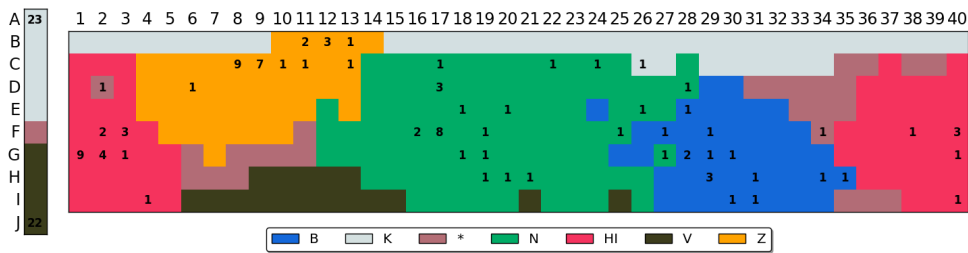
Model	QF	RP
Bayes	**1.02**	**15.72**
Like	1.21	33.40
Proto	1.27	32.43
Exemp	1.08	28.98
Chroma	1.44	59.71

Legend: B K * N HI V Z

Language 82 - Paya
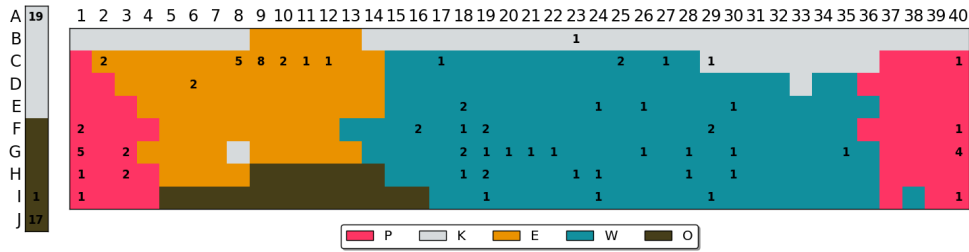
Model	QF	RP
Bayes	**0.84**	**21.41**
Like	1.42	40.32
Proto	1.55	50.27
Exemp	1.05	37.24
Chroma	1.53	88.70

Legend: P, K, E, W, O

Language 83 - Podopa
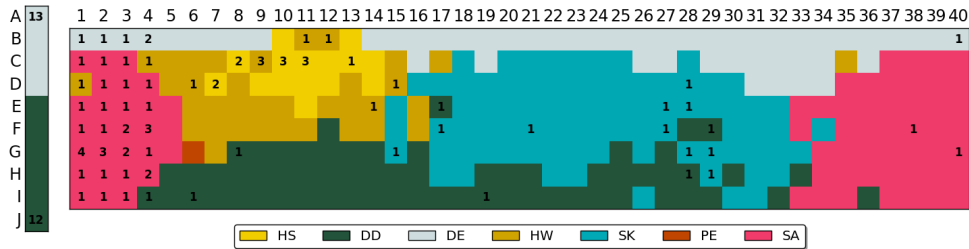
Model	QF	RP
Bayes	**1.56**	**49.43**
Like	1.66	68.16
Proto	1.68	72.85
Exemp	1.68	63.88
Chroma	1.85	92.45

Legend: HS, DD, DE, HW, SK, PE, SA

Language 84 - Saramaccan
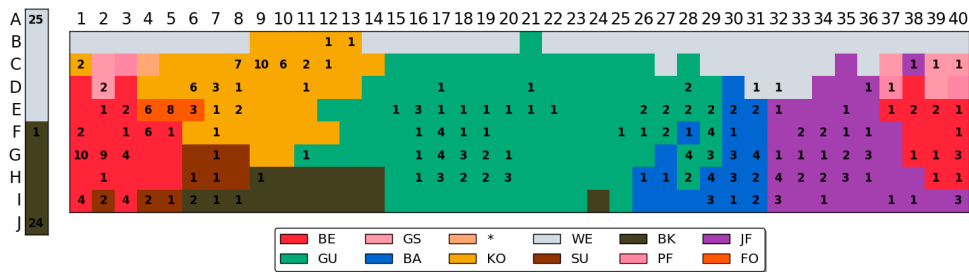
Model	QF	RP
Bayes	**0.98**	**18.36**
Like	1.60	28.60
Proto	1.74	35.42
Exemp	1.34	26.94
Chroma	1.41	50.13

Legend: BE, GS, *, WE, BK, JF, GU, BA, KO, SU, PF, FO

Language 85 - Seri
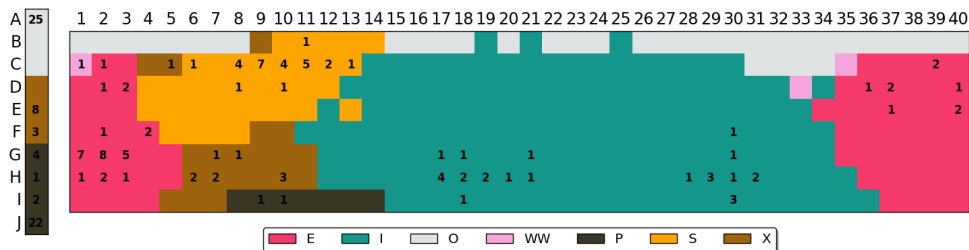
Model	QF	RP
Bayes	**1.05**	**22.33**
Like	1.11	34.96
Proto	1.12	41.43
Exemp	1.06	35.55
Chroma	1.54	70.14

Legend: E, I, O, WW, P, S, X

Language 86 - Shipibo
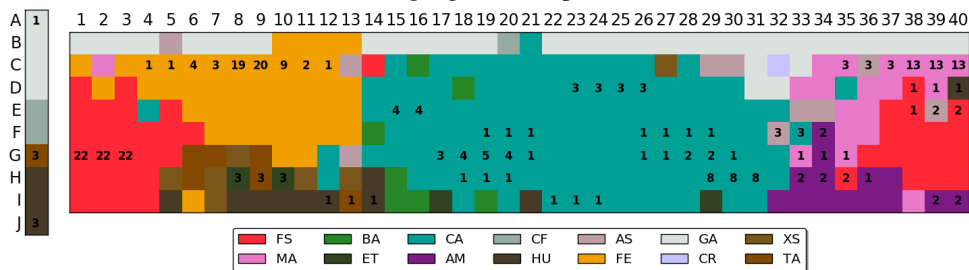
Model	QF	RP
Bayes	**1.21**	**34.23**
Like	1.73	39.27
Proto	1.85	56.96
Exemp	1.80	38.53
Chroma	1.87	83.69

Legend: FS, BA, CA, CF, AS, GA, XS, MA, ET, AM, HU, FE, CR, TA

Language 87 - Sirionó
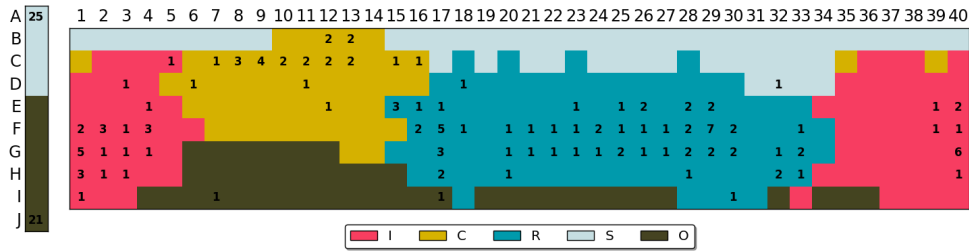
Model	QF	RP
Bayes	**1.26**	**27.21**
Like	1.47	45.16
Proto	1.62	53.77
Exemp	1.37	41.13
Chroma	1.76	87.99

Language 88 - Slave
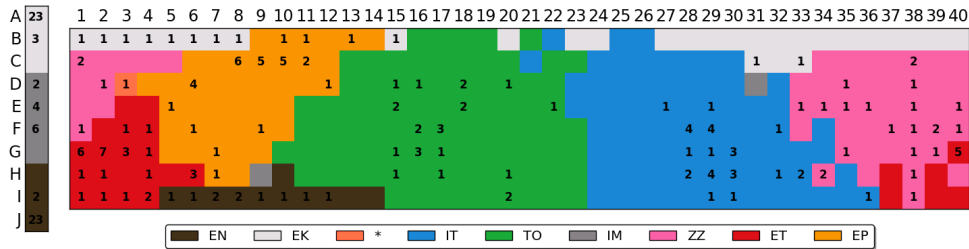
Model	QF	RP
Bayes	**0.98**	**23.40**
Like	1.18	35.54
Proto	1.22	40.34
Exemp	1.18	34.23
Chroma	1.55	54.69

Language 89 - Sursurunga
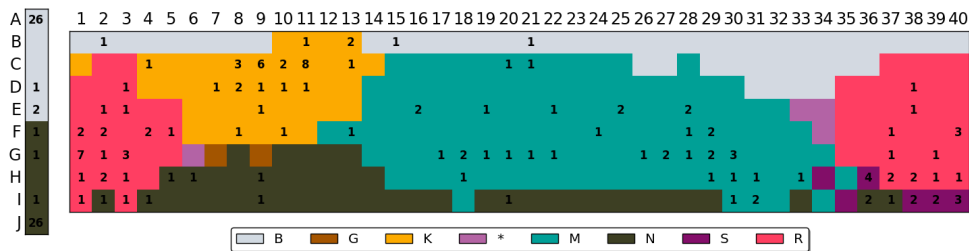
Model	QF	RP
Bayes	1.09	**25.65**
Like	1.17	36.15
Proto	1.32	46.63
Exemp	**1.04**	34.70
Chroma	1.59	82.78

Language 90 - Tabla
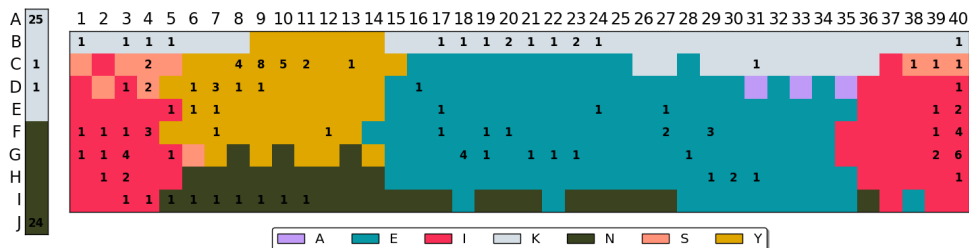
Model	QF	RP
Bayes	1.06	**21.86**
Like	1.30	38.22
Proto	1.45	49.31
Exemp	**1.04**	35.52
Chroma	1.49	70.80

Language 91 - Tacana
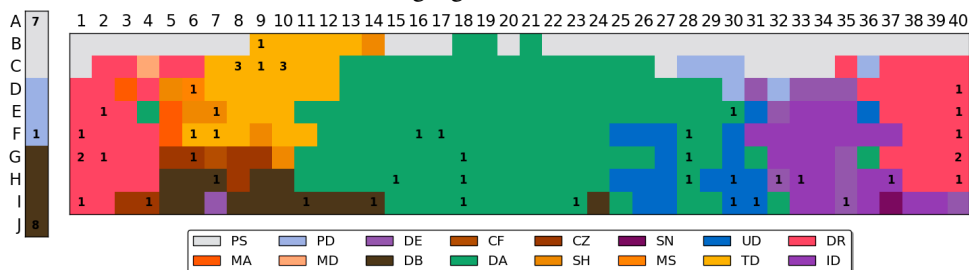
Model	QF	RP
Bayes	**0.99**	**23.51**
Like	1.26	41.30
Proto	1.22	38.89
Exemp	1.25	37.33
Chroma	1.53	62.32

