Menzerath's law: Is it just regression toward the mean?

Jiří Milička[1*]

[1] Institute of the Czech National Corpus, Charles University
[*] Corresponding author's email: jiri@milicka.cz

DOI: https://doi.org/10.53482/2023_55_409

ABSTRACT

The study revisits the Menzerath's Law, which articulates the inverse relationship between the length of constructs and the mean length of their constituents. This relationship is famously modelled by Gabriel Altmann's model, which combines power and exponential relations. His formulas have been widely used to describe this relationship across linguistics and biology, however, there is no satisfactory explanation for his model. Therefore, the paper proposes shifting our perspective to examine directly the relationship between the number of constituents in a construct and the number of subconstituents in the same construct. This relationship may be explained by a simple model based on linear regression, which leads to a hyperbolic model of the Menzerath's Law. This approach is successful for several datasets, but insufficient for others.

Menzerath's Law, Menzerath-Altmann Law, MAL, regression to mean, linear model

1 Introduction

Menzerath's Law describes the relationship between the length of text segments and the mean length of their subsegments (i.e. constructs and their constituents), a principle that applies to various levels and has been confirmed in numerous languages. The law is named after Paul Menzerath, who was the first to notice the peculiar relation between the length of a syllable and its duration,[1] as well as between the length of a word and the mean length of its syllables.[2]

The law is also known as the Menzerath-Altmann Law (MAL), in honor of Gabriel Altmann. Altmann developed models of this relationship, popularized the concept among quantitative linguists, and most significantly, recognized that the model applied to more than just syllables and phonemes. He described the generalized form as "the longer a language construct the shorter its components (constituents)" (Altmann, 1980, p. 1).

[1] "... a sound is the shorter the longer the whole in which it occurs" (Menzerath, 1928, p. 104), as translated in Altmann (1980, p. 1). Also formulated as: "... the more sounds in a syllable the smaller its relative length" (Menzerath, 1928, p. 104), translation by Altmann (1980, p. 1).

[2] "The relative number of sounds in the syllable decreases as the number of syllables in the word increases, or said differently: the more syllables in a word, the shorter (relatively) it is"(Menzerath, 1954, p. 100), translation by Altmann (1980, p. 1).

To be more specific, the relationship is between the number of constituents in a construct and the *mean* number of subconstituents within these constituents. Altmann's model (1980, p. 3) expresses this relationship by the equation

$$\bar{L}_{n-1,n-2} = aL_{n,n-1}^{b} e^{cL_{n,n-1}}. \tag{1}$$

In this equation, n refers to the level of constructs (e.g. words), while $n-1$ denotes the level of constituents (e.g. syllables), and $n-2$ denotes the level of subconstituents (e.g. phonemes). Hence the term $L_{n,n-1}$ represents the length of the construct in terms of its constituents (e.g. the number of syllables in a word). Meanwhile, $\bar{L}_{n-1,n-2}$ represents the mean length of the constituent in terms of its subconstituents, for example the mean number of phonemes in a syllable.

Many studies have found that when assuming $c = 0$, the abbreviated form of the model provides a satisfactory fit for the data:

$$\bar{L}_{n-1,n-2} = aL_{n,n-1}^{b}. \tag{2}$$

To give an example, the following formula describes the Menzerath-Altmann law on phoneme-syllable-word level:

$$\bar{L}_{syllable,phoneme} = aL_{word,syllable}^{b}. \tag{3}$$

It has been discovered that the model can be applied to almost any conceivable method of text segmentation — phonemes, morphemes (Gerlach, 1982; Milička, 2014; Pelegrinová et al., 2021), words, phrases (Mačutek et al., 2021; Mačutek et al., 2017), clauses (Buk and Rovenchak, 2008) and sentences (Milička, 2015; Motalová, 2022), but also to non-human language — geladas (Gustison et al., 2016; Semple et al., 2022) and there were also various attempts at applying MAL outside linguistics, mostly biology (Altmann, 2014; Altmann and Schwibbe, 1989; Semple et al., 2022).

Since real-world datasets of these relationships tend to be noisy, numerous models, not just Altmann's, fit the empirical data. However, very few models have actually been tested, and even those that have been published, typically bear some connection to the original Altmann models (Buk and Rovenchak, 2007; Kułacka and Mačutek, 2007; Mačutek and Rovenchak, 2011). A hyperbolic model (4, Figure 1) has successful fit to many datasets on various levels of segmentation and languages (Milička, 2014):

$$\bar{L}_{n-1,n-2} = \frac{a}{L_{n,n-1}} + b. \tag{4}$$

Actually, already Menzerath himself modelled the relation as a hyperbolic one, however, it is a bit obfuscated. In his book from 1954 he does not analyze the MAL relationship, but the relationship of number of syllables in a word and mean number of phonemes in word. I.e. the dependent variable is not

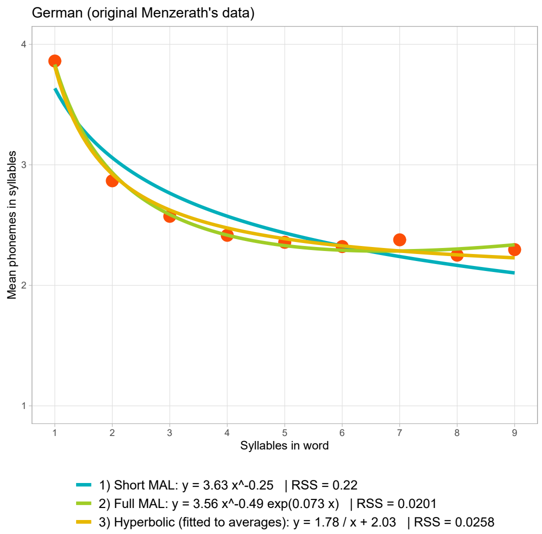


Figure 1: Three models of Menzerath's relation fitted to the original Menzerath's data on the phoneme-syllable-word level (Menzerath, 1954, p. 96). Residual sum of squares is reported as it scales with the main objective of the fitting function.

