Complex and Social Networks
Assignment 4 - Non-linear regression on dependency trees

Daniel Benedí García
Fernando Martín Canfrán

January, 2024

Contents

1 Introduction

In the field of linguistics and natural language processing, syntactic dependency trees are a key element for understanding the grammar of sentences and documents. These intricate hierarchical structures, with nodes representing words or tokens and edges indicating syntactic dependencies, provide valuable insights into language structure and syntax. An example can be seen in Figure 1. An important part of dependency tree analysis is to examine the correlation between the number of nodes (words or tokens) in the tree and the average length of its edges. This exploration can uncover significant linguistic patterns and be applied to various tasks such as machine translation, sentiment analysis, and text summarization.

This assignment delves into the intriguing world of dependency trees by providing collections of syntactic dependency trees from diverse languages. The primary objective is to employ non-linear regression techniques

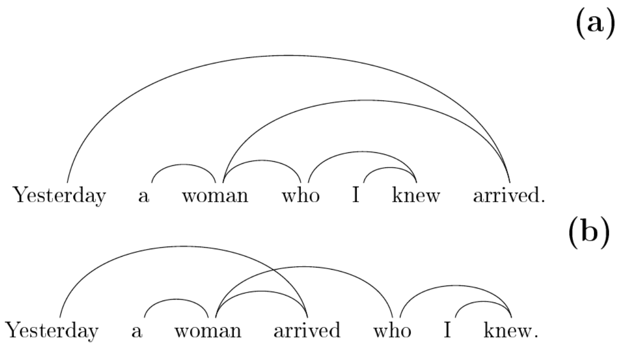


Figure 1: Example of dependency trees. (a) A sentence without crossings. (b) An alternative ordering yielding one crossing. Extracted from [1]

to model and understand the dependencies between the number of vertices, n, in a tree and the mean length, $\langle d \rangle$, of its edges.

Non-linear regression models offer a powerful toolset for addressing the complexities inherent in linguistic data. By fitting these models to our dependency tree data, we can capture and quantify the non-linear relationships that may exist between the number of vertices and the mean edge length. Throughout this assignment, we will examine various non-linear regression models, assess their suitability for the given data, and interpret the insights gained from these models.

2 Results

The results of the quality of the fit of each of the non-linear models tested for each of the languages are contained on Tables 8, 6 and 7. The quality of the fit of a model was measured with two types of metrics: s, residual standard error; and AIC (Akaike information criterion).

Both metrics are functions that quantify the discrepancy between the actual data and the model function. Consequently, the lower the value of these error functions, the more accurately a regression model fits. In particular, Table 8 contains the s, residual standard error values and table 6 the AIC values obtained for each of the models, in each of the languages.

Table 7 contains for each of the languages the differences in Δ AIC of each of the models with respect to the model that gives the best AIC.

Table 9 contains the values of the parameters that give the best fit for each model and for each of the languages.

Finally, for each of the languages two kind of plots were elaborated, one with the data samples and the different model curves and another with the data samples and only the curve of the selected as best model according to the error metrics. All these plots can be found in the Appendices A and B, respectively.

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	22.599	1.022	0.996	0.971	1.091	0.958	1.006	0.965	0.964	1.073	1.060
Basque	6.380	0.477	0.476	0.535	0.524	0.481	0.477	0.482	0.931	0.519	0.488
Catalan	16.624	0.424	0.417	0.445	0.482	0.404	0.419	0.411	0.475	0.472	0.521
Chinese	6.219	1.050	1.023	0.961	1.119	0.974	1.036	0.990	0.964	1.106	1.220
Czech	14.600	4.497	4.182	4.439	4.792	4.233	4.325	4.199	4.223	4.647	4.476
English	14.642	0.828	0.832	0.953	0.951	0.834	0.833	0.833	1.427	0.912	0.838
Greek	14.958	0.864	0.863	0.901	0.893	0.867	0.864	0.867	0.949	0.889	0.872
Hungarian	10.376	0.833	0.836	1.093	1.301	0.841	0.836	0.841	1.317	1.045	0.846
Italian	15.185	0.771	0.741	0.782	0.843	0.731	0.750	0.730	0.859	0.820	0.735
Turkish	8.885	0.380	0.384	0.505	0.403	0.378	0.383	0.384	0.854	0.397	0.378

Table 1: Residual standard error for each model

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	1090.844	348.789	343.564	337.472	364.490	335.264	346.055	337.043	336.620	361.473	360.419
Basque	276.863	59.999	60.872	70.595	67.913	62.666	61.012	62.701	118.112	68.111	64.666
Catalan	814.121	110.801	108.285	120.996	135.333	103.241	109.350	106.574	134.331	132.297	153.161
Chinese	274.712	126.245	125.080	119.838	131.616	121.833	126.102	123.254	121.014	131.641	141.704
Czech	723.596	517.324	505.521	516.025	528.527	508.621	511.453	507.224	508.202	524.104	519.399
English	724.102	219.578	221.388	245.239	243.884	222.683	221.509	222.513	317.254	237.543	224.499
Greek	711.355	221.851	222.721	230.065	227.511	224.541	222.881	224.493	240.015	227.840	226.488
Hungarian	610.872	203.171	204.834	248.298	275.524	206.772	204.811	206.811	279.376	240.936	208.752
Italian	705.677	199.994	194.329	203.435	215.157	192.881	196.287	192.775	220.399	211.555	194.727
Turkish	412.776	54.528	56.489	87.728	61.155	55.821	56.420	57.479	148.680	60.547	56.832

Table 2: Akaike information criterion (AIC) of each model

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	755.580	13.525	8.300	2.209	29.227	0.000	10.791	1.780	1.356	26.209	25.155
Basque	216.864	0.000	0.873	10.596	7.914	2.667	1.013	2.702	58.113	8.112	4.667
Catalan	710.880	7.560	5.044	17.755	32.092	0.000	6.110	3.333	31.090	29.056	49.920
Chinese	154.874	6.407	5.242	0.000	11.778	1.995	6.264	3.416	1.176	11.803	21.866
Czech	218.075	11.803	0.000	10.504	23.006	3.100	5.932	1.703	2.681	18.583	13.878
English	504.524	0.000	1.809	25.660	24.306	3.105	1.931	2.934	97.676	17.965	4.921
Greek	489.504	0.000	0.870	8.214	5.660	2.690	1.030	2.642	18.164	5.989	4.637
Hungarian	407.701	0.000	1.663	45.127	72.353	3.601	1.640	3.640	76.205	37.765	5.581
Italian	512.902	7.219	1.554	10.660	22.382	0.106	3.512	0.000	27.624	18.779	1.952
Turkish	358.249	0.000	1.961	33.200	6.628	1.293	1.892	2.952	94.152	6.019	2.304

Table 3: AIC differences

	Model																						
	1	2		3		4	5	1+		2+			3+			4+		5+					
Language	b	a	b	a	c	a	a	b	c	b	d	a	b	d	a	c	d	a	d				
Arabic	0.351	0.432	0.487	1.873	0.007	0.816	1.097	0.174	0.005	0.385	-0.477	0.013	1.129	1.596	8.336	0.003	-6.764	1.055	-0.977	1.481	0.177	0.002	-0.767
Basque	0.437	0.623	0.487	1.600	0.022	0.951	0.702	0.425	0.003	0.455	-0.154	0.409	0.575	0.372	19.900	0.002	-18.580	1.083	-0.412	0.744	0.410	0.003	-0.052
Catalan	0.357	0.630	0.409	1.800	0.009	0.817	0.938	0.251	0.004	0.373	-0.193	0.164	0.651	0.954	17.862	0.001	-16.367	0.940	-0.474	1.156	0.251	0.001	-0.512
Chinese	0.459	0.363	0.663	1.289	0.030	0.990	1.246	0.017	0.030	0.509	-0.490	0.002	2.031	1.637	0.335	0.054	1.299	1.285	-0.920	0.319	0.640	-0.004	0.308
Czech	0.525	0.044	1.170	2.495	0.011	1.298	0.000	2.659	-0.014	0.629	-2.611	0.026	1.274	0.584	55.278	0.002	-55.380	2.712	-5.369	0.000	2.580	-0.015	1.533
English	0.457	0.680	0.473	2.527	0.009	1.131	0.824	0.402	0.002	0.460	-0.051	0.322	0.609	0.795	40.647	0.001	-38.789	1.453	-1.216	0.243	0.677	-0.001	0.950
Greek	0.364	0.619	0.420	1.805	0.010	0.829	0.718	0.360	0.001	0.383	-0.216	0.322	0.536	0.553	14.571	0.002	-13.241	0.969	-0.528	0.245	0.604	-0.001	0.676
Hungarian	0.594	0.615	0.612	2.948	0.015	1.715	0.577	0.637	-0.001	0.599	-0.115	0.664	0.598	-0.123	37.671	0.002	-36.046	2.620	-3.342	0.472	0.689	-0.001	0.203
Italian	0.373	0.454	0.503	1.832	0.010	0.845	0.749	0.318	0.004	0.409	-0.449	0.079	0.837	1.065	47.040	0.001	-45.684	1.085	-0.899	0.149	0.677	0.001	0.893
Turkish	0.413	0.735	0.419	1.810	0.015	0.924	0.539	0.568	-0.006	0.417	-0.037	1.613	0.286	-1.189	11.701	0.003	-10.423	1.027	-0.345	0.121	1.006	-0.015	0.767

Table 4: Parameters of each fitted model

3 Discussion

3.1 Differences of the fit between null model and alternative models

After having fitted all of the models for each of the languages, one of the aspects that we can be wondering about is whether these more complicated alternative models show any difference with respect to the basic linear model: $\langle d \rangle = f(n) = \frac{n}{3} + \frac{1}{3}$ proposed as null model.

By visually checking the fit of all the models together in the appendix A, it can be observed that for all of the languages we can observe a significative difference between the fits of the null hypothesis model (painted in gray color) and the alternative models. The null hypothesis model overestimates by far the value of the function $\langle d \rangle$ for all the n values, except when $n = 2$, that $\langle d \rangle = 1$, which holds for all the languages.

Similarly, when analyzing the error metric tables, both in the AIC (Tables 6 and 7) and in the residual standard error case (Table 8), the same phenomena can be seen for all of the languages. In the case of the residual standard error metric it was obtained around one order of magnitude more of error value for the null hypothesis model with respect to the alternative models errors. Similarly, way larger values of the AIC metric error were obtained for the null hypothesis model.

3.2 Best function fit discussion

One of the questions that arises when visually checking the model plots and the fit of the selected as best model is if this is really the function that provides the best fit for the data.

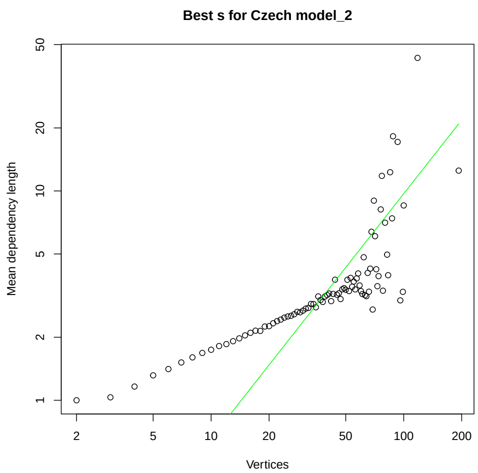
It is clear that in terms of the error metrics that were proposed, it is the best fit. But when visually checking, for some of the languages, it can be seen that there are other models that apparently could provide a better fit.

Examining Figure 2, one could think that the model is not accurately fitted because the model is far from the mean length when the number of vertices is low (Figure 2a). This can be attributed to the fact that the densities of entries for the number of vertices are not uniform, thus making it more likely to fit better in areas with a higher concentration of entries. Moreover, when the number of vertices is high, the error is greater than when it is low, so our method will attempt to reduce the error for a larger number of vertices.

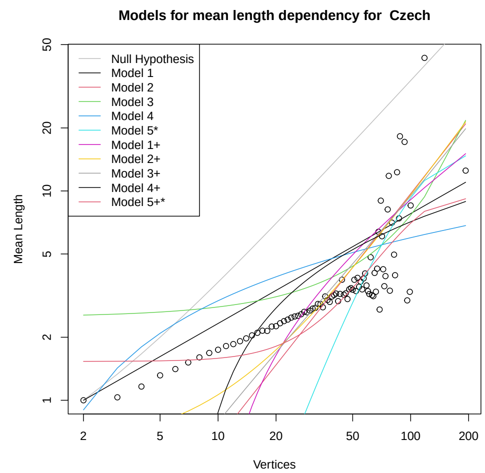
There could several reasonable explanations: either none of the models are good to explain the relation between the number of vertices and the mean length or our optimising method is not appropriate. In Appendix E, we show the results of a possible solution in which we weight the samples with the following function $\omega(n) = \frac{1}{n}$. We observe that all the models fit the lower values of the number of vertices better; however, the model with the smallest AIC for Czech still has the same issue, though there is an improvement.We were unable to determine if our weight function $\omega(n)$ was suitable or if we still lacked an adequate model. Our intuition, however, suggests that the weight function could be further enhanced.

3.3 Extent to which languages resemble or differ

When visually checking the best models in the appendix B, we could separate into two groups: those languages for which the model resembles the behaviour and those for which it differs. The models resemble to Basque, English, Greek, Hungarian, and Turkish. The models differ for Arabic, Catalan, Chinese, Czech, and Italian. We could observe the fact that was mentioned in the previous section that the models fail to resemble the



(a) Model with smallest AIC for Czech.



(b) Models fitted for language Czech.

Figure 2: Models fitted for language Czech

languages when the mean length is really spread for large values of number of vertices. We tried to explain this split using the table 5, but there is no clear explanation.

Language	N	μ_n	σ_n	$\mu_{\langle d \rangle}$	$\sigma_{\langle d \rangle}$
Arabic	4099	26.997	20.655	2.168	0.932
Basque	2932	11.336	6.529	1.961	0.691
Catalan	15052	25.573	13.618	2.319	0.702
Chinese	54198	6.247	3.311	1.445	0.484
Czech	25015	16.434	10.724	2.020	0.875
English	18779	24.046	11.223	3.051	0.895
Greek	2951	22.820	14.382	2.201	0.813
Hungarian	6424	21.660	12.566	3.878	1.780
Italian	4142	18.411	13.347	1.972	0.768
Turkish	6016	11.104	8.290	1.843	0.820

Table 5: Summary of the properties of the degree sequences

3.4 Conclusions

To summarize, it is concluded that the alternative hypothesis models are a good approach to fit the $\langle d \rangle$ distribution, in comparison with the null hypothesis model.

In addition, it was observed that model 1 was the one providing the best fit for many languages, in particular for half of them, this is due to the fact that the distribution of the values for most of them is linear in the double logarithmic scale, which is what the model 1 and model 2 approach better.

Finally, we suggest a further improvement of the weighted function approach described in 3.2.

4 Methods

4.1 Checking the validity of the metrics

In order to be considered valid, an initial validity check of the data samples was performed, so that only samples satisfying these two conditions were taken:

$$4 - 6/n \leq \langle k^2 \rangle \leq n - 1$$

and

$$\frac{n}{8(n-1)}\langle k^2 \rangle + \frac{1}{2} \leq \langle d \rangle \leq n - 1$$

4.2 Selection of initial values for the non-linear regression models

For a certain language, the corresponding parameters for each of the models were obtained using the `nls` R non-linear regression optimiser, for which some initial values of these parameters had to be provided. The selection of the initial values for a certain model adjustment can be really important to be able to obtain a solution or, if obtained, to obtain it faster.

As was described in the statement, to obtain good initial values for a certain model, we applied a double logarithmic transformation and then a linear regression on that transformation. This was the method followed for the "non-additive" models (1, 2, 3 and 4). However, for the "additive" models (1+, 2+, 3+ and 4+) this same approach had to be slightly modified, due to the presence of the additive term, which makes it not possible to apply linear regression on the double logarithmic transformation. For that, the approach followed for these models consisted of giving an initial estimate value for the additive term (d_0), and then applying linear regression on the double logarithmic transformation of $f(n) - d_0$ and the rest of the model without the additive term (*note that in our case* $f(n) = \langle d \rangle$).

An example of the initial values acquisition process for model 2+ ($f(n) = an^b + d$) is provided here.

1. Take an initial estimate for d, d_0. For example, $d_0 = min(\langle d \rangle) - \epsilon$.
2. Rewrite the model as: $f(n) - d_0 = an^b$.
3. Apply the double logarithmic transformation: $\log(f(n) - d_0) \approx \log(an^b) = \log a + b \log n$
4. Apply linear regression on $\log(f(n) - d_0) \sim \log n$.
5. From the linear regression, we obtain $\log a$ as the intercept and b as the slope of that model. Therefore, we have the initial values $a = e^{\log a}$, b, and $d = d_0$.

4.3 Homocesdasticity Assumption

The non-linear regression method is predicated on the assumption of homoscedasticity within the $\langle d \rangle$ data samples. Homoscedasity is the homogeneity of variance, which signifies uniform variances among the data samples. We can observe plotting the data samples (Appendix D) that the data do not have this property. To properly validate this assumption, a Breusch-Pagan test was performed on the linear regression model involving the variables $\langle d \rangle$ and n.

For all languages under examination, it was established that the homoscedasticity assumption did not hold. Consequently, the values of $\langle d \rangle$ were aggregated for each number of vertices, denoted as n, using the R function `aggregate`. With this processed data set, the non-linear regression R method `nls` could be effectively employed, even in cases where the homoscedasticity assumption was not satisfied.

4.4 Convergence problem of model 3+

Despite using the method explained in Section 4.2, we observed that the `nls` optimiser had problems stabilising the solution which minimises the standard residual error for the languages.

In order to understand the fundamental reason of the problem, we could apply how non-linear least squares (NLS) works. NLS tries to minimise the sum of residuals, which happens when the gradient is equal to zero. To obtain the gradient, we compute the partial derivative of our sum of residuals.

$$
\begin{aligned}
S &= \sum_{i=1}^{m}(y_i f(x_i; a, c, d))^2 = \sum_{i=1}^{m}(y_i(ae^{cx_i} + d))^2 \\
\nabla S &= \begin{pmatrix} \frac{\partial S}{\partial a} \\ \frac{\partial S}{\partial c} \\ \frac{\partial S}{\partial d} \end{pmatrix} \\
\frac{\partial S}{\partial a} &= 2\sum_{i=1}(y_i - ae^{c_x i} - d)(-e^{cx_i}) \\
\frac{\partial S}{\partial c} &= 2\sum_{i=1}(y_i - ae^{c_x i} - d)(-ax_i e^{cx_i}) \\
\frac{\partial S}{\partial d} &= 2\sum_{i=1}(y_i - ae^{c_x i} - d)(-1)
\end{aligned}
$$

One could notice that the roots of our gradient are not easy to obtain and that there are many. These factors can give insight to better understand the instability of our solver and why it does not converge to a stable solution. In fact, we can observe that $d = 0$ helps converge and that is the reason why Model 3 does not have this problem.

Our proposed solution uses the fact that, although it does not converge, every iteration it reduces the sum of residuals. Therefore, we will accept the last solution found as if it were optimal. We show that the solution is suboptimal with an asterisk in the legend of the figures in the appendix A. It is remarkable that there are some languages for which `nls` was able to converge, we could not find a reasonable explanation.