Language 92 - Tarahumara (Central)
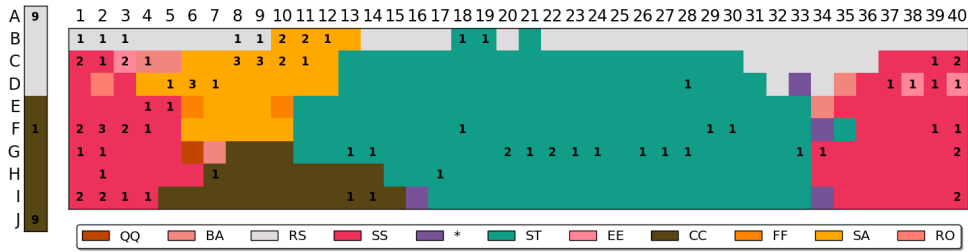
Model	QF	RP
Bayes	**1.14**	**28.93**
Like	1.20	34.77
Proto	1.37	36.78
Exemp	1.32	33.34
Chroma	2.03	70.93

Language 93 - Tarahumara (Western)
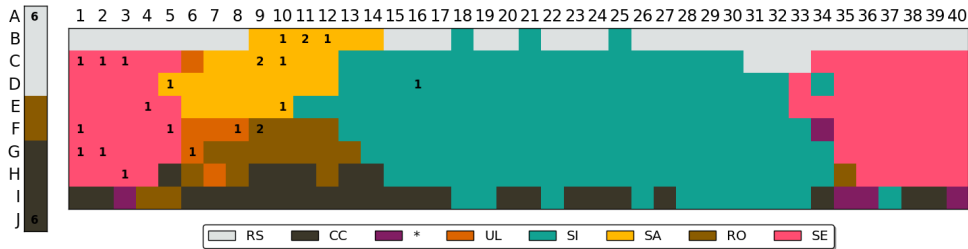
Model	QF	RP
Bayes	1.56	38.93
Like	**1.02**	**32.13**
Proto	1.09	41.61
Exemp	1.06	36.38
Chroma	1.46	51.02

Language 94 - Tboli
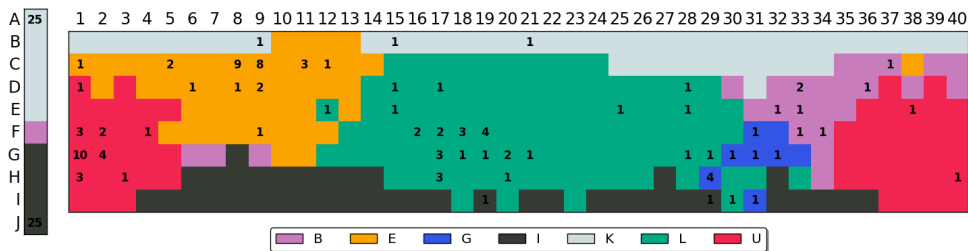
Model	QF	RP
Bayes	**0.76**	**18.55**
Like	1.17	40.32
Proto	1.22	44.16
Exemp	1.04	34.83
Chroma	1.63	74.44

Language 95 - Teribe
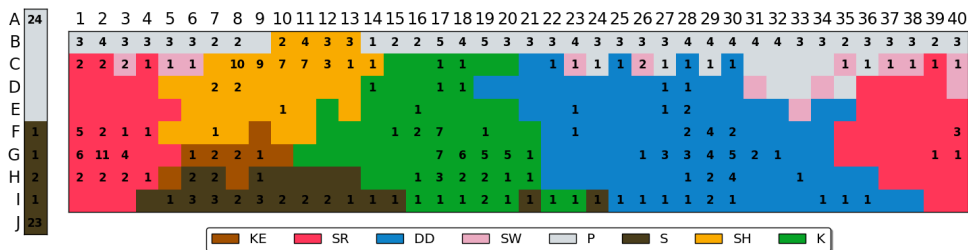
Model	QF	RP
Bayes	1.89	**26.48**
Like	1.82	37.32
Proto	**1.54**	37.74
Exemp	1.58	32.58
Chroma	3.52	71.63

Language 96 - Ticuna
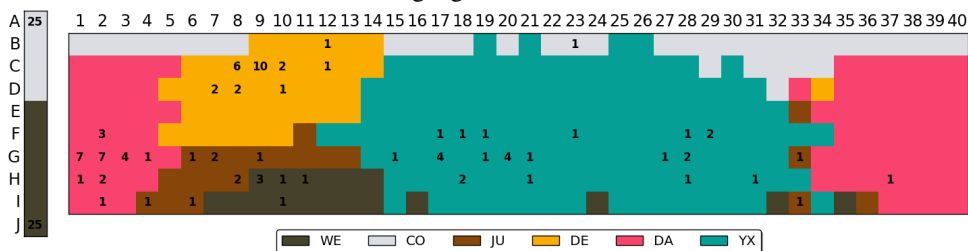
Model	QF	RP
Bayes	**0.78**	**16.86**
Like	1.23	36.39
Proto	1.44	46.55
Exemp	1.22	35.00
Chroma	1.69	79.22

Language 97 - Tifal
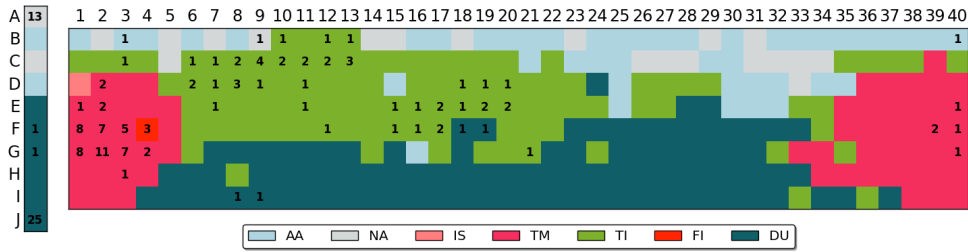
Model	QF	RP
Bayes	**0.97**	**39.95**
Like	2.04	78.00
Proto	1.93	85.13
Exemp	1.99	69.90
Chroma	2.06	85.02

Legend: AA, NA, IS, TM, TI, FI, DU

Language 98 - Tlapanec
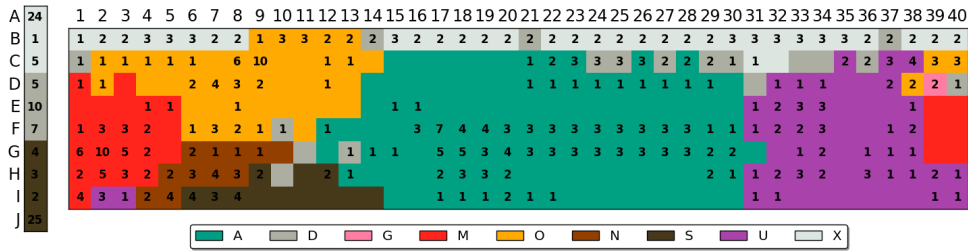
Model	QF	RP
Bayes	1.50	**25.05**
Like	1.31	34.77
Proto	1.32	39.24
Exemp	**1.23**	34.32
Chroma	3.11	73.09

Legend: A, D, G, M, O, N, S, U, X

Language 99 - Tucano
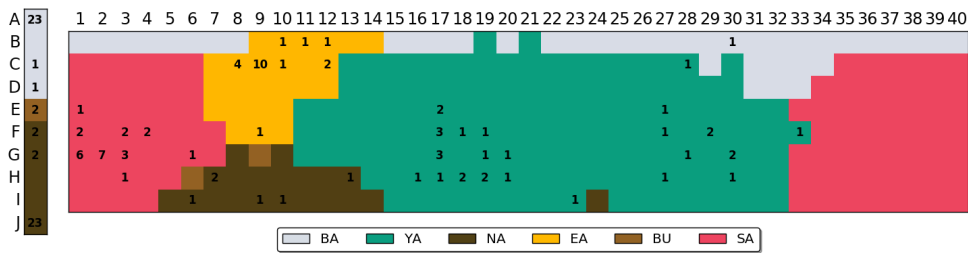
Model	QF	RP
Bayes	**0.94**	**18.00**
Like	1.18	41.03
Proto	1.44	46.19
Exemp	1.13	37.40
Chroma	1.52	80.45

Legend: BA, YA, NA, EA, BU, SA

Language 100 - Vagla
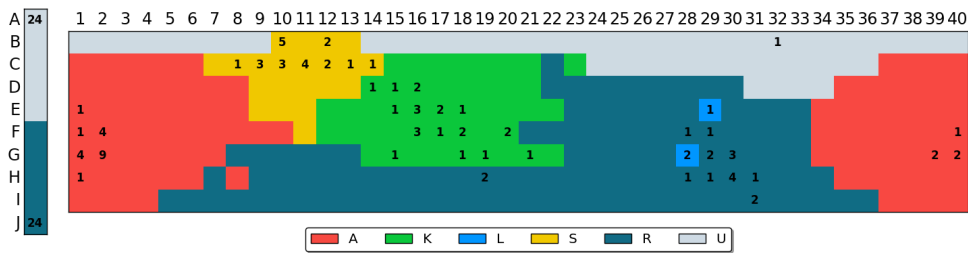
Model	QF	RP
Bayes	**0.98**	**14.44**
Like	1.42	36.53
Proto	1.44	33.81
Exemp	1.34	31.65
Chroma	1.69	71.22

Legend: A, K, L, S, R, U

Language 101 - Vasavi
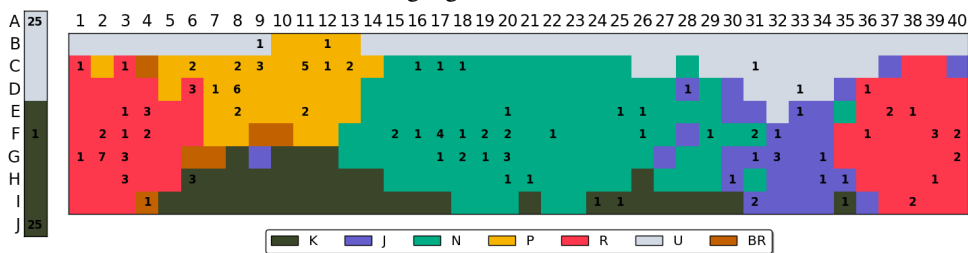
Model	QF	RP
Bayes	**0.73**	**25.24**
Like	1.55	48.87
Proto	1.55	65.55
Exemp	1.09	45.16
Chroma	1.59	93.12

Legend: K, J, N, P, R, U, BR

Language 102 - Waorani (Auca, Huao)