$\bar{L}_{n-1,n-2}$ but $\bar{L}_{n,n-2}$, in this case it means that the dependent variable is mean number of phonemes in word instead in syllables.

But it does not matter, since the mean number of phonemes in a word can be calculated as the mean number of phonemes in syllable times number of syllables. To be more general, $\bar{L}_{n,n-2} = \bar{L}_{n-1,n-2}L_{n,n-1}$. This means we get the model for this relationship by multiplying both sides of the equation 4 by length of construct $L_{n,n-1}$. This multiplication makes the hyperbolic model linear:

$$\bar{L}_{n-1,n-2}L_{n,n-1} = \frac{aL_{n,n-1}}{L_{n,n-1}} + bL_{n,n-1}$$
$$\bar{L}_{n,n-2} = a + bL_{n,n-1}. \tag{5}$$

Figure 2 shows the actual Menzerath's data (page 108): the empirical dataset looks fairly linear, so Menzerath used linear regression to model it.

The linear model can be interpreted easily and straightforwardly: as the length of a construct increases in terms of its constituents, its length in terms of subconstituents also increases at a steady and consistent rate. The more syllables a word has, the more phonemes it contains in proportion. This relationship seems to align with our intuition, except for the parameter a. Menzerath refers to this parameter as an *inexplicable additive constant* (*unarklärliche additive Konstante*) and feels that it requires an explanation (Menzerath, 1954, p. 111). In order to provide this explanation, Menzerath suggests that each word contains one *core syllable* (*Kernsilbe*), which is longer than the other syllables in the word. For instance, monosyllabic words are composed of only the core syllable, which is why they are relatively long.

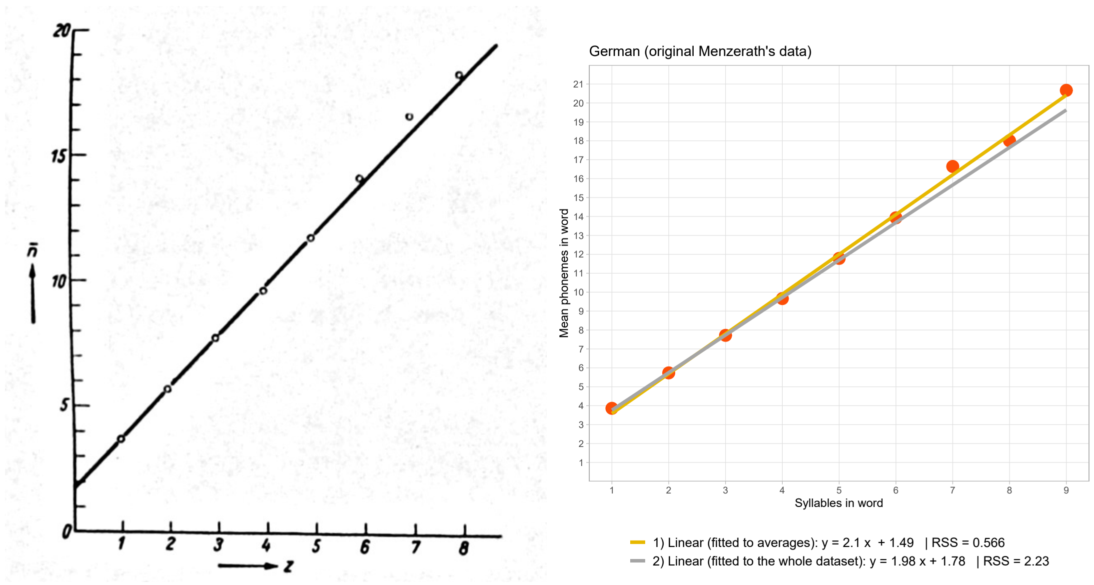


Figure 2: Original Menzerath's linear model (reprint from Menzerath (1954, p. 108)), right is recreation of the data points from his dataset (p. 96) . His model $y = 2x + 1.9$ (ibid.) do not seem to match none of our linear models, presumably because he excluded the last data point. Residual sum of squares (RSS) is calculated in respect for averaged points.

Bisyllabic words, in contrast, consist of one core syllable and one ordinary syllable.

This explanation appears plausible and aligns with our experience, even when considering other units: there are core morphemes in words (root or base morphemes), and core words in clauses (the vast majority of clauses contain a predicate).[3] If we really try, we would be able to find something like core clauses in sentences etc. . .

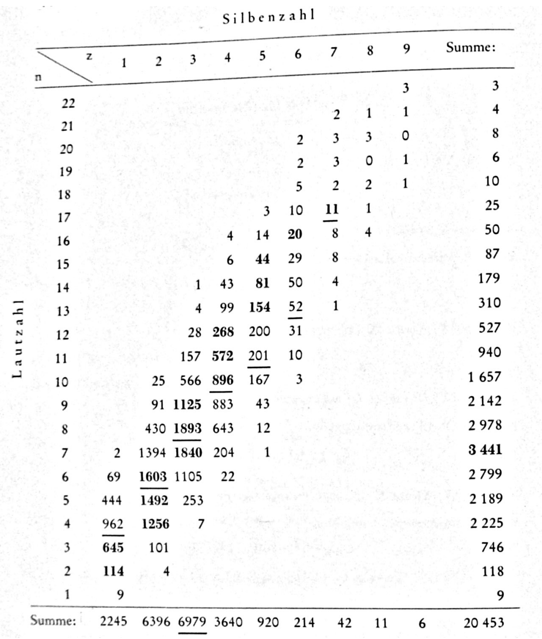
The paper by Milička (2014), which further develops the same formula, bases its explanation of the constant a upon Reinhard Köhler's idea of structure information. Köhler posits that this information is stored in constituents besides constructs (Köhler, 1984). A year later, a more generalized approach was presented in Milicka's PhD thesis (2015), in which the parameter a was examined from the perspective of the Theory of Communication.

