Menzerath's Law: The whole is greater than the sum of its parts[1]

Jiří Milička

Abstract

Reinhard Köhler (1984) proposed an idea that the linguistic constructs which have to be processed by the human parser consist of plain information (that is needed to be communicated) and the structure information, and that this can explain Menzerath's law. Our paper assumes that the amount of plain information and the amount of the structure information are mutually independent. A new model of the nested structure of text and Menzerath's law can be based on this assumption. A formula derived from the model is successfully tested and the results are compared to the classical Menzerath-Altmann law.

Introduction

The average length of constituents depends on the length of their mother construct. This simple hypothesis, established by Paul Menzerath (Menzerath 1928; Menzerath 1954), was popularized and further developed by Gabriel Altmann (1980) into the well-known Menzerath-Altmann's Law (MAL):[2]

$$L_{n-1} = a_n {L_n}^{-b_n} + e^{c_n L_n} \tag{1}$$

Mostly abbreviated to:

$$L_{n-1} = a_n {L_n}^{-b_n} \tag{2}$$

Since this time this relation was investigated and tested on many linguistic levels and in many languages and even on some non-linguistic data (e.g. Wilde — Schwibbe 1989, Kaumanns — Schwibbe 1989, most recently Baixeries et al. 2013). It has been used to study phenomena on the supra-sentencial level (Hřebíček 1990, 1992) and fractal structures of text (e.g. Hřebíček 1994, Andres 2010). The law is a good example of the importance of the quantitative linguistic methodology, since it clearly shows that the "independent language subsystems" are in fact interconnected by relationships which are hard to detect by a qualitative research.

---

[1] *This is an Author's Original Manuscript of an article whose final and definitive form, the Version of Record, has been published in the Journal of Quantitative Linguistics 2014/2/21 ©Taylor & Francis, available online at:* http://www.tandfonline.com/10.1080/09296174.2014.882187

[2] Where L_n refers to the length of a construct on the level n measured in its direct constituents, L_{n-1} refers to the average length of its constituent measured in its direct subconstituents; a_n, b_n and c_n are text dependent parameters.

Altmann’s formula is widely acknowledged but as its parameters are difficult to interpret[3] and some discrepancies with measured data can be found (they are discussed in the next section), this articles aims to establish a new formula for Menzerath's Law, supported by a linguistic explanation, with easy to interpret parameters and fitting the data at least as successfully as the previous one.

Motivation

Paul Menzerath (1952: p. 108) observes that the mean number of phonemes in a word is in a linear relation to the number of syllables of the word. Also Gabriel Altmann (Altmann et al. 2002: 43-46) observes that the relation between the length of the canonical word forms (CWF) measured in syllables and CWF length measured in phonemes follows a linear function:[4]

(3)

$$y = ax + b$$

However, in this paper Gabriel Altmann still models the relation between CWF length in syllables and syllables length in phonemes by his classical MAL formula.

Nine years later Mačutek and Rovenchak (2011) pointed out this contradiction (which is caused by the fact that the mean CWF length in phonemes equals mean length of syllables in phonemes multiplied by the number of syllables), but “theoretical considerations” lead them to reject the linear model and to derive a new one, directly from the classical MAL formula even though the new one does not fit the data better than the linear one.[5]

Taking into account that they considered the advantages of the linear model, it is really pity that they did not try the opposite procedure — reformulating and proper testing Menzerath's law derived from the linear model:

(4)