Legend: W, WT, O, WA, N
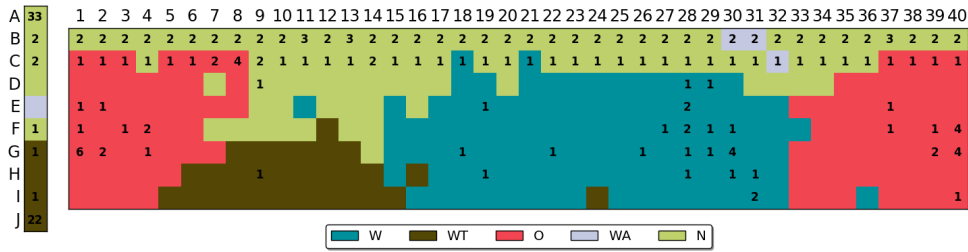
Model	QF	RP
Bayes	1.77	**37.64**
Like	**1.60**	55.90
Proto	1.94	55.56
Exemp	1.68	42.69
Chroma	3.42	111.59

Language 103 - Walpiri

Legend: A, D, K, M, WJ, U, W, Y
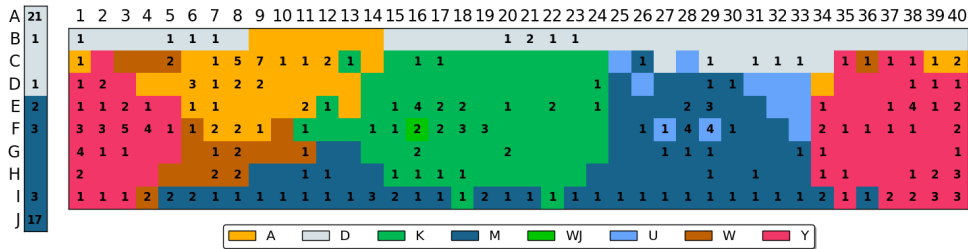
Model	QF	RP
Bayes	1.15	**48.09**
Like	1.25	52.95
Proto	1.37	56.05
Exemp	**1.14**	51.34
Chroma	2.39	82.77

Language 104 - Wobé

Legend: S, K, L
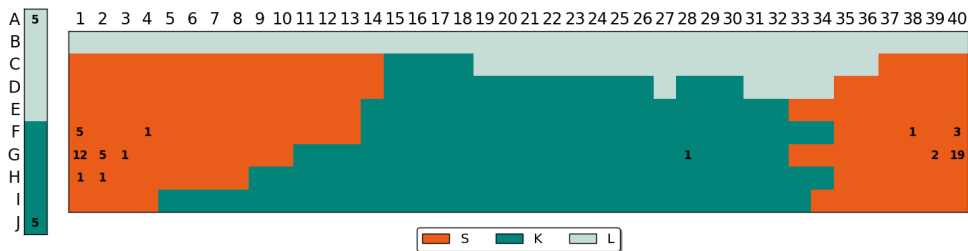
Model	QF	RP
Bayes	2.20	57.67
Like	2.17	61.34
Proto	2.13	57.82
Exemp	**1.79**	**42.31**
Chroma	2.30	110.96

Language 105 - Yacouba

Legend: P, Z, T
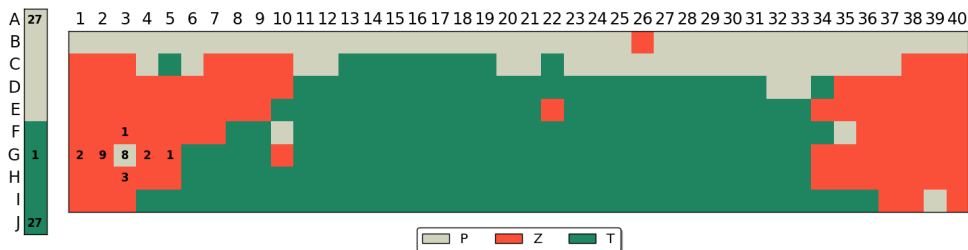
Model	QF	RP
Bayes	**1.12**	**24.77**
Like	1.30	65.11
Proto	1.45	53.79
Exemp	1.20	49.73
Chroma	1.83	135.72

Language 106 - Yakan

Legend: A, B, G, I, K, J, BI, P, S, T, KU, PO
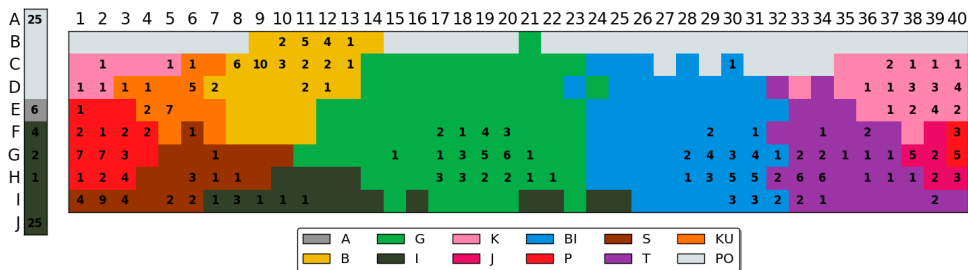
Model	QF	RP
Bayes	**1.19**	**15.38**
Like	1.42	24.01
Proto	1.47	27.46
Exemp	1.33	23.06
Chroma	1.74	42.57

Language 107 - Yaminahua

Model	QF	RP
Bayes	**1.00**	**30.31**
Like	1.62	51.68
Proto	1.76	66.49
Exemp	1.44	49.80
Chroma	2.03	90.45

Legend: FI, CH, OS, DA, OX

Language 108 - Yucuna

Model	QF	RP
Bayes	**0.67**	**21.56**
Like	1.58	45.99
Proto	1.56	60.67
Exemp	1.50	44.19
Chroma	1.37	58.48

Legend: IP, CA, *, QA, IA, JE, JA

Language 109 - Yupik

Model	QF	RP
Bayes	1.35	**29.81**
Like	1.27	37.07
Proto	1.30	44.07
Exemp	**1.26**	36.15
Chroma	2.98	82.57

Legend: E, K, Q, S, T, Y

Language 110 - Zapotec

Model	QF	RP
Bayes	**0.84**	**14.25**
Like	1.27	27.47
Proto	1.18	29.39
Exemp	1.17	25.14
Chroma	1.53	57.11

