Menzerath-Altmann law

Ján Mačutek, Bratislava, Slovakia
George Mikros, Athens, Greece
Andrij Rovenchak, Lviv, Ukraine
Valentin Vydrin, Russia/France

Menzerath-Altmann law (MA law) - short history

- Menzerath (1954) – the longer a word (measured in number of syllables), the shorter its syllables (measured in number of phonemes)

- Altmann (1980), two generalizations (in two directions)
  - – first, not only for words and syllables, but also for other language units (clause – word, sentence - clause)
  - – second, monotonicity is not required, the mean size of constituents is a function of the size od the construct

Mathematical formulation of MA law

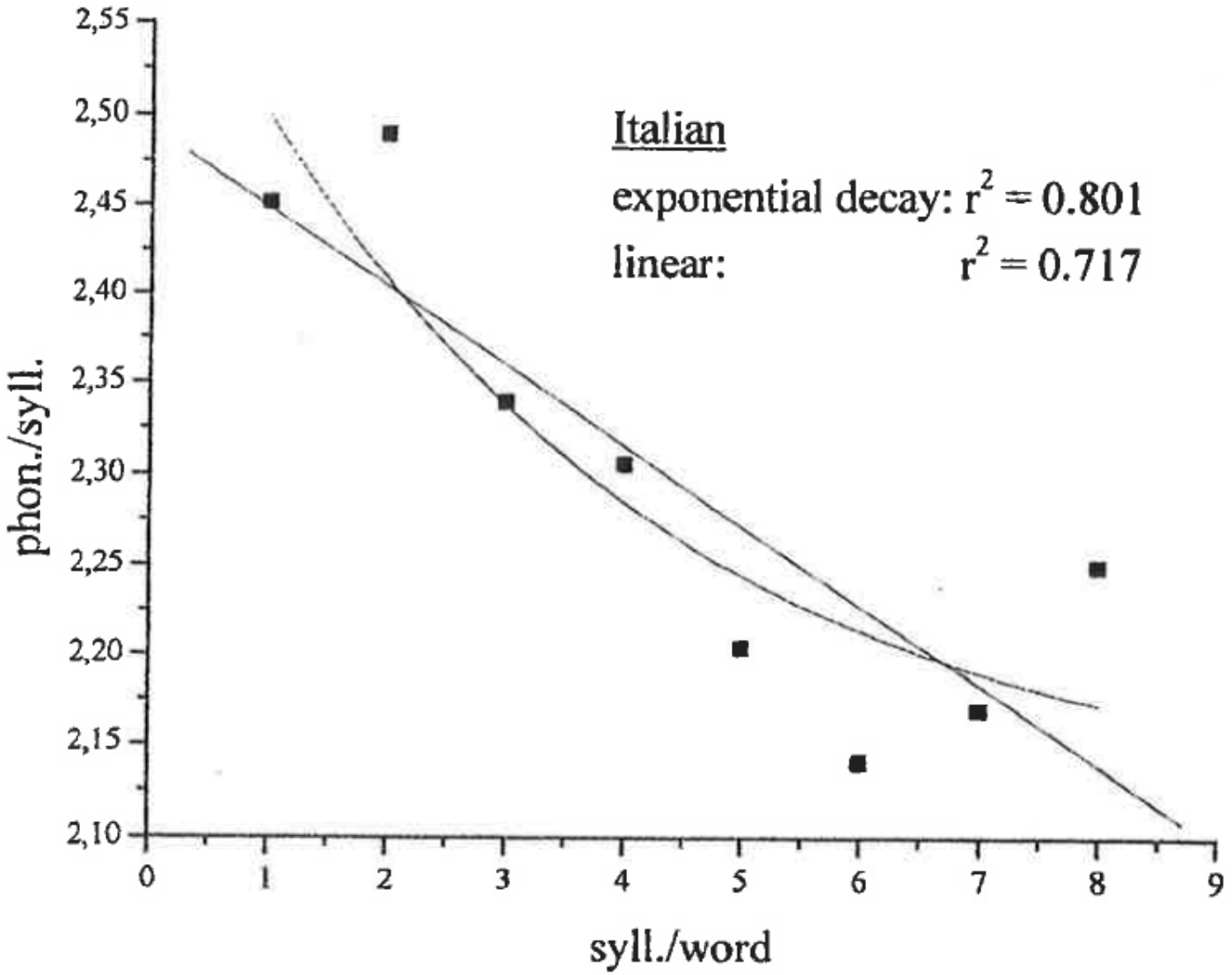
(1) $y(x) = ax^b$

- parameter *b* is negative, decreasing function

(2) $y(x) = ax^b e^{cx}$

- this function can attain its maximum not only for x=1, but also in other points
- y – mean size of constituent, x – construct size

Formula (1) is in many aspects more simple and „nicer“, but it does not fit data sufficiently well in some cases..