$$L_{n-1} = a_n + \frac{b_n}{L_n}$$

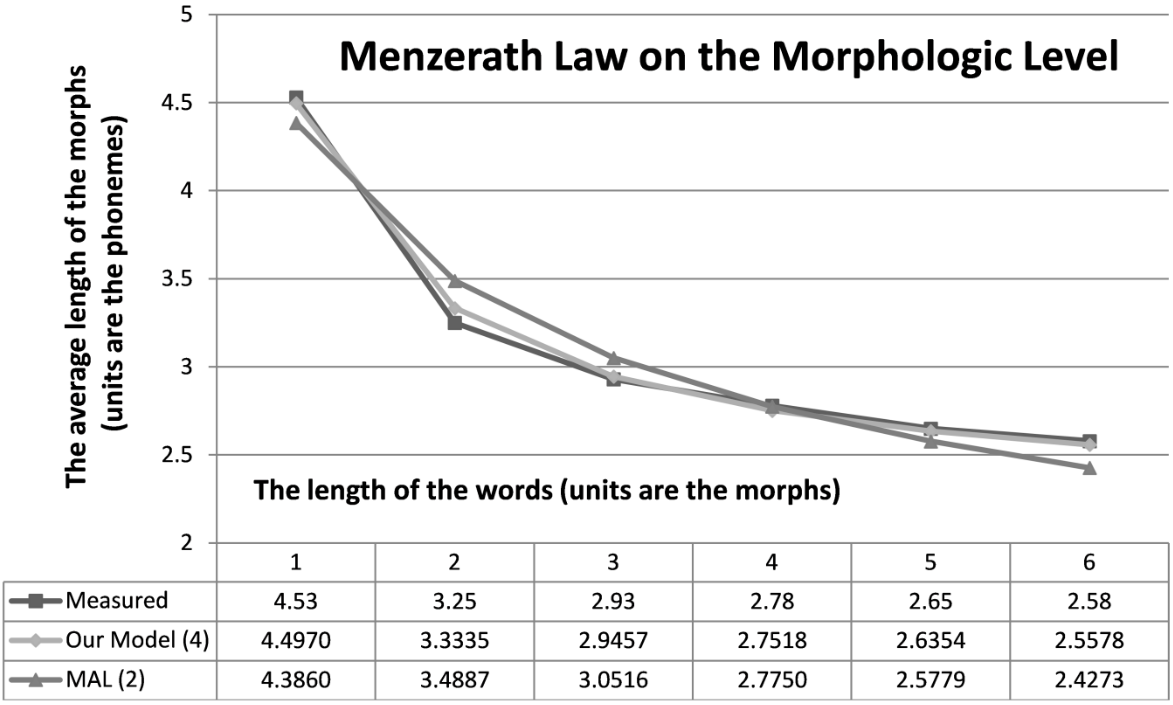
They would possibly find out that the new formula fits not only their data, but also other datasets, like Reiner Gerlach’s observations on the words-morphs-phoneme level as well as the classical one:

---

[3] Despite of some attempts (e.g. Hřebíček 1990, Cramer 2005, Kelih 2010, Kułacka 2010).

[4] Where x refers to the length of the CWF measured in syllables, y refers to the length of the CWF measured in phonemes, a and b are parameters.

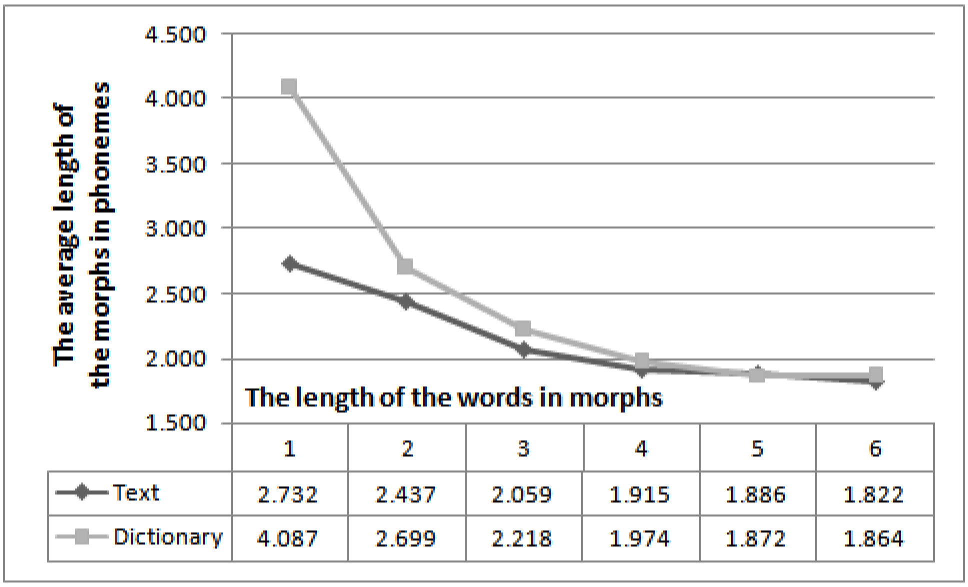
[5] Altmann's original data (for CWF in Indonesian, as quoted by Mačutek and Rovenchak) can be fitted by the linear function (a=1.654; b= 1.946; R^2= 0.9966), by the formula derived from the full MAL ($L_{n-1} = 2.409{L_n}^{0.535} + L_n e^{0.02623 L_n}$; R^2= 0.9961)); by the formula derived from the abbreviated MAL ($L_{n-1} = 3.115{L_n}^{0.8154}$; R^2= 0.9946)) and by the formula that is derived from an extended version of MAL (see the original paper by Mačutek and Rovenchak) ($L_{n-1} = L_n + 2.201{L_n}^{0.6771}$; R^2= 0.9951).



