

Agnieszka Kułacka<sup>1</sup>  
Uniwersytet Wrocławski  
Wrocław

## BADANIA NAD PRAWEM MENZERATHA–ALTMANNA

### Wstęp

Prawidłowość nosząca obecnie nazwę prawa Menzeratha–Altmanna (prawo MA) została po raz pierwszy dostrzeżona w 1899 roku przez francuskiego lingwistę Antoine’a Grégoire’a, który

[...] badał długość sylab. Pokazał on, że samogłoska ‘a’ w francuskim słowie ‘pâtisserie’ jest znacznie krótsza niż w słowach ‘pâte’ lub ‘pâté’. Zjawisko to nazywane jest tendencją ścieśniania<sup>2</sup> (Cramer, 2005: 41–42).

W 1904 r. Ernst A. Meyer i później w 1910 r. Leonce Roudet zajmowali się tym zjawiskiem w fonetyce eksperymentalnej (por. Cramer 2005: 41–42). Jednakże inni językoznawcy nie znali tej prawidłowości aż do momentu, kiedy została ona opisana przez Paula Menzeratha w 1928 r. i sformułowana dla fonologii. Menzerath odkrył dwie zależności:

(a) im krótszy jest dźwięk tym dłuższa konstrukcja, w której się pojawia (Menzerath, 1928: 104)

oraz

(b) trzysylabowe słowo z ośmioma głoskami trwa relatywnie dłużej niż trzysylabowe słowo z 9 lub więcej głoskami (Menzerath, 1928: 104).

---

<sup>1</sup> Autorka pozostaje pod opieką naukową prof. dra hab. Adama Pawłowskiego.

<sup>2</sup> Tłumaczenia cytatów: A. Kułacka.

Menzerath odkrył później, że im więcej słowo ma głosek, tym są one krótsze (Altmann 1980). W 1954 roku Menzerath uogólnił to statystyczne prawo fonologiczne dla całego systemu językowego. Ta zależność pomiędzy składnikiem a konstrukcją językową, nazywaną prawem MA, może być opisana w następujący sposób (Menzerath, 1954, cyt. za Hammerl, Sambor 1993: 23): *im dłuższa konstrukcja językowa tym krótsze jej składniki*.

Gabriel Altmann (1980) podał wzór opisujący prawo MA. W tym modelu założył, że szybkość zmian długości składnika ( $y'$ ) jest wprost proporcjonalna do jego średniej długości ( $y$ ) i odwrotnie proporcjonalna do długości konstrukcji językowej ( $x$ ). Otrzymał on następujące równanie różniczkowe:

$$(1) \quad y' = \frac{b \cdot y}{x}$$

gdzie  $b$  jest pewnym współczynnikiem. Długość składnika i konstrukcji może być mierzona za pomocą różnych jednostek.

Po scałkowaniu obustronnie i przeprowadzeniu odpowiednich działań, Altmann otrzymał następujący wzór dla danych ciągłych:

$$(2) \quad y = ax^b$$

gdzie  $y$  jest średnią długością składnika,  $x$  jest długością konstrukcji językowej, a  $a$  i  $b$  są pewnymi współczynnikami.

\* \* \*

W 2006 r. Agnieszka Kułacka i Jan Maćutek odnieśli się do wzoru (2) zaproponowanego przez Gabriela Altmanna i przedyskutowali ograniczenia jego zastosowania do danych dyskretnych. Porównali wyniki badań nad prawem MA w polskiej składni, kiedy stosuje się wzór dla danych ciągłych i dyskretnych (Kułacka, Maćutek, 2007). Zaproponowali następujący wzór, który otrzymali z wzoru (1) po zastosowaniu metody równań rekurencyjnych:

$$(3) \quad \forall n \in N \setminus \{1\} \quad y(n) = y(1) \prod_{i=2}^n \left( \frac{b}{i-1} + 1 \right)$$

gdzie  $n$  jest liczbą składników w danej konstrukcji językowej,  $y(n)$  jest średnią długością składnika w konstrukcji o  $n$  liczbie składników, a  $b$  jest pewnym współczynnikiem.

Od odkrycia prawa MA badania wykonywano w wielu dziedzinach językoznawstwa (tab. 1.).