References

[1] R. Ferrer-i-Cancho. "A stronger null hypothesis for crossing dependencies". In: *EPL (Europhysics Letters)* 108.5 (Dec. 2014), p. 58003. DOI: `10.1209/0295-5075/108/58003`. URL: `https://doi.org/10.1209%2F0295-5075%2F108%2F58003`.

A Graphs of all the models for each language

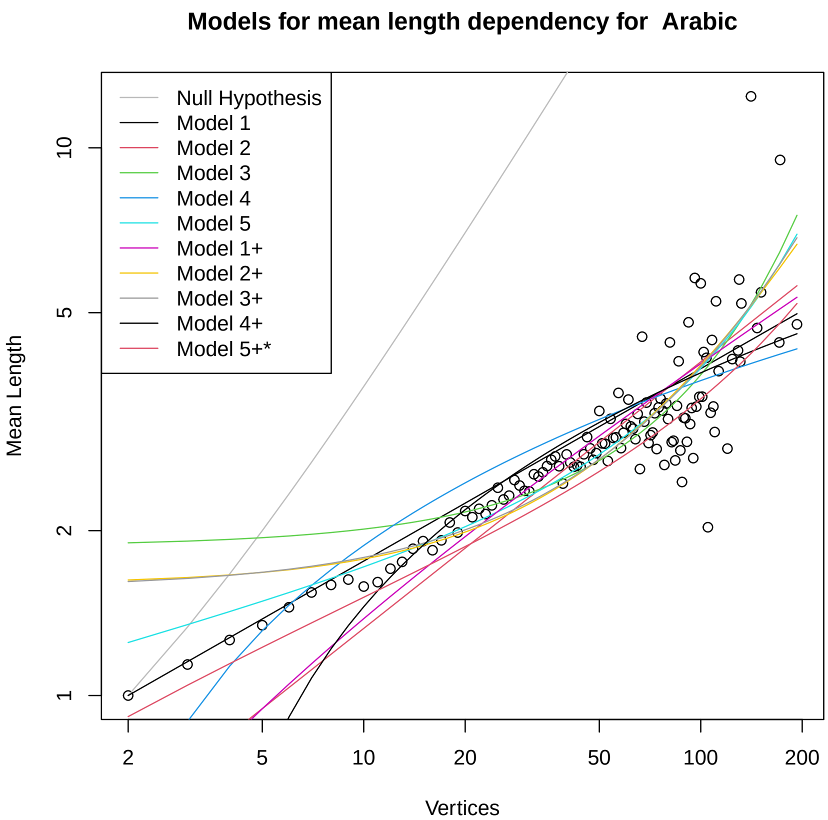


Figure 3: All models fitted for Arabic

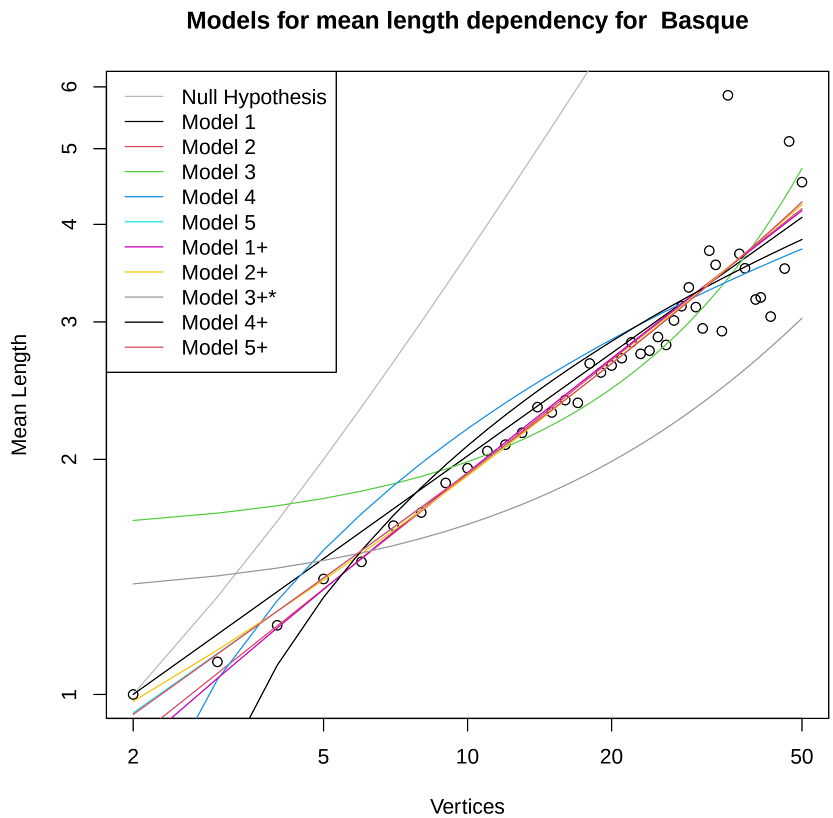


Figure 4: All models fitted for Basque

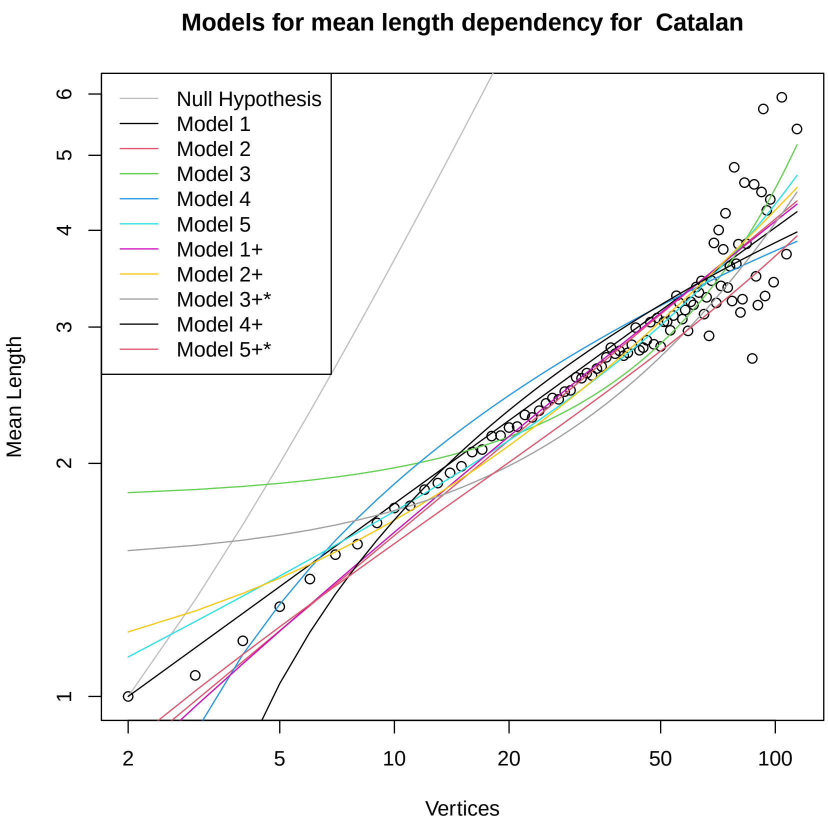


Figure 5: All models fitted for Catalan

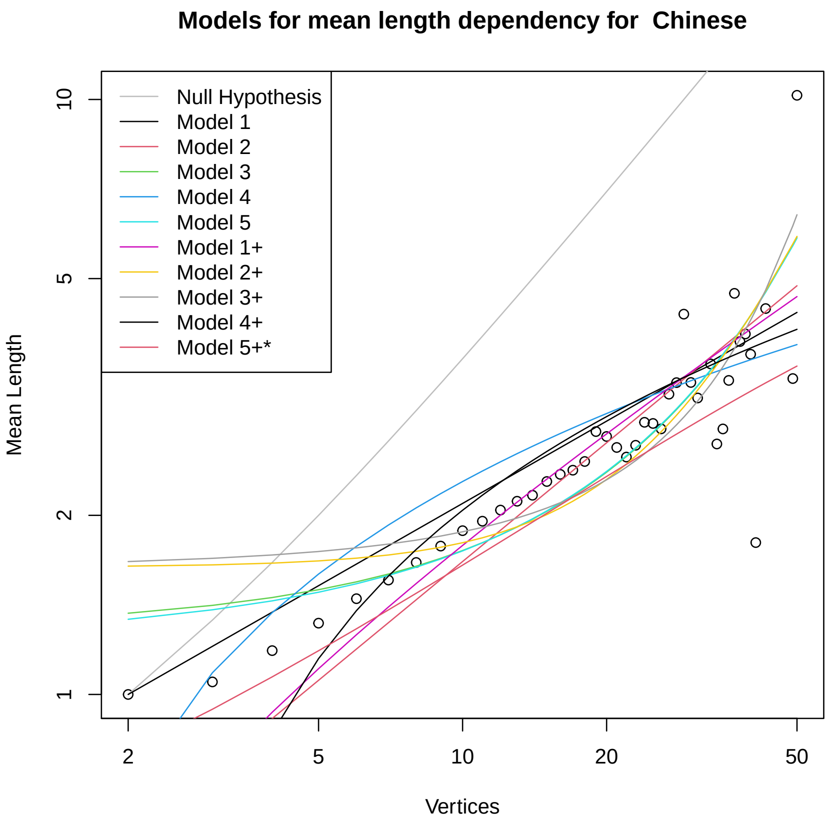


Figure 6: All models fitted for Chinese

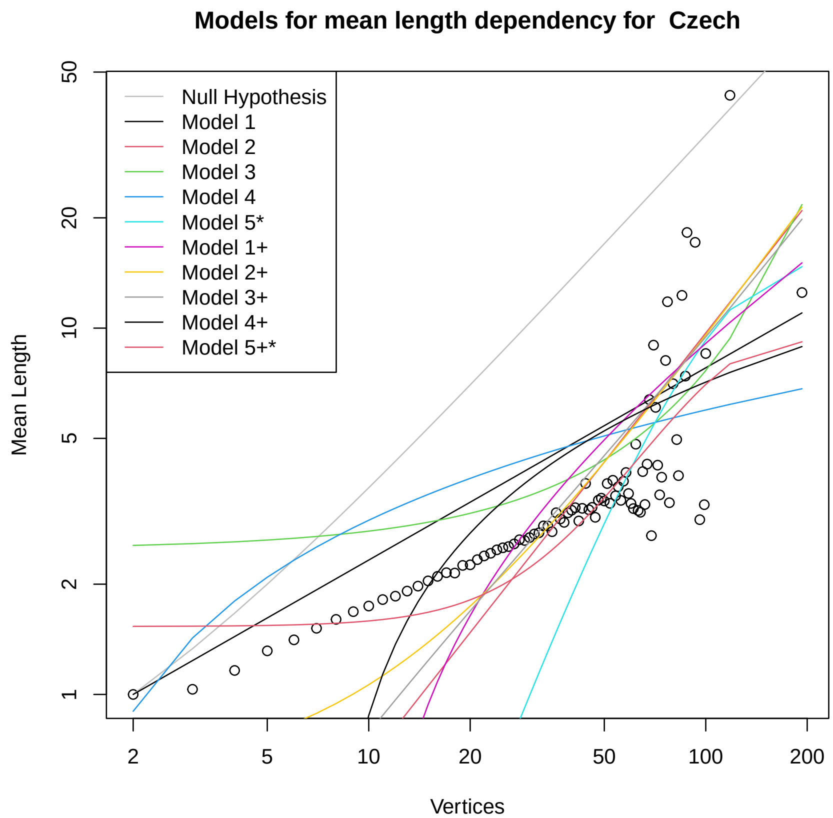


Figure 7: All models fitted for Czech

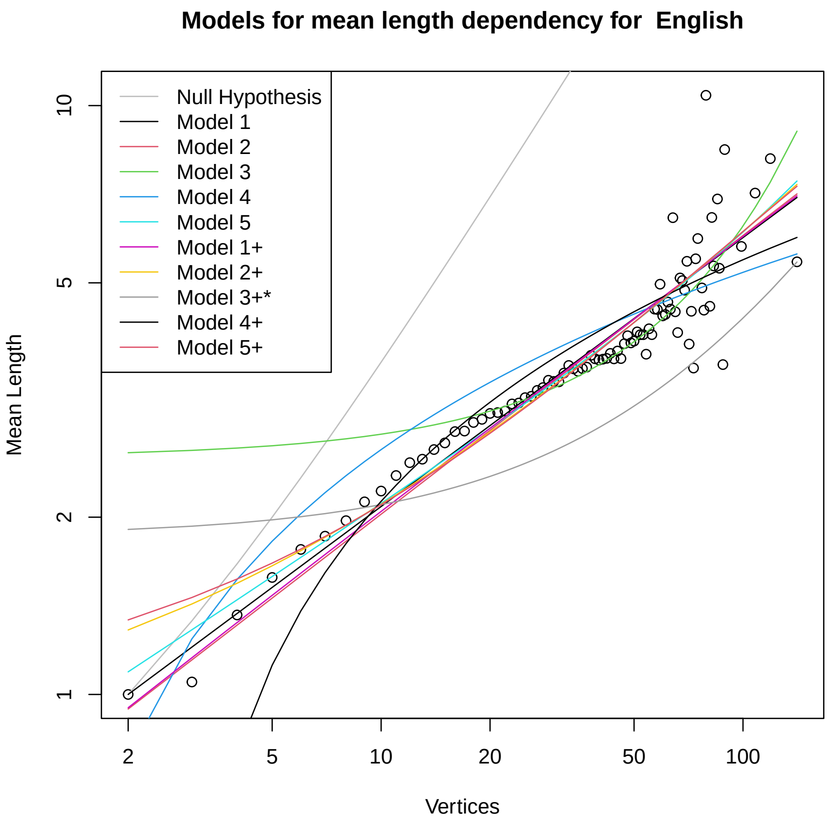


Figure 8: All models fitted for English

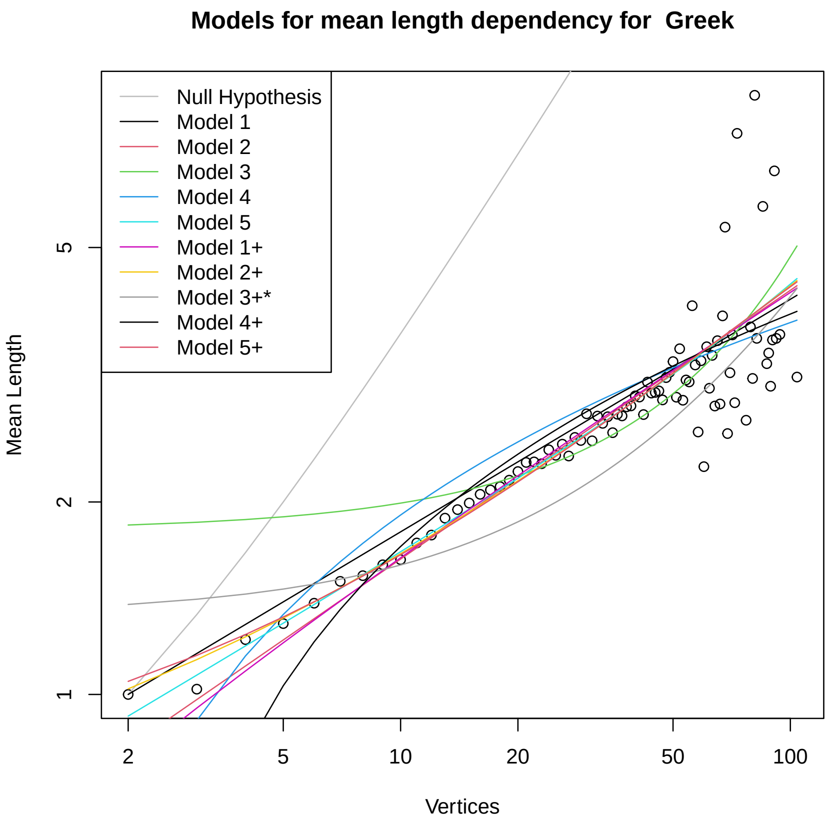


Figure 9: All models fitted for Greek

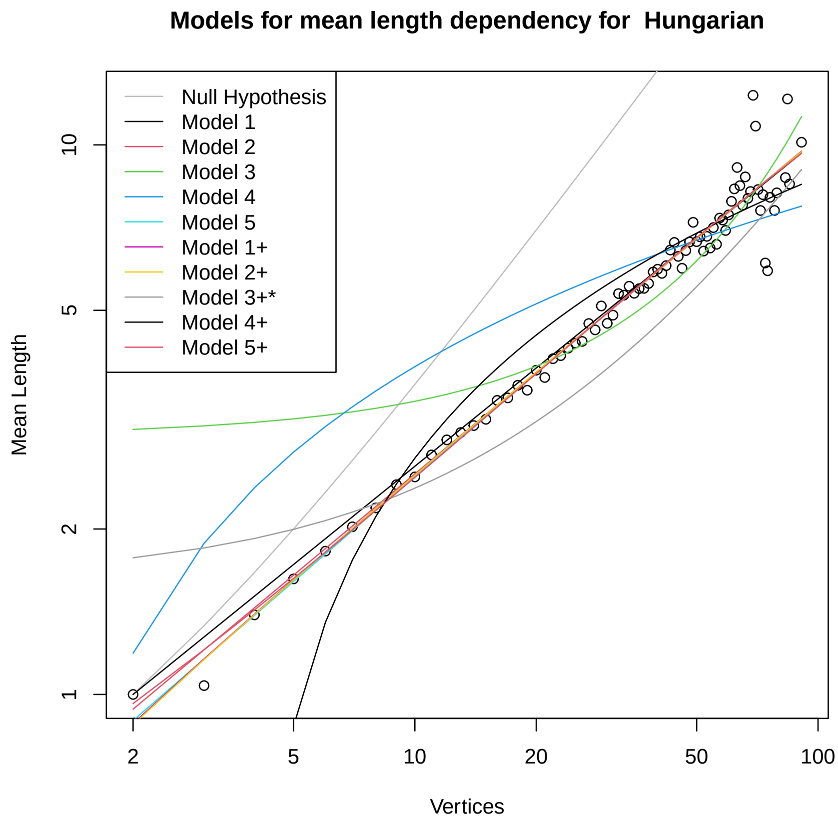


Figure 10: All models fitted for Hungarian

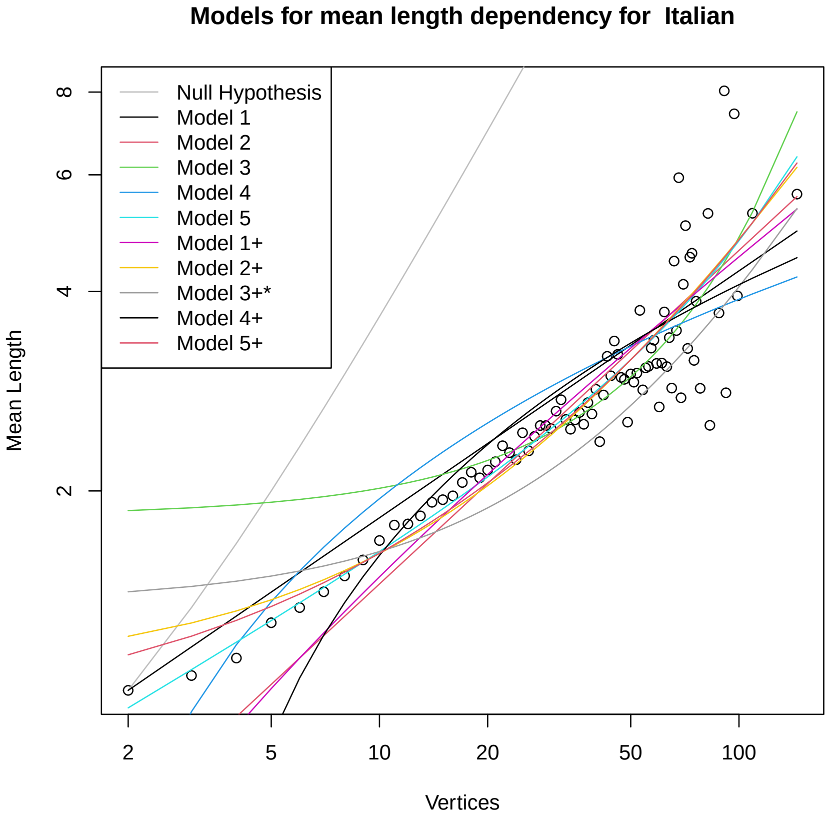