	1	2	3	4	5	6
Measured	4.53	3.25	2.93	2.78	2.65	2.58
Our Model (4)	4.4970	3.3335	2.9457	2.7518	2.6354	2.5578
MAL (2)	4.3860	3.4887	3.0516	2.7750	2.5779	2.4273

Fig 1. Menzerath's Law on the morphologic level as measured by Reiner Gerlach (1982) on *Augst's Lexikon zur Wortbildung*. Fitted by the formula 2 ($L_{n-1} = 4.386L_n{}^{-0.3302};\ R^2 = 0.9546$) and 4 ($L_{n-1} = 2.17 + \frac{2.327}{L_n}; R^2 = 0.9963$). The full formula 1 can be also used ($L_{n-1} = 4.079L_n{}^{-0.5832}e^{0.1005L_n}; R^2 = 0.9951$).

Both these datasets have something in common: They were not measured on texts, but on dictionaries (not on the tokens but on the lists of types). The following image illustrates that the relation measured on tokens differs from the relation measured on list of types substantially:



	1	2	3	4	5	6
Text	2.732	2.437	2.059	1.915	1.886	1.822
Dictionary	4.087	2.699	2.218	1.974	1.872	1.864

Fig 2. Menzerath's Law on the morphologic level (words-morphs-phonemes), measured on the Czech novel *Krysař* by Viktor Dyk. The data that were measured on the dictionary data (list of the word types extracted from the text) can be fitted by the MAL (2) with $R^2 = 0.9491$ and by our new formula (4) with $R^2 = 0.9818$,

whereas the data measured on the original text can be fitted by the MAL (2) with $R^2 = 0.9708$ and by our new formula (4) with $R^2 = 0.9634$. The author of the paper is grateful to Zuzana Komrsková who did all these tedious segmentations.

Due to the lack of the data on the phoneme-morpheme-word level, this hypothesis has been tested on Greek texts on the phonemes-syllables-words level (Mikros, Milička 2014). The differences between the data measured on texts and the data measured on sets of their types (dictionaries) turned to be significant on the 5% significance level for the words containing one to three syllables.

This leads us to study the phenomena more deeply and to formulate a more general model which would explain the difference.

The model and its explanation

We will proceed from a very promising attempt at explaining the MAL by Reinhard Köhler which was published three decades ago in Glottometrika 6 (Köhler 1984, the idea is further developed in Köhler 1989). The paper suggests that every construct contains not only its plain information, but also the so-called *Strukturinformation* (the structure information). According to the paper, the amount of the structure information is dependent on the length of the construct and this should lead to the abbreviated form of the classical MAL (2).

The notion of the *structure information (SI)* is treated very abstractly here, and we will not concretize it as well.[6] It is fair to admit that the real content of the *structure information* is not important to our future consideration, the kind reader can also imagine markers of some features of the construct, e.g. the markers of the borders of the construct (not necessarily placed on the beginning and the ending) or, when being fond of noisy channels, redundancy. The only crucial point is that the whole is greater than its parts, or, the construct is greater than the sum of its constituents that contain the plain information. In other words, some spare subconstituents are somewhat distributed among the constituents of a construct. The idea can be visualised as follows:

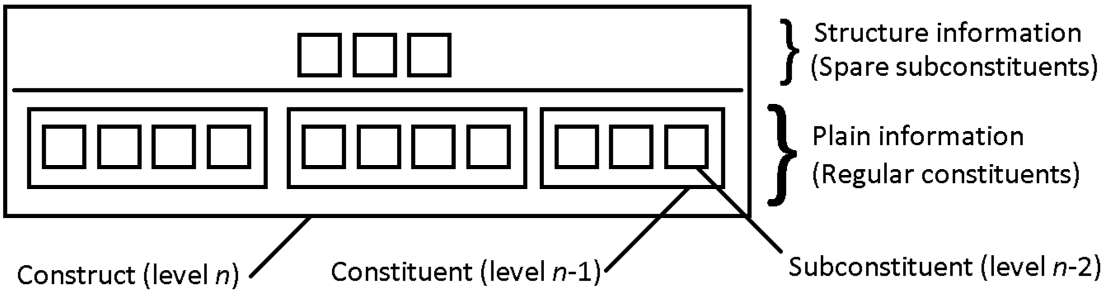


Fig 3. The visualisation of the structure of the construct.