Legend: C, B, G, BL, P, R, W, Y

5   Model Predictions for WCS languages and individual speakers

We present model predictions for a select number of languages, including speaker-level analyses. In particular, we focus on the six languages from which the example categories (a)-(f) of Figure 3 in the main text were drawn. For each language we first provide a language-level analysis including model predictions. As in the previous section, we display the naming data as mode maps, showing terms used by a plurality of the speakers. The number of focus hits per color chip, aggregated over speakers in the language, are overlaid on top, for both empirical and predicted foci. We then provide example individual speaker-level analyses, presenting empirical and predicted focus data for two speakers of each language.
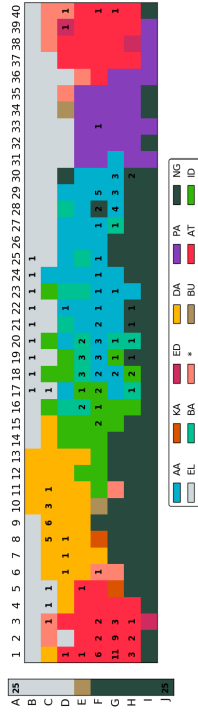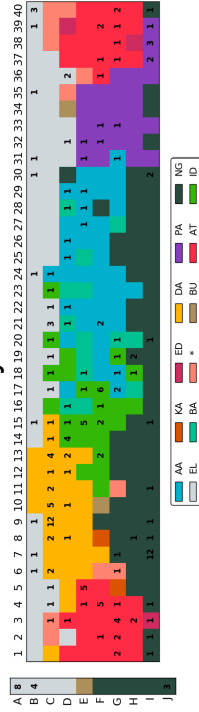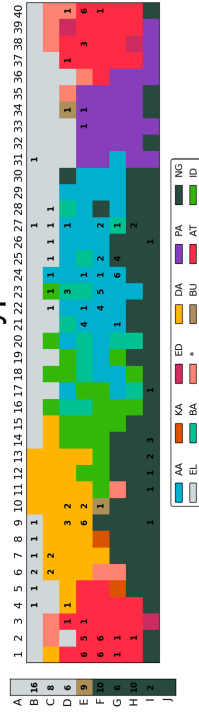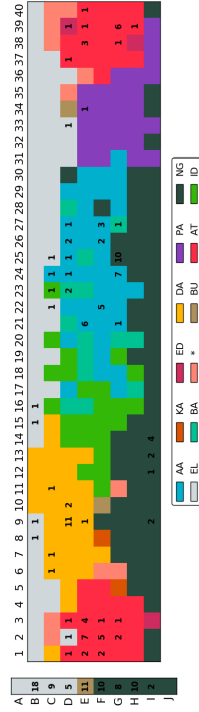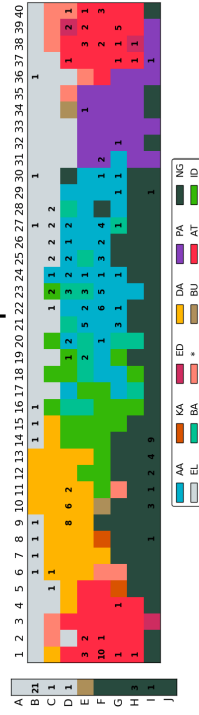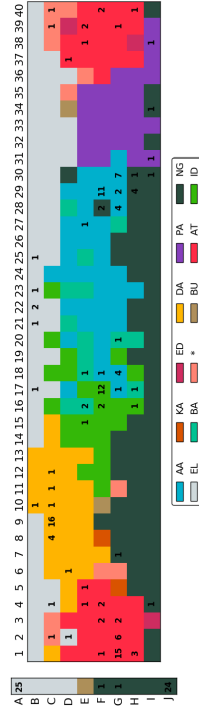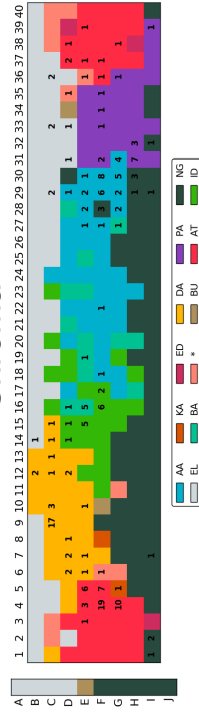
Figure 9: Model predictions for the Agta language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
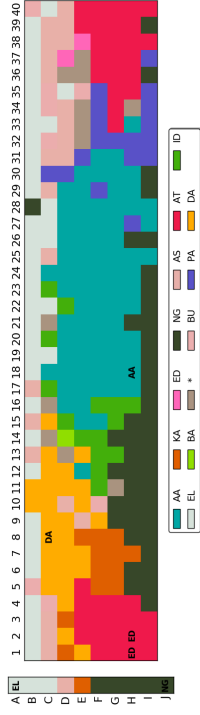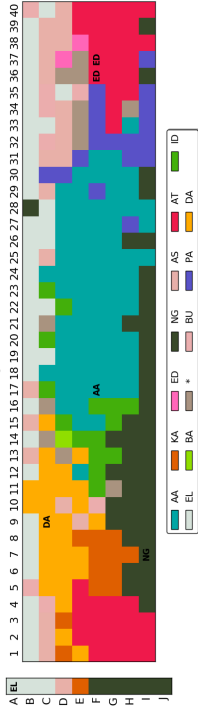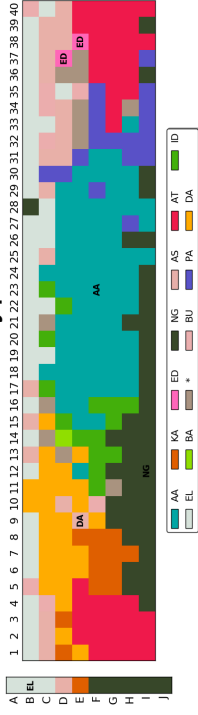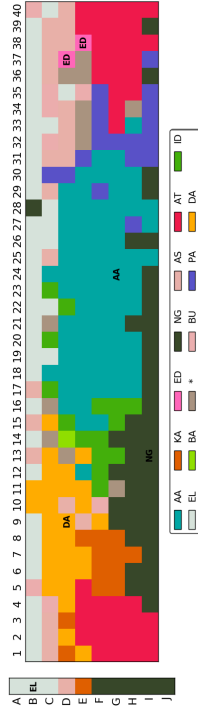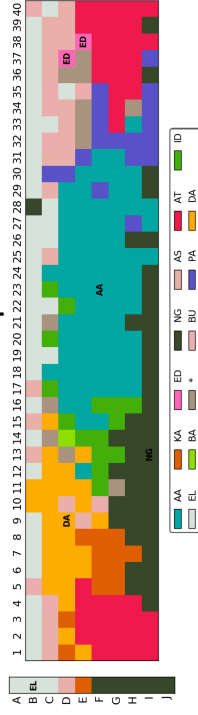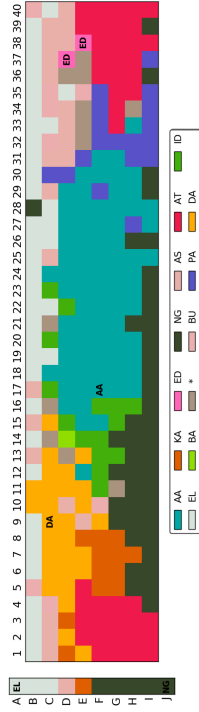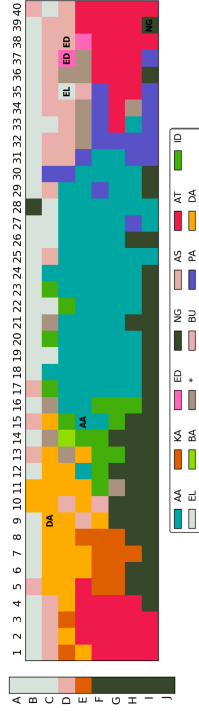
Figure 10: Model predictions for speaker 10 of the Agta language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