However, it seems that no explanation is actually necessary in this instance, since the parameter a can be interpreted through the concept of Galtonian regression to the mean.

2 Paul Menzerath Meets Francis Galton

We are fortunate that Menzerath not only shared the means of the construct lengths but also provided the entire joint distribution, i.e. a table which states how many words of certain lengths he found. For

[3] Actually, predicates are typically short while they consist of high-frequency verbs. This observation aligns with the finding that at the syllable-word-clause level, Menzerath's law is inverted, showing an increasing function, as seen in Figure 7 of Wang and Chen, 2022, Figure 7.



Silbenzahl

Lautzahl n \ z	1	2	3	4	5	6	7	8	9	Summe:
22									3	3
21							2	1	1	4
20						2	3	3	0	8
19						2	3	0	1	6
18						5	2	2	1	10
17					3	10	**11**	1		25
16				4	14	**20**	8	4		50
15				6	**44**	29	8			87
14			1	43	**81**	50	4			179
13			4	99	**154**	52	1			310
12			28	**268**	200	31				527
11			157	**572**	201	10				940
10		25	566	**896**	167	3				1 657
9		91	**1125**	883	43					2 142
8		430	**1893**	643	12					2 978
7	2	1394	**1840**	204	1					**3 441**
6	69	**1603**	1105	22						2 799
5	444	**1492**	253							2 189
4	962	**1256**	7							2 225
3	**645**	101								746
2	**114**	4								118
1	9									9
Summe:	2245	6396	6979	3640	920	214	42	11	6	20 453

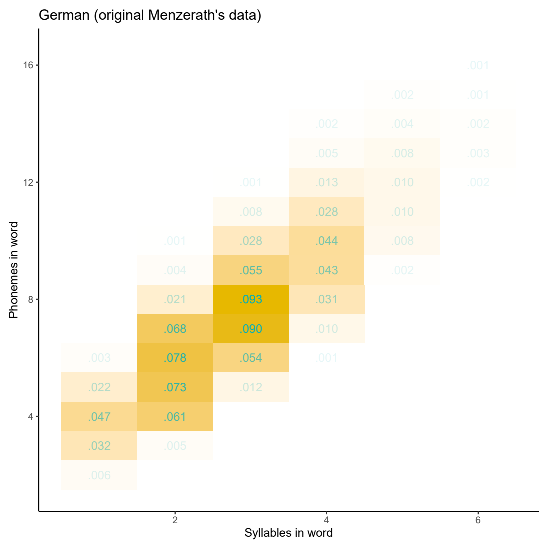


Figure 3: Original Menzerath's joint distribution of his dataset (reprint from Menzerath (1954, p. 96)). The chart on right represents the same data.

example he found 101 words that contain 2 syllables and 3 phonemes, 1893 words with 3 syllables and 8 phonemes etc., the complete distribution can be viewed in the table reprinted in Figure 3 (Menzerath, 1954, p. 96). These data make it possible to directly reanalyze his findings.

In order to get parameters of a linear model, the line is shifted and rotated until "discrepancy" between the line and the data points is as small as possible. There are several metrics of this "discrepancy". The most favourite method for fitting is the *least squares* method, where the metric is sum of squared vertical distances between the line and the data points. Gabriel Altmann used this metric in his seminal paper on the topic (Altmann, 1980) and as far as I know everybody who fitted his model to Menzerathian relation did so.

In studies of Menzerath's law, the method of the least squares has traditionally been used to model the means, not the complete joint distribution (meaning the dataset as shown in the Figure 3), and I followed this tradition in the models presented so far in this study (Figures 1 and 2), with the exception of the yellow line in Figure 2, where the whole dataset was used. As can be observed, the two lines in the right chart of the Figure 2 are quite similar — it does not matter much, whether the model is fitted to the averages or to the whole dataset of joint distribution. This is because the least squares method inherently targets central values of the dependent variable.[4]

[4]By the way, this means we can use Galton's estimators to determine the parameters of the hyperbolic model for the Menzerath's law. To obtain these parameters, we need the correlation between the two variables as well as the mean and standard deviation of the marginal distributions. Consequently, the parameters of the hyperbolic model are straightforward to interpret.

Fitting the entire dataset with a linear model in this manner is useful for highlighting the regression toward the mean, a statistical artifact produced by averaging the values of the dependent variable (i.e., "vertically"). The line is actually called *regression* because of this. The regression toward the mean was famously discovered by Francis Galton who noticed that short people tend to have offsprings who are relatively taller than they are, and, surprisingly, also tall people have offsprings who are on average shorter than they are (Galton, 1886). This phenomenon actually resembles Paul Menzerath's observation that short words, when measured in syllables, do not appear as short when their length is counted by the number of phonemes. Conversely, words with a large number of syllables have relatively fewer phonemes on average. Such a phenomenon manifests whenever two variables are imperfectly correlated. The number of syllables in a word *is* imperfectly correlated with the number of phonemes in that word. The question is, whether the imperfect correlation can explain the parameter a completely.

Averaging the values, as we do in case of Menzerath's law, does not respect the way how the data points actually originated. We do not know which stochastic process best models the data's origin, but we can be sure that the random processes did not take place solely in the vertical direction. That is to say, it is not as if the independent variable was predetermined and all the "errors" can be attributed solely to the dependent variable. The dependent-independent dichotomy is in this case just a technical characteristic. We regard the number of phonemes in a word as being dependent on the number of syllables in the same word just for historical reasons, it is not as if some *Genius of the Language* first determined the word's length in syllables and then selected the appropriate number of phonemes to match it. The evolutionary process was presumably very chaotic and many phenomena had some effects on both variables. This situation actually mirrors Galton's data on height inheritance — there is some shared genetic material, which forms the basis for correlation, however, the actual heights of both the ancestor and their offspring are influenced by a multitude of other stochastic events, which affect both variables independently.