Figure 11: All models fitted for Italian

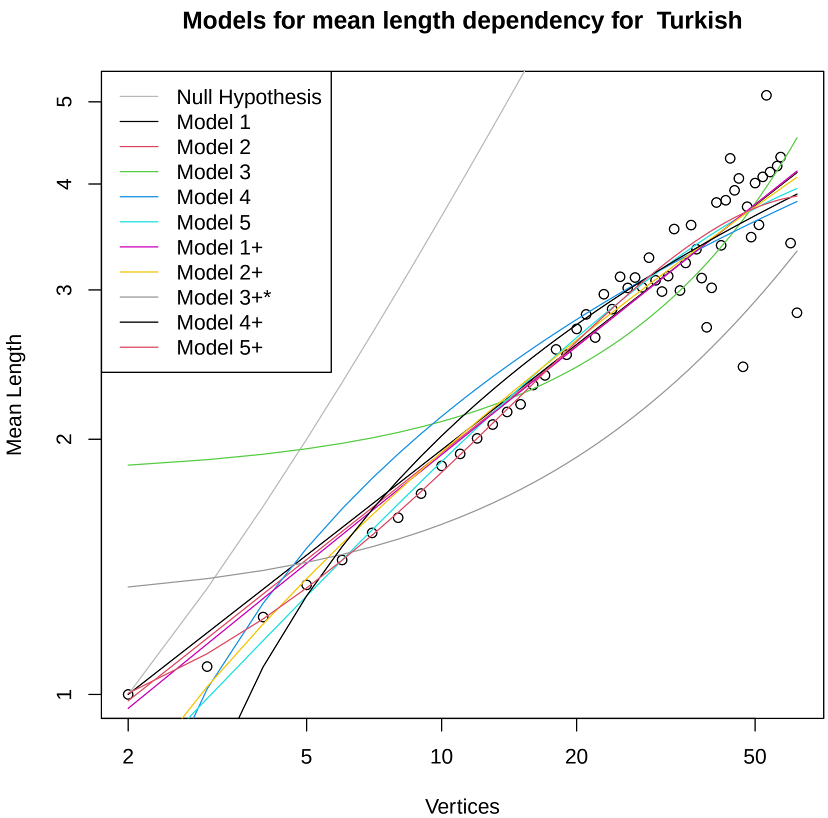


Figure 12: All models fitted for Turkish

B The best model by AIC for each language

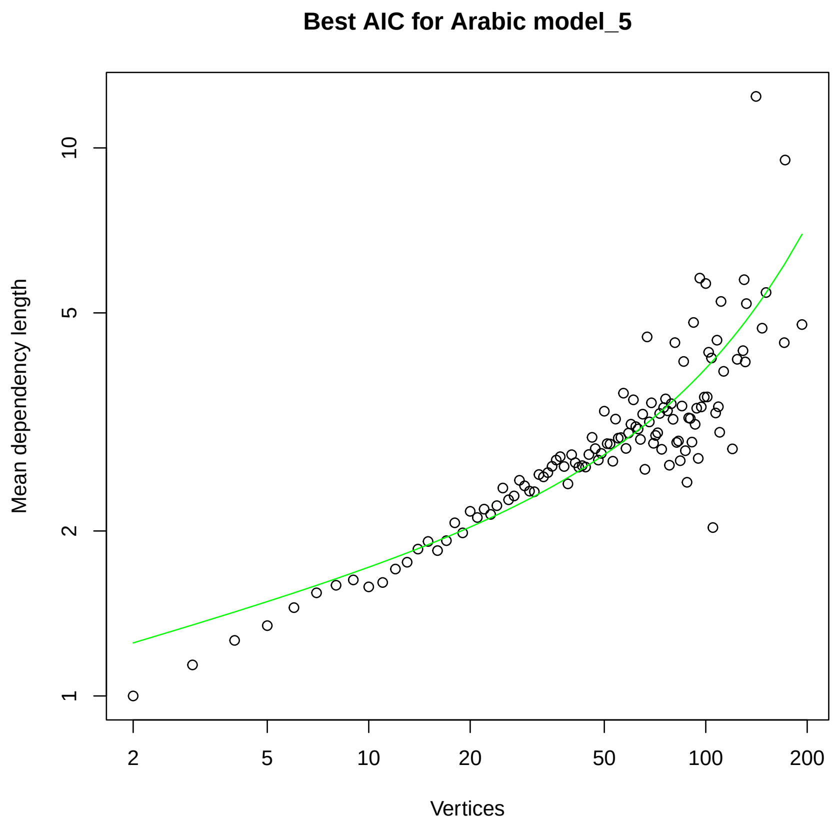


Figure 13: The best model for Arabic

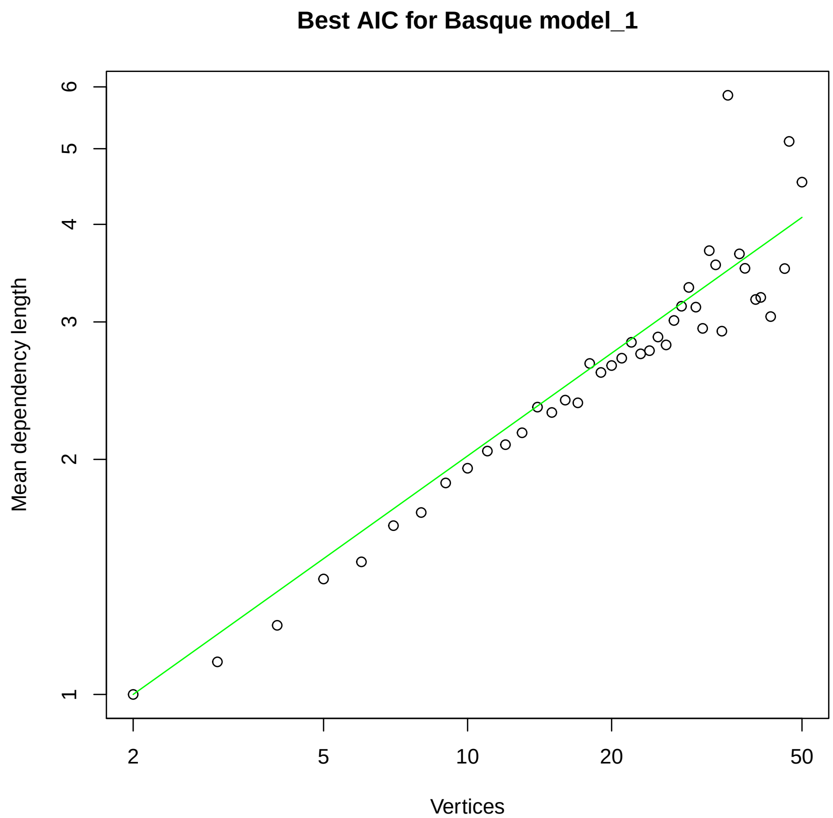


Figure 14: The best model for Basque

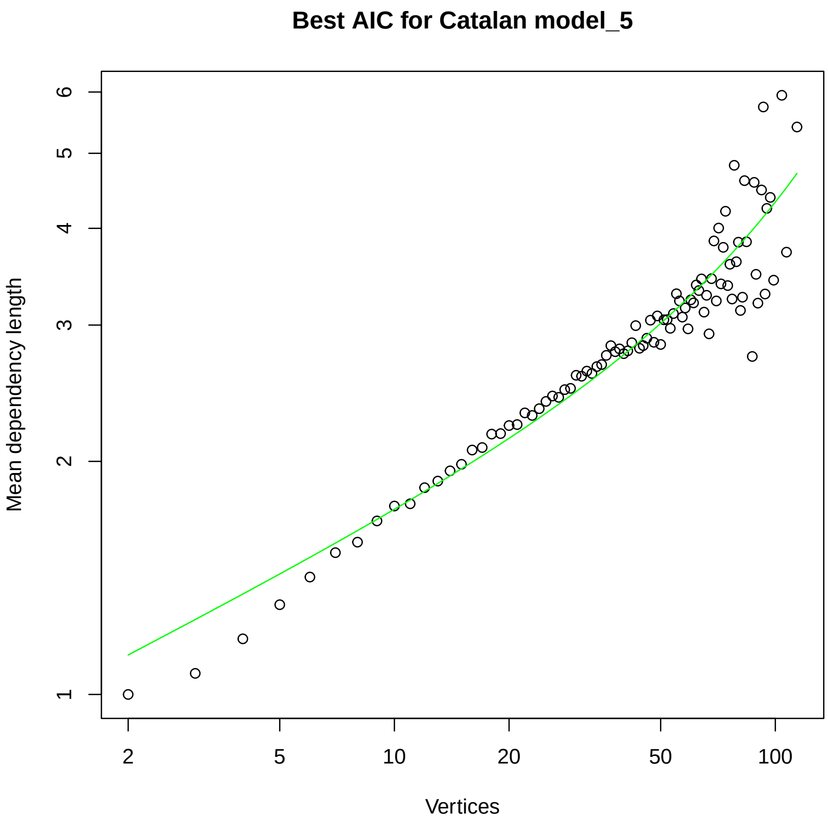


Figure 15: The best model for Catalan

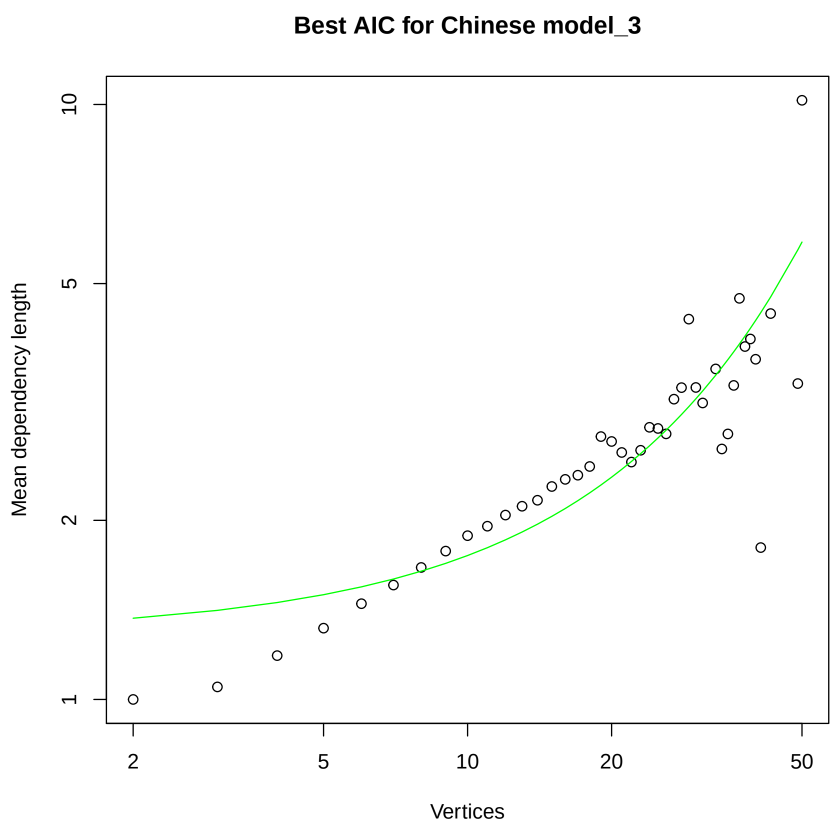


Figure 16: The best model for Chinese

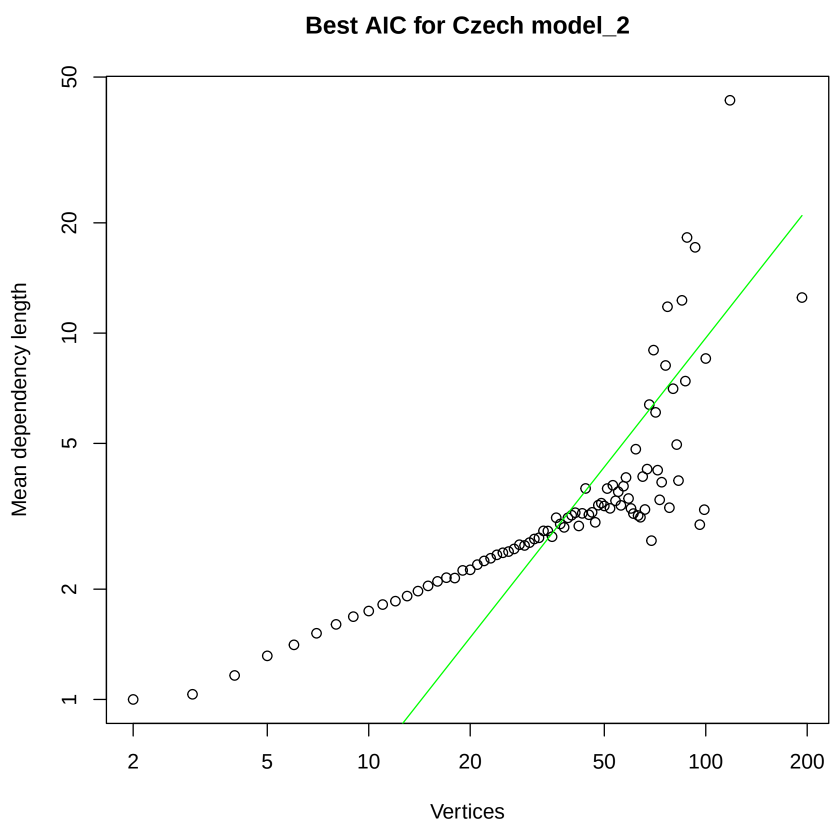


Figure 17: The best model for Czech

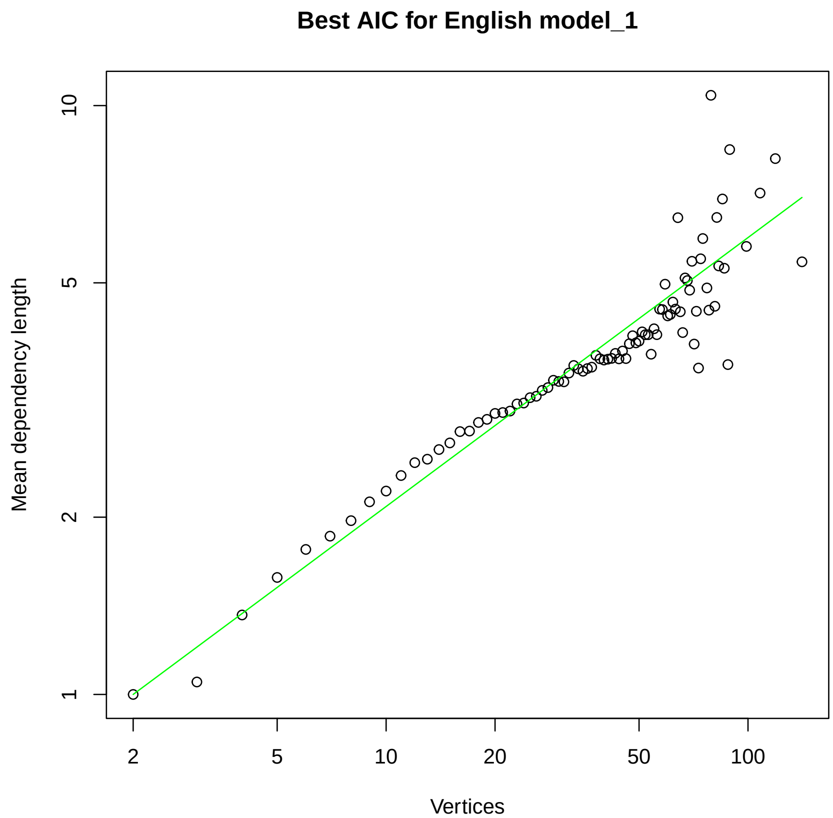


Figure 18: The best model for English

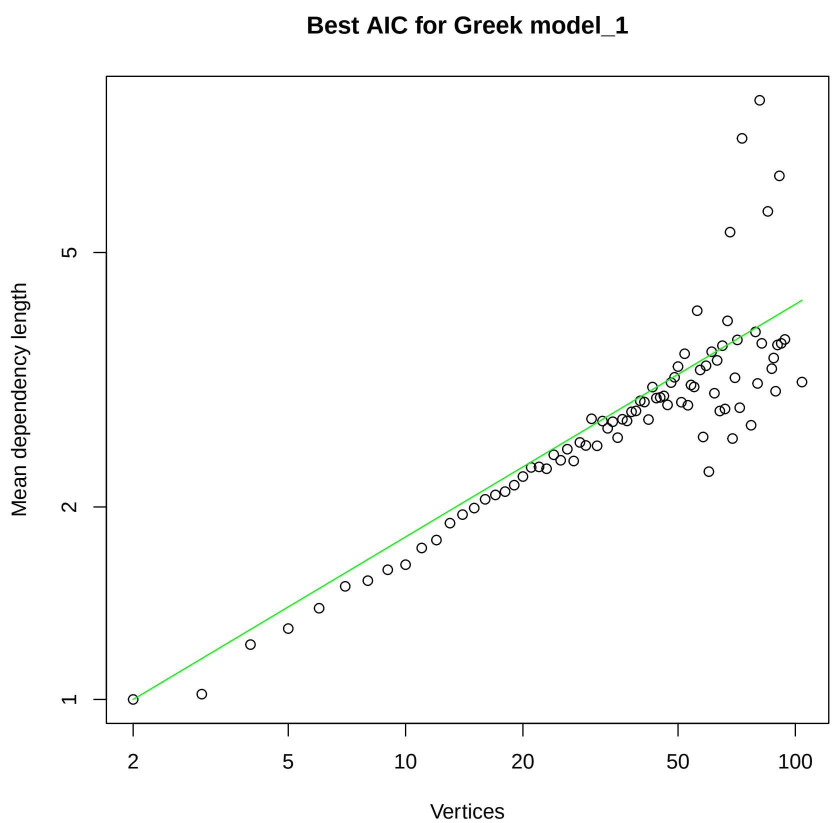


Figure 19: The best model for Greek

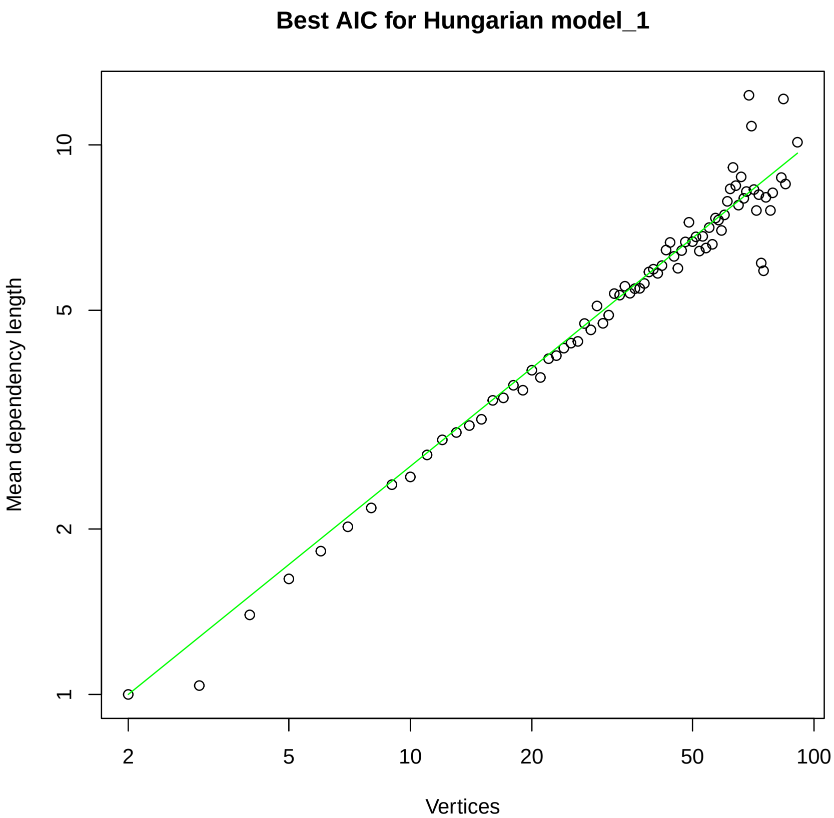


Figure 20: The best model for Hungarian

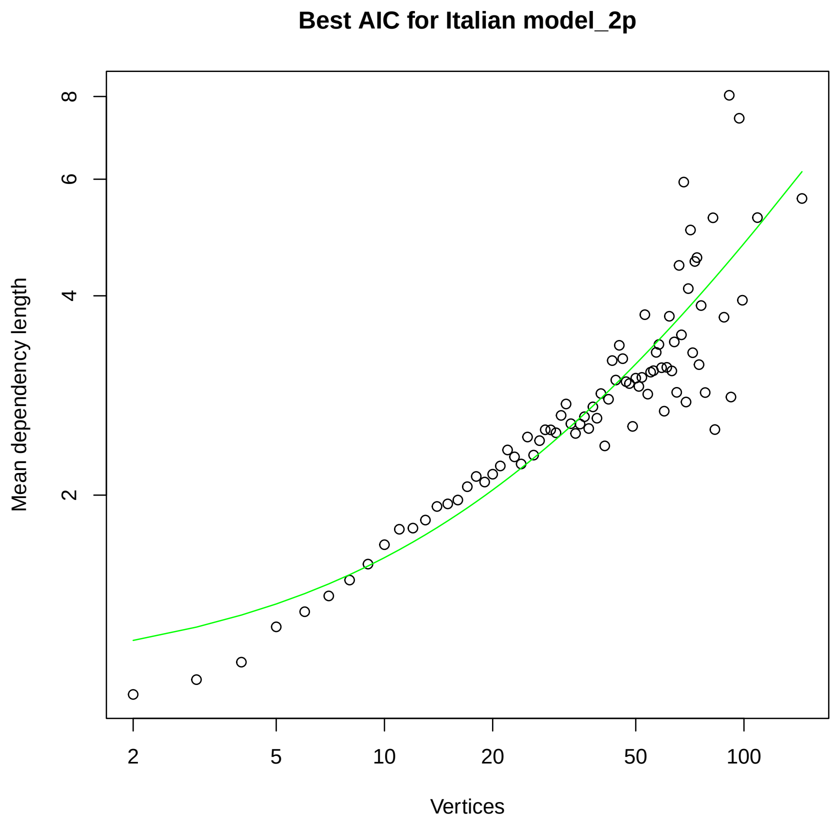


Figure 21: The best model for Italian

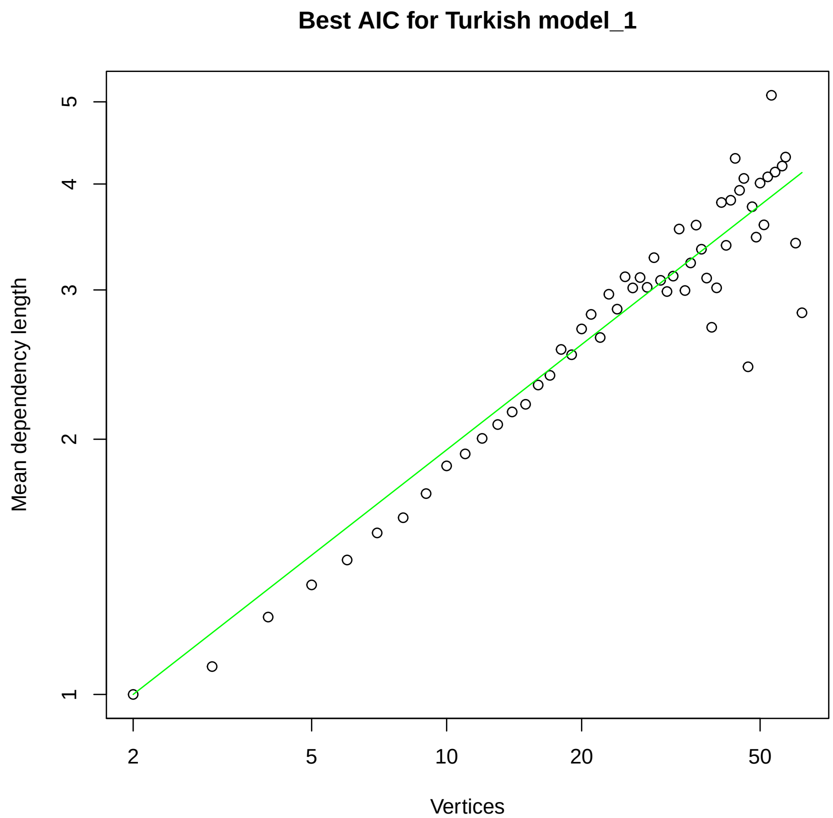


Figure 22: The best model for Turkish

C The best model by s for each language

