

# CSN - Lab 4 - Non-linear regression on dependency trees

Eloy Marín, Sonia Petrini

November 10, 2021

## 1 Introduction

The goal of this lab is to fit non-linear functions to syntactic dependency trees from 10 different languages. Specifically, we focused on modeling the degree second moment  $\langle k^2 \rangle$  distribution as a function of the number of vertices (words) of a sentence,  $n$ . To do so, we implemented an ensemble of models, composed of a null model, some *base models* and their counterparts assuming an offset (*+ models*):

- Model 0:  $f(n) = (1 - \frac{1}{n})(5 - \frac{6}{n})$
- Model 1:  $f(n) = (\frac{n}{2})^b$
- Model 2:  $f(n) = an^b$
- Model 3:  $f(n) = ae^{cn}$
- Model 4:  $f(n) = a \log n$
- Model 1+:  $f(n) = (\frac{n}{2})^b + d$
- Model 2+:  $f(n) = an^b + d$
- Model 3+:  $f(n) = ae^{cn} + d$
- Model 4+:  $f(n) = a \log n + d$

## 2 Results

First, we compute the summary of our data, reporting some descriptive statistics of the sentences composing the tree and their second moment degree sequences. **Table 1** shows the summary for all languages, where  $N$  is the number of sentences,  $\mu_n$  and  $\sigma_n$  are the mean and standard deviation of  $n$ , and  $\mu_x$  and  $\sigma_x$  are the mean and standard deviation of  $\langle k^2 \rangle$ . Then, the results shown in the following tables are obtained after fitting the whole ensemble of models to our data. For each model and each language, **Table 2** and **Table 3** show the values of the parameters yielding the best fit, **Table 4** shows the models' residual standard error, **Table 5** shows the AIC values (with the best model marked in blue), and **Table 6** shows the AIC differences w.r.t the best model.

Table 1: Summary

Language	$N$	$\mu_n$	$\sigma_n$	$\mu_x$	$\sigma_x$
<i>Arabic</i>	4108	26.96	20.65	4.16	1.27
<i>Basque</i>	2933	11.34	6.53	4.14	1.09
<i>Catalan</i>	15053	25.57	13.62	4.96	0.82
<i>Chinese</i>	54238	6.25	3.31	3.22	1.07
<i>Czech</i>	25037	16.43	10.72	4.29	1.3
<i>English</i>	18779	24.05	11.22	5.17	0.8
<i>Greek</i>	2951	22.82	14.38	4.6	1.07
<i>Hungarian</i>	6424	21.66	12.57	5.96	1.71
<i>Italian</i>	4144	18.41	13.35	4.34	1.17
<i>Turkish</i>	6030	11.1	8.28	3.76	0.93

Table 2: Estimated coefficients of base models

	1: $(\frac{n}{2})^b$	2: $an^b$		3: $ae^{cn}$		4: $a \log n$
Language	b	a	b	a	c	a
<i>Arabic</i>	0.445	2.708	0.148	4.121	0.002	1.224
<i>Basque</i>	0.614	2.228	0.257	3.635	0.011	1.619
<i>Catalan</i>	0.504	2.762	0.176	4.353	0.004	1.422
<i>Chinese</i>	0.579	2.398	0.207	3.606	0.009	1.485
<i>Czech</i>	0.517	2.404	0.216	4.348	0.004	1.460
<i>English</i>	0.525	2.742	0.192	4.501	0.004	1.510
<i>Greek</i>	0.506	2.668	0.178	4.177	0.004	1.402
<i>Hungarian</i>	0.610	2.615	0.266	4.933	0.007	1.923
<i>Italian</i>	0.511	2.502	0.200	4.219	0.004	1.429
<i>Turkish</i>	0.531	2.546	0.175	3.722	0.006	1.376

Table 3: Estimated coefficients of + models

	1+: $(\frac{n}{2})^b + d$		2+: $an^b + d$			3+: $ae^{cn} + d$			4+: $a \log n + d$	
Language	b	d	a	b	d	a	c	d	a	d
<i>Arabic</i>	0.289	2.266	20.000	0.031	-17.745	24.042	0.001	-20.000	0.703	2.122
<i>Basque</i>	0.449	1.938	20.000	0.053	-18.539	23.441	0.002	-20.000	1.219	1.244
<i>Catalan</i>	0.340	2.452	20.000	0.040	-17.854	24.241	0.001	-20.000	0.916	1.949
<i>Chinese</i>	0.383	1.998	20.000	0.042	-18.169	23.476	0.002	-20.000	0.947	1.676
<i>Czech</i>	0.380	2.150	20.000	0.047	-18.364	24.113	0.001	-20.000	1.086	1.405
<i>English</i>	0.366	2.500	20.000	0.044	-17.936	24.347	0.001	-20.000	1.024	1.842
<i>Greek</i>	0.341	2.324	20.000	0.039	-17.924	24.065	0.001	-20.000	0.896	1.895
<i>Hungarian</i>	0.476	2.710	20.000	0.068	-18.623	24.666	0.002	-20.000	1.679	0.901
<i>Italian</i>	0.364	2.207	20.000	0.044	-18.196	24.051	0.001	-20.000	1.005	1.592
<i>Turkish</i>	0.337	2.098	20.000	0.036	-17.926	23.627	0.001	-20.000	0.799	1.944

