

Menzerathov-Altmanov zákon očami matematika

Ján Mačutek

Katedra aplikovanej matematiky a štatistiky

Univerzita Komenského v Bratislave

Krátko o histórii MA zákona

- Menzerath (1954) – čím dlhšie slovo (merané počtom slabík), tým kratšie slabiky (merané počtom foném)
- Altmann (1980)
 - prvé zovšeobecnenie: pre rôzne úrovne (napr. klauza - slovo), čím väčší celok, tým menšie sú jeho časti
 - druhé zovšeobecnenie: nevyžaduje sa monotónnosť, priemerná veľkosť konštituentov (častí) je funkciou veľkosti konštruktú (celku)

Matematická formulácia MA zákona

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

- parameter b je záporný, klesajúca funkcia

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

- funkcia môže (a nemusí) mať maximum inde ako v bode 1

Tvar (1) je v mnohých ohľadoch jednoduchší a „krajší“, ale pre niektoré dáta nestačí.



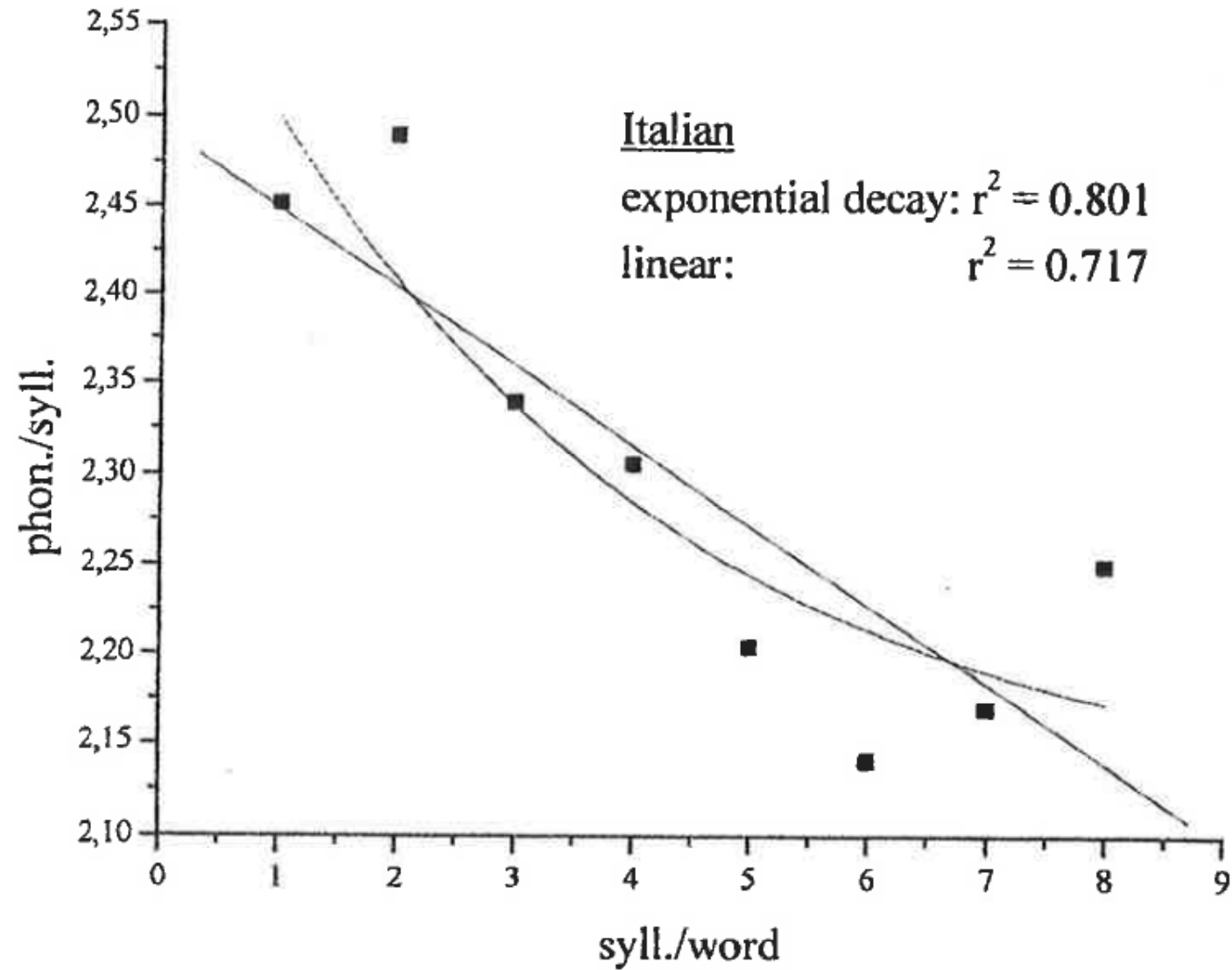
141%



Nástroje

Vyplniť a podpísať

Poznámka



Je matematický model pre MA zákon dobrý?