Therefore, let us fit the linear model using a method that accounts for errors in both vertical and horizontal directions, i.e. the method aiming to minimize the sum of squared distances between the line and the data points. We are interested in the Euclidean distance between the line and the data points. The metric is called *total least squares* (and the method is called *orthogonal fitting*).

Let us look at the difference between the blue and red lines in the Figure 4. The blue line represents the classical least squares regression, while the red one shows the linear model fitted by minimizing the total least squares. The linear model represented by the blue line has a notably pronounced parameter a (commonly referred to as the intercept). This intercept nearly disappears when we fit the linear model orthogonaly, diminishing to a value almost 30 times smaller. Suddenly, instead of two phonemes, the size is a mere 0.07 phonemes.

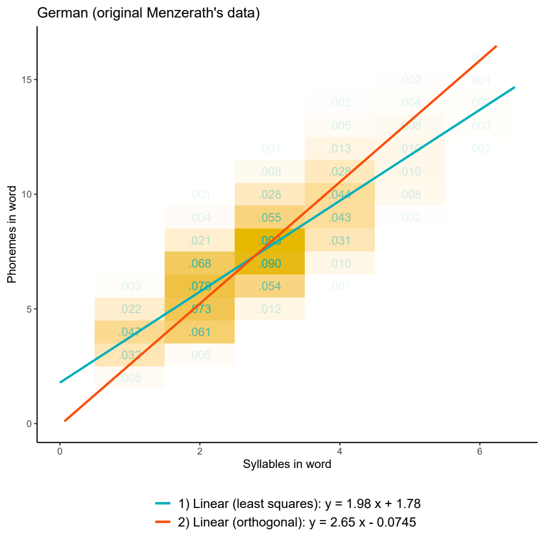


Figure 4: Linear model fitted to the original Menzerath's data (Menzerath, 1954, p. 96). Least squares fitting method and total least squares (orthogonal) method are put here in contrast.

While this result might be coincidental, we will dedicate the remainder of this study to empirically exploring this phenomenon across different texts and levels.

3 Material

It is still debatable whether the Menzerath's Law should be applied to tokens or types (Stave et al., 2021) and the difference between the two is quite pronounced (Mikros and Milička, 2014). Menzerath himself used data from a dictionary, indicating that his measurements were based on types. Gabriel Altmann (1980) also used dictionaries in his research on the topic, as did other pioneers in the field. However, many subsequent studies have applied the MAL directly to tokens. Since tokens cannot be considered independent trials, the statistical analysis and potential explanations are more complex than for types. Therefore, I prefer to use types, but for the sake of completeness, I will also present the results for tokens to illustrate the importance of this consideration.

Since we need to analyze the entire joint distribution, we can only use datasets where this distribution is available, e.g. Mikros and Milička (2014) and Milička (2014, 2015). Consequently, the number of tests for this hypothesis is limited; however, all the necessary scripts are available online so that the study can be replicated and repeated on other texts.[5]

[5]The archive is available at http://milicka.cz/kestazeni/MenzerathRegression.zip. The archive also includes the datasets that were used in the study. The first column in each table contains the actual forms of the given construct, for example a word tokens. The second column contains the number of its subconstituents, such as the number of phonemes. The third column contains the number of its constituents, such as the number of syllables. While the first column can be left empty, doing so will

4 Results

The first dataset used in this analysis allows for the examination of Menzerath's Law at the phoneme-syllable-word level for Greek blog posts, making it comparable to the original Menzerath's dataset discussed in previous sections. This dataset comes from Mikros and Milička (2014), although only a subset of the expansive dataset was employed. As illustrated in Figure 5, the difference between the joint distribution measured by types and tokens is relatively minor: In both cases, the intercept left by the orthogonal fitting is approximately one tenth of the parameter b, which is slightly higher than what was observed in the German data.

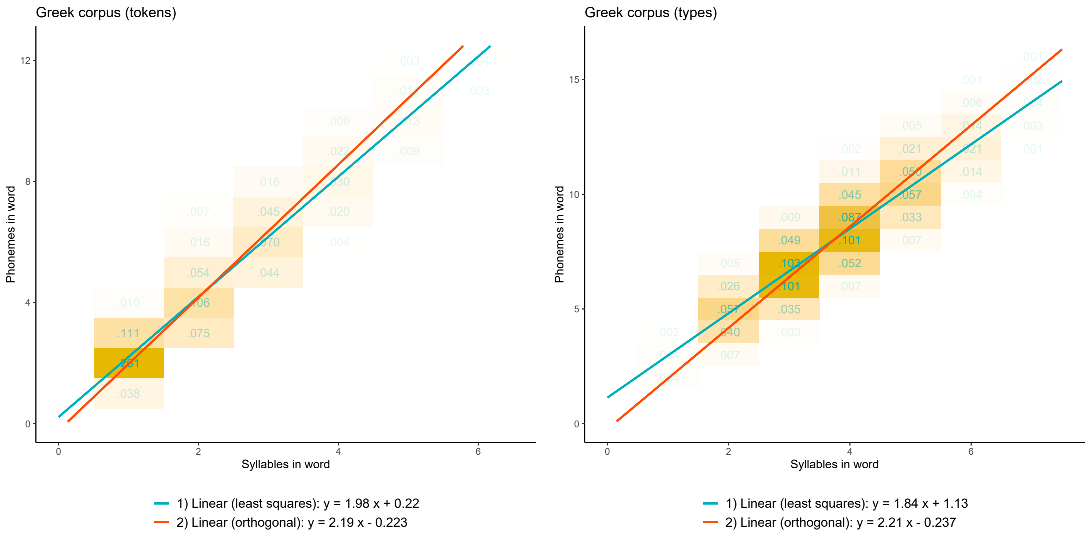


Figure 5: Phoneme — syllable — word level (Greek data, series of blog posts).