The spare SI subconstituents are in the real text incorporated into the regular constituents (possibly according to some patterns, which may be level-dependent):

---

[6] We will resist the temptation luring us to imagine that (some) synsemantic words are (part of) the structure information on the sentence level and (some) affixes are (part of) the structure information on the level of clauses etc. It could lead us to false assumptions.

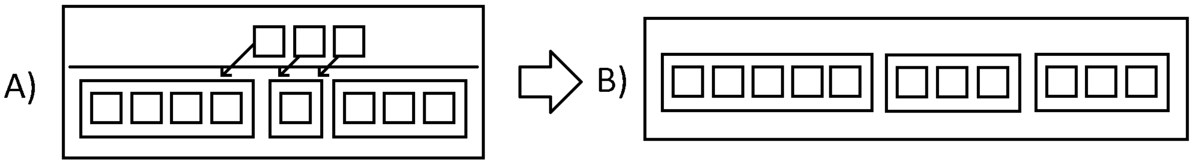


Fig 4. How the spare SI constituents can be incorporated into regular constituents. The structure B) depicts what we observe when measuring lengths of constructs and constituents. The structure A) can be only reconstructed.

Unlike Köhlers paper, we will assume that the amount of the structure information is *independent* of the number of constituents in the construct. The assumption enables us to interpret formula (4) by claiming that a_n is the average length of a constituent that contains the plain information and b_n is the average length of the structure information.

The assumption sounds acceptable when considering that some features of a construct, which should be marked either for a short one either for a long one, could exist. Some analogies are e.g. header of the internet packet, checksum, salutation and signature in the letter… And the validity of the assumption can be tested by fitting the real data by the subsequent formula, which has been done in the previous chapter.

Formula (4) assumes that the considered construct is isolated. That is why it works well with the dictionaries (lists of types), where the natural structure of the text is destroyed and the influence of higher levels is deformed. This explanation predicts that the performance of the model would be worse when trying to fit the data measured on a real text.

Let us imagine the construct in its context (as a constituent of a higher level construct):

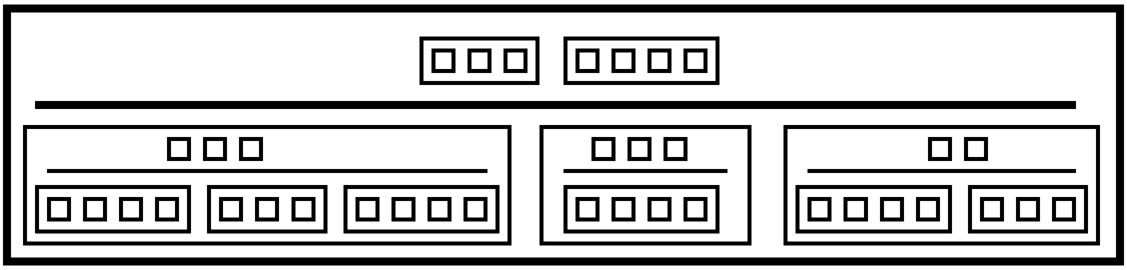


Fig 5. The visualisation of the structure in its context of the higher level constructs.

The higher level structural information subconstituents also merges with regular subconstituents as is depicted below:

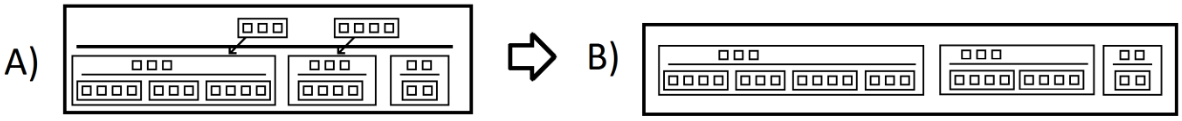


Fig 6. The visualisation of the merging in the structure within its context of the higher level constructs.

This means that the construct under consideration consists of regular constituents and possibly of a higher level structural information. The only exceptions are constructs that consist of only one constituent. E.g. constructs with two constituents could develop either from a construct with two regular constituents or from a construct with one regular constituent and one higher level SI constituent added, but construct with one constituent could develop only from a construct with one constituent (the regular one). I.e. we do not expect the existence of constructs with no regular constituents.