**Tab. 1.** Konstrukcje językowe i ich komponenty zbadane w kontekście prawa Menzeratha–Altmanna (źródło: Strauss, Altmann 2007)

Konstrukcja językowa	Jednostka miary konstrukcji	Zmienna zależna
hreb <sup>3</sup>	zdanie	długość zdania
zdanie	zdanie składowe	długość zdania składowego
zdanie	wyraz	długość wyrazu
zdanie	sylaba	czas trwania sylaby
zdanie	sylaba	długość sylaby
jednostka rytmiczna	sylaba	czas trwania sylaby
słowo	morf	długość morfu
słowo	sylaba	długość sylaby
słowo	dźwięk	długość dźwięku
sylaba	dźwięk	długość dźwięku
słowa i sylaby	słowa i sylaby	długość sylaby
znaki chińskie	grafem	złożoność grafemu

W tabeli podaje się badania dotyczące hipotezy bardziej ogólnej niż prawo MA; średnia długość składnika konstrukcji językowej jest funkcją długości tej konstrukcji (Strauss, Altmann 2007).

## 1. Weryfikacja empiryczna prawa

W tej części artykułu zademonstruję metodę badań nad prawem MA. Jest ona nieco odmienna od stosowanej dotychczas przez językoznawców. W tabeli 2 podano dane zebrane z pierwszych 11 rozdziałów polskiego tłumaczenia *Kodu Leonarda da Vinci* Dana Browna. Przyjęto następujące oznaczenia:  $x$  jest liczbą zdań podrzędnych w zdaniu złożonym,  $y_E$  – średnią długością zdania podrzędnego danych empirycznych,  $n$  – liczbą zdań  $x$  zdaniami podrzędnymi,  $y_T$  – długością zdań składowych znalezionej po zastosowaniu formuły (3).

**Tab. 2.** Dane empiryczne i teoretyczne dla rozdziałów P-R11

$x$	$y_E$	$n$	$y_T$
1	5,8664	711	5,8664
2	5,4780	499	5,4206
3	5,5000	260	5,2146
4	4,9932	110	5,0825

<sup>3</sup> Hreb jest jednostką tekstu wprowadzoną przez Ludka Hřebička, czeskiego językoznawcę zajmującego się badaniem języka perskiego i tureckiego. Hreb zdaniowy jest zbiorem zdań pewnego tekstu, w których występuje dany leksem. Rozważmy tekst *Ala ma kota. Kot ma długi ogon. Mysz się boi kota*. Do hrebu *kot* należą wszystkie zdania. Do hrebu *mieć* należy zdanie pierwsze i drugie.

6	4,8529	17	
7	5,7857	4	
8	5,8125	2	
9	6,2778	2	

Podwójna linia oddziela dane, które zostały zastosowane do obliczeń danych teoretycznych. Zdecydowałam się korzystać z 98% zdań, w tym wypadku zdań z 5 lub mniej składnikami. Długość zdań składowych mierzona była liczbą wyrazów. Można zaobserwować w 2 kolumnie tablicy, że prawo MA działa. Im więcej zdań składowych, tym są one średnio krótsze. By formalnie to potwierdzić, wykonam test z wykorzystaniem współczynnika korelacji rangowej Spearmana.

Znajdę wyżej wspomniany współczynnik i zweryfikuję hipotezę dotyczącą dodatniej korelacji pomiędzy liczbą zdań składowych ( $x$ ) a ich średnią długością ( $y_E$ ).

Rozważmy hipotezę zerową  $H_0$  i jej alternatywę  $H_1$ :

$H_0$ :  $\rho = 0$  – nie ma korelacji pomiędzy  $x$  i  $y_E$ ;

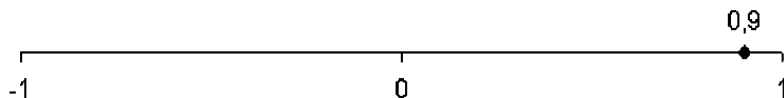
$H_1$ :  $\rho > 0$  – jest dodatnia korelacja, co oznacza, że prawo MA działa.

W tabeli 3 podano rangi dla  $x$  i  $y_E$ .

**Tab. 3.** Rangi dla  $x$  i  $y_E$

Ranga $x$	1