Menzerath's law can also be observed at the morpheme level, much like at the syllable level (Pelegrinová et al., 2021). Indeed, in many languages, morphemes typically equate to a single syllable. Therefore, I have incorporated several datasets that include the morpheme level, taken from the PhD dissertation by Milička (2015, Appendix C).

The first dataset allows for the exploration of the Menzerath's law at the phoneme-morpheme-word level in Czech text, specifically the novella *Krysař* by Viktor Dyk. The segmentation was done by Zuzana Komrsková and was initially published in Milička (2014). This dataset exemplifies that the hypothesis of zero intercept holds true for types rather than tokens. On tokens, the absolute intercept remains very large, but when looking at types,the absolute intercept is extremely small, even smaller than in the original Menzerath's dataset (Figure 6).

Let us stay at the phoneme-morpheme-word level. The next dataset is also taken from Milička (2014)

cause the scripts to operate solely on tokens rather than on types.

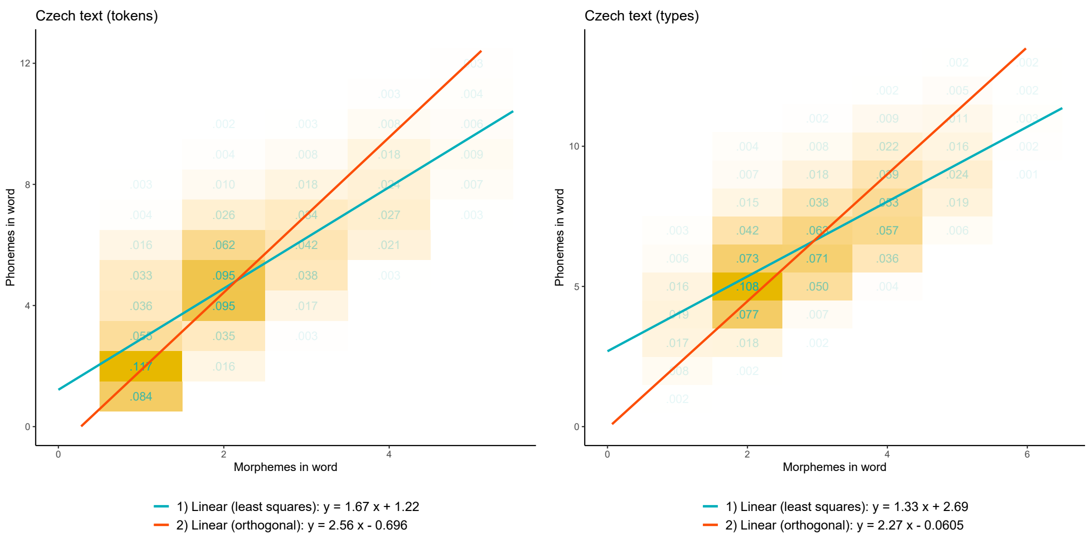


Figure 6: Phoneme — morpheme — word level (Czech data, short novel *Krysař* by Viktor Dyk).

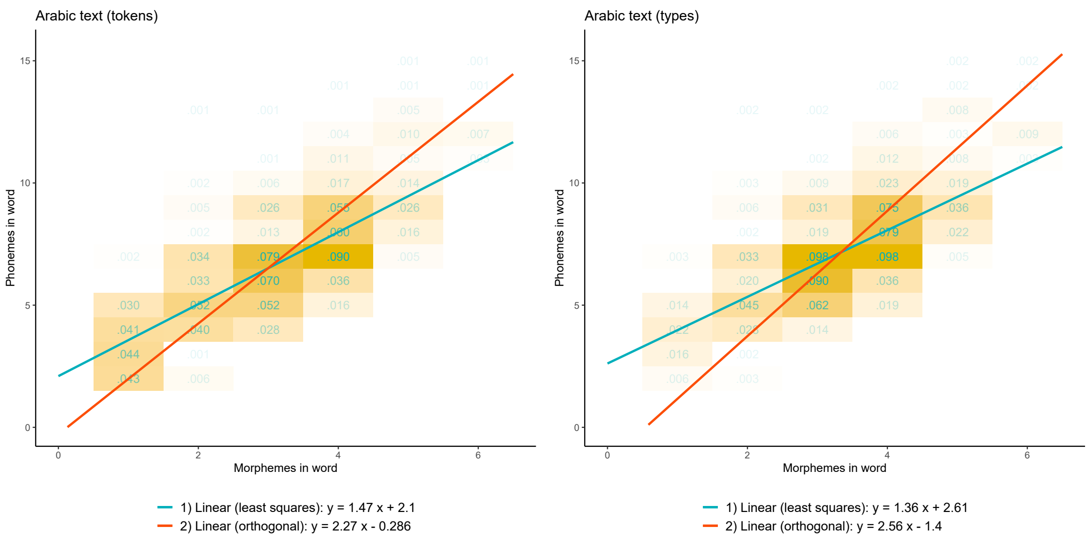


Figure 7: Phoneme — morpheme — word level (Arabic data, part of *Kalīla wa-Dimna* by Ibn al-Muqaffa').

and it is based on a chapter from the famous Arabic book *Kalīla wa Dimna* by Abdallāh ibn Muqaffa'. Interestingly, unlike the previous case, the hypothesis of zero intercept holds better for tokens than for types. I do not have a good explanation for this. However, Arabic nonconcatenative morphology differs greatly from Czech morphology, so I would not be surprised if the stochastic principles behind them also differ. The results can be seen in Figure 7.