Is mathematical model for MA law good?

- Criteria for a good model:

1) goodness of fit – curve (probability distribution, function,…) is close enough to data

2) it can be deduced from theory

3) interpretable (and, in the ideal case, interpreted) parameters

Then the differential equation (1) has the general solution

$$y = Ce^{a_0 x} \prod_{i=1}^{k_1} (x - b_{1i})^{a_{1i}} \cdot$$

$$\cdot \exp\left(\sum_{j \geq 2} \sum_{i=1}^{k_j} \frac{a_{ji}}{(1 - c_j)(x - b_{ji})^{c_j - 1}} \right) + d. \quad (1a)$$

Let us consider some special cases of (1).

Is mathematical model for MA law good?

- Criteria for a good model:

1) goodness of fit – curve (probability distribution, function,…) is close enough to data

2) it can be deduced from theory

3) interpretable (and, in the ideal case, interpreted) parameters

How to interpret parameters?

- Possibilities:

1. deductively – interpretation follows from a more general theory

2. inductively – we notice some regularities, dependencies, correlations, …

The first possibility would be nice…but we are far away from it.

Interpretation of parameters - problems

- huge heterogeneity of data and of approaches

1. different languages – we don’t know how important is the influence of language, how important the influence of language unit, …
2. Sometimes we move only across language levels (clause in words – word in syllables – syllable in phonemes), sometimes we measure sizes in„non-linguistic“ units (word in syllables – syllable in milliseconds)
3. different texts (written/spoken, different genres, ...)

A way towards an interpretation?

- investigation of several levels in one text (eg. text aggregates – sentences – clauses – (phrases) - words – syllables/morphemes – graphemes/phonemes)

- analysis of many texts from one language

Other problems/questions (1/4)

- Large constructs are usually omitted, because they occur rarely and their properties display a high degree of variability.

- How do large constructs behave?

- It seems that up from some size they begin to “live a non-menzerathian“ life. Is it really so? If yes, why?

Other problems/questions (2/4)

- MA law expresses a relation between a size of the construct and the mean size of its constituents.

- What does variability of constituent sizes look like? Is it the same for all construct sizes? If yes, why? If not, why? If not, how does it behave, which influences are important?

Other problems/questions (3/4)

- MA law can be observed (almost certainly) also in random/randomized texts, but parameter values in the „normal“ and random/randomized texts differ significantly.

- Is the difference the same for all language units? Have we the same “degree of freedom” when we “compose” a sentence from clauses and when we “compose” a word from syllables? Or are we “forced” to achieve some particular parameter values for some language units?

Other problems/questions (4/4)

- Which relations between language units can be modelled by MA law?

- There is a relation between the length of a word and the number of its meanings (longer words have fewer meanings). This relations was considered to be one case of MA law. Where are constructs and constituents? A word is not composed of its meanings!

- Or do we want to understand MA law more generally – is it any relation satisfying the MA law mathematical formula? Is it a relation between the size of a language unit and its another property? But then the Zipf law is a special case of the MA law...

Intermezzo

- word length motif – the longest non-decreasing sequence of word lengths (in syllables here)

- 2 2 1 3 1 1 2 2 4 2 1 1 3 2

- 2 2 | 1 3 | 1 1 2 2 4 | 2 | 1 1 3 | 2

Some preliminary results

- formula $y(x) = ax^b e^{cx}$
- y – mean size of constituent, x – construct size

Ukrainian (word length in syllables – mean syllable length in phonemes)

a = 2.54, b = 0.15, c = -0.08

Ukrainian (motif length in words – mean word length in syllables)

a = 2.44, b = -0,08, c = 0.02

Other preliminary results

- tendency – if we go upwards in language unit hierarchy, parameter b is getting smaller, parameter c increases its value

- the same tendency observed in Modern Greek and in Bamana (language from West Africa)