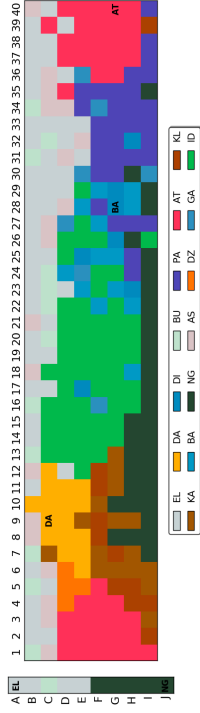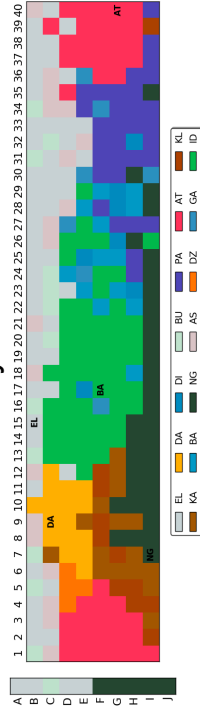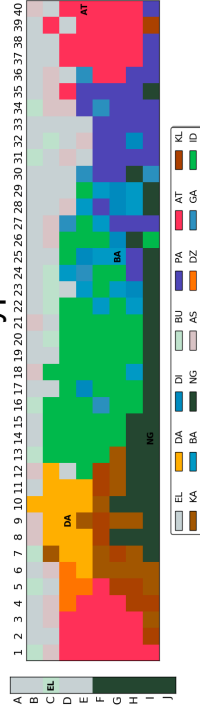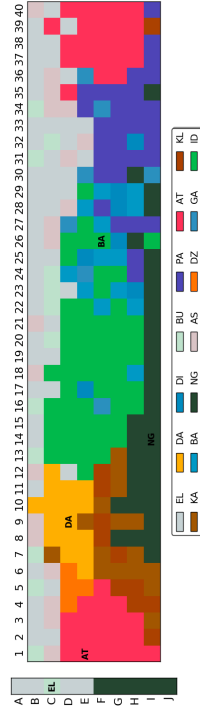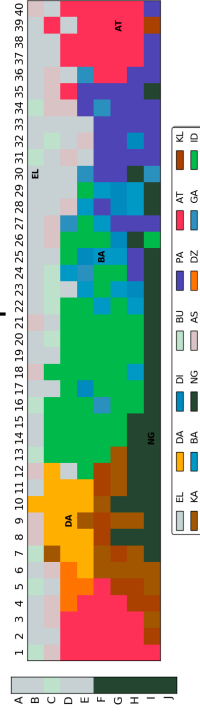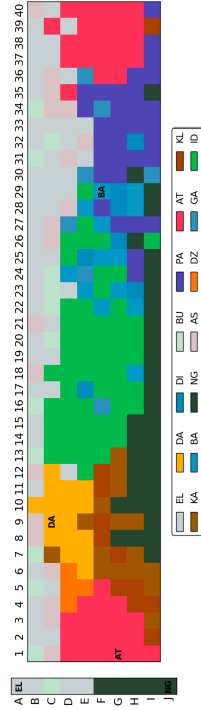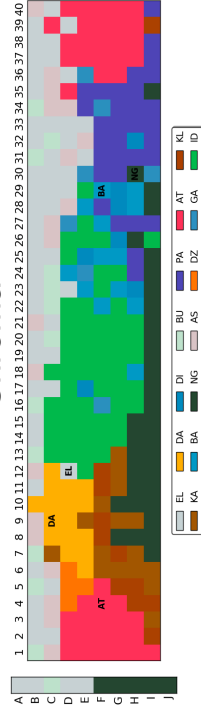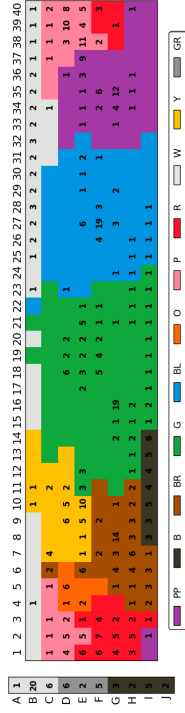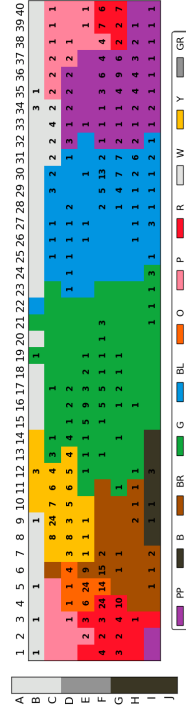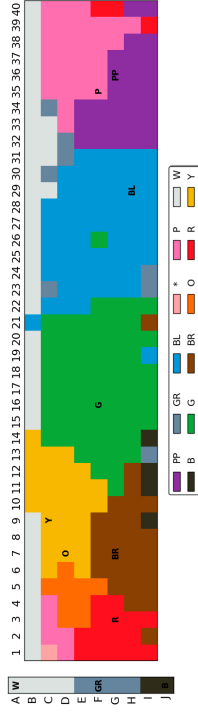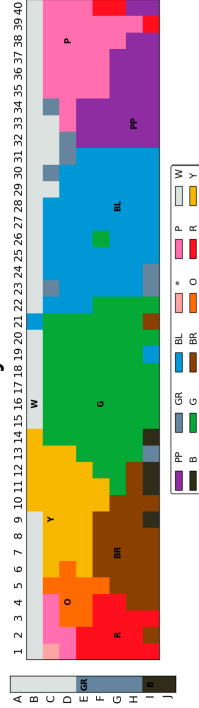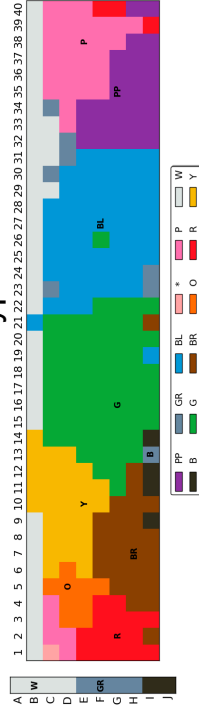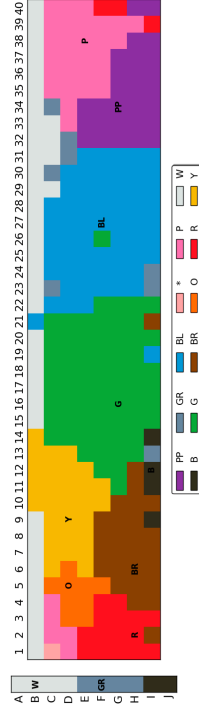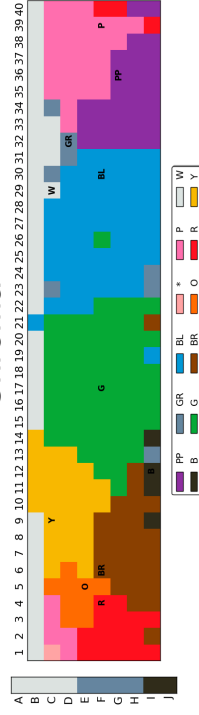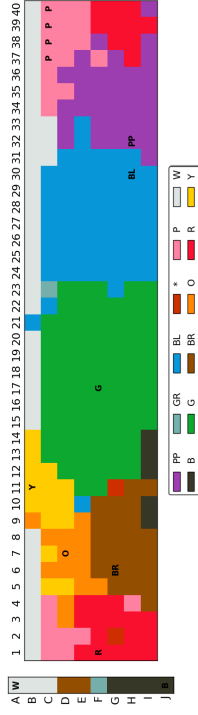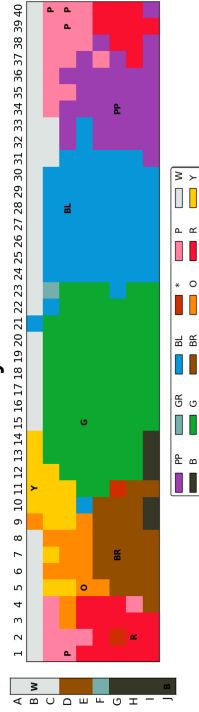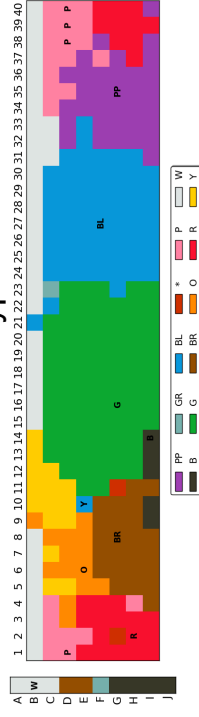
Figure 11: Model predictions for speaker 15 of the Agta language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
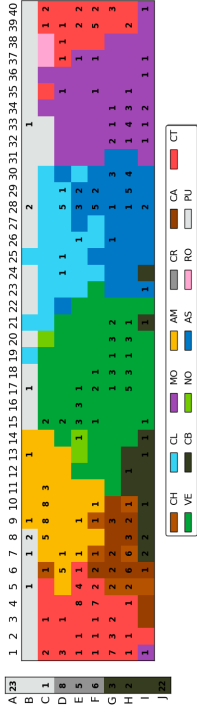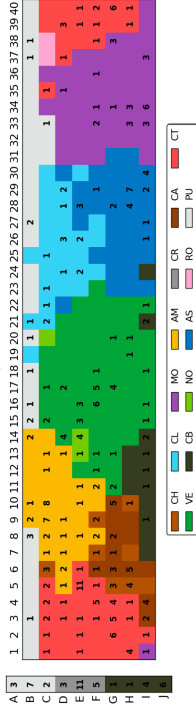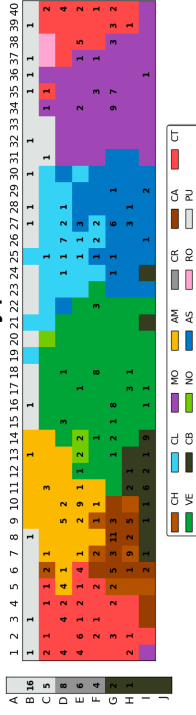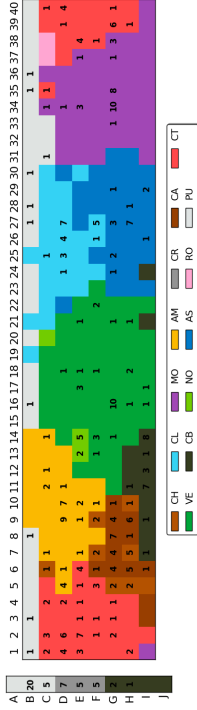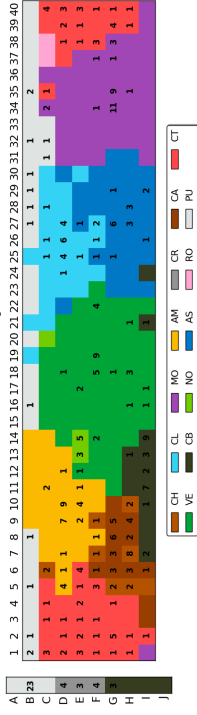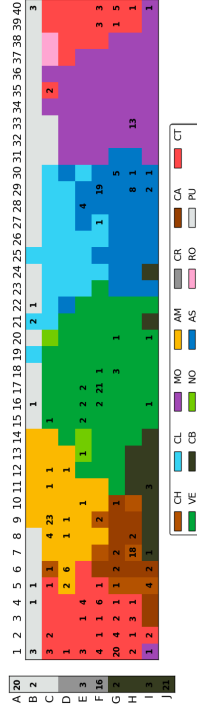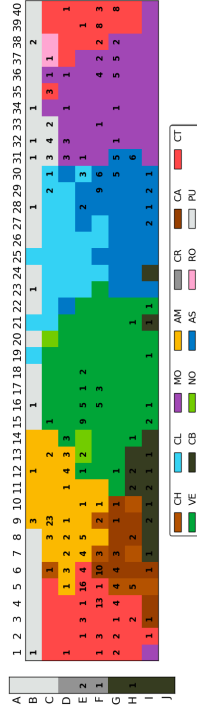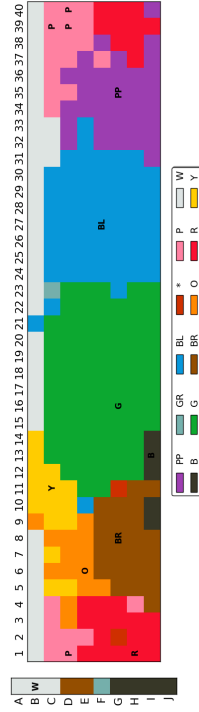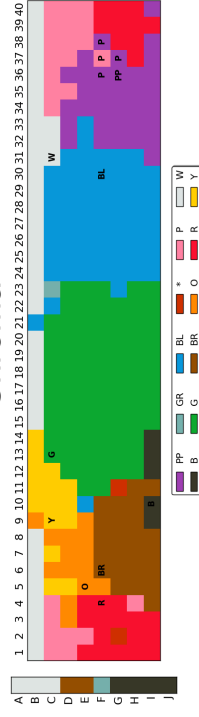
Figure 12: Model predictions for the Chiquitano language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
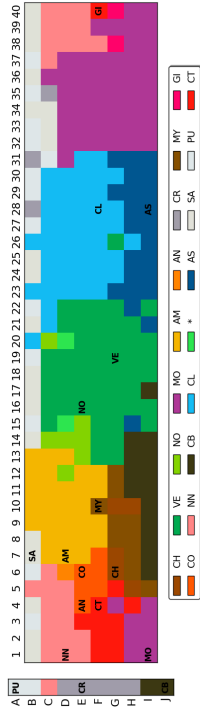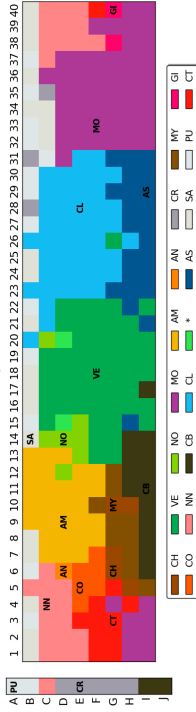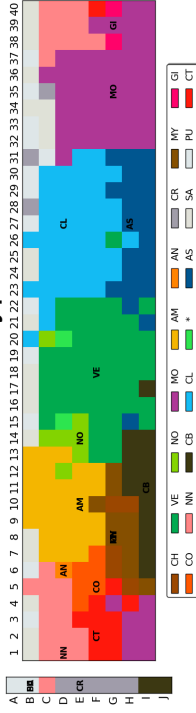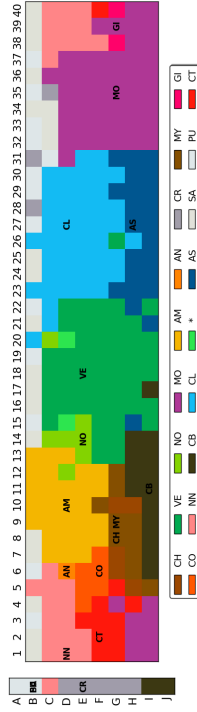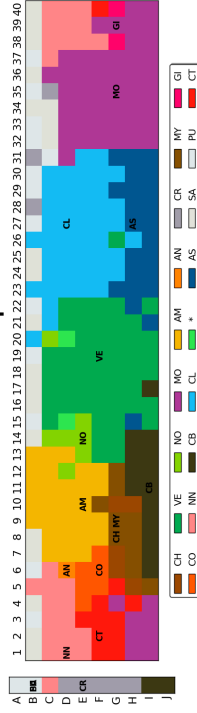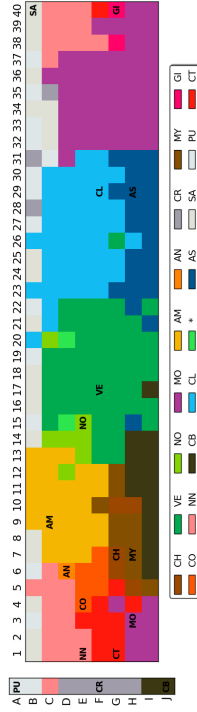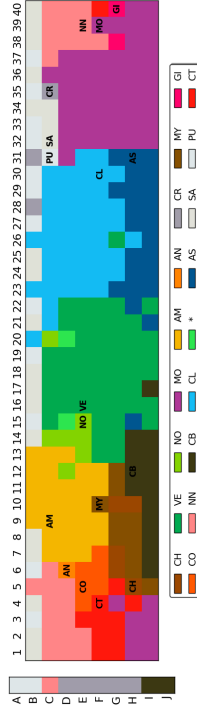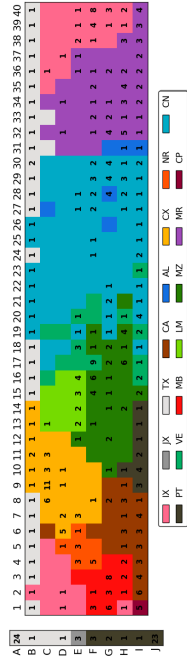