Both Czech and Arabic texts were further analyzed at the morpheme-word-clause (Figures 8 and 9) and word-clause-sentence levels Figure 8. The word-clause-sentence level was not explored in the Arabic text due to insufficient data. These three datasets were only examined in terms of tokens, yet, this is likely to have a minimal impact on the results, given that clauses and sentences recur less frequently compared to words.

In all cases, the intercept was found to be negative, with its absolute value always lower than the parameter b, indicating a positive value even for the shortest construct, which makes sense. It is possible that the negative intercepts in these models are due to artifacts arising from the discrete nature of the joint distribution. Or the assumption that the type-token distinction is not necessary at the clause and sentence levels are false. However, it is also plausible that the stochastic principles underlying these datasets differ greatly from those shaping the joint distributions at the word level. Words are pre-processed units that have been shaped over the centuries of the evolution of language, while clauses and sentences are shaped by the capabilities of a single human brain.[6] In fact, it is interesting that the data produced by these two vastly different processes are not more divergent.

Therefore it may be the case that the negative intercepts are inherent results of the stochastic processes involved. The combination of the negative intercept and the regression toward the mean explains the existence of datasets, where the Menzerath's relation manifests as an increasing function (Buk and Rovenchak, 2008).

The last dataset examines the words-phrase-clause level and is sourced from Mačutek et al. (2017). Unlike words, clauses and sentences, the phrases were not delimited by the speakers; they were defined as chunks of text that depend on a predicate (for a more detailed definition, see the cited paper). Therefore, their segmentation relies on the linguistic annotation of the corpus they come from — the Prague Dependency Treebank 3.0 (Bejček et al., 2013). This dataset shows what the result looks like when it is negative (see Figure 10). Meanwhile, the hyperbolic model itself can be fitted well to the data.

[6] Here I describe a an overall trend rather than a strict rule, there are some creative aspects in morphology — nonce words (occasionalisms) do exist, especially in some registers. On the other side large parts of clauses or even sentences can be formulaic multi-word expressions whose structure is given beforehand.

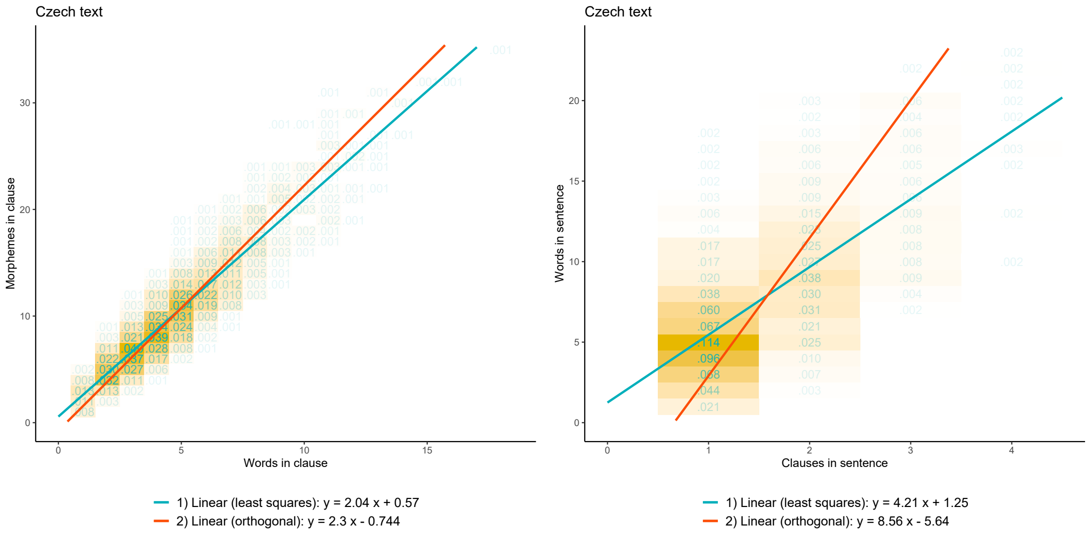


Figure 8: Morpheme — word — clause level and word — clause — sentence level (Czech data, short novel *Krysař* by Viktor Dyk).

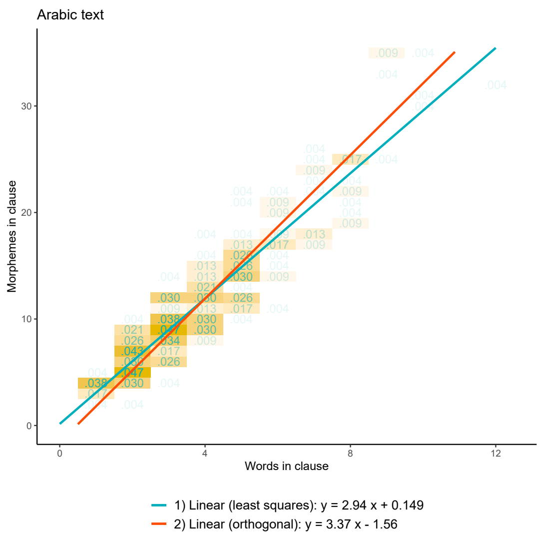


Figure 9: Morpheme — word — clause level (Arabic data, part of *Kalīla wa-Dimna* by Ibn al-Muqaffaʿ).