Thus the formula for the constructs consisting of one constituent is straightforward:

$$L_{n-1} = a_n + b_n \qquad\qquad for\ L_n = 1 \tag{5}$$

But for the longer constructs it will be slightly more complicated, as we cannot assume that the SI constituent would have the same average length as the regular one. The average length of the SI constituents is represented by the symbol a'_n.

Let us make an assumption that the probability that the SI constituents of the higher level merges with the construct under consideration does not depend on the length of the construct. This means that the probability p_n that the construct contains a higher level SI constituent is:

$$p_n = \frac{b_{n+1}}{a_{n+2}} \tag{6}$$

And the formula for the constructs consisting of more than one constituent is:

$$L_{n-1} = \frac{b_n}{L_n} + a_n \frac{L_n - p_n}{L_n} + a'_n \frac{p_n}{L_n} \qquad\qquad for \quad L_n > 1 \tag{7}$$

The whole formula for all L_n can be expressed in more concise form with three arguments.

$$L_{n-1} = a_n + \frac{b_n}{L_n} + \frac{c_n \min(1, L_n - 1)}{L_n} \tag{8}$$

where the parameter c_n can be interpreted as

$$c_n = (a'_n - a_n) \frac{b_{n+1}}{a_{n+2}}$$

Of course, the influence of even higher levels cannot be excluded; our model thus cannot be more than an approximation. Let us test the formula to make sure that our assumptions are consistent with the data and that we did not reduce some important part of the reality.

Evaluating the models

Subjectivity of any segmentation of a text should be taken into account. The segmentation is affected by the definition of the segment (according to some linguistic framework), the researcher's interpretation of the definition (which can be ambiguous, especially when applying to many different languages) and precision of the segmentation method (even double independent manual segmentation is not hundred percent correct). Thus, hypotheses related to a somehow segmented text are not only relevant to the text itself but also to the researcher's concepts and abilities. The deviations of the measured data from the model thus can be caused either by random deviations, either by an inappropriate segmentation. Or by the inappropriate or incomplete model, of course.

The data were fitted with Eureqa by Newtonian (Schmidt et al. 2009), minimizing R^2. The values were not weighted according to their distributions, but values of lower frequency than 20 were omitted.

Source:	Gerlach 1982, Table 3, German			
Level:	Lemma -- morpheme -- phoneme			Dictionary
	a	b	c	R^2
New (4)	2.17	2.327	N/A	0.9963
MAL (2)	4.386	0.3302	N/A	0.9546
New (8)	2.259	2.271	- 0.274	0.9999
MAL (1)	4.079	0.5832	0.1005	0.9951

Source:	Polikarpov 2000, Table 6, Russian			
Level:	Lemma -- morpheme -- phoneme			Dictionary
	a	b	c	R^2
New (4)	1.62	2.893	N/A	0.9949
MAL (2)	4.292	0.3956	N/A	0.9323
New (8)	2.885	1.685	-0.325	0.9981
MAL (1)	4.165	0.687	0.08678	0.9975

Source:	Own data, Czech novel Krysař by V. Dyk			
Level:	Word -- morph -- phoneme			Text
	a	b	c	R^2
New (4)	1.685	1.119	N/A	0.9281
MAL (2)	2.759	0.2409	N/A	0.9708
New (8)	1.492	1.24	0.5927	0.9919
MAL (1)	2.149	0.1148	-0.5023	0.9778

Source:	Own data, Czech novel *Válka s Mloky* by K. Čapek			
Level:	Word -- syllable -- phoneme			Text
	a	b	c	R^2
New (4)	2.251	0.2366	N/A	0.4680
MAL (2)	2.493	-0.05092	N/A	0.7042
New (8)	2.13	0.2821	0.4898	0.9769
MAL (1)	2.502	-0.02022	-0.02179	0.8276

For the syllable level, another rationale for our model shall be found. It can be more technical, e.g. higher probability of initial consonant clusters for all words and closing consonant clusters for disyllabic words and longer ones.

Source:	Geršič et al. 1980, German, Experimental data			
Level:	Word -- syllable -- time			N/A
	a	b	c	R^2
New (4)	174.4	76.35	N/A	0.6732
MAL (2)	250.6	0.1946	N/A	0.8101
New (8)	131.7	108.2	109.5	0.9978
MAL (1)	279.6	-0.1331	-0.1419	0.9318

Clause -- word -- morph level was not found.