Figure 13: Model predictions for speaker 10 of the Chiquitano language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
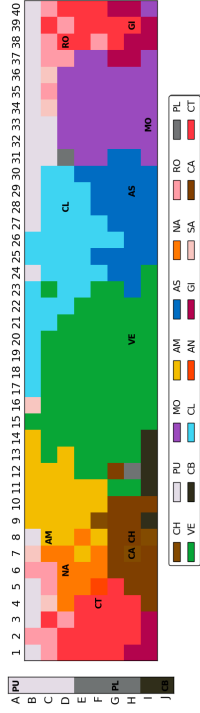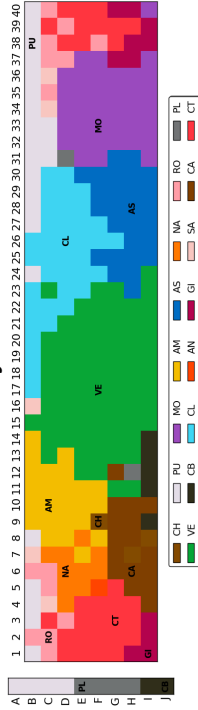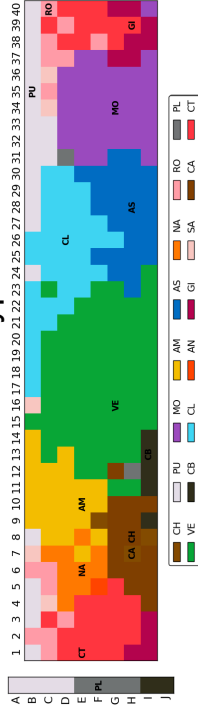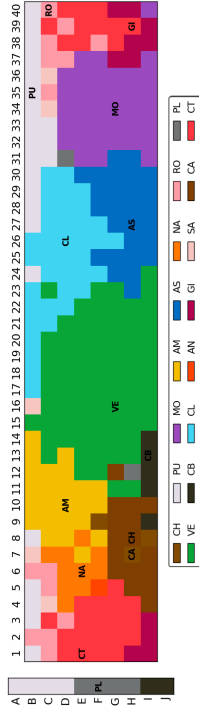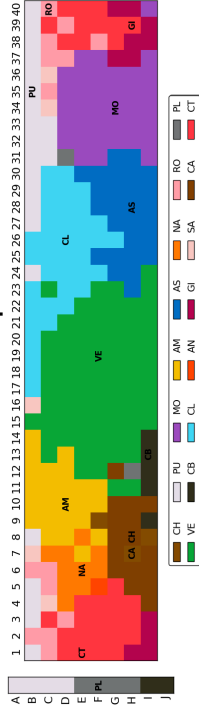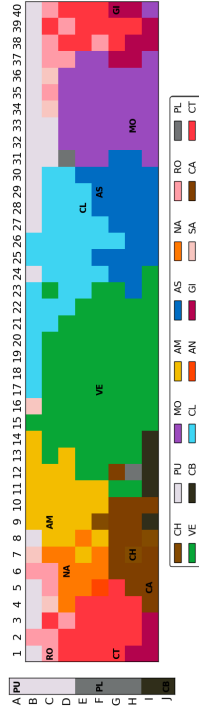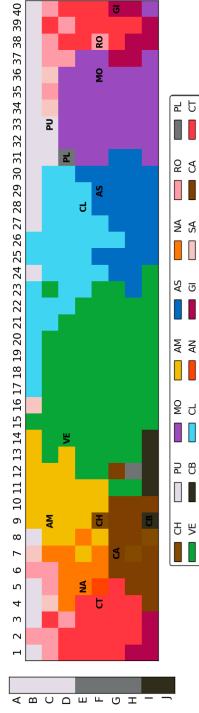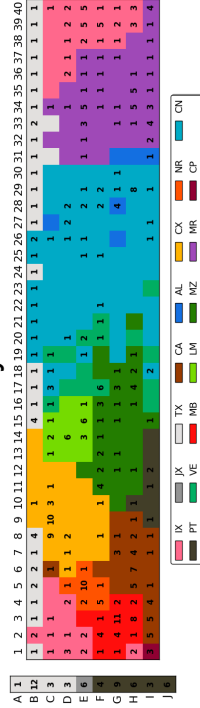
Figure 14: Model predictions for speaker 15 of the Chiquitano language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
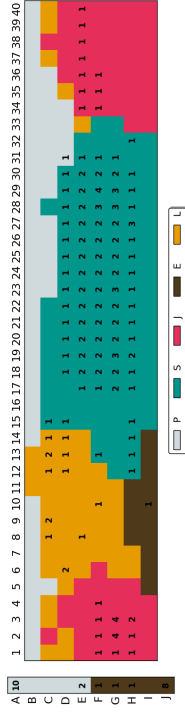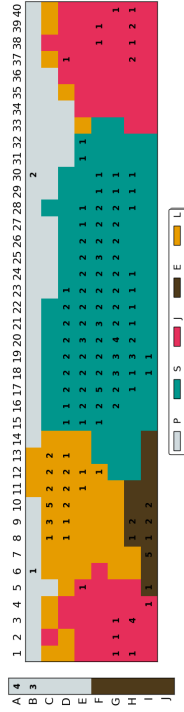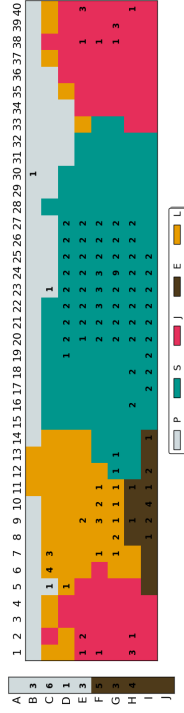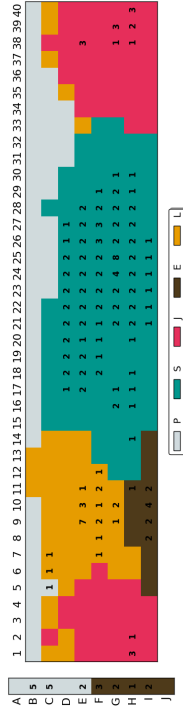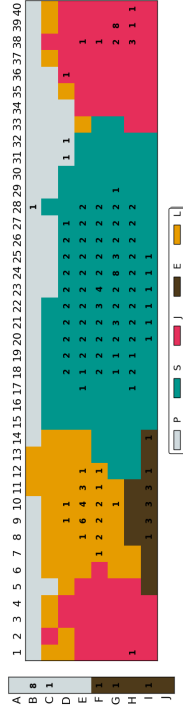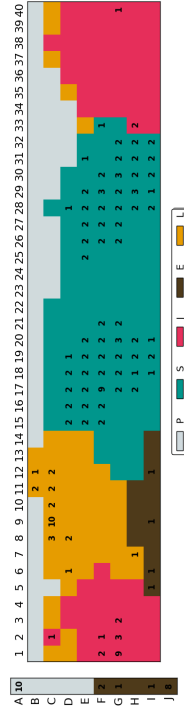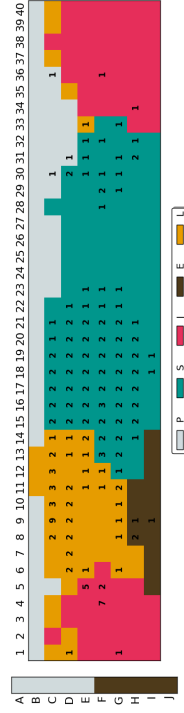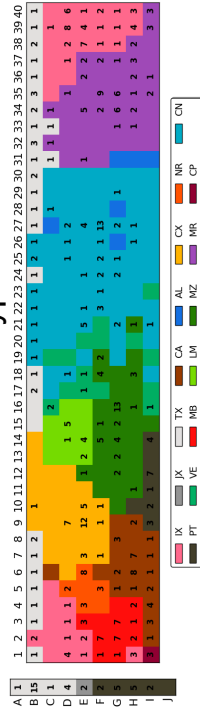
Figure 15: Model predictions for the Jicaque language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
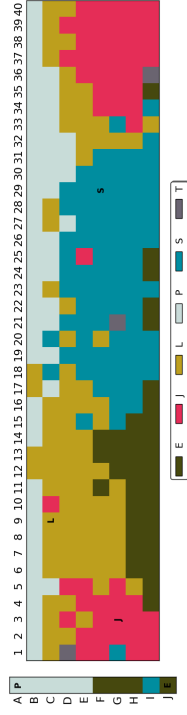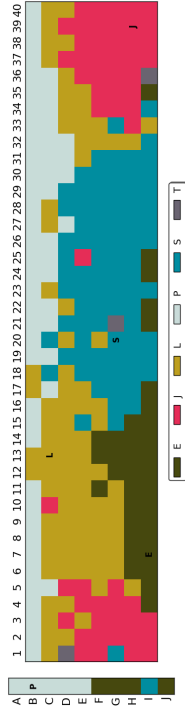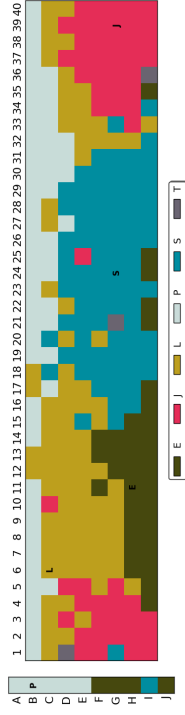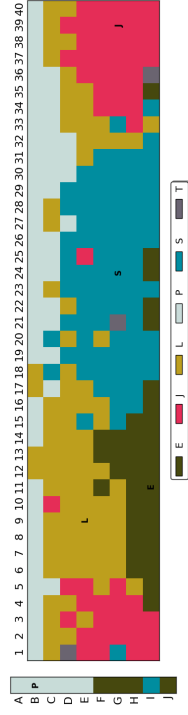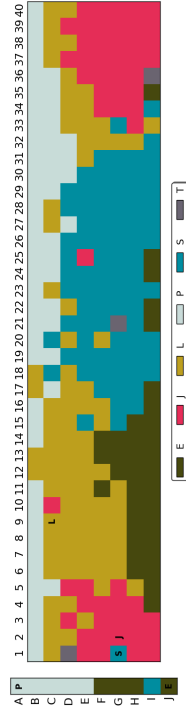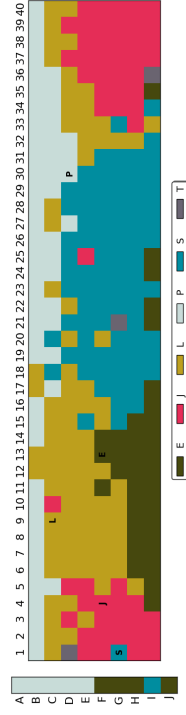
Figure 16: Model predictions for speaker 4 of the Jicaque language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
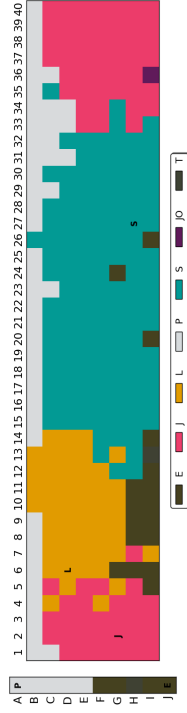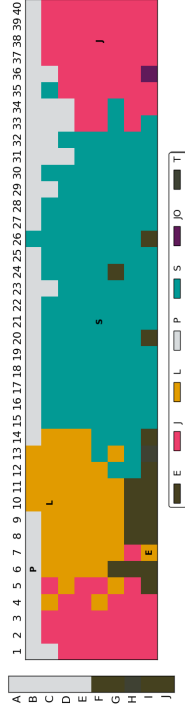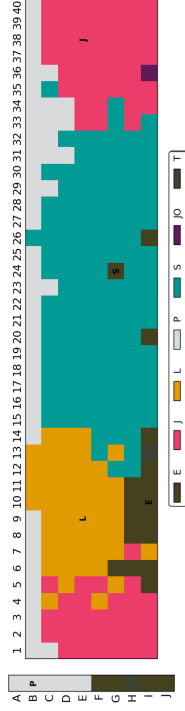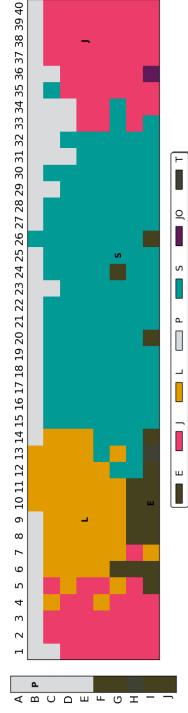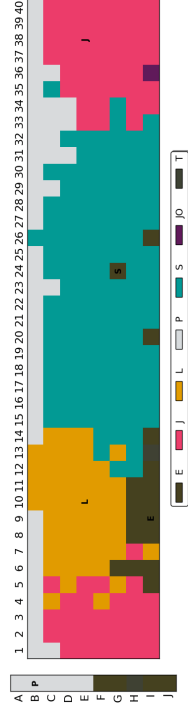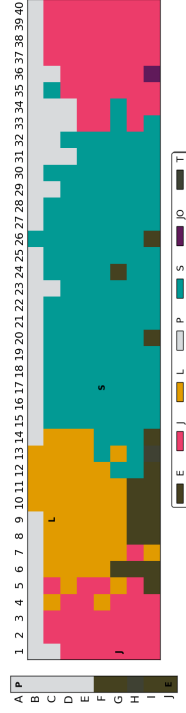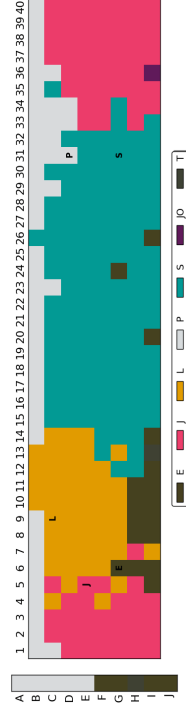