Table 4: Residual standard errors

Language	0	1	2	3	4	1+	2+	3+	4+
<i>Arabic</i>	0.439	1.075	0.425	0.549	0.624	0.447	0.407	0.543	0.401
<i>Basque</i>	0.672	1.107	0.472	0.694	0.506	0.527	0.414	0.672	0.393
<i>Catalan</i>	0.797	1.166	0.386	0.603	0.553	0.437	0.342	0.589	0.327
<i>Chinese</i>	0.274	1.107	0.461	0.654	0.587	0.508	0.415	0.643	0.397
<i>Czech</i>	0.968	1.182	0.557	0.775	0.619	0.589	0.533	0.739	0.523
<i>English</i>	1.089	1.266	0.478	0.684	0.609	0.521	0.444	0.664	0.431
<i>Greek</i>	0.628	1.138	0.401	0.605	0.563	0.447	0.360	0.592	0.346
<i>Hungarian</i>	2.463	1.472	0.648	0.936	0.603	0.720	0.589	0.901	0.568
<i>Italian</i>	0.812	1.152	0.493	0.702	0.583	0.532	0.460	0.679	0.448
<i>Turkish</i>	0.160	1.055	0.392	0.562	0.588	0.432	0.356	0.554	0.342

Table 5: AICs

Language	0	1	2	3	4	1+	2+	3+	4+
<i>Arabic</i>	143.857	357.923	137.915	199.106	228.428	150.272	128.855	197.174	123.922
<i>Basque</i>	87.803	130.731	60.083	92.399	64.897	69.278	49.911	90.673	44.598
<i>Catalan</i>	230.788	304.972	93.892	179.255	161.774	117.259	71.551	175.708	61.748
<i>Chinese</i>	12.398	130.708	58.096	87.426	77.463	66.247	50.293	86.974	45.584
<i>Czech</i>	234.955	269.499	144.181	199.572	160.709	153.550	137.503	192.620	133.561
<i>English</i>	266.776	294.290	123.686	186.842	165.475	138.848	111.661	182.641	105.548
<i>Greek</i>	164.017	266.147	89.860	159.728	146.578	108.180	72.625	156.915	64.842
<i>Hungarian</i>	377.875	295.549	163.658	223.204	150.890	180.611	149.043	217.977	142.257
<i>Italian</i>	207.791	268.249	125.122	185.016	152.587	138.009	114.055	180.331	108.629
<i>Turkish</i>	-45.288	170.891	58.941	100.071	104.169	69.919	48.859	99.405	43.475

Table 6:  $\Delta AIC$ 's

Language	0	1	2	3	4	1+	2+	3+	4+
<i>Arabic</i>	19.935	234.001	13.993	75.184	104.506	26.350	4.933	73.252	-
<i>Basque</i>	43.205	86.133	15.485	47.800	20.299	24.680	5.313	46.075	-
<i>Catalan</i>	169.040	243.224	32.144	117.507	100.026	55.512	9.804	113.960	-
<i>Chinese</i>	-	118.310	45.698	75.028	65.065	53.849	37.895	74.576	33.186
<i>Czech</i>	101.394	135.938	10.620	66.011	27.148	19.990	3.942	59.059	-
<i>English</i>	161.229	188.742	18.139	81.294	59.928	33.301	6.113	77.093	-
<i>Greek</i>	99.175	201.305	25.018	94.886	81.737	43.338	7.783	92.073	-
<i>Hungarian</i>	235.619	153.292	21.401	80.947	8.633	38.355	6.786	75.720	-
<i>Italian</i>	99.162	159.620	16.493	76.387	43.958	29.380	5.427	71.702	-
<i>Turkish</i>	-	216.179	104.228	145.358	149.456	115.206	94.146	144.692	88.762

Finally, we can extract the best model obtained for each language and plot it against the real data to visualize the goodness of the fit. Below we show the resulting figures: in red we show the best model, and in blue the validity bounds for second degree moments.

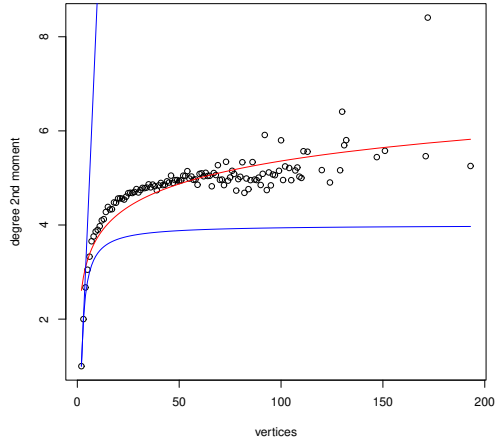


Figure 1: Arabic (4+)

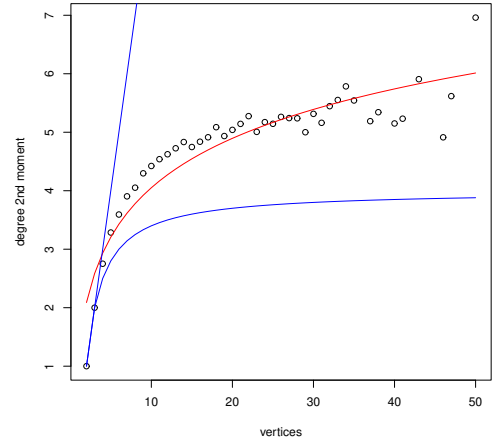


Figure 2: Basque (4+)