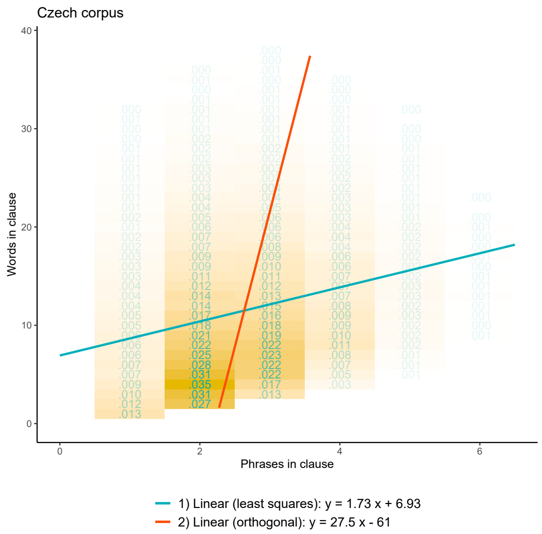


Figure 10: Word — phrase — clause level (Czech data, PDT corpus).

5 Conclusion

The main idea I propose in this study is that the Menzerath's law is a consequence of the features of the relation between the number of constituents in a construct and the number of subconstituents in the same construct. Several datasets suggest that the relation can be fairly simple with the number of constituents being directly proportional to the number of subconstituents, additionally there are some random processes scattering the data points around.

Revisiting the original question posed in the title — is the Menzerath's law just the regression toward the mean? There is a journalistic adage stating that whenever a title contains a question, the answer to that question is negative, or at least partially so. In this case I take issue with the term *just*.

I do not think the Menzerath's law is *just* a regression toward the mean in the sense that this interpretation would render the entire phenomenon obsolete and unworthy of further study.

We tend to understand a linear relation as a default but this inclination is rather magical thinking. Linearity does not mean that the process involved in the relation does not need to be studied. Besides, we need to characterize this random process. It would be worthwhile to turn our attention to the joint distribution itself, and possibly to the marginal distributions as well. When studying the phenomenon directly, the appropriate model of Menzerath's law would follow, with the advantage that this time we will have an explanation of the model and interpretation of its parameters.

Even if we find that the stochastic process involved is fairly simple, it does not mean, that the phenomenon is trivial.[7] We need to find out which real world phenomenon works in the way corresponding with the

[7]Even not *mathematically trivial*, see Ferrer-i-Cancho et al. (2014). This stochastic process is yet to be found, Meyer

stochastic process so that it is plausible not only as a mechanism creating the data, but also as a model of reality. Moreover, one stochastic process usually does not explain the data completely, there are typically residuals left for the others to analyze.

Therefore I do not think the Menzerath's law is *just* regression toward the mean in its second sense either, meaning that it explains everything. Even some datasets presented in this paper do not fit the idea completely.

The question is, whether we can find something to generalize beyond the hyperbolic model, since the attractiveness of MAL as a research topic mostly dwells in its generality. The detailed stochastic processes on various linguistic levels may differ wildly and they may not be suitable for generalization. It may be the case that the original vague Paul Menzerath's hypothesis on decreasing function is actually the only idea that can be generalized to all the datasets on various language levels and in various human and non-human languages. And this vague idea might be adequately represented by a hyperbolic model.

In any case, for practical use for Menzerath's parameters, I suggest checking whether it might be more advantageous to use some parameters of the marginal distribution models and the correlation metric between the two variables instead. For instance instead of fitting the MAL parameters on the phoneme-syllable-word for stylometric text classification (Chen and Liu, 2022), it may be simpler to directly use the mean number of syllables in words,[8] the mean number of phonemes in words, and the correlation coefficient between the number of syllables in a word and the number of phonemes in the same word.

Acknowledgments

I am grateful to all those who inspired me to write this paper. If I have inadvertently borrowed some of your ideas, please be generous. The notion of Menzerath's law as a linear regression took root in my mind during some discussions with Miroslav Kubát in Trier, 2013. In 2014, Damian Blasi confided in me the idea of writing several articles asserting that MAL is merely a statistical artifact, but then he moved on to more fruitful topics. I do not know what he had in mind but I guess it was something more sophisticated. In 2019 I discussed the topic with Vladimír Matlach and Tereza Motalová. Also, I probably picked up some arguments during several heated discussions with Radek Čech and Ján Mačutek.

Language models by OpenAI (GPT-3.5 and GPT-4) have been used to improve the paper stylistically.

This output was supported by the NPO "Systemic Risk Institute" number LX22NPO5101, funded by

(2007) and Torre et al. (2021) found very simple stochastic models that produce a decreasing function resembling the classic Menzerathian curve, but they failed to fit their models to empirical data.

[8] If the word length distributions can be successfully modeled by 1-displaced Poisson distribution — and it often does, see Chebanow (1947) and Grzybek (2007) — then the mean word length is the only parameter needed.

European Union — Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES).

References

Altmann, G. (1980). Prolegomena to Menzerath's law. In R. Grotjahn (Ed.), *Glottometrika 2,* pp. 1–10. Studienverlag Dr. N. Brockmeyer.

Altmann, G. (2014). Bibliography: Menzerath's law. *Glottotheory*, *5*(1), pp. 121–123. https://doi.org/doi:10.1515/ glot-2014-0008.

Altmann, G., Schwibbe, M. H. (1989). *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Georg Olms Verlag.

Bejček, E., Hajičová, E., Hajič, J., Jínová, P., Kettnerová, V., Kolářová, V., Mikulová, M., Mírovský, J., Nedoluzhko, A., Panevová, J., Poláková, L., Ševčíková, M., Štěpánek, J., Zikánová, Š. (2013). Prague Dependency Treebank 3.0 [LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University]. http://hdl.handle.net/11858/00-097C-0000-0023-1AAF-3.

Buk, S., Rovenchak, A. (2007). Statistical parameters of Ivan Franko's novel Perekhresni stežky (The Cross-Paths). In: Grzybek, P., Köhler, R. (Eds.). *Exact methods in the study of language and text. dedicated to Gabriel Altmann on the occasion of his 75th birthday,* pp. 39–48. De Gruyter Mouton. https://doi.org/doi:10.1515/ 9783110894219.39.