Figure 17: Model predictions for speaker 8 of the Jicaque language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
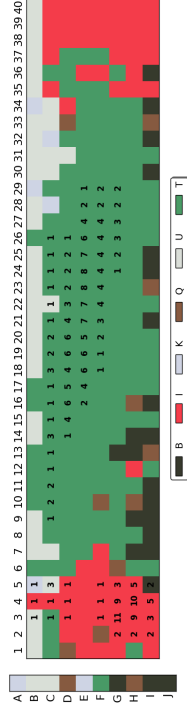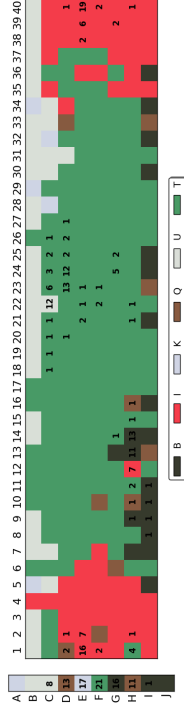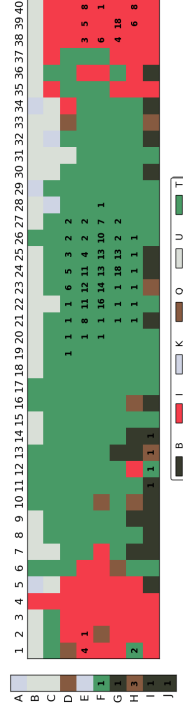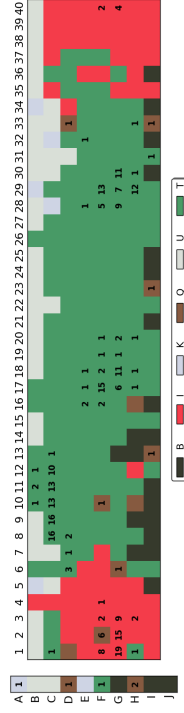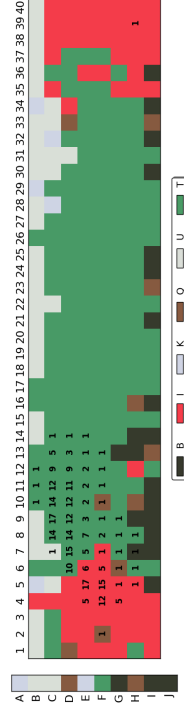
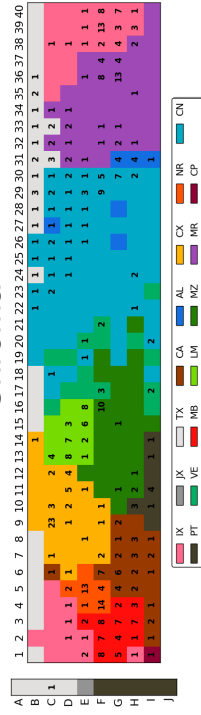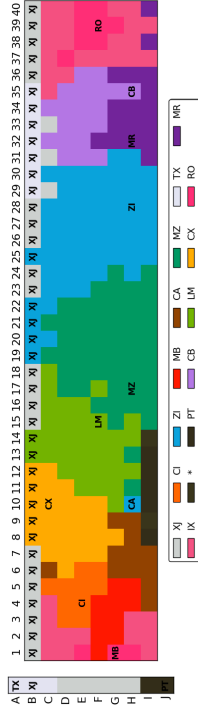Figure 18: Model predictions for the Karajá language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
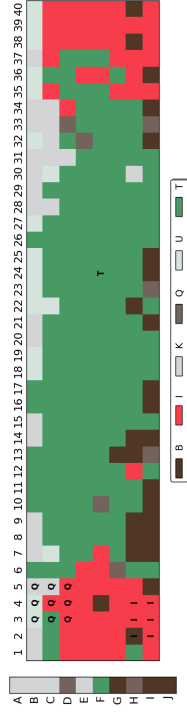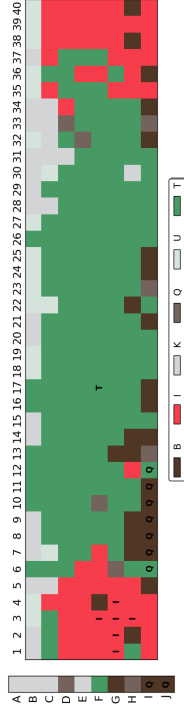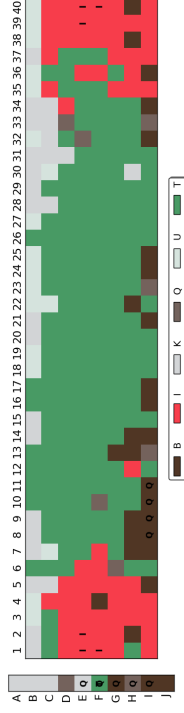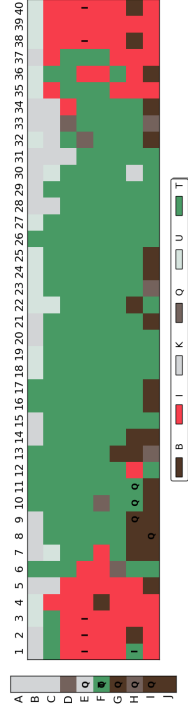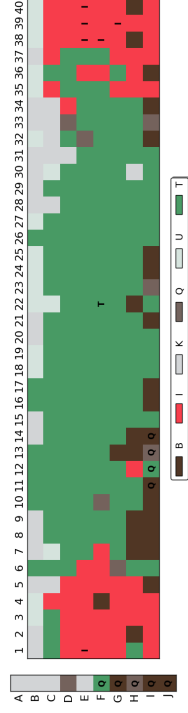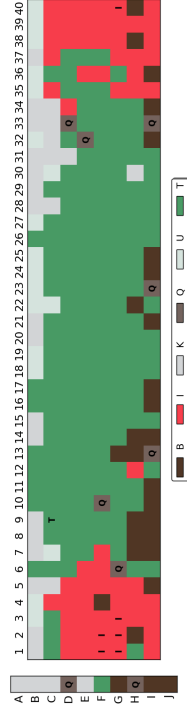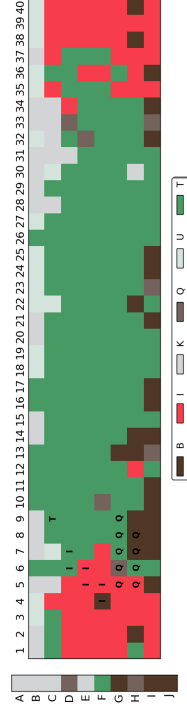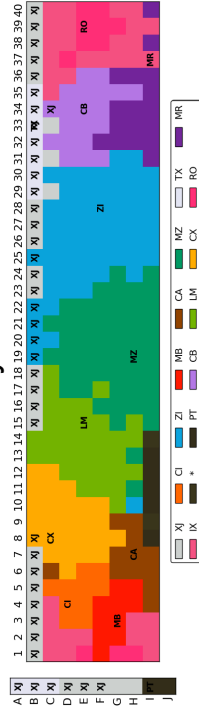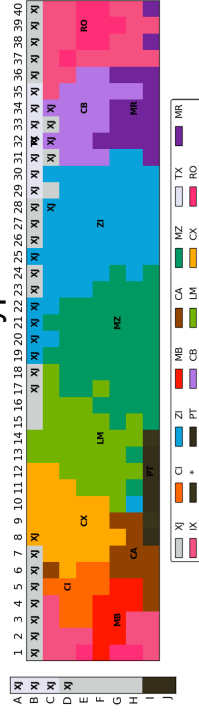
Figure 19: Model predictions for speaker 1 of the Karajá language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
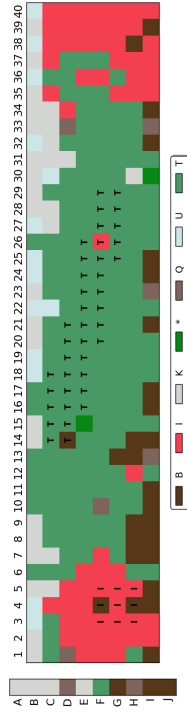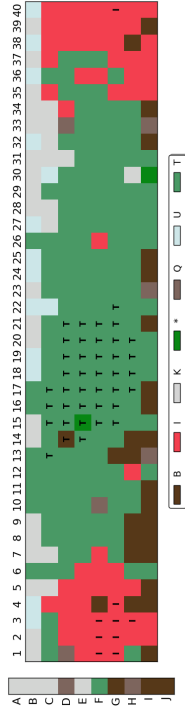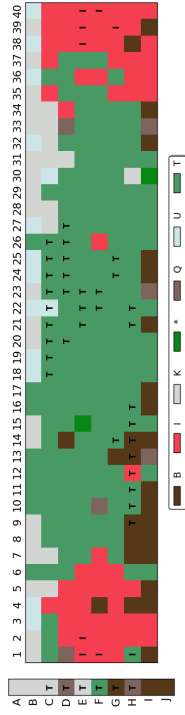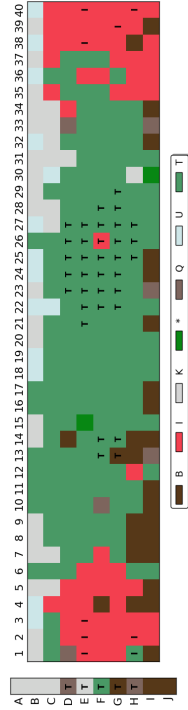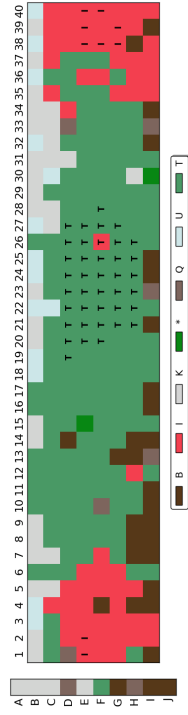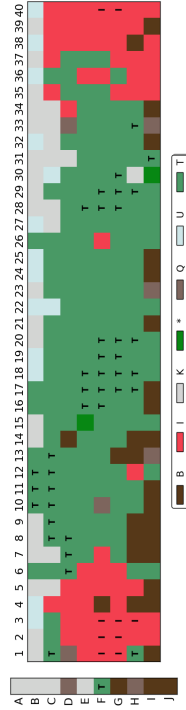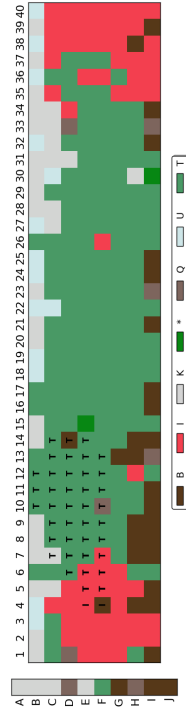
Figure 20: Model predictions for speaker 9 of the Karajá language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
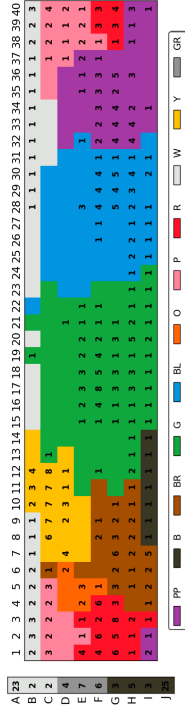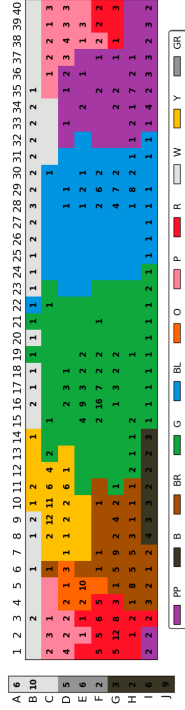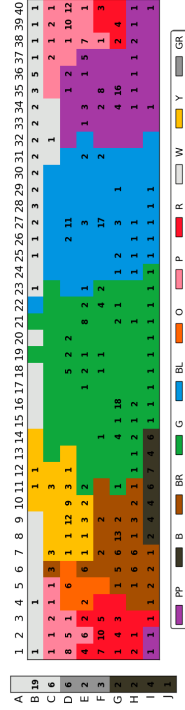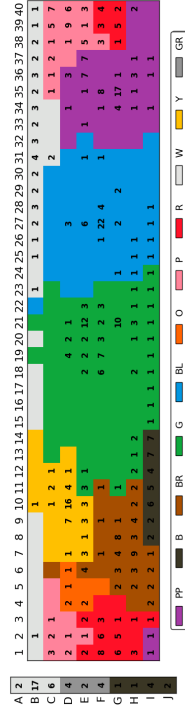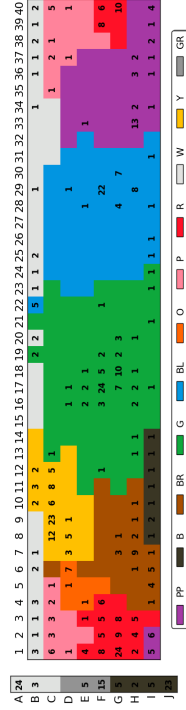