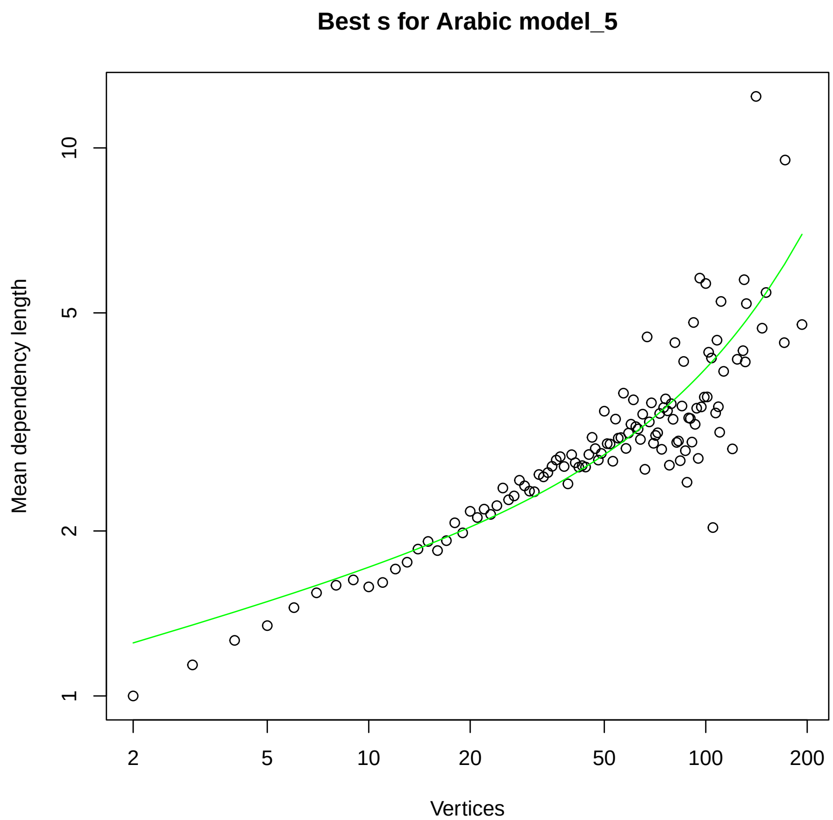


Figure 23: The best model for Arabic

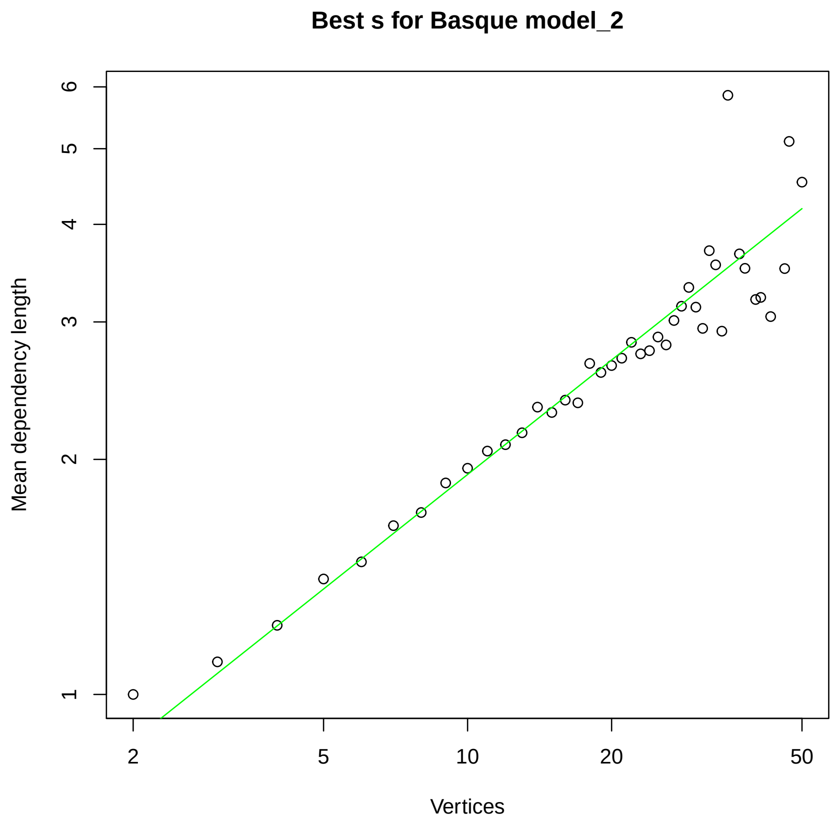


Figure 24: The best model for Basque

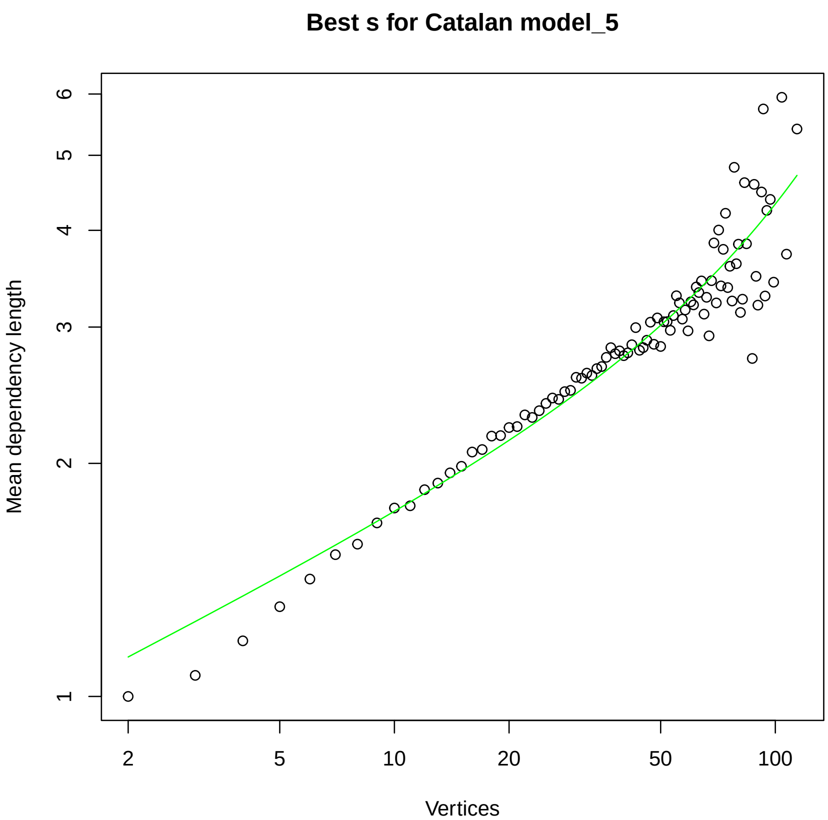


Figure 25: The best model for Catalan

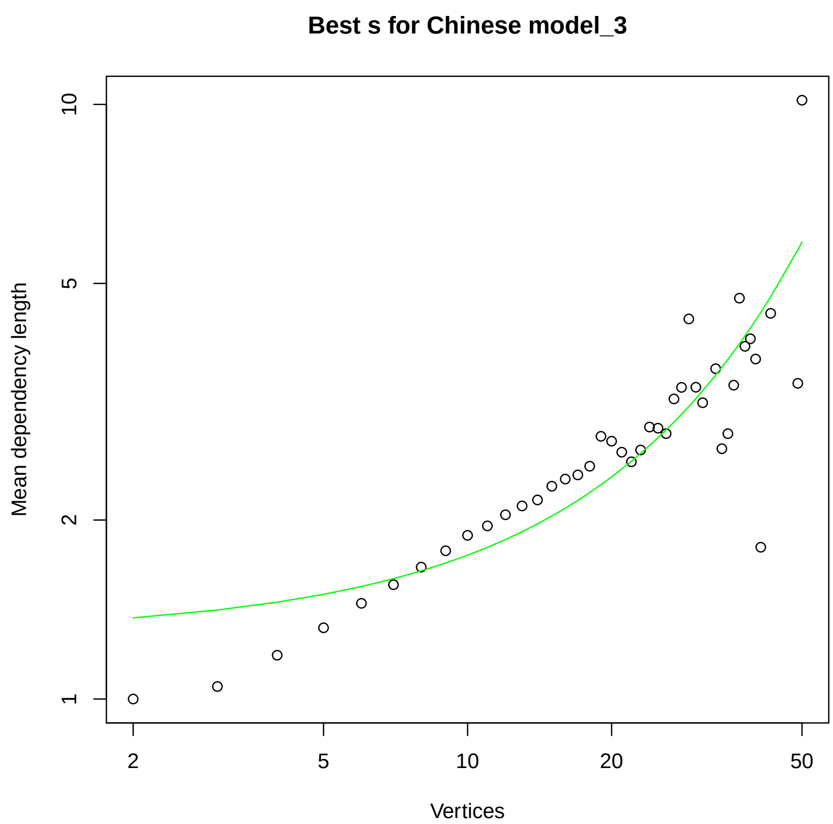


Figure 26: The best model for Chinese

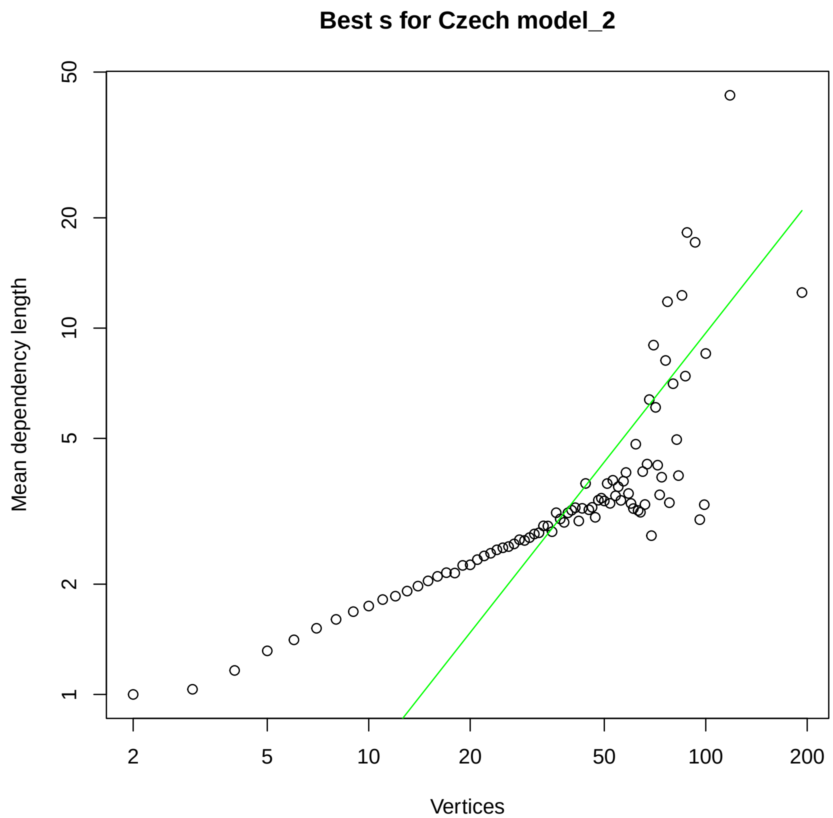


Figure 27: The best model for Czech

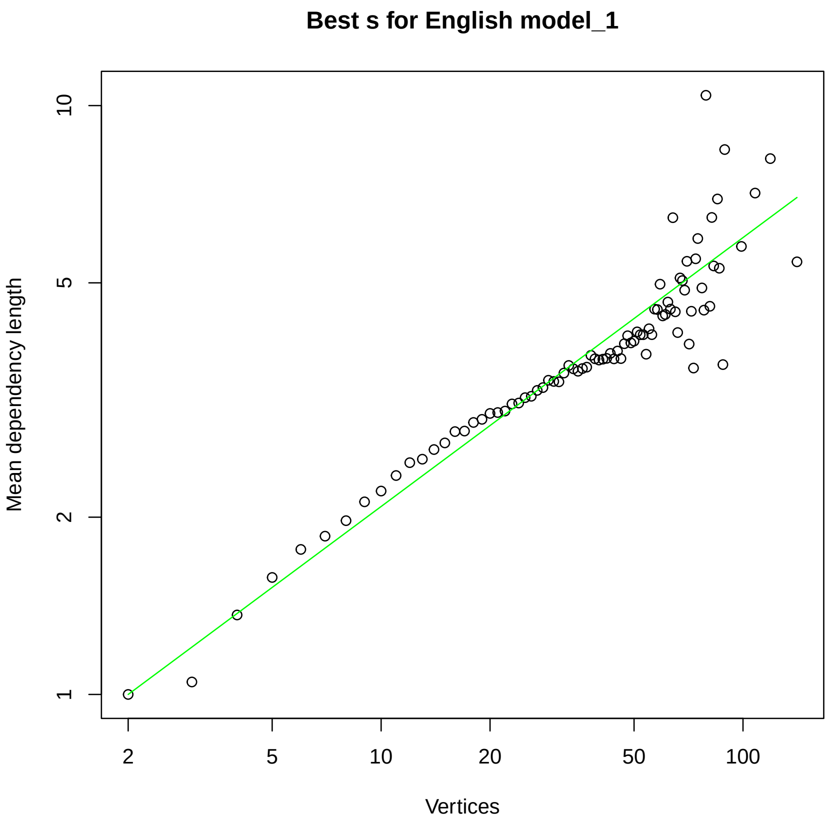


Figure 28: The best model for English

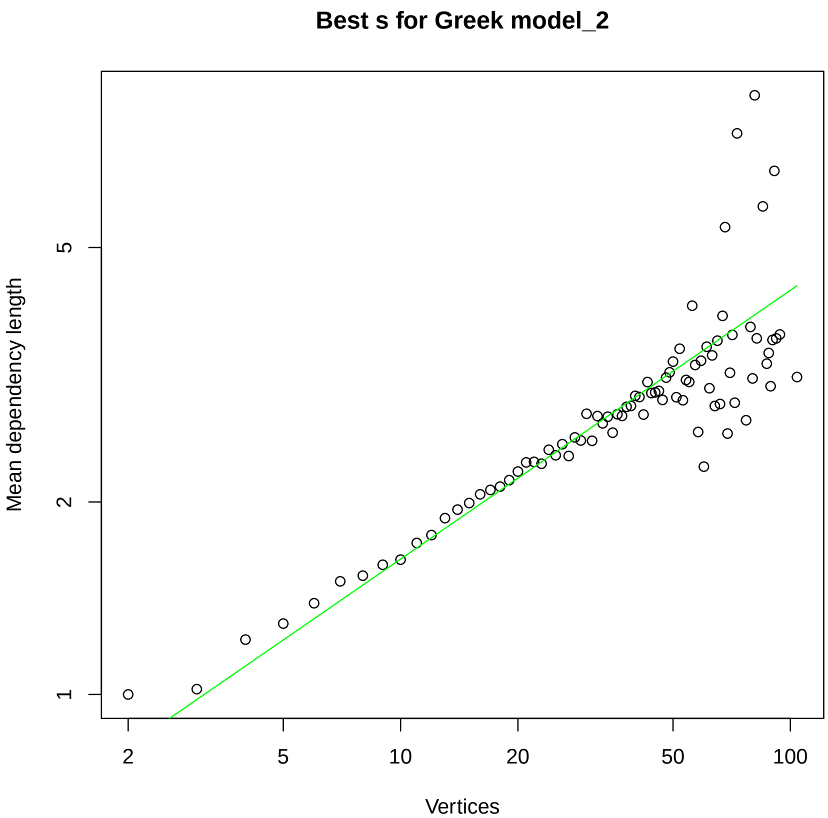


Figure 29: The best model for Greek

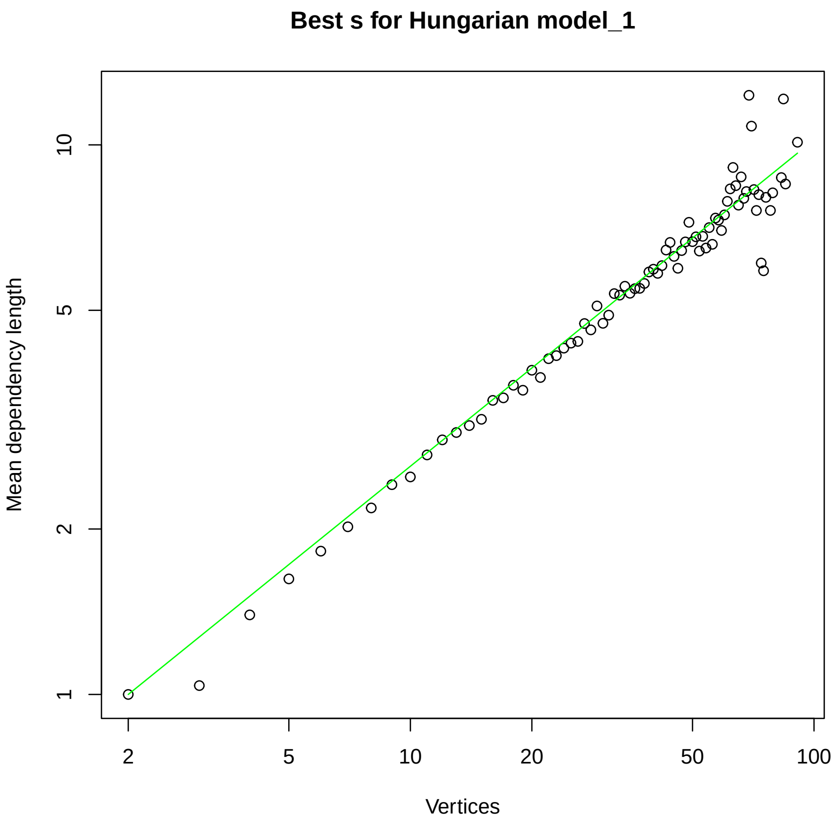


Figure 30: The best model for Hungarian

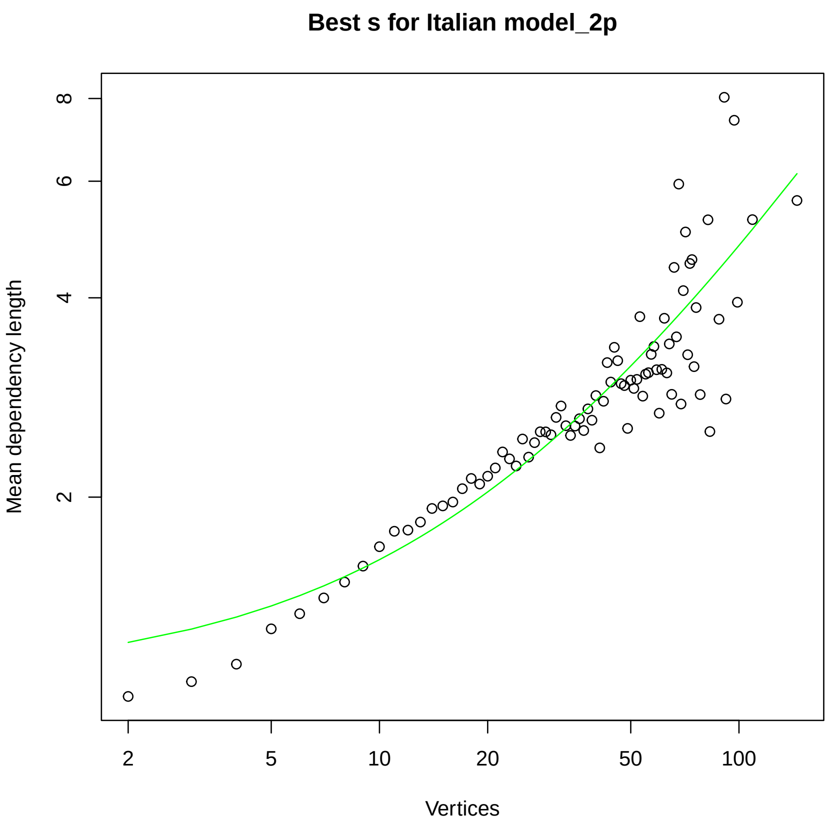


Figure 31: The best model for Italian

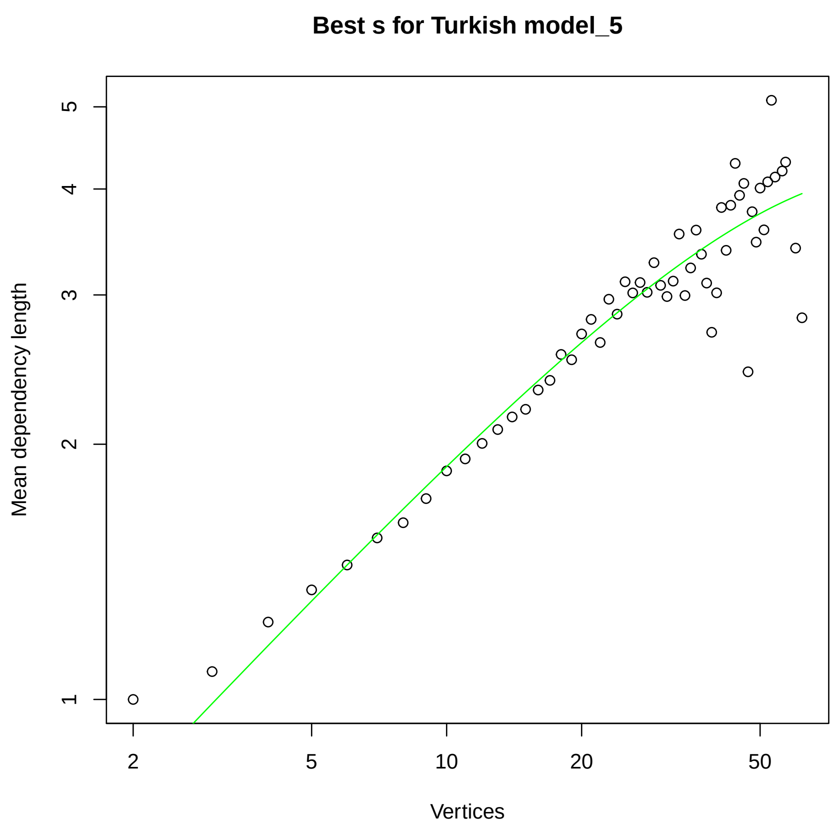


Figure 32: The best model for Turkish

D Data Samples provided for each language

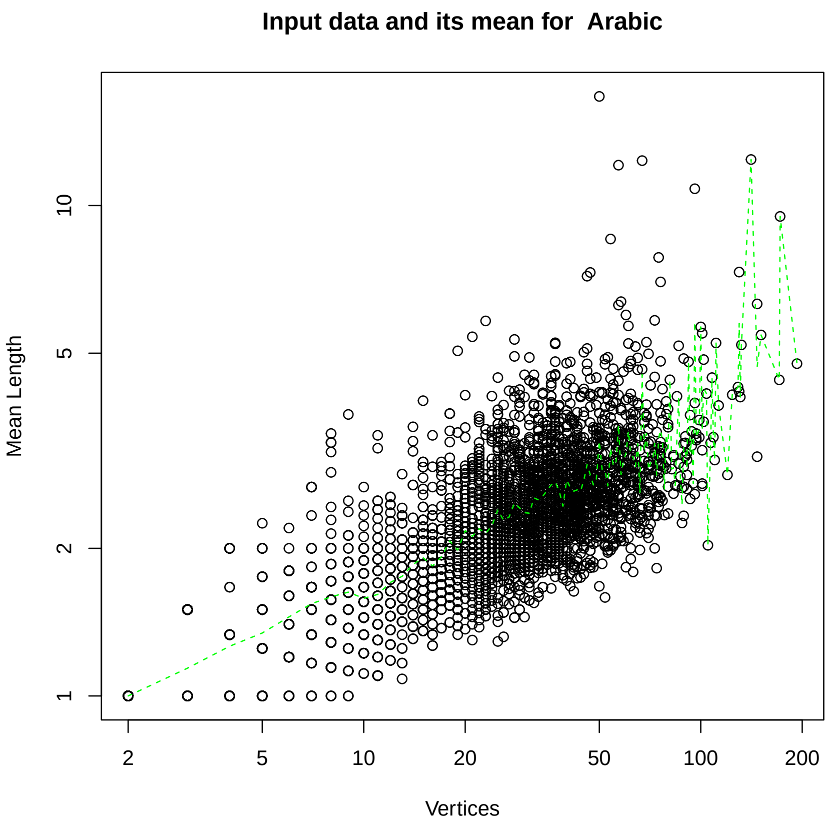


Figure 33: Data Samples for Arabic

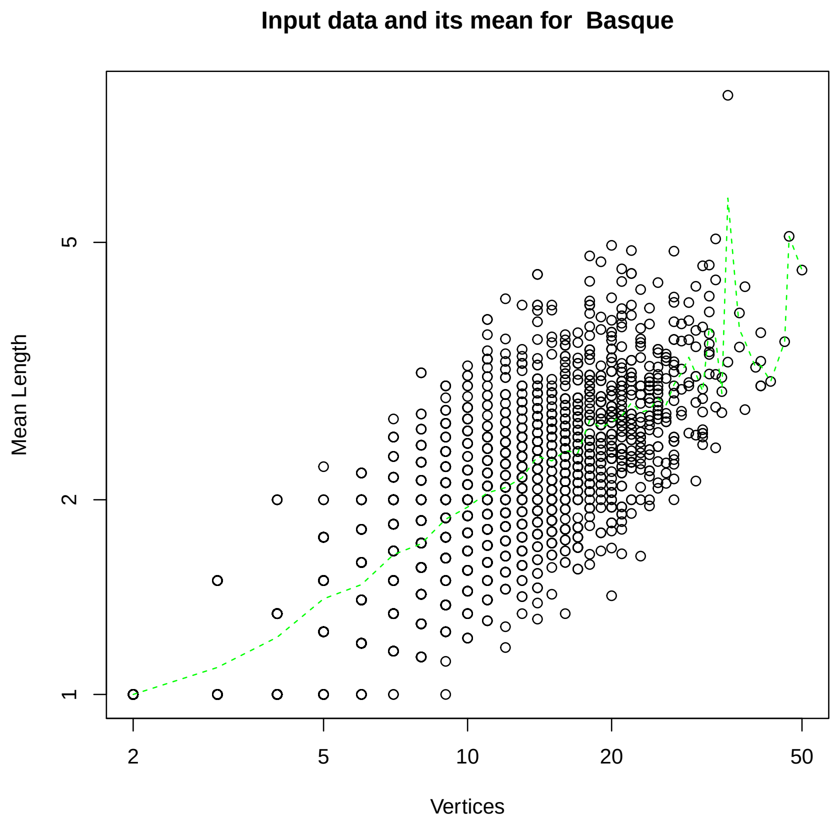


Figure 34: Data Samples for Basque

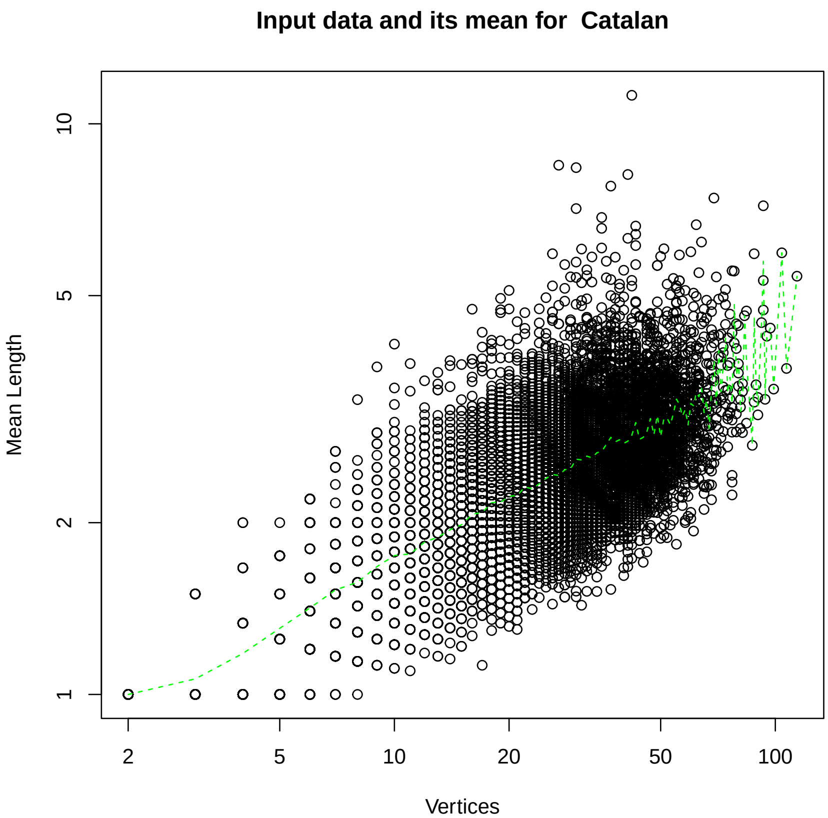