Buk, S., Rovenchak, A. (2008). Menzerath–Altmann law for syntactic structures in Ukrainian. *Glottotheory*, *1*(1), pp. 10–17. https://doi.org/10.1515/glot-2008-0002.

Chebanow, S. (1947). On conformity of language structures within the Indoeuropean family to poisson's law. *Comptes rendus de l'Academie de science de l'URSS*, *55*(2), pp. 99–102.

Chen, H., Liu, H. (2022). Approaching language levels and registers in written Chinese with the Menz-erath–Altmann Law. *Digital Scholarship in the Humanities*, *37*(4), pp. 934–948. https://doi.org/10.1093/llc/fqab110.

Ferrer-i-Cancho, R., Hernández-Fernández, A., Baixeries, J., Dębowski, Ł., Mačutek, J. (2014). When is Menzerath-Altmann law mathematically trivial? A new approach. *Statistical Applications in Genetics and Molecular Biology*, *13*(6), pp. 633–644. https://doi.org/10.1515/sagmb-2013-0034.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, pp. 246–263.

Gerlach, R. (1982). Zur Überprüfung des Menzerathschen Gesetzes im Bereich der Morphologie. In: Lehfeldt, W., Strauss, U. (Eds.). *Glottometrika 4,* pp. 95–102. Studienverlag Dr. N. Brockmeyer.

Grzybek, P. (2007). History and methodology of word length studies. In: Grzybek, P. (Ed.). *Contributions to the science of text and language: Word length studies and related issues*, pp. 15–90. Springer Netherlands. https://doi.org/10.1007/978-1-4020-4068-9_2.

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19), pp. E2750–E2758. https : //doi.org/10.1073/pnas.1522072113.

Köhler, R. (1984). Zur Interpretation des Menzerathschen Gesetzes. In J. Boy , R. Köhler (Eds.), *Glottometrika 6,* pp. 177–183.

Kułacka, A., Mačutek, J. (2007). A discrete formula for the Menzerath-Altmann law*. *Journal of Quantitative Linguistics*, *14*(1), pp. 23–32. https://doi.org/10.1080/09296170600850585.

Mačutek, J., Čech, R., Courtin, M. (2021). The Menzerath-Altmann law in syntactic structure revisited. *Proceedings of the Second Workshop on Quantitative Syntax (Quasy, SyntaxFest 2021)*, pp. 65–73. https://aclanthology. org/2021.quasy-1.6.

Mačutek, J., Čech, R., Milička, J. (2017). Menzerath-Altmann law in syntactic dependency structure. *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, pp. 100–107. https://aclanthology. org/W17-6513.

Mačutek, J., Rovenchak, A. A. (2011). Canonical word forms: Menzerath-Altmann law, phonemic length and syllabic length. In: Kelih, E., Levickij, V., Matskulyak, Y. (Eds.). *Issues in quantitative linguistics 2*, pp. 136–147. RAM Verlag.

Menzerath, P. (1928). Über einige phonetische Probleme. *Actes du premier Congres international de linguistes, à La Haye*, pp. 104–105.

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes* (Vol. 3). F. Dümmler.

Meyer, P. (2007). Two semi-mathematical asides on Menzerath-Altmann's law. In: Grzybek, P., Köhler, R. (Eds.). *Exact methods in the study of language and text. dedicated to Gabriel Altmann on the occasion of his 75th birthday,* pp. 449–460. De Gruyter Mouton. https://doi.org/doi:10.1515/9783110894219.449.

Mikros, G., Milička, J. (2014). Distribution of the Menzerath's law on the syllable level in Greek texts. In: Altmann, G., Čech, R., Mačutek, J., Uhlířová, L. (Eds.). *Empirical approaches to text and language analysis,* pp. 180–189. RAM-Verlag.

Milička, J. (2014). Menzerath's law: The whole is greater than the sum of its parts. *Journal of Quantitative Linguistics*, *21*(2), pp. 85–99. https://doi.org/10.1080/09296174.2014.882187.

Milička, J. (2015). Teorie komunikace jakožto explanatorní princip přirozené víceúrovňové segmentace textů [The theory of communication as an explanatory principle for the natural multilevel text segmentation]. Charles University, Faculty of Arts, Prague. (PhD Thesis) https://dspace.cuni.cz/handle/20.500.11956/77988.

Motalová, T. (2022). Menzerath-Altmann law in Chinese. Palacký University, Faculty of Arts, Olomouc. (PhD Thesis) https://theses.cz/id/vqk2ml/motalova_menzerath_altmann_law_in_chinese.pdf.

Pelegrinová, K., Mačutek, J., Čech, R. (2021). The Menzerath-Altmann law as the relation between lengths of words and morphemes in Czech. *Journal of Linguistics / Jazykovedný časopis*, *72*(2), pp. 405–414. https://doi.org/10. 2478/jazcas-2021-0037.

Semple, S., Ferrer-i-Cancho, R., Gustison, M. L. (2022). Linguistic laws in biology. *Trends in Ecology & Evolution*, *37*(1), pp. 53–66. https://doi.org/10.1016/j.tree.2021.08.012.

Stave, M., Paschen, L., Pellegrino, F., Seifart, F. (2021). Optimization of morpheme length: A cross-linguistic assessment of Zipf's and Menzerath's laws. *Linguistics Vanguard*, *7*(s3: 20190076), pp. 1–11. https://doi.org/10.1515/ lingvan-2019-0076.

Torre, I. G., Dębowski, Ł., Hernández-Fernández, A. (2021). Can Menzerath's law be a criterion of complexity in communication? *PLOS One*, *16*(8), e0256133. https://doi.org/10.1371/journal.pone.0256133.

Wang, Y., Chen, H. (2022). The Menzerath-Altmann law on the clause level in English texts. *Linguistics Vanguard*, *8*(1), pp. 331–346. https://doi.org/doi:10.1515/lingvan-2022-0048.