Figure 21: Model predictions for the Kriol language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
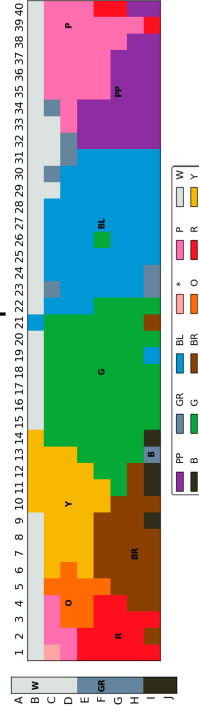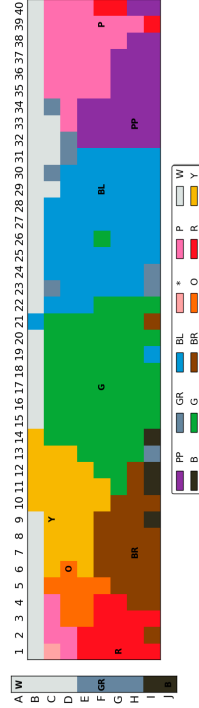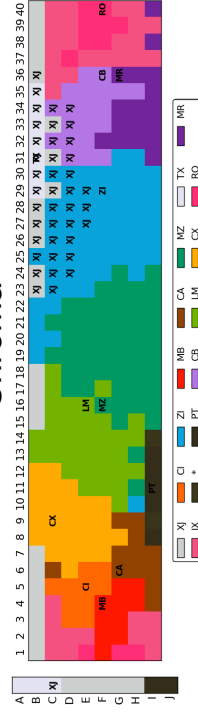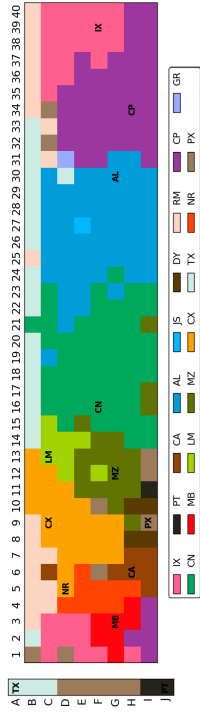
Figure 22: Model predictions for speaker 13 of the Kriol language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
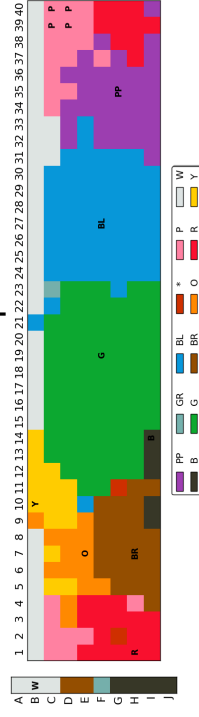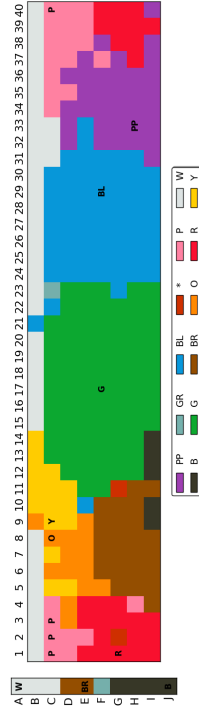
Figure 23: Model predictions for speaker 16 of the Kriol language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
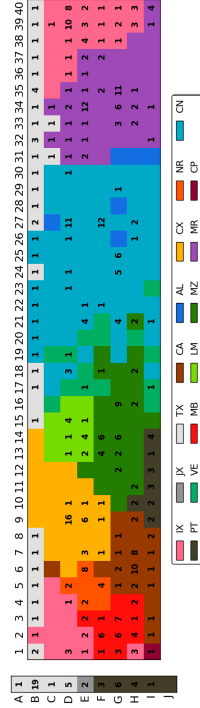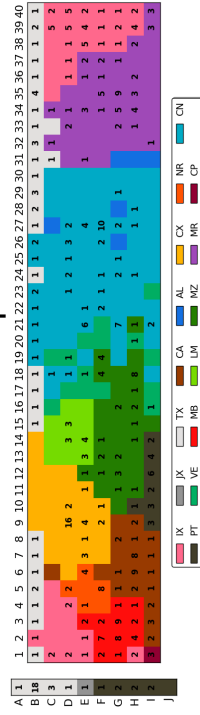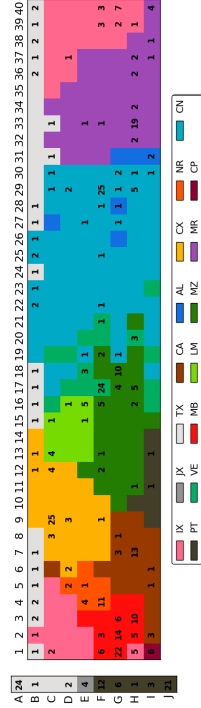
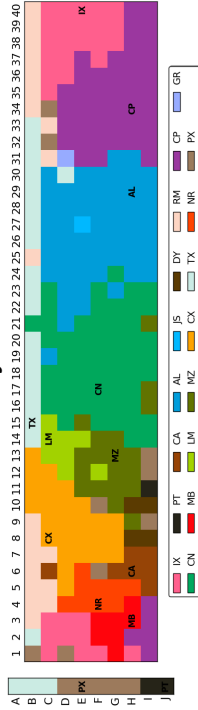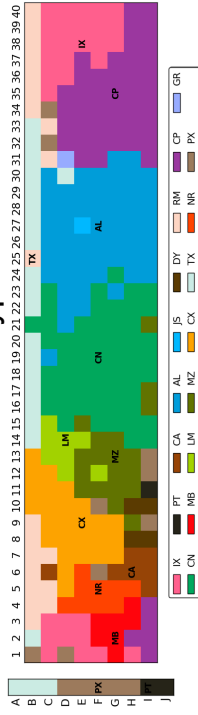Figure 24: Model predictions for the Mazahua language. Each figure displays the mode-map naming data overlaid with a particular focus distribution, aggregated over all speakers in the language.
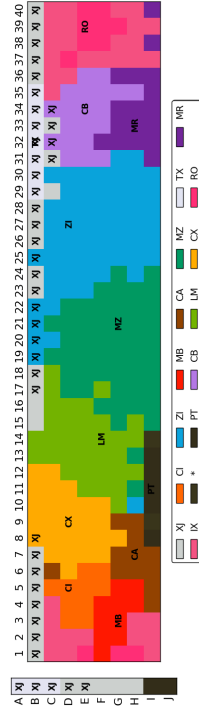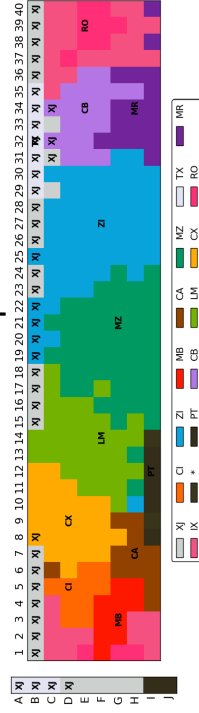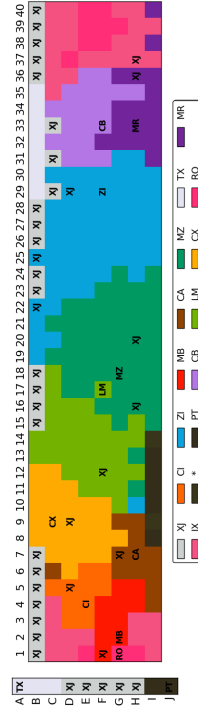
Figure 25: Model predictions for speaker 10 of the Mazahua language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.
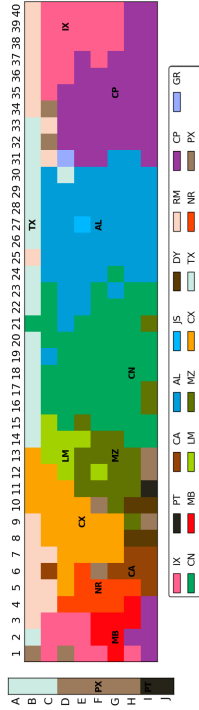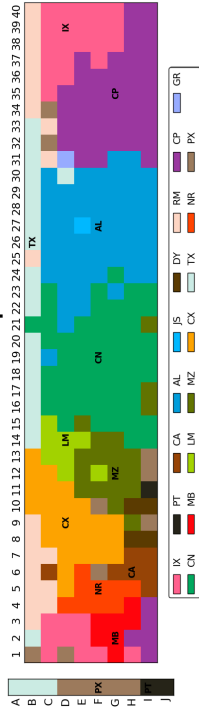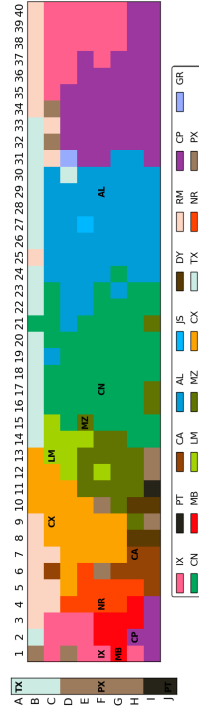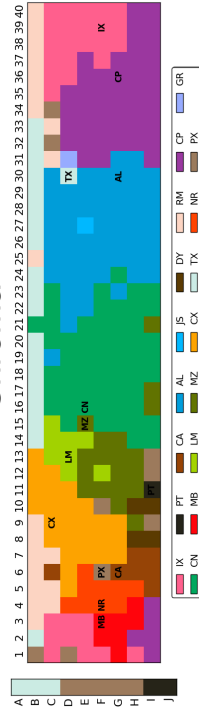
Figure 26: Model predictions for speaker 22 of the Mazahua language. Each figure displays this speaker's individual naming data overlaid with a particular focus distribution.