Figure 35: Data Samples for Catalan

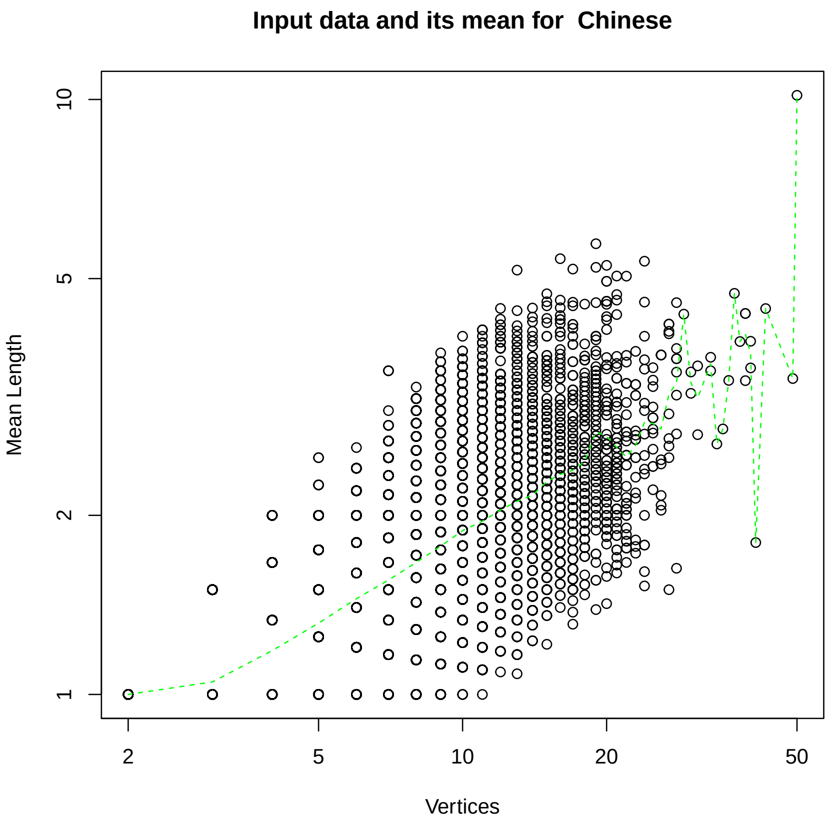


Figure 36: Data Samples for Chinese

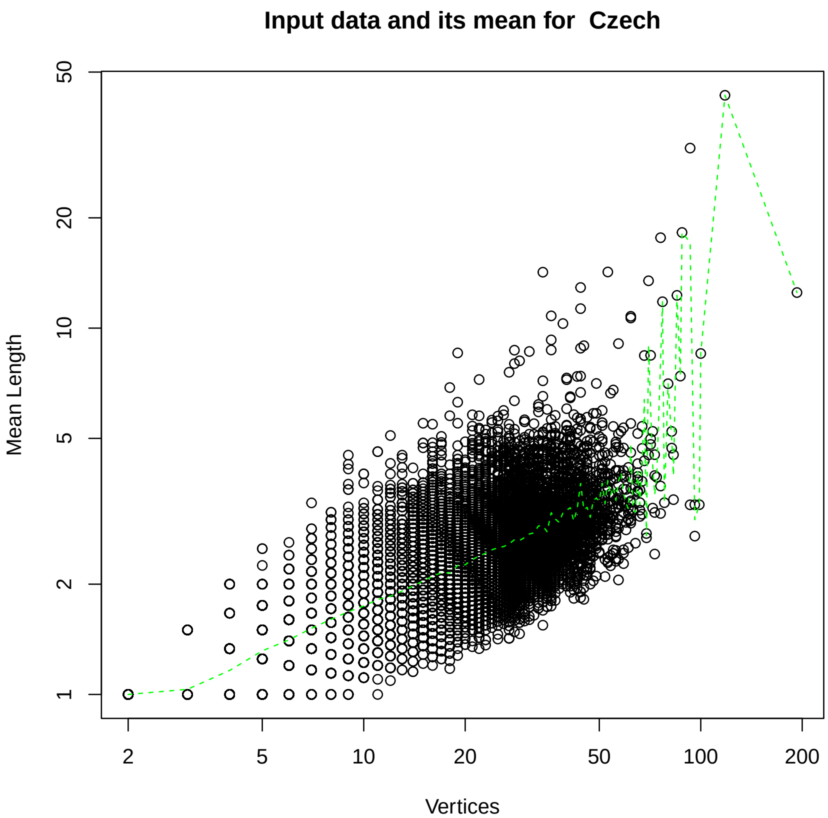


Figure 37: Data Samples for Czech

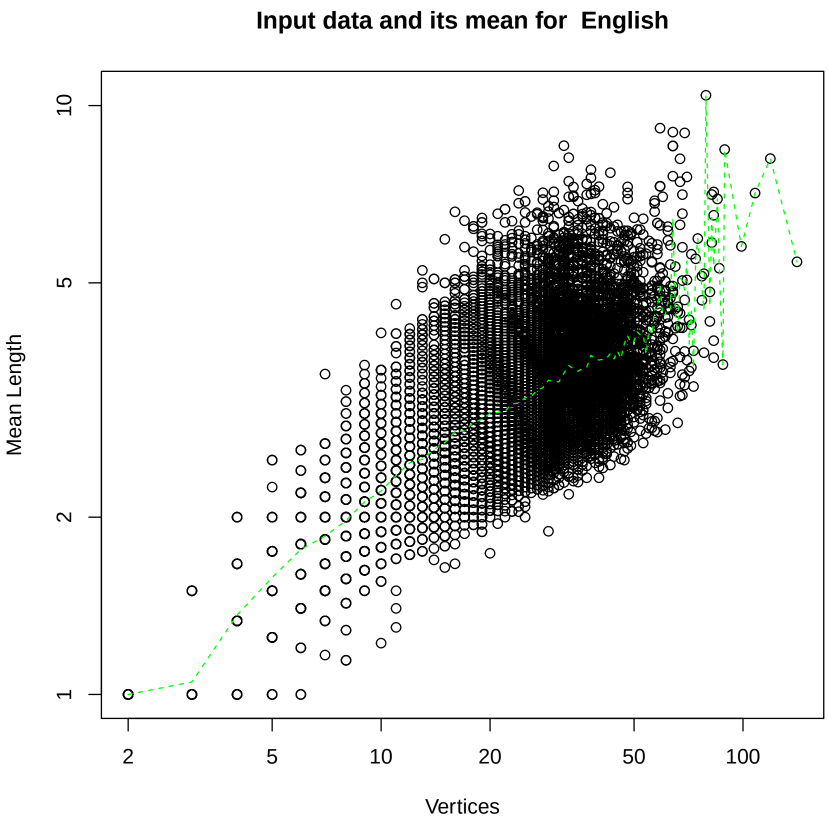


Figure 38: Data Samples for English

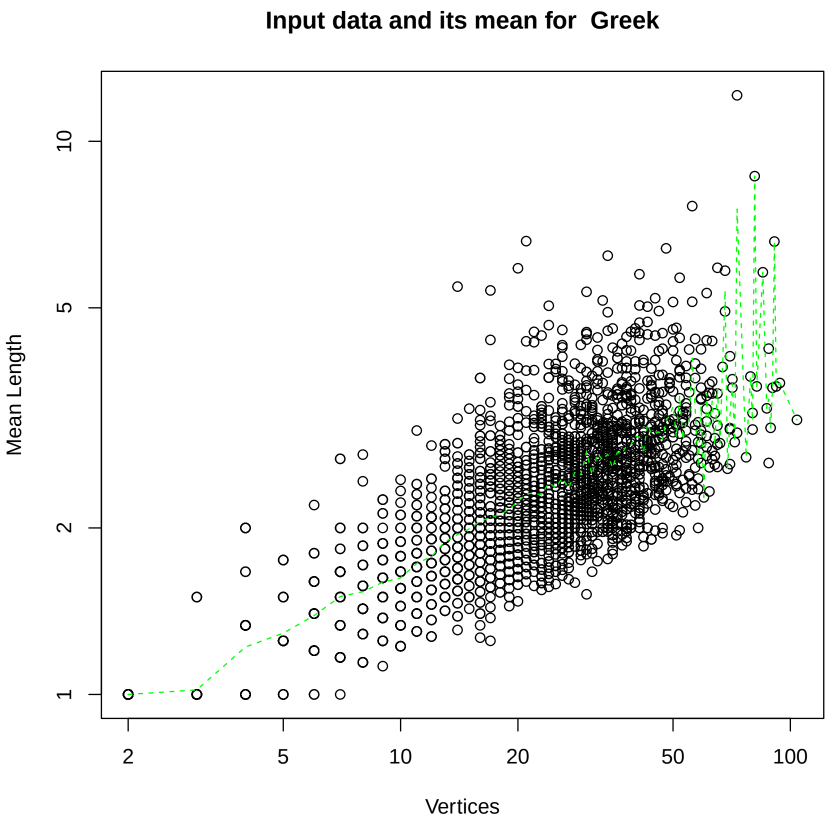


Figure 39: Data Samples for Greek

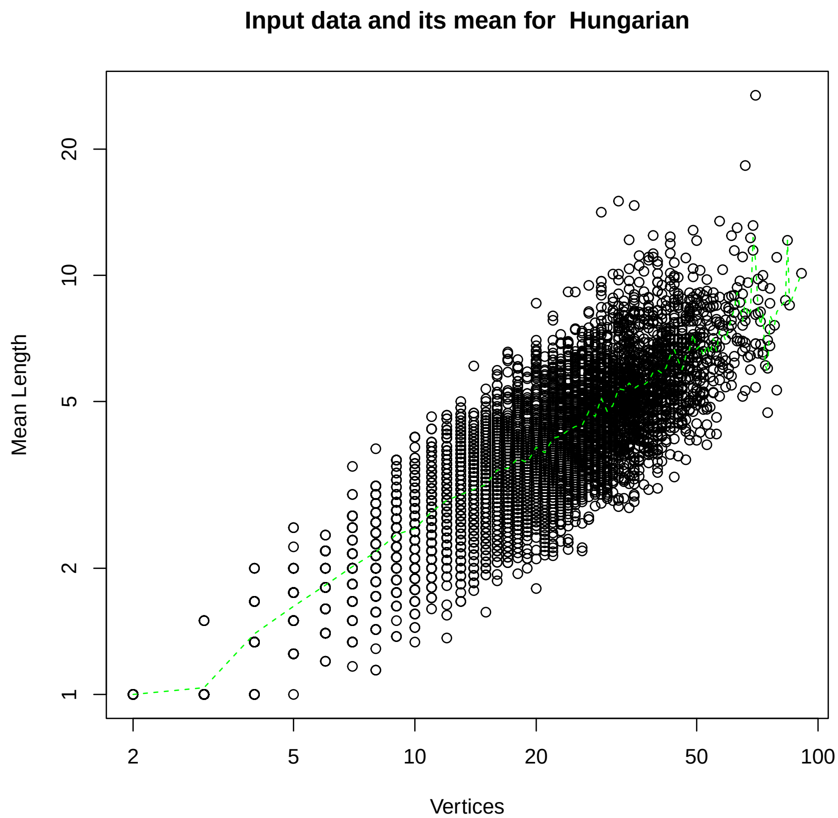


Figure 40: Data Samples for Hungarian

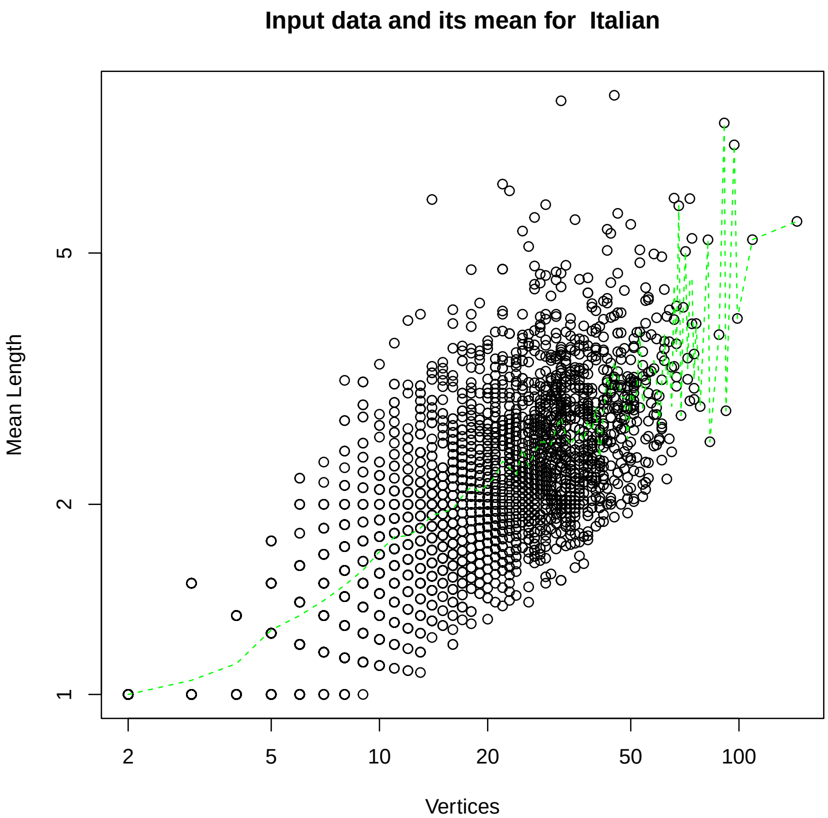


Figure 41: Data Samples for Italian

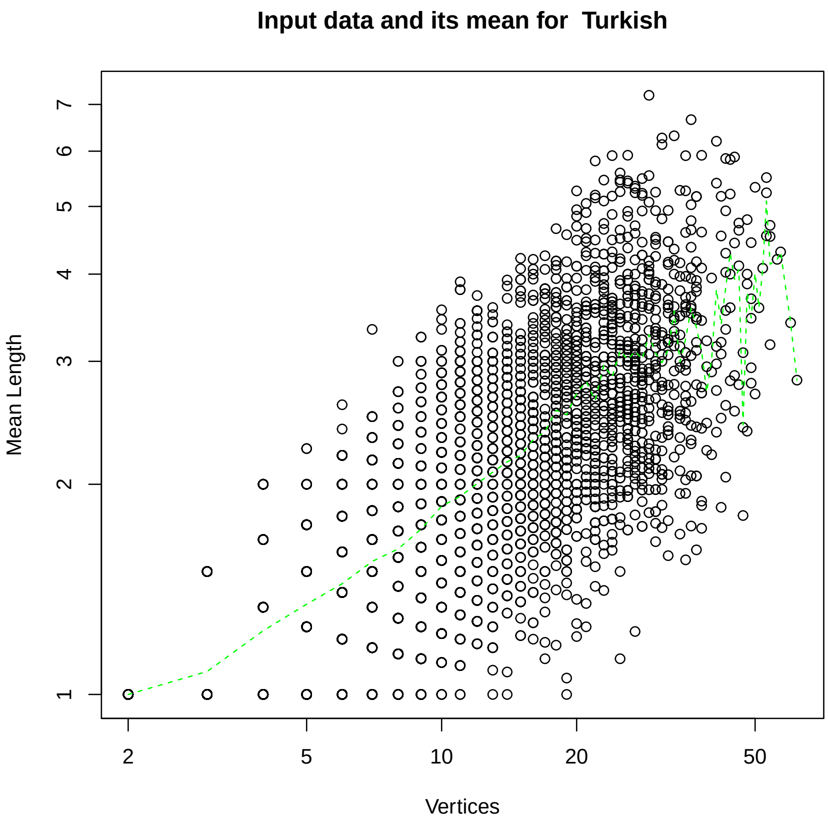
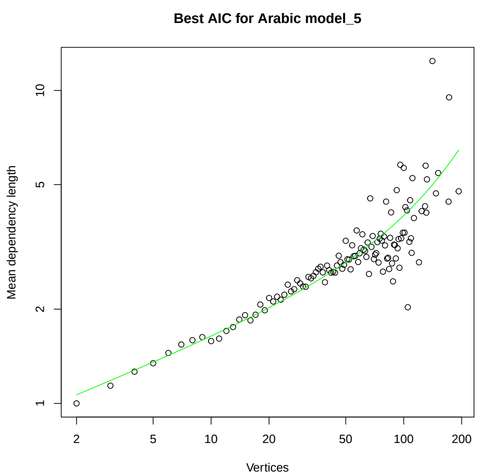
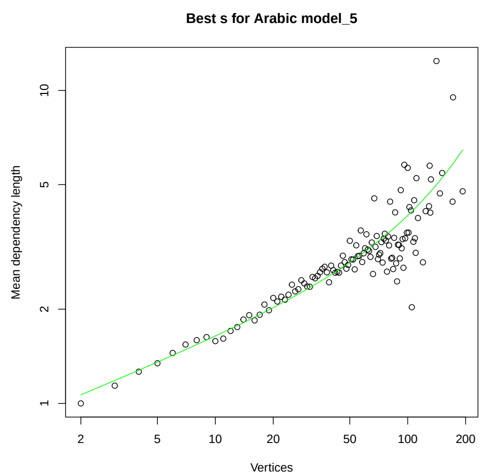


Figure 42: Data Samples for Turkish

E Weighted Nonlinear Least Squares



(a) Model with smallest AIC for Arabic.



(b) Models with smallest s Arabic.

Figure 43: Models fitted for language Arabic

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	1090.844	235.935	235.521	275.158	268.201	224.634	236.397	231.094	274.438	266.068	241.632