Randomized texts

- random permutation of sequence of phonemic length of words
- random permutation of sequence of syllabic length of words

- MA law still valid…or even “more valid”, both on level of words in syllables – syllables in phonemes and on the level of motifs in words – words in syllables

- same tendency in all three languages (Ukrainian, Modern Greek, Bamana)

Preliminary result (speculation)

- MA law (maybe also other language laws) would be valid also for randomized texts

- their validity is caused by distribution of sizes of language units (e.g. distribution of word length)

- but we speak in such a way that parameters of language laws differ significantly from randomized texts, language as we use it has some non-random “structure”

References

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

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

Altmann, G., & Schwibbe, M.H. (1989). *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen.* Hildesheim: Olms.

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

Benešová, M., Faltýnek, D., & Hadwiger Zámečník, L. (2015). Menzerath-Altmann law in differently segmented texts. In A. Tuzzi, M. Benešová & J. Mačutek (Eds.), *Recent Contributions to Quantitative Linguistics* (pp. 27–40). Berlin, Boston: de Gruyter.

Cramer, I.M. (2005). Das Menzerathsche Gesetz. In R. Köhler, G. Altmann & R.G. Piotrowski (Eds.), *Quantitative Linguistics. An International Handbook* (pp. 659–688). Berlin, New York: de Gruyter.

Fickermann, I., Markner-Jäger, B., & Rothe, U. (1984). Wortlänge und Bedeutingskomplexität. In J. Boy & R. Köhler (Eds.), *Glottometrika 6* (pp. 115–126). Bochum: Brockmeyer.

Gerlach, R. (1982). Zur Überprüfung des Menzerath'schen Gesetzes im Bereich der Morphologie. In W. Lehfeldt & U. Strauss (Eds.), *Glottometrika 4* (pp. 95–102). Bochum: Brockmeyer.

Geršić, S., & Altmann, G. (1980). Laut - Silbe - Wort und das Menzerathsche Gesetz. In Wodarz, H.W. (Ed.), *Frankfurter phonetische Beiträge III* (pp. 115–123). Hamburg: Buske.

Grzybek, P. (1999). Randbemerkungen zur Korrelation von Wort- und Silbenlänge im Kroatischen. In B. Tošović (Ed.), *Die grammatischen Korrelationen. GraLiS 1999* (pp. 67–77). Graz: Institut für Slawistik der Karl-Franzens-Universität.

Heups, G. (1983). Untersuchungen zum Verhältnis von Satzlänge zu Clauselänge am Beispiel deutscher Texte verschiedener Textklassen. In R. Köhler & J. Boy (Eds.), *Glottometrika 5* (pp. 113–133). Bochum: Brockmeyer.

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

Kelih, E. (2010). Parameter interpretation of Menzerath law: evidence from Serbian. In P. Grzybek, E. Kelih & J. Mačutek (Eds.), *Text and Language. Structures, Functions, Interrelations, Quantitative Perspectives* (pp. 71 –79). Wien: Praesens.

Köhler, R. (1982). Das Menzeratsche Gesetz auf Satzebene. In W. Lehfeldt & U. Strauss (Eds.), *Glottometrika 4* (pp. 103–113). Bochum: Brockmeyer.

Köhler, R. (2015). Linguistic motifs. In G. Mikros & J. Mačutek (Eds.), *Sequences in Language and Text* (pp. 89–108). Berlin, Boston: de Gruyter.

Mačutek, J., & Mikros, G.K. (2015). Menzerath-Altmann law for word length motifs. In G.K. Mikros & J. Mačutek (Eds.), *Sequences on Language and Text* (pp. 125–131). Berlin, Boston: de Gruyter.

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

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

Mikros, G., & Milička, J. (2014). Distribution of the Menzerath's law on the syllable level in Greek texts. In G. Altmann, R. Čech, J. Mačutek & L. Uhlířová (Eds.), *Empirical Approaches to Text and Language Analysis* (pp. 180-189). Lüdenscheid: RAM-Verlag.

Sambor, J. (1984). Menzerath's law and the polysemy of words. In J. Boy & R. Köhler (Eds.), *Glottometrika 6* (pp. 94–114). Bochum: Brockmeyer.

Teupenhayn, R. & Altmann, G. (1984). Clause length and Menzerath's law. In J. Boy & R. Köhler (Eds.), *Glottometrika 6* (pp. 127–138). Bochum: Brockmeyer.