- Kritériá, ktoré má spĺňať dobrý model:
 - 1) „goodness of fit“ – krivka (rozdelenie pravdepodobnosti,...) sa „v dostatočnej miere podobá na dáta“
 - 2) dá sa odvodiť z teórie
 - 3) interpretovateľnosť parametrov

Then the differential equation (1) has the general solution

$$y = Ce^{a_0 x} \prod_{i=1}^{k_1} (x - b_{1i})^{a_{1i}} \cdot \exp \left(\sum_{j \geq 2} \sum_{i=1}^{k_j} \frac{a_{ji}}{(1 - c_j)(x - b_{ji})^{c_j - 1}} \right) + d. \quad (1a)$$

Let us consider some special cases of (1).

Je matematický model pre MA zákon dobrý?

- Kritériá, ktoré má spĺňať dobrý model:

- 1) „goodness of fit“ – krivka (rozdelenie pravdepodobnosti,...) sa „v dostatočnej miere podobá na dáta“
- 2) dá sa odvodiť z teórie
- 3) interpretovateľnosť parametrov

Ako interpretovať parametre?

- Možnosti:

1. Deduktívne – interpretácia vyplýva z teórie.
2. Induktívne – všimneme si isté pravidelnosti, závislosti,...

Prvá možnosť – bolo by to pekné, ale ešte tam nie sme...

Problémy s interpretáciou parametrov

- treba si všimnúť veľkú heterogenitu dát aj prístupov
1. rôzne jazyky - nevieme, čo je vplyv jazyka, čo vplyv zvolenej jazykovej jednotky, ...
 2. niekedy sa pohybujeme len po stupňoch hierarchie jazykových jednotiek (slovo v slabikách – slabika vo fonémach), inokedy sa meria v „nejazykových“ jednotkách (slovo v slabikách – trvanie slabiky v milisekundách)
 3. rôzne typy textov (písaný/hovorený, rôzne žánre,...)

Cesta k interpretácii?

- preskúmať čo najviac úrovní na tom istom texte (napr. textové agregáty – vety – klauzy – slová – slabiky/morfémy – grafémy/fonémy)
- analyzovať mnoho textov z jedného jazyka a hľadať závislosti hodnôt parametrov od iných vlastností
- potom to isté urobiť pre viacero jazykov

Ďalšie problémy/otázky/podnety (1/4)

- Veľké konštrukty sa zvyknú vynechať, lebo ich je málo a ich vlastnosti majú veľkú variabilitu.
- Ako sa správajú veľké konštrukty?
- Zdá sa, že niektoré od istého bodu začínajú „žiť nemenzerathovsky“. Je to naozaj tak? Ak áno, prečo?

Ďalšie problémy/otázky/podnety (2/4)

- MA zákon hovorí o vzťahu medzi veľkosťou konštruktú a priemernou veľkosťou jeho konštituentov.
- Ako vyzerá variabilita veľkostí konštituentov? Je rovnaká pre všetky veľkosti konštruktú? Ak áno, prečo? Ak nie, prečo? Ak nie, ako sa správa, čo je ovplyvňuje?

Ďalšie problémy/otázky/podnety (3/4)

- MA zákon (skoro iste) platí aj pre znáhodnené/náhodné texty, ale parametre v „normálnych“ a znáhodnených/náhodných textoch sa od seba (skoro iste) významne líšia.
- Je rozdiel medzi hodnotami parametrov pre také texty rovnaký pre všetky jazykové jednotky? Máme napr. pri tvorení viet z kláuz a pri tvorení slov zo slabík rovnakú „mieru slobody“, alebo sme na niektorých úrovniach viac „donútení dosiahnuť“ isté hodnoty parametrov?

Ďalšie problémy/otázky/podnety (4/4)

- Čo všetko sa dá nazvať MA zákonom?
- Aj vzťah medzi dĺžkou slova a počtom jeho významov bol označený ako MA zákon.
- Chceme MA zákon chápať ešte všeobecnejšie, ako vzťah medzi veľkosťou jazykovej jednotky a nejakou jej inou vlastnosťou? Potom aj Zipfov zákon je MA zákon...

References

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