Basque	276.863	32.845	32.396	64.283	49.796	33.756	32.756	33.646	89.041	50.941	35.646
Catalan	814.121	35.646	32.216	148.302	98.362	28.594	33.100	31.855	154.736	90.916	35.415
Chinese	274.712	89.864	88.606	88.801	97.953	85.676	89.439	87.615	87.605	99.951	102.902
Czech	723.596	428.590	421.061	424.942	439.635	416.160	426.906	411.535	413.922	440.122	418.808
English	724.102	152.683	154.417	247.931	184.732	155.759	154.457	156.371	257.720	185.313	183.950
Greek	711.355	156.227	155.661	193.590	173.208	156.703	156.055	156.800	205.467	173.253	158.695
Hungarian	610.872	143.952	143.565	274.453	269.788	145.511	143.633	145.551	265.868	241.818	147.481
Italian	705.677	134.719	128.651	173.938	160.282	122.408	130.492	124.636	175.543	161.160	124.370
Turkish	412.776	20.871	19.327	96.241	49.615	20.250	19.339	21.317	110.629	50.554	17.680

Table 6: Akaike information criterion (AIC) of each model

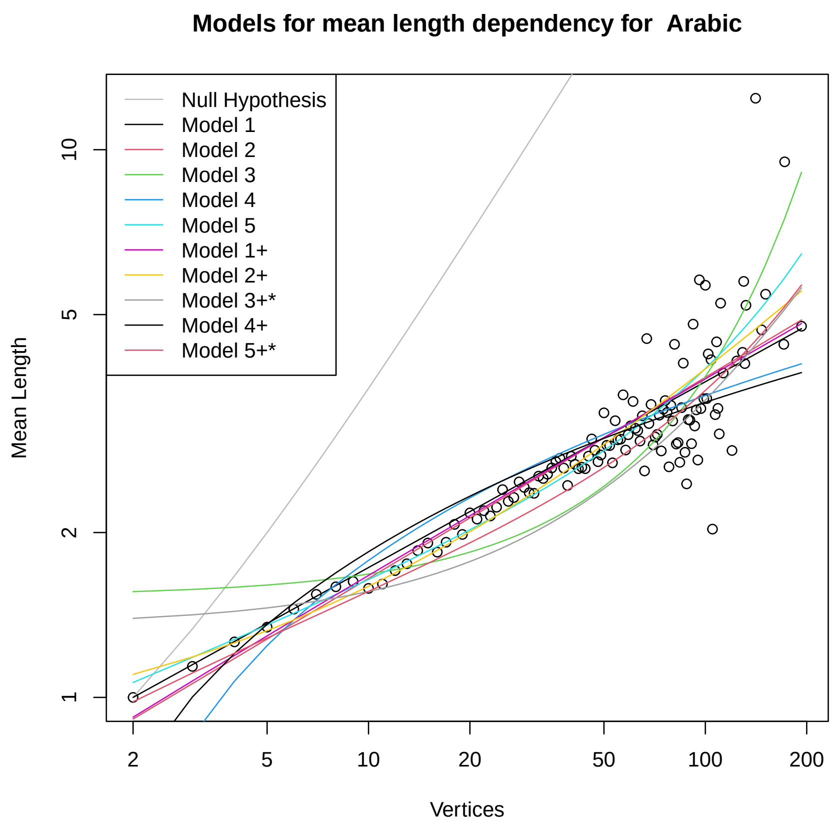
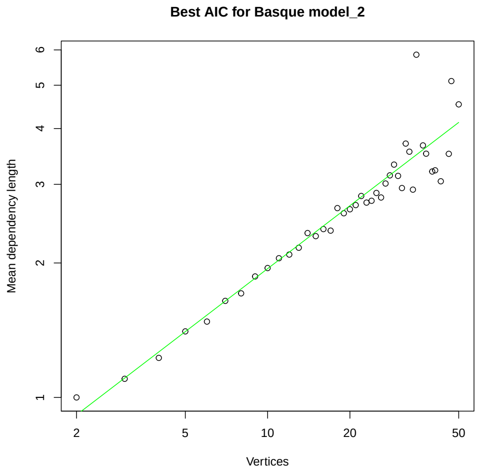


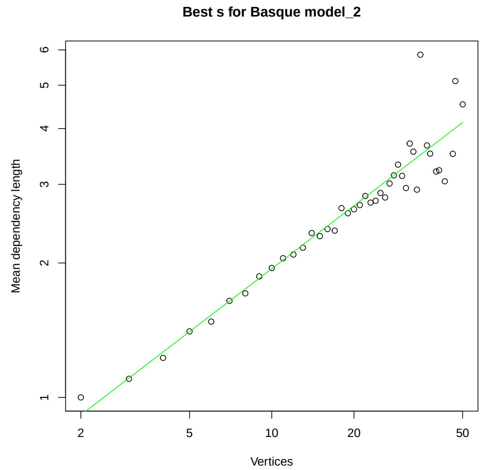
Figure 44: All models for language Arabic.