Source:	Altmann et al. (1989), Table 5.1a, p. 38, German			
Level:	Sentence -- clause -- word			Text
	a	b	c	R^2
New (4)	7.09	4.772	N/A	0.9620
MAL (2)	11.57	0.2285	N/A	0.9659
New (8)	6.662	4.988	1.528	0.9868
MAL (1)	11.36	0.3231	0.0335	0.9799

Source:	Hřebíček (1995)[7], Table 3.1.1, p. 33, Turkish			
Level:	Hreb -- sentence -- word			Text
	a	b	c	R^2
New (4)	13.43	1.662	N/A	0.0457
MAL (2)	14.93	0.04319	N/A	0.0578
New (8)	13.14	1.531	2.022	0.0630
MAL (1)	14.95	0.04867	0.001111	0.0580

Source:	Bohn (2002), Table 1, p. 165 , Chinese			
Level:	Components -- strike -- writing effort			Dictionary
	a	b	c	R^2
New (4)	1.048	1.059	N/A	0.9926
MAL (2)	1.993	0.2624	N/A	0.9433
New (8)	1.033	1.058	0.08241	0.9942
MAL (1)	1.994	0.4331	0.04549	0.9966

Source:	Bohn (2002), Table 2, p. 165 , Chinese			
Level:	Character -- component -- strike			Dictionary
	a	b	c	R^2
New (4)	2.687	2.206	N/A	0.9294
MAL (2)	4.793	0.2695	N/A	0.9884
New (8)	2.354	2.375	1.19	0.9972
MAL (1)	4.833	0.2287	-0.01463	0.9903

---

[7] Only rough trend can be observed in Hřebíček's data.

Source:	Motalová et al. (2013), Table 20, p. 67 , Chinese			
Level:	Character -- component -- strike			Text
	a	b	c	R^2
New (4)	1.137	3.725	N/A	0.9615
MAL (2)	4.758	0.583	N/A	0.9790
New (8)	0.8362	3.877	1.073	0.9815
MAL (1)	4.841	0.526	-0.02235	0.9797

Source:	Bohn (2002), Table 3, p. 166 , Chinese			
Level:	Lemma -- character -- component			Dictionary
	a	b	c	R^2
New (4)	2.043	0.6731	N/A	0.9620
MAL (2)	2.699	0.155	N/A	0.9659
New (8)	1.952	0.751	0.1871	0.9777
MAL (1)	2.625	0.2208	0.03207	0.9730

Source:	Bohn (2002), Table 4, p. 166 , Chinese			
Level:	Clause -- word -- character			Corpus
	a	b	c	R^2
New (4)	1.427	0.8607	N/A	0.9528
MAL (2)	1.942	0.08737	N/A	0.7374
New (8)	1.435	0.9014	- 0.1613	0.9611
MAL (1)	2.088	0.153	0.004844	0.8423

Source:	Bohn (2002), Table 5, p. 169 , Chinese			
Level:	Clause -- word -- character			Text
	a	b	c	R^2
New (4)	1.374	0.6273	N/A	0.9202
MAL (2)	1.945	0.1577	N/A	0.8955
New (8)	1.372	0.628	0.007981	0.9203
MAL (1)	1.929	0.2409	0.02633	0.9191

Source:	Bohn (2002), Table 6, p. 170 , Chinese			
Level:	Sentence -- clause -- word			Corpus
	a	b	c	R^2
New (4)	6.124	5.233	N/A	0.9288
MAL (2)	10.84	0.2318	N/A	0.9751
New (8)	5.747	5.171	2.243	0.9727
MAL (1)	10.86	0.2661	0.008528	0.9779

The nested structure of the language and the General model

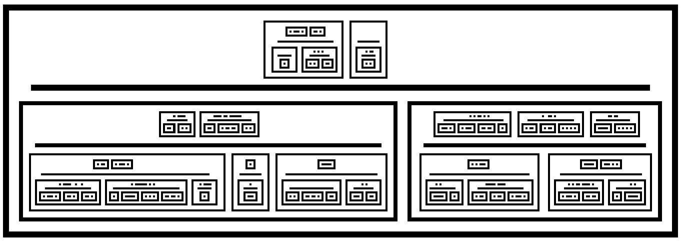
The further development of the model is in progress. We can go onwards in the following two directions:

A) We can contest the assumption that the amount of the structural information and the average length of the regular constituent are independent of the length of the construct/constituent. It would lead us to replacing parameters *a*, *a'* and *b* by some functions.