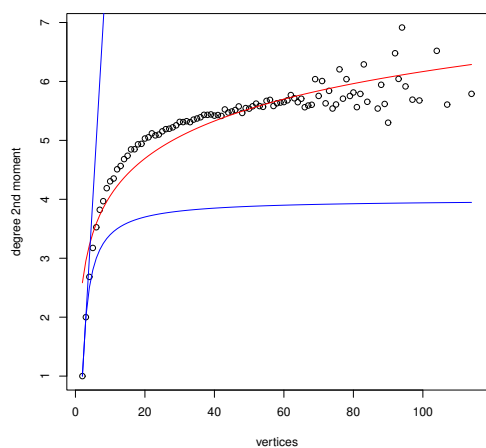


Figure 3: Catalan (4+)

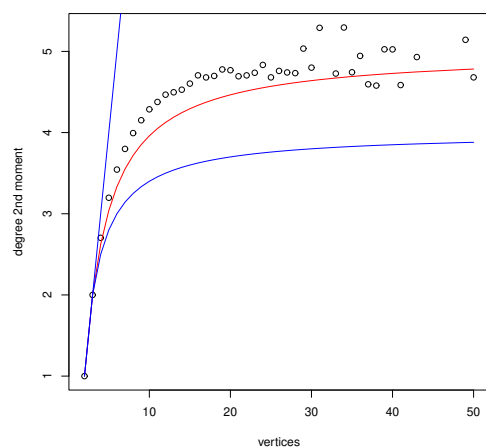


Figure 4: Chinese (0)

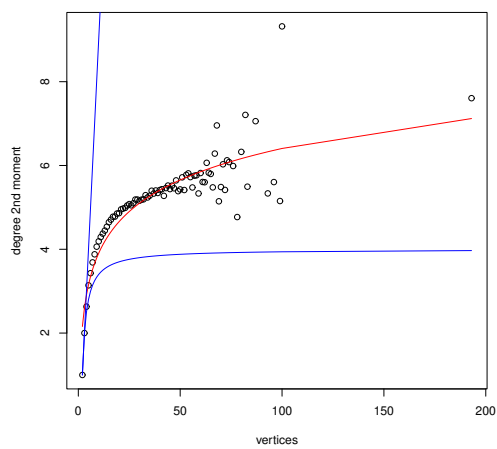


Figure 5: Czech (4+)

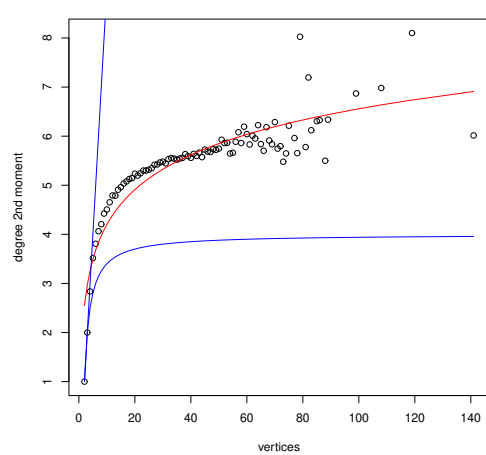


Figure 6: English (4+)

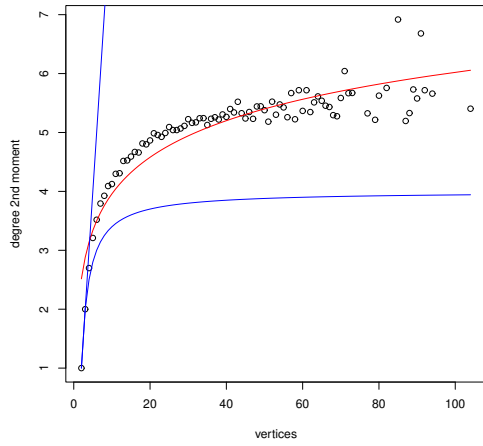


Figure 7: Greek (4+)

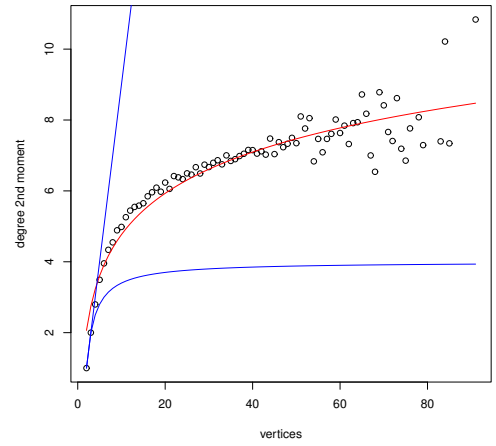


Figure 8: Hungarian (4+)

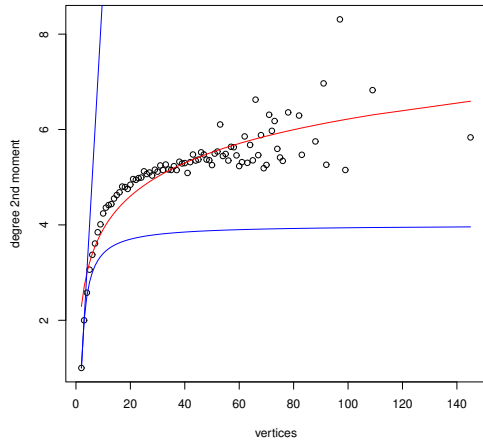


Figure 9: Italian (4+)

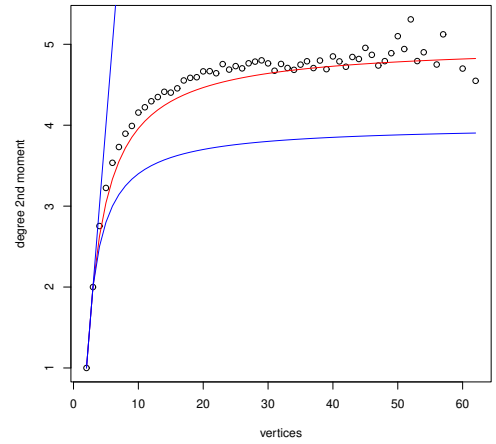


Figure 10: Turkish (0)

### 3 Methods

All the plots displayed in this section are related to the Czech tree for illustrative purposes. The considerations are however generally valid, with a degree of between-language variability.