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	866.211	11.301	10.888	50.525	43.567	0.000	11.764	6.460	49.805	41.434	16.998
Basque	244.467	0.449	0.000	31.887	17.400	1.360	0.360	1.250	56.646	18.545	3.250
Catalan	785.527	7.052	3.621	119.708	69.767	0.000	4.506	3.261	126.141	62.322	6.820
Chinese	189.036	4.188	2.931	3.126	12.277	0.000	3.763	1.939	1.929	14.275	17.227
Czech	312.061	17.055	9.527	13.408	28.100	4.625	15.371	0.000	2.388	28.587	7.274
English	571.419	0.000	1.734	95.248	32.049	3.076	1.774	3.688	105.037	32.630	31.267
Greek	555.693	0.566	0.000	37.928	17.546	1.042	0.393	1.138	49.806	17.592	3.033
Hungarian	467.306	0.387	0.000	130.887	126.223	1.945	0.068	1.986	122.302	98.252	3.916
Italian	583.268	12.310	6.243	51.530	37.874	0.000	8.084	2.227	53.135	38.752	1.962
Turkish	395.096	3.191	1.647	78.561	31.935	2.570	1.659	3.636	92.949	32.874	0.000

Table 7: AIC differences



(a) Model with smallest AIC for Basque.



(b) Models with smallest s Basque.

Figure 45: Models fitted for language Basque

Language	0	1	2	3	4	5	1+	2+	3+	4+	5+
Arabic	22.599	0.044	0.044	0.052	0.050	0.042	0.044	0.043	0.051	0.050	0.045
Basque	6.380	0.044	0.043	0.063	0.054	0.044	0.044	0.044	0.084	0.054	0.044
Catalan	16.624	0.023	0.022	0.041	0.031	0.022	0.022	0.022	0.042	0.030	0.022
Chinese	6.219	0.087	0.085	0.085	0.096	0.081	0.086	0.083	0.083	0.097	0.098
Czech	14.600	0.225	0.214	0.219	0.239	0.207	0.221	0.202	0.204	0.239	0.209
English	14.642	0.047	0.047	0.080	0.056	0.047	0.047	0.047	0.084	0.056	0.055
Greek	14.958	0.050	0.049	0.061	0.055	0.049	0.049	0.049	0.065	0.054	0.049
Hungarian	10.376	0.051	0.050	0.112	0.110	0.050	0.050	0.050	0.106	0.092	0.051
Italian	15.185	0.044	0.043	0.056	0.052	0.041	0.043	0.041	0.056	0.052	0.041
Turkish	8.885	0.030	0.030	0.058	0.039	0.030	0.030	0.030	0.066	0.039	0.029

Table 8: Residual standard error for each model

	Model																						
	1	2		3		4	5	1+		2+			3+			4+		5+					
Language	b	a	b	a	c	a	a	b	c	b	d	a	b	d	a	c	d	a	d				
Arabic	0.339	0.708	0.367	1.531	0.009	0.773	0.888	0.252	0.003	0.348	-0.080	0.200	0.601	0.799	35.108	0.001	-33.756	0.700	0.232	1.812	0.143	0.003	-1.030
Basque	0.432	0.661	0.468	1.312	0.028	0.922	0.707	0.420	0.003	0.445	-0.091	0.424	0.567	0.345	18.722	0.002	-17.646	0.875	0.118	0.423	0.567	-0.000	0.346
Catalan	0.351	0.720	0.374	1.456	0.012	0.794	0.787	0.324	0.002	0.359	-0.070	0.476	0.449	0.342	17.843	0.001	-16.533	0.723	0.218	1.882	0.170	0.003	-1.183
Chinese	0.442	0.559	0.530	1.199	0.032	0.927	0.813	0.263	0.017	0.469	-0.192	0.109	0.929	0.828	15.477	0.005	-14.535	0.931	-0.010	0.894	0.253	0.011	-0.179
Czech	0.471	0.158	0.864	1.703	0.015	1.013	0.567	0.412	0.008	0.515	-0.600	0.007	1.550	1.223	8.684	0.007	-7.779	1.238	-0.677	0.000	2.360	-0.007	1.328
English	0.456	0.754	0.447	1.862	0.013	1.063	0.794	0.420	0.001	0.453	0.031	0.817	0.432	-0.092	34.841	0.001	-33.341	1.114	-0.155	3.937	0.128	0.002	-3.421
Greek	0.358	0.685	0.393	1.423	0.014	0.800	0.747	0.345	0.002	0.370	-0.107	0.402	0.494	0.409	17.421	0.001	-16.162	0.741	0.176	0.633	0.381	0.002	0.132
Hungarian	0.593	0.605	0.617	2.078	0.020	1.509	0.594	0.626	-0.000	0.598	-0.103	0.621	0.611	-0.031	43.430	0.002	-42.110	1.940	-1.277	0.540	0.652	-0.001	0.084
Italian	0.359	0.632	0.416	1.434	0.013	0.797	0.771	0.308	0.004	0.378	-0.159	0.202	0.644	0.697	31.214	0.001	-30.029	0.755	0.126	1.137	0.235	0.004	-0.400
Turkish	0.412	0.686	0.439	1.399	0.021	0.900	0.645	0.479	-0.002	0.422	-0.082	0.709	0.433	-0.032	14.405	0.003	-13.249	0.866	0.094	0.141	0.957	-0.013	0.721

Table 9: Parameters of each fitted model

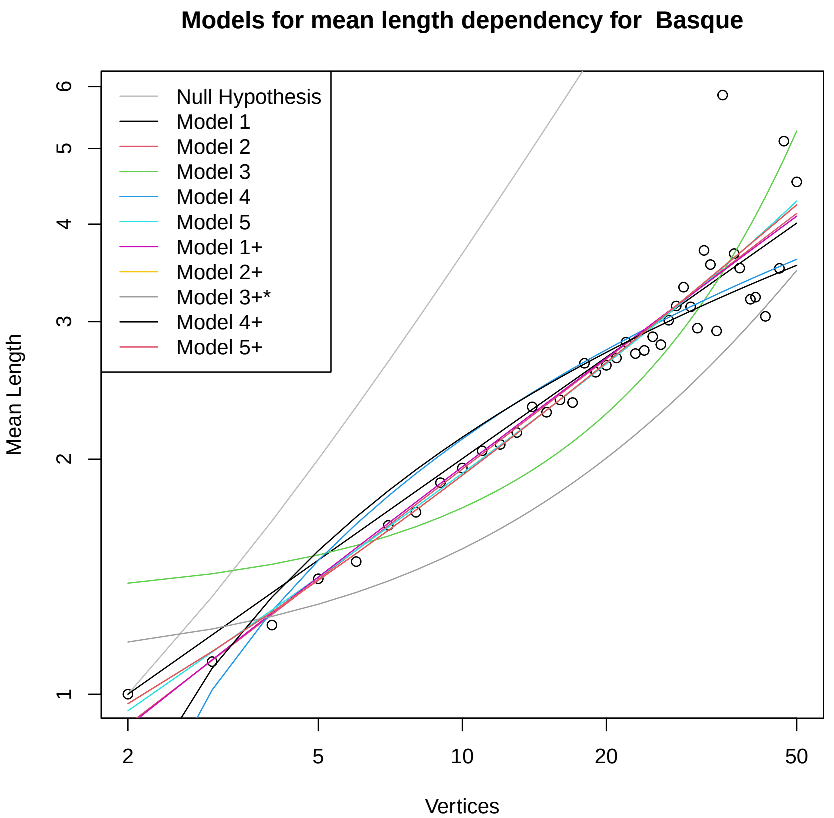
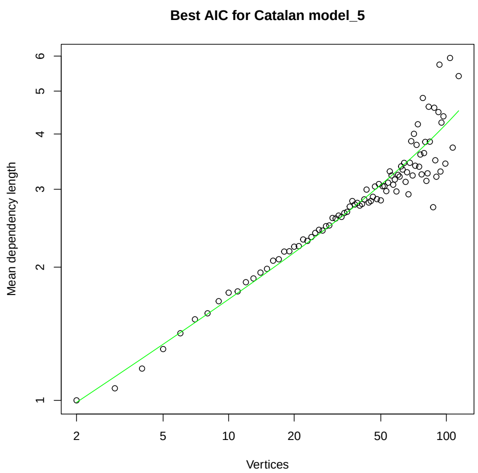
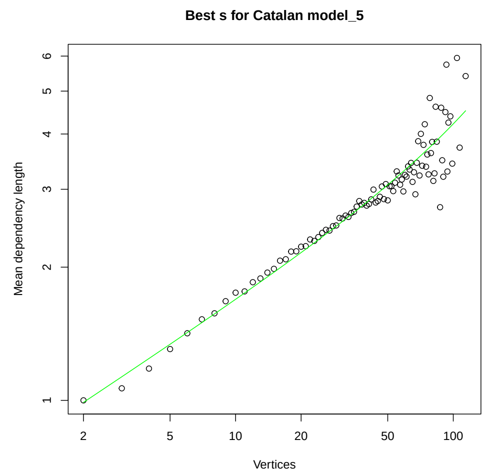


Figure 46: All models for language Basque.



(a) Model with smallest AIC for Catalan.



(b) Models with smallest s Catalan.

Figure 47: Models fitted for language Catalan

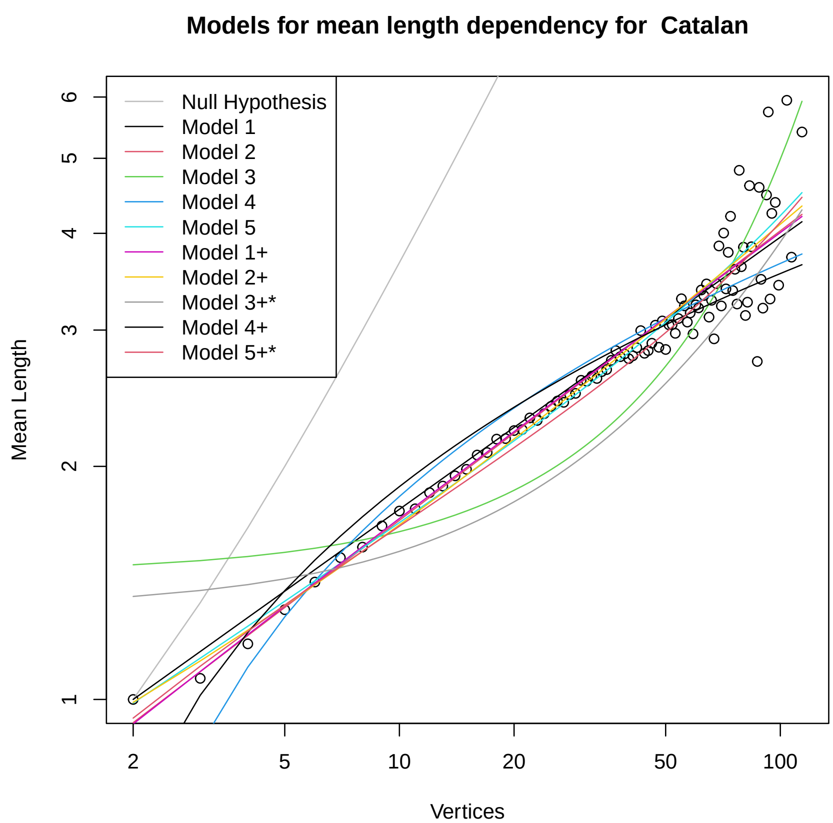
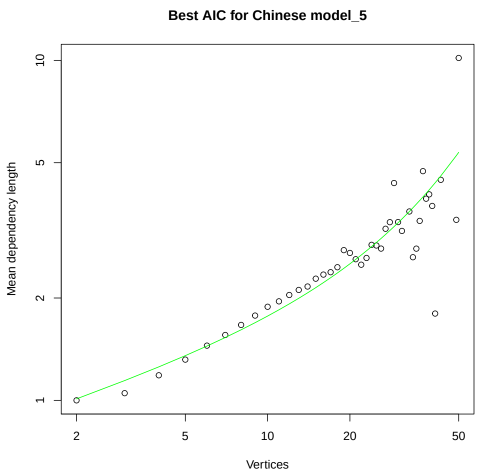
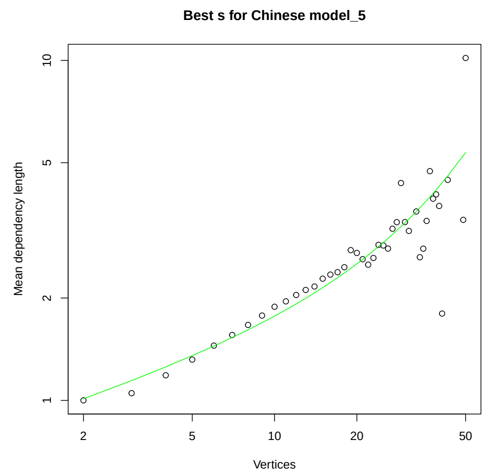


Figure 48: All models for language Catalan.



(a) Model with smallest AIC for Chinese.



(b) Models with smallest s Chinese.

Figure 49: Models fitted for language Chinese

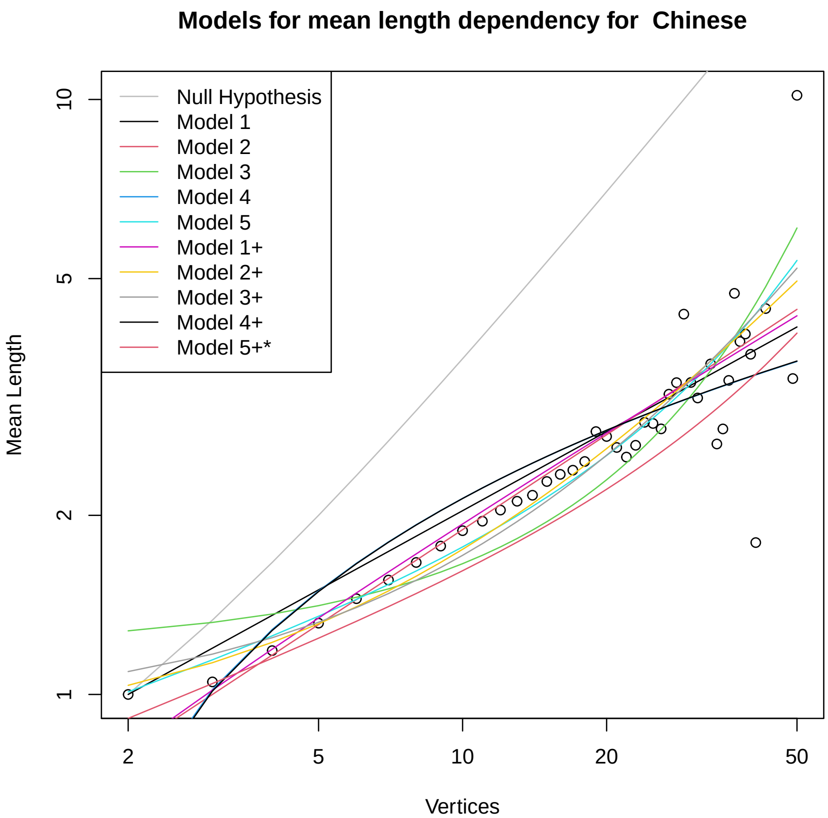
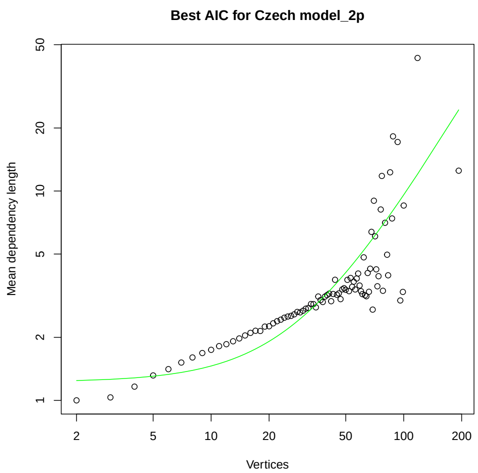
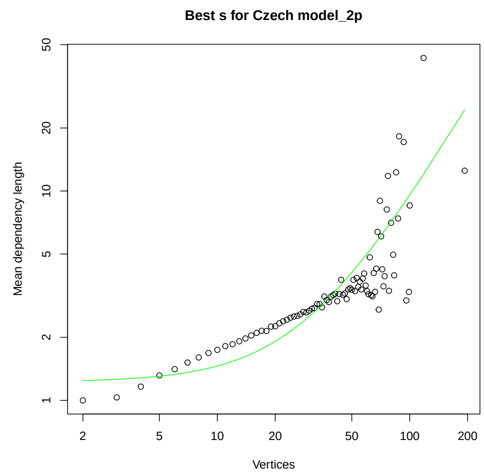


Figure 50: All models for language Chinese.



(a) Model with smallest AIC for Czech.



(b) Models with smallest s Czech.

Figure 51: Models fitted for language Czech

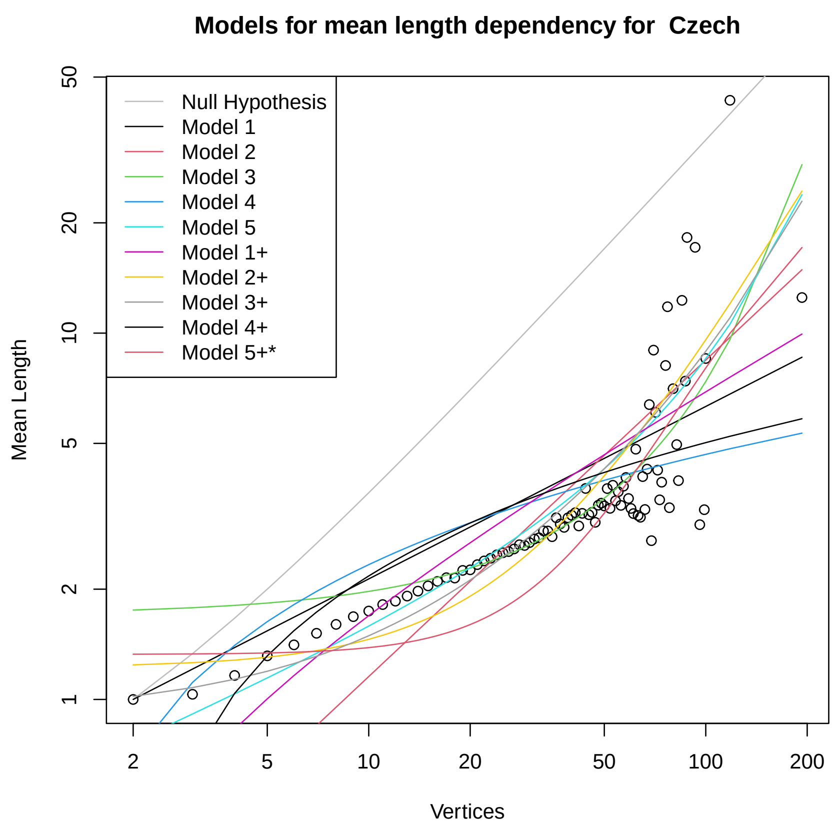
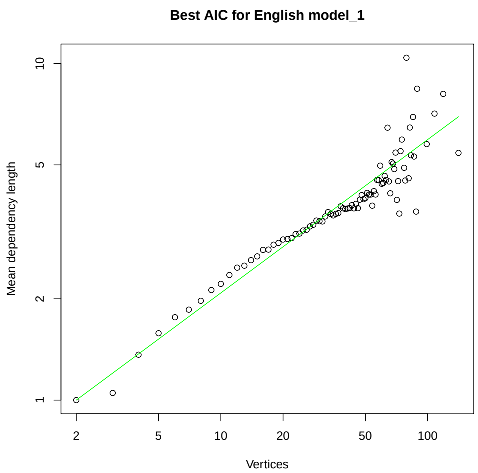
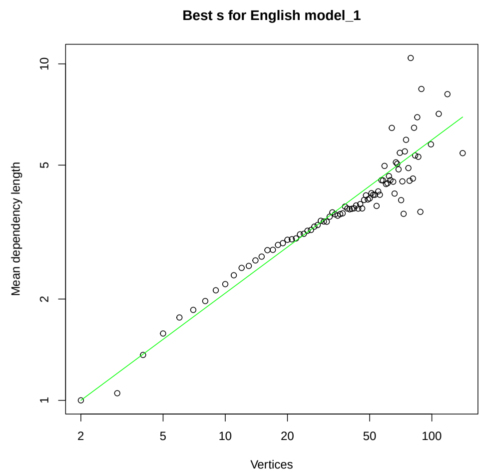


Figure 52: All models for language Czech.