B) The higher levels can be taken into account.

The second way seems to be more interesting, as using parameters taken from higher linguistic levels (a_{n+4}, b_{n+3} etc.) connects those levels and makes the hierarchy even more coherent than Menzerath's Law itself. In order to support the readers' imagination, a larger schema of the nested structure can be found here:

Fig 7. The visualisation of the structure in its larger context of the higher level constructs. (§§ will be extended according to the resolution of the journal print).



The figures stated in the previous chapter show that even the formula (4) gives plausible results in many cases, while the extended model (8) can be fitted with higher R^2 in all cases (but further testing is needed to find out, whether the differences are significant or not). But even a minimalistic model that assumes equality of the amount of the structure information and the average length of the regular constituent is plausible in many cases.

$$L_{n-1} = a_n \left(1 + \frac{1}{L_n}\right) \tag{9}$$

In general, the more parameters, the better results can be achieved and the question is, what is enough for our purposes — when trying to explain Menzerath's law, whatever number of well interpretable parameters is acceptable, but when trying to predict values, compare texts or language levels or deduct some complicated consequences, one might consider using less complex versions of the model.

Conclusion

The relation between the number of constituents in the construct (L_n) and the average number of subconstituents in the constituents (L_{n-1}) can be successfully described by the following formula:

$$L_{n-1} = a_n + \frac{b_n}{L_n} + \frac{c_n \min(1, L_n - 1)}{L_n}$$

where the parameter c_n can be interpreted as

$$c_n = (a'_n - a_n)\frac{b_{n+1}}{a_{n+2}}$$

Index n represents the language level, parameter a can be interpreted as the average length of a constituent that contains the plain information, b can be interpreted as the average length of the structure information and a' can be interpreted as the average length of a constituent that contains structure information of a higher level construct.

The model is based on the following 5 assumptions:

1) Constructs consist of regular constituents (abbreviated as RC, they bare the plain information) and spare subconstituents (which can contain the structure information, therefore the abbreviation SI).

2) The average length of the RC is independent of the length of the construct.

3) The average length of the SI is independent of the length of the construct.

4) The probability that the higher level SI is inserted into the construct is independent of the length of the construct.

5) The influence of higher levels than $n + 2$ level is not needed to be taken into account.

Additional assumption that no higher levels need to be taken into account produces the following abbreviated formula:

$$L_{n-1} = a_n + \frac{b_n}{L_n}$$

that gives satisfactory results in some cases, especially when the data are measured using lists of types instead of tokens.

This is not the first attempt to reformulate the classical Menzerath-Altmann Law[8] and hopefully also not the last one. The general model, as outlined in the previous chapter, enables endless variations and, of course, it is not the only possible one.

As the model needs to be tested and further developed, the author of the paper welcomes any remarks (critical or positive ones), suggestions and datasets — segmented texts or measured figures related to Menzerath's Law.

---

[8] Gabriel Altmann (2002) noticed that the popular abbreviated form of MAL (2) converges to zero and upgraded it by adding constant 1, which improves its fit to measured data but still leads to theoretically unrealistic consequences, e.g. very long compound words consisting of many morphs would have average morph lengths close to 1. According to our model, the average length of morphs converges to the average length of regular morphs without "spare phonemes" bearing the structural information, which seems to be acceptable. Another attempt adjusts the classical MAL formula for discrete data, to be found in (Mačutek, Kułacka 2007).

References

Altmann, G. (1980). Prolegomena to Menzerath's law. In R. Grotjahn (Ed.), Glottometrika 2. 1–10. Bochum: Brockmeyer.

Altmann, G., Bagheri, D., Goebl, H., Köhler, R., Prün, C. (2002). Einführung in die quantitative Lexikologie. Peust & Gutschmidt, Göttingen.

Andres, J. (2010). On a Conjecture about the fractal structure of language. Journal of Quantitative Linguistics 17 (2). 101–122.

Bohn, H. (2002). Untersuchungen zur chinesischen Sprache und Schrift. In Köhler, R. (Ed.) Korpuslinguistische Untersuchungen zur quantitativen und systemtheoretischen Linguistik. Url: http://ubt.opus.hbz-nrw.de/volltexte/2004/279/pdf/05_bohn.pdf. [cit.: 8/8/2013] 127 - 177.