### 3.1 Extreme values

The first step of our analysis consists in dealing with extreme observations. From the preliminary plotting we can observe the presence of some sentences with abnormally high degree second moment. These observations could belong to a different distribution, or may also derive from an annotation error. We display an example of this in **Figure 11**, where we plot the degree second moment of Czech, and a threshold at 4 standard deviations above the mean. The rule of thumb is generally to use 2 or 3 standard deviations, but as the variance itself is inflated by the presence of outliers, and we don't want to modify the data excessively, we choose a stricter criterion. This procedure also allows us to partly account for the non constant variance across different sentence lengths - an issue discussed below - since outliers have a great influence on variance, and they mainly occur for high values of  $n$ . In the Appendix in section 6 we show the tables with the differences in estimated coefficients, residual standard error, and AIC values after outliers removal.

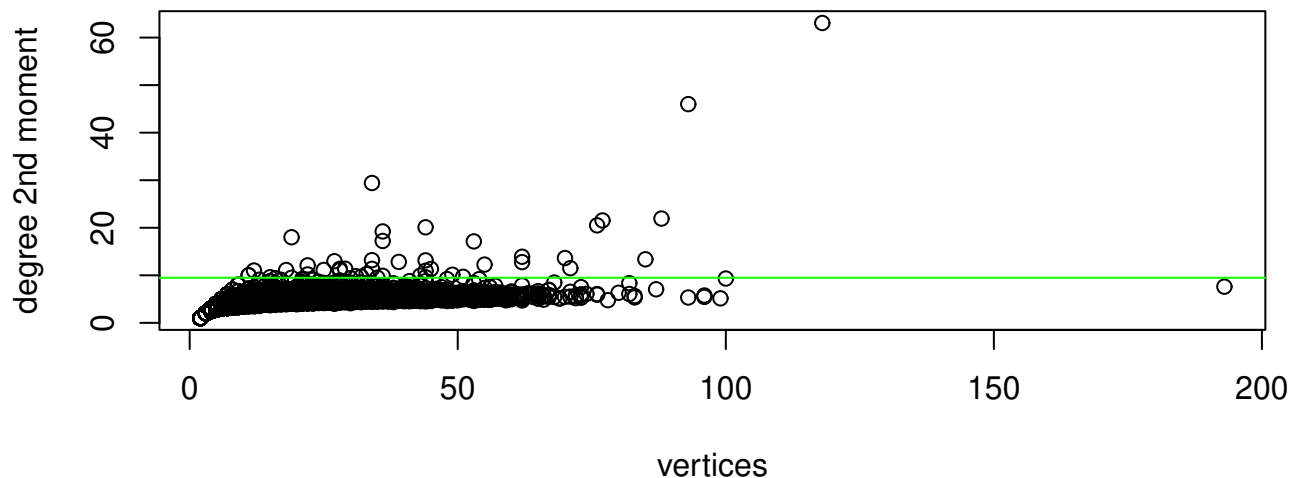


Figure 11: Czech data with threshold at  $\text{mean} + 4 \cdot \text{sd}$

### 3.2 Heteroscedasticity

As the linear least squares model, also the non linear least squares regression assumes normally distributed residuals. However, this assumption is often not verified, especially with respect to constancy of the variance across all values of the explanatory variable. When first fitting the models to the original data, we notice indeed that even after outliers removal heteroscedasticity is present, as **Fig.12** and **Fig.13** show. The residuals are centered around 0, but the variance is evidently changing across the X axis. When this assumption does not hold, the confidence intervals found for the coefficients are not reliable anymore. Since they are computed using the estimated standard deviation, and this is erroneously considered to be constant, the computation itself will be wrong, as it should instead adapt to the different x values.

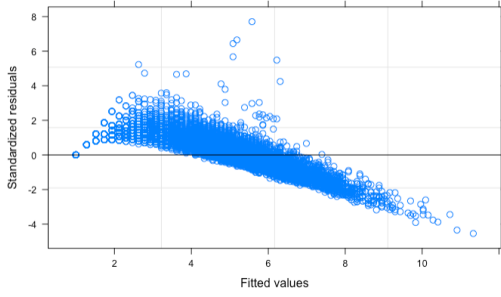


Figure 12: Standardized residuals in model 1 regression for Catalan.

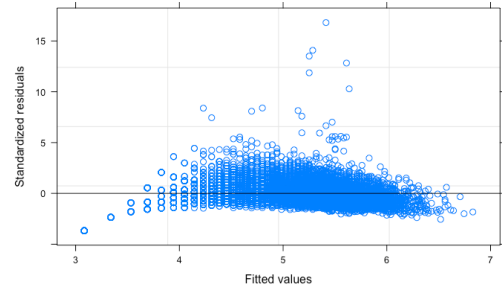


Figure 13: Standardized residuals in model 2 regression for Catalan.

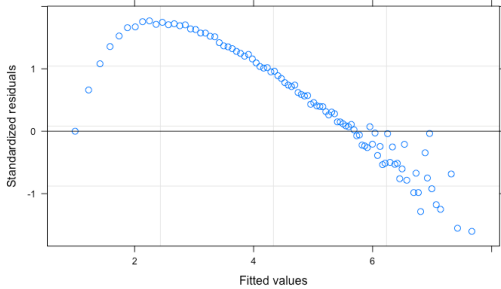


Figure 14: Standardized residuals in model 1 regression for aggregated Catalan.

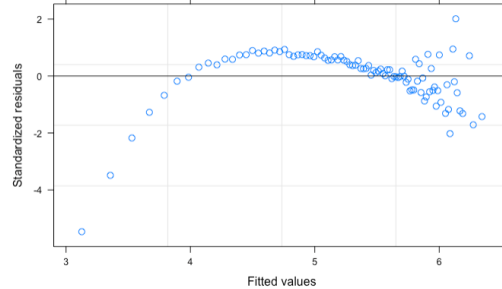


Figure 15: Standardized residuals in model 2 regression for aggregated Catalan.

### 3.3 The influence of long sentences

One way to account for this non-constant variance is to average the data over the vertices, as this always results in a smoothing of the curve. Thus, we proceed with data aggregation, but as we can see in **Figure 14** and **Figure 15**, if this works very well for short sentences, *long sentences* still show a *larger variability* in the residual. The retained large variability is very likely to be due to the small number of sentences longer than 50-60 words in the data: as shown in **Figure 16**, the boxplot of the distribution is skewed towards short sentences, while long ones are rare.

This led us to suspect that long sentences could have a heavy influence on the fitted models, while not being representative of the real distribution; thus, after running the non linear regression we analyze the *DFBETAS*, a measure of the impact of a single observation on the estimated coefficients. This is obtained through a Jackknife procedure, by evaluating the model after the removal of each observation one by one. In **Figure 17** we can see how the most influential observations are indeed the ones linked to larger number of vertices: the degree second moment of the sentence with almost 200 vertices has an impact of more than 80% on the coefficient estimation for the Czech network. A possible extension of this work could involve fitting a different model for long sentences, or down-weighting these observations when running the regressions.



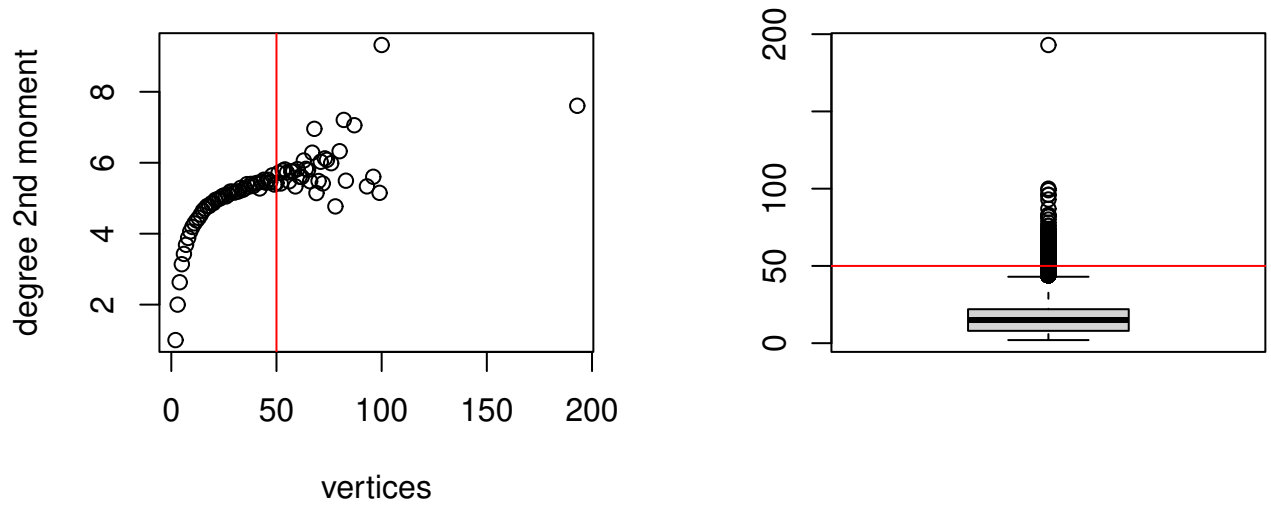


Figure 16: Aggregated data and boxplot for Czech

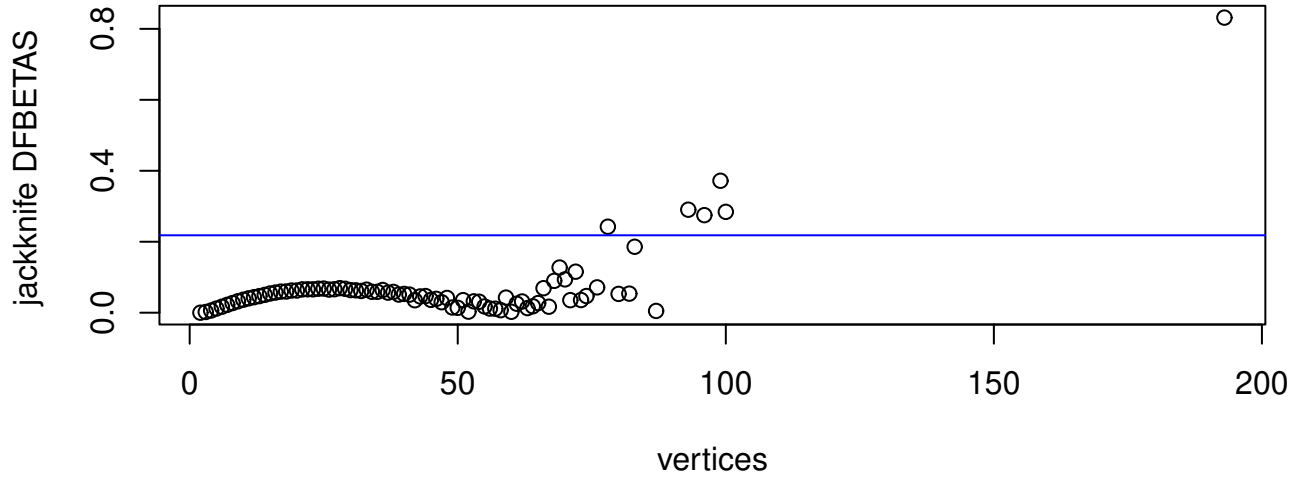


Figure 17: DFBETAS values obtained through jackknife procedure, with the usual choice of a sample-dependent threshold at  $\frac{2}{\sqrt{n}}$  for Czech

### 3.4 Initializing `nls()`

In order to get the starting values for the coefficient’s estimation, we first fit a log-log or log-linear regression of the degree’s second moment over the length of the sentence, depending on the analyzed model. Initializing `nls()` with these values yields to almost a halving of convergence time.

#### 3.4.1 Advanced models initialization

Some advanced models required a more careful choice of the initial parameters used to call `nls()` in order to be able to fit our data. Specifically, models 2+ and 3+ did not work using the default `nls()` algorithm, Gauss-Newton, so we used the *port* algorithm, that corresponds to the *NL2SOL* algorithm [4], setting carefully some upper and lower bounds for the coefficients to optimize.

### 3.5 Robustness of Confidence Intervals

To get appropriate confidence intervals for the coefficients, and assess the impact of heteroscedasticity on their estimation, we relied on `nlsvcovhc()` [1], a modified version of the R function `vcov()` allowing one to compute the asymptotic covariance matrix without assuming homoscedasticity. This function (available in the script `fun.R`) leverages the popular *Huber-White sandwich* estimator [2] on a simulated linear regression, and extends its scope to a non-linear framework. To use `nlsvcovhc()` we fitted the non-linear model with `nlsLM()` [3].

## 4 Discussion

### 4.1 Main results

As Table 6 shows, there is a difference between the quality of the fit from model 0 (null hypothesis) to more advanced models like 2+ and 4+, indicating that real syntactic dependency trees are far from being random. Nevertheless, *Chinese* and *Turkish* seem to behave differently from the other languages, because model 0 is the one that performs the best fit. Both languages have really small sentence lengths and degree second moment compared to the others - as Table 1 shows - which could be one of the main reasons that leads them to be closer to random trees and, therefore, being best fitted by the null hypothesis. This is particularly true for Chinese, which has the largest dataset available, but the second shortest mean sentence length. Overall, while models 2+ and 4+ have very similar AIC values, the latter turns out to have the best fitting for almost every language, excluding the two previously mentioned (*Chinese* and *Turkish*), which nevertheless show the lowest loss of information (AICs). Thus, the relation that seems to best model the degree second moment in the majority of syntactic dependency one-dimensional networks is a logarithmic one (with an offset).

Remarkably, when testing the significance of syntactic dependency networks with respect to two Null Hypotheses in a previous project, these two languages showed the lowest *p-value* with respect to the hypothesis of a switching model graph. *Turkish* was also the only language with a *p-value* non significant at a 95% confidence level with respect to the hypothesis of a randomly generated graph. Thus, it appears like these two languages exploit different syntactic structures, and have also different building blocks, as sentences are remarkably shorter comparing to other languages.

## 4.2 Quality of fitting

All languages have a reasonably good fit, as the residual standard errors are small enough compared to the average values of degree second moment: for instance, model 4+ leads an average error of 0.401 (**Table 4**) for *Arabic*, meaning that given the expected value of 4.16 (**Table 1**), our model could instead yield a 3.759 or 4.561, which are fairly good estimates. The achievement of this good performance is certainly partly due to the outlier removal procedure.

This is more true for certain languages, but as we can see from **Table 9** and **Table 10** in the Appendix in section 6, both the residual standard errors and consequently the AIC values are lower for every language after outliers are excluded. In particular, while in the *Turkish* tree no outliers were identified, for *Czech* this procedure drastically enhanced the fitting. For instance, in **Table 8** we can see how the  $a$  parameter in model 2+ reached a level similar to that of other languages, but only after removing outlying observations.

## 4.3 Heteroscedasticity

Regarding the variability of the variance, which is deeply discussed in section 3, we detected that homoscedasticity does not hold, and we thus decided to account for this issue by fitting the models at an aggregated level. Although this smooths the problem, the influence of the large variability of long sentences' second degree moment limits the overall performance of the models, as it does not allow to confidently estimate the optimal parameters. However, heteroscedasticity seems to be intrinsically belonging to the degree second moment distribution of syntactic dependency networks, as sentences with more than 50-60 words are uncommon in every language. This leads to difficulties in correctly estimating the behaviour on the tail of the distribution, due to the large variability, but also to issues related to non-constant variance.

To evaluate the impact of heteroscedasticity on the coefficients' estimation, we compared the standard deviations produced by the traditional computation, and the ones obtained with the *sandwich* estimator. In **Figure 18** we show the estimated value of  $c$  in model 3+, the traditional CI in orange (computed as  $1.96 * (sd/\sqrt{n})$ ), and the robust CI in red. As expected, the robust bands are wider, and the bigger the difference between the two intervals, the less reasonable is the homoscedasticity assumption. This issue is particularly relevant when the coefficient is close to 0: if the confidence interval (CI) includes 0, then we cannot be sure that the value we estimated is actually different from it. This is the case for 3+ :  $c$ , the coefficient of the sentence length in the exponential model with a set-off, which shows very low values, as displayed in **Table 3**.

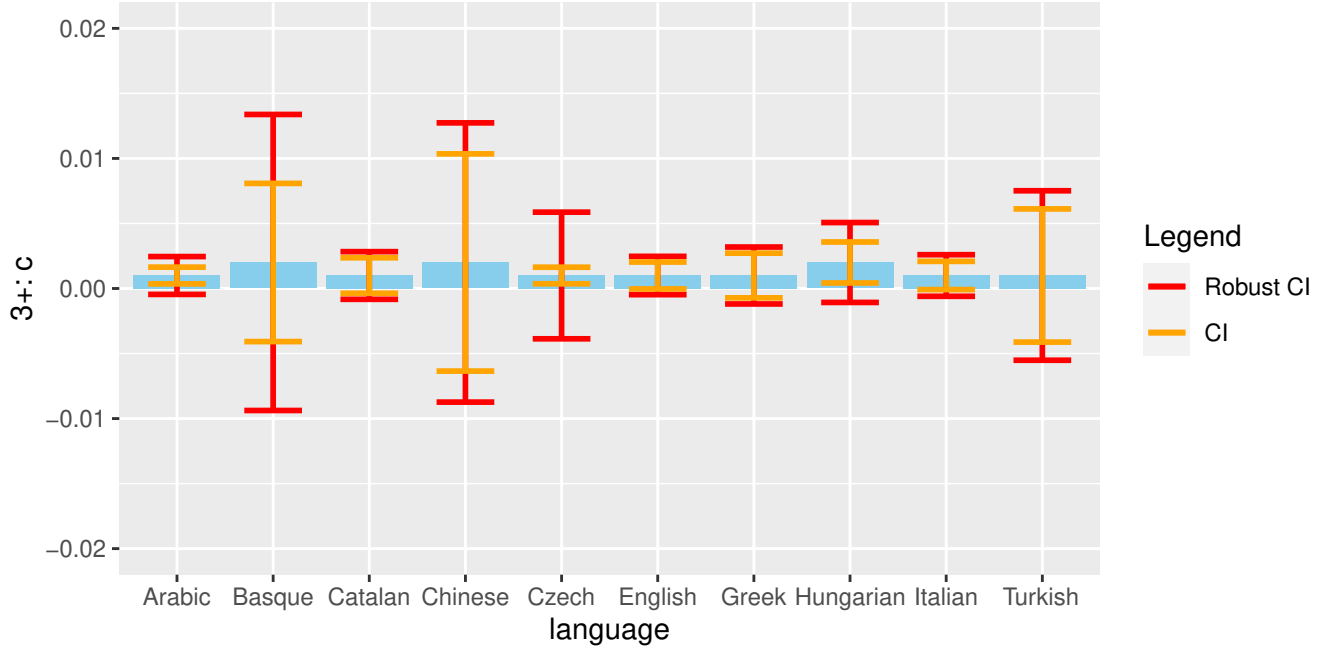


Figure 18: Estimated  $c$  for model 3+, with the original Confidence Interval and the robust ones.

As we can see in **Figure 18**, English has a wider traditional CI compared to Czech, but for the latter there is a noticeable difference between the robust and non robust ones, while for English the intervals are very similar. Thus, the estimation for Czech is very sensitive to the different variability observed in long sentences; indeed, from the summary in **Table 1** we notice how Czech has a lower mean sentence length compared to English, and it shows a greater variability. While for the latter the standard deviation is roughly 46% of  $\mu_n$ , for Czech this ratio raises to 65%, leading it to having a greater heteroscedasticity.

## 5 Conclusions

After extracting some descriptive properties from each language and fitting all networks with different models, we can conclude that most languages resemble and therefore are best fitted with the same model, namely a *logarithmic model with an offset*. Yet, two languages (*Chinese* and *Turkish*) behave in a remarkably different way, obtaining the best fit with the Null model, i.e. the one following the distribution found in uniformly random labelled trees. As broadly discussed, heteroscedasticity and extreme outlying values can have a great impact on the models estimations and performance, hence they are both issues that should to be handled carefully when optimizing the quality of the fit. In conclusion, future work could involve a deeper analysis of long sentences, as compared to short ones they are not so common and they seem to behave differently. Thus, the developing of dependency links within them could be driven by different mechanisms.

## 6 Appendix: handling extreme values

In this section we report the tables with the differences in estimates between the outlier free data and the original data. For each language we removed the following number of outliers:

- *Arabic*: 7
- *Basque*: 2
- *Catalan*: 24
- *Chinese*: 11
- *Czech*: 43
- *English*: 52
- *Greek*: 8
- *Hungarian*: 25
- *Italian*: 7
- *Turkish*: 0

Table 7: Difference in coefficients of base models

	1 : $(\frac{n}{2})^b$	2: $an^b$		3: $ae^{cn}$		4: $a \log n$
Language	b	a	b	a	c	a
<i>Arabic</i>	-0.006	0.160	-0.020	0.046	-0.001	-0.025
<i>Basque</i>	-0.001	0.002	-0.001	-0.006	-0.001	-0.004
<i>Catalan</i>	-	-0.007	0.001	-0.008	-	-0.001
<i>Chinese</i>	-0.001	-0.002	-	-0.003	-	-0.001
<i>Czech</i>	-0.108	2.056	-0.555	-0.085	-0.005	-0.491
<i>English</i>	-0.002	0.016	-0.003	-0.019	-	-0.011
<i>Greek</i>	-0.007	0.128	-0.018	0.039	-0.001	-0.027
<i>Hungarian</i>	-0.009	0.108	-0.018	-0.006	-0.001	-0.047
<i>Italian</i>	-	-0.015	0.001	-0.017	-	-0.003
<i>Turkish</i>	-	-	-	-	-	-

Table 8: Difference in coefficients of + models

	1+: $(\frac{n}{2})^b + d$		2+: $an^b + d$			3+: $ae^{cn} + d$			4+: $a \log n + d$	
Language	b	d	a	b	d	a	c	d	a	d
<i>Arabic</i>	-0.022	0.119	-	-0.004	0.228	0.069	-	-	-0.086	0.246
<i>Basque</i>	-0.001	-0.001	-	-	0.010	-0.003	-	-	-0.008	0.012
<i>Catalan</i>	-	-0.008	-	-	-0.008	-0.008	-	-	0.001	-0.008
<i>Chinese</i>	-	-0.003	-	-	-0.002	-0.002	-	-	-	-0.002
<i>Czech</i>	-0.259	2.576	19.933	-1.036	-20.607	1.419	-0.002	-	-1.910	5.372
<i>English</i>	-0.004	0.003	-	-0.001	0.038	-0.012	-	-	-0.022	0.046
<i>Greek</i>	-0.021	0.092	-	-0.004	0.198	0.061	-	-	-0.087	0.219
<i>Hungarian</i>	-0.017	0.069	-	-0.004	0.231	0.034	-	-	-0.123	0.280
<i>Italian</i>	0.001	-0.017	-	0.001	-0.013	-0.019	-	-	-	-0.012
<i>Turkish</i>	-	-	-	-	-	-	-	-	-	-

Table 9: Difference in residual standard error

Language	0	1	2	3	4	1+	2+	3+	4+
<i>Arabic</i>	-0.360	-0.110	-0.296	-0.246	-0.200	-0.285	-0.308	-0.247	-0.309
<i>Basque</i>	-0.013	-0.003	-0.004	-0.006	-	-0.004	-0.003	-0.005	-0.003
<i>Catalan</i>	-0.003	-0.005	-0.003	-0.002	-0.003	-0.002	-0.003	-0.003	-0.002
<i>Chinese</i>	-0.002	-0.002	-0.001	-0.001	-0.002	-0.001	-0.002	-0.001	-0.001
<i>Czech</i>	-6.268	-5.157	-5.796	-5.729	-5.992	-5.782	-5.820	-5.651	-6.059
<i>English</i>	-0.063	-0.024	-0.056	-0.050	-0.036	-0.053	-0.059	-0.050	-0.060
<i>Greek</i>	-0.287	-0.078	-0.255	-0.193	-0.172	-0.236	-0.275	-0.195	-0.279
<i>Hungarian</i>	-0.274	-0.148	-0.324	-0.263	-0.326	-0.302	-0.349	-0.266	-0.355
<i>Italian</i>	-0.010	-0.011	-0.007	-0.008	-0.007	-0.008	-0.006	-0.008	-0.006
<i>Turkish</i>	-	-	-	-	-	-	-	-	-

Table 10: Difference in AIC

Language	0	1	2	3	4	1+	2+	3+	4+
<i>Arabic</i>	-144.880	-26.352	-128.250	-90.222	-68.775	-119.543	-136.223	-91.806	-138.450
<i>Basque</i>	-1.637	-0.236	-0.728	-0.745	-0.124	-0.714	-0.755	-0.700	-0.767
<i>Catalan</i>	-0.876	-0.739	-1.231	-0.823	-1.019	-1.165	-1.378	-0.895	-1.403
<i>Chinese</i>	-0.547	-0.128	-0.223	-0.145	-0.184	-0.204	-0.262	-0.157	-0.271
<i>Czech</i>	-365.089	-308.238	-434.937	-383.681	-424.440	-426.077	-442.591	-388.484	-451.793
<i>English</i>	-9.797	-3.266	-19.582	-12.375	-10.176	-17.223	-22.095	-12.646	-22.928
<i>Greek</i>	-66.679	-14.483	-85.536	-49.562	-47.484	-74.281	-98.141	-50.866	-102.231
<i>Hungarian</i>	-17.113	-15.443	-65.662	-40.066	-69.963	-56.755	-75.367	-41.859	-78.576
<i>Italian</i>	-2.157	-1.575	-2.303	-1.852	-1.838	-2.347	-2.272	-2.037	-2.224
<i>Turkish</i>	-	-	-	-	-	-	-	-	-

## References

- [1] <https://matloff.wordpress.com/2015/06/07/heteroscedasticity-in-regression-it-matters/>
- [2] White, Halbert (1980), "A *Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*", *Econometrica*, 48 (4): 817–838
- [3] <https://www.rdocumentation.org/packages/minpack.lm/versions/1.2-1/topics/nlsLM>
- [4] [https://people.sc.fsu.edu/~jburkardt/f77\\_src/nl2sol/nl2sol.html](https://people.sc.fsu.edu/~jburkardt/f77_src/nl2sol/nl2sol.html)