(a) Model with smallest AIC for English.



(b) Models with smallest s English.

Figure 53: Models fitted for language English

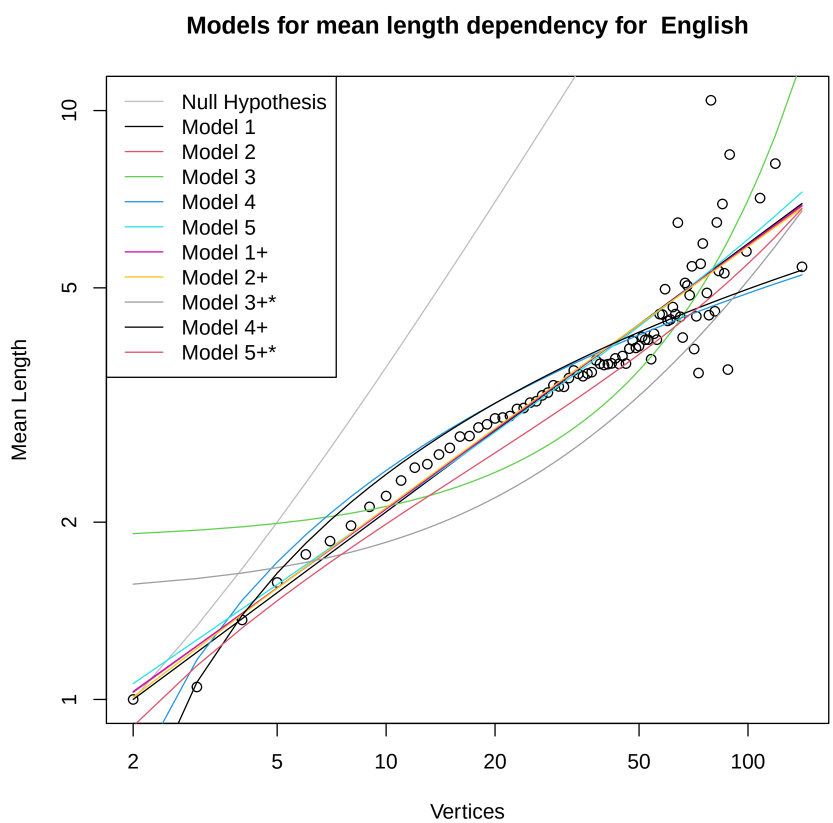
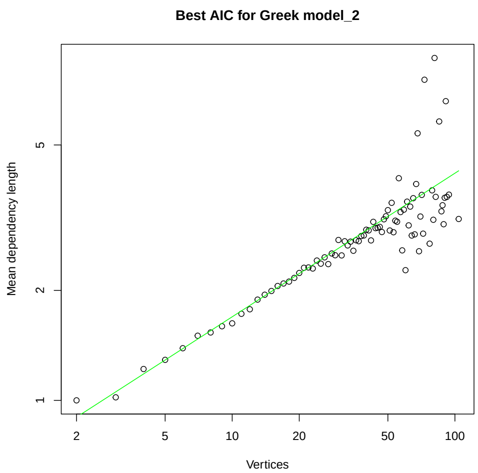
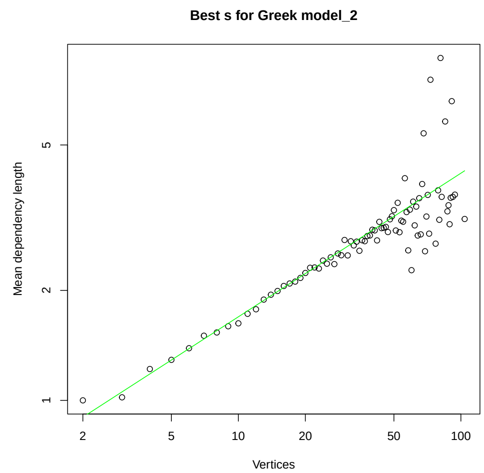


Figure 54: All models for language English.



(a) Model with smallest AIC for Greek.



(b) Models with smallest s Greek.

Figure 55: Models fitted for language Greek

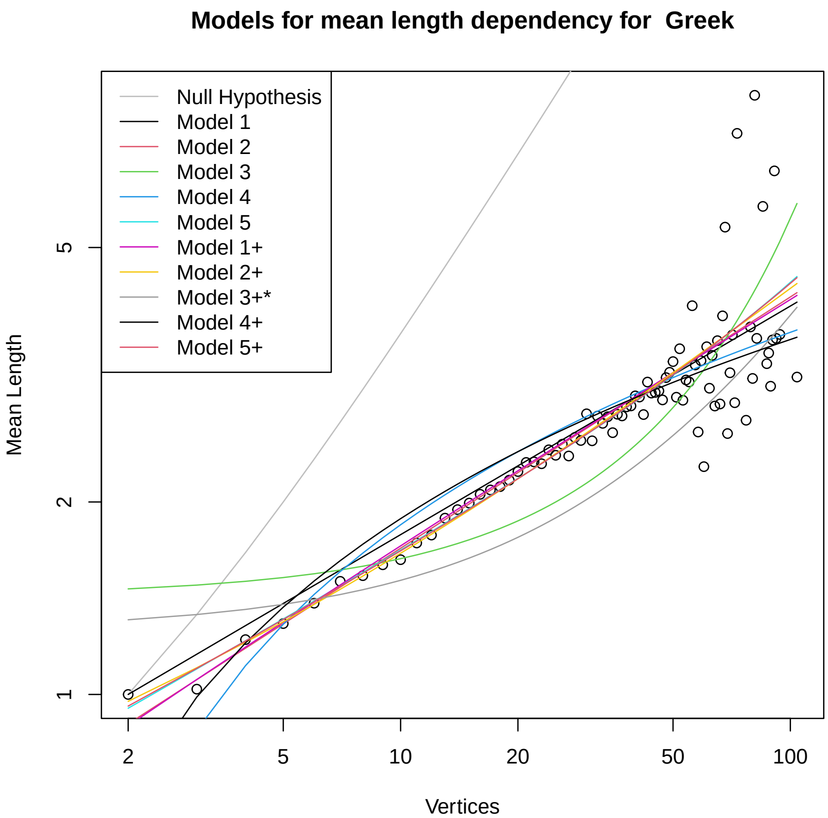
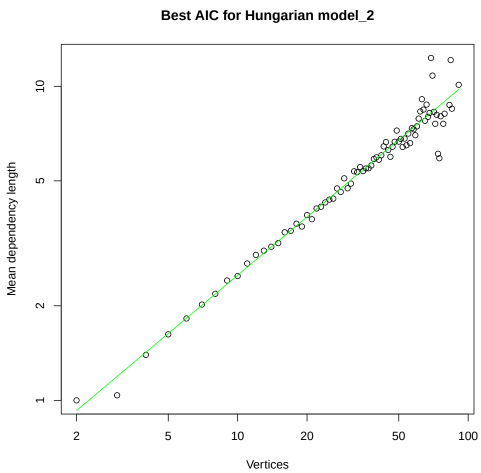
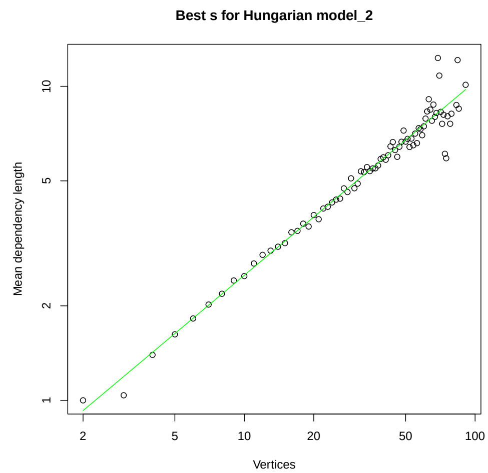


Figure 56: All models for language Greek.



(a) Model with smallest AIC for Hungarian.



(b) Models with smallest s Hungarian.

Figure 57: Models fitted for language Hungarian

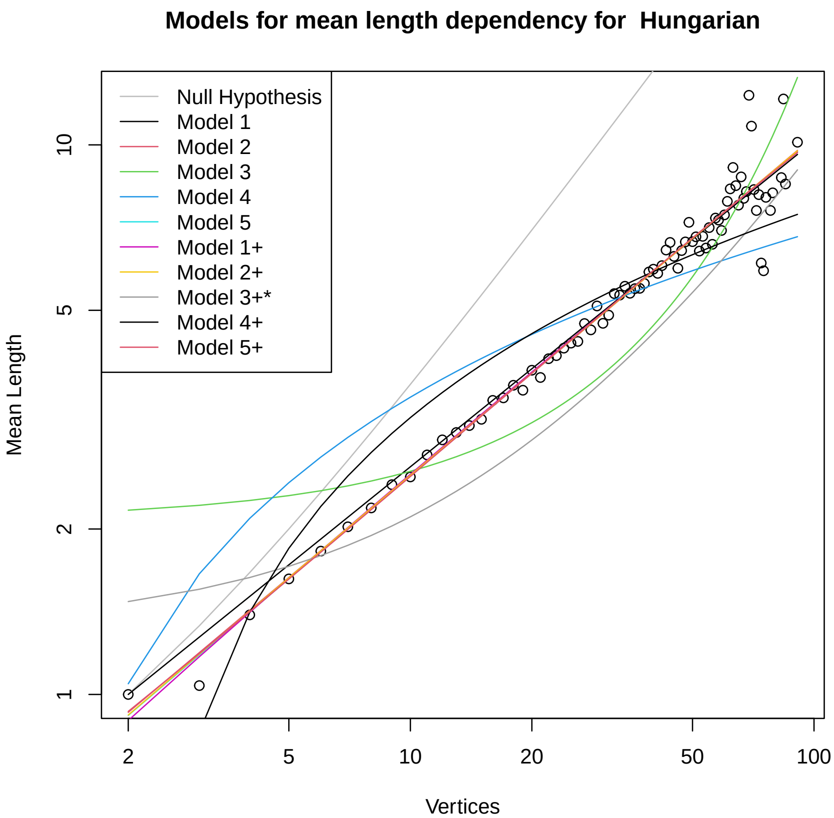
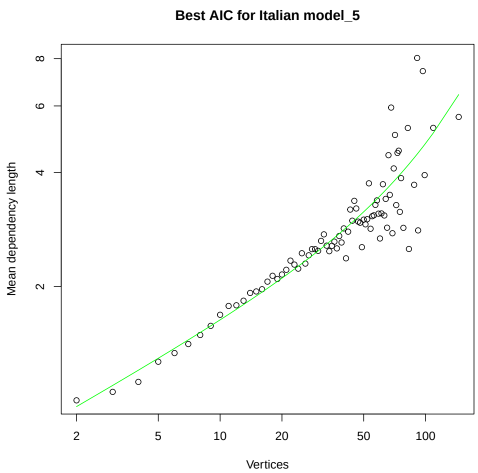
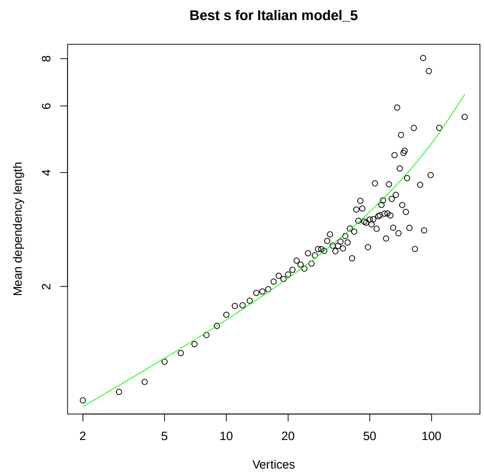


Figure 58: All models for language Hungarian.



(a) Model with smallest AIC for Italian.



(b) Models with smallest s Italian.

Figure 59: Models fitted for language Italian

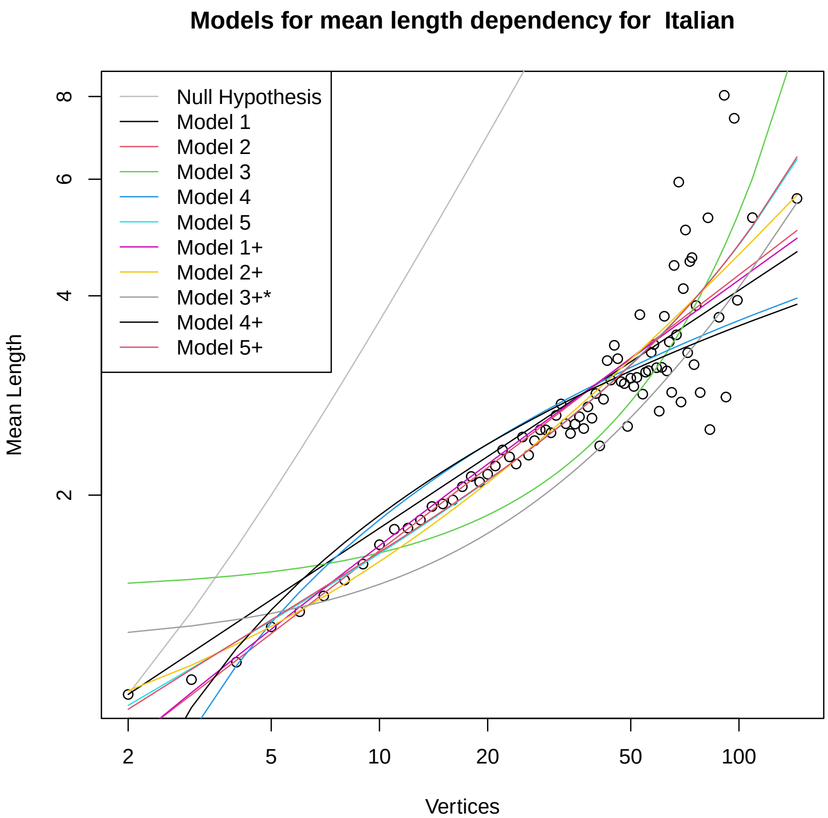
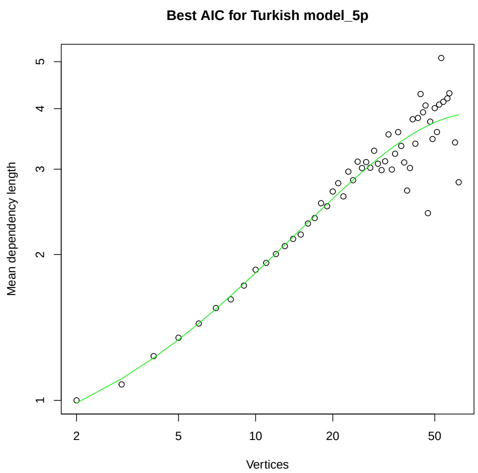
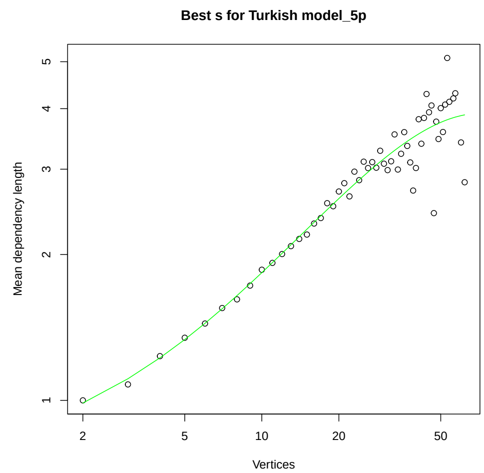


Figure 60: All models for language Italian.



(a) Model with smallest AIC for Turkish.



(b) Models with smallest s Turkish.

Figure 61: Models fitted for language Turkish

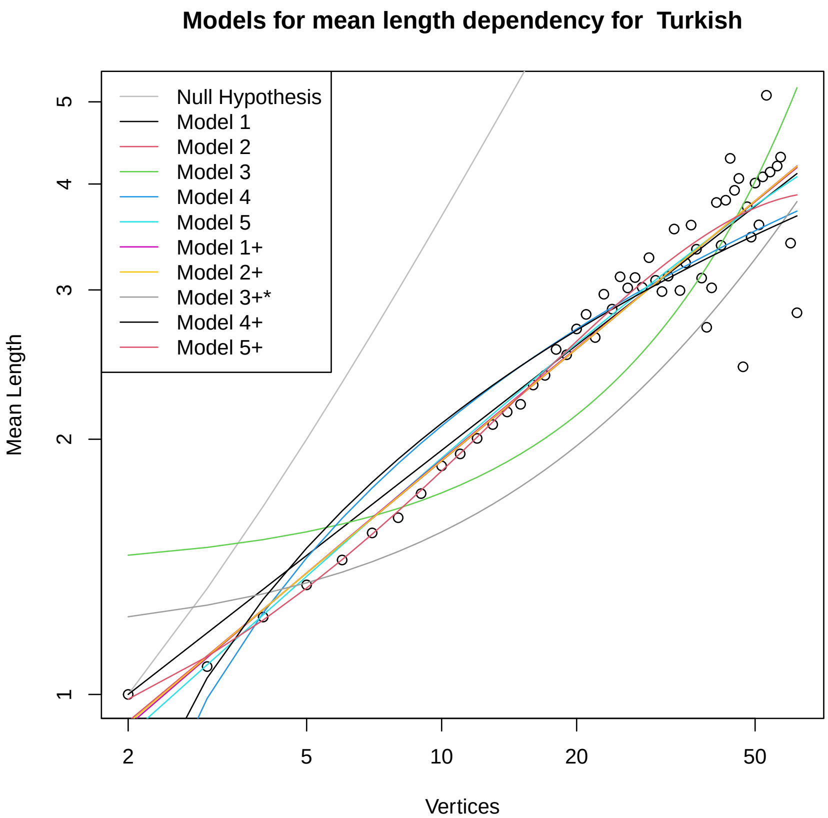


Figure 62: All models for language Turkish.