Cramer, I. M. (2005). The Parameters of the Altmann-Menzerath Law. Journal of Quantitative Linguistics (1). 41-55.

Gerlach, R. (1982). Zut Überprüfung des Menzerath'schen Gesetzes im Bereich der Morphologie. In W. Lehfeld, U. Strauss (Eds.), Glottometrika 4. 95-102. Bochum: Brockmeyer.

Geršič, S. Altmann. G. (1980). Laut-Silbe-Wort und das Menzerathsche Gesetz. Frankfurter Phonetische Beitrage 3, 115-128.

Hřebíček, L. (1990). The constants of Menzerath-Altmann’s Law. In R. Hammerl (Ed.), Glottometrika 12. 61–71. Bochum: Brockmeyer.

Hřebíček, L. (1992). Text in comunication: Supra-sentence structures. Brockmeyer, Bochum.

Hřebíček, L. (1994). Fractals in language. Journal of Quantitative Linguistics 1. 82-86.

Hřebíček, L. (1995). *Text Levels. Language Constructs, Constituents and the Menzerath-Altmann Law*. Wissenschaftlicher Verlag Trier.

Baixeries, J., Hernández-Fernández, A., Forns, N., Ferrer-i-Cancho, R. (2013). The Parameters of the Menzerath-Altmann Law in Genomes. Journal of Quantitative Linguistics 20/2. 94-104.

Kaumanns, W, Schwibbe, M. H. (1989). Strukturmerkmale von Primatensozietäten unter dem Gesichtspunkt der Menserathschen Regel. In Altmann, G. Schwibbe, H. (eds.): Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Georg Olms Verlag, Hildesheim 1989. 99-108.

Kelih, E. (2010). Parameter Interpretation of Menzerath’s Law: Evidence from Serbian. In:Grzybek, P.; Kelih, E.; Mačutek, J. (eds.) (2010): Text and Language. Structures · Functions · Interrelations · Quantitative Perspectives. Wien: Praesens, 71–78.

Köhler, R. (1984). Zur Interpretation des Menzerathschen Gesetzes. In R. Köhler, J. Boy (Eds.), Glottometrika 6. 177-183. Bochum: Brockmeyer.

Köhler, R. (1989). Das Menzerathsche Gesetz als Resultat des Sprachverarbeitungsmechanismus. In Altmann, G. Schwibbe, H. (eds.): Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Georg Olms Verlag, Hildesheim 1989. 109-112.

Kułacka, A. (2010). The Coefficients in the Formula for the Menzerath-Altmann Law. Journal of Quantitative Linguistics Vol. 17 No. 4. 257-268.

Kułacka, A., Mačutek, J. (2007). A Discrete Formula for the Menzerath-Altmann Law. [in:] Journal of Quantitative Linguistics. Vol.14. 23-32.

Mačutek, J., Rovenchak, 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. 136-147. Lüdenscheid: RAM-Verlag.

Menzerath, P. (1928). Über einige phonetische probleme. In: Actes du premier Congres International de Linguistes; Sijthoff: Leiden, 1928.

Menzerath, P. (1954). Die Architektonik des deutschen Wortschatzes. Bonn: Dümler.

Mikros, G., Milička, J. (2014) Distribution of the Menzerath's Law on the Syllable Level in Greek texts. Forthcoming.

Motalová, T., Spáčilová, L. (2013) Aplikace Menzerath-Altmannova zákona na současnou psanou čínštinu (An Application of the Menzerath-Altmann Law to Contemporary Written Chinese; master thesis). [cit. 8/8/2013] Url: http://theses.cz/id/n8j0z5/DP_Motalova_Spacilova-2013.pdf.

Polikarpov A. A. (2000). Chronological Morphemic and Word-Formational Dictionary of Russian: Some System Regularities for Morphemic Structures and Units. In: Linguistische ArbeitsBerichte 75. Leipzig. 201-212.

Schmidt M., Lipson H. (2009). Distilling Free-Form Natural Laws from Experimental Data. In Science, Vol. 324, no. 5923. 81 - 85.

Wilde, J., Schwibbe, M. H. (1989). Organisationsformen von Erbinformation Im Hinblick auf die Menzerathsche Regel. In Altmann, G. Schwibbe, H. (eds.): Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Georg Olms Verlag, Hildesheim 1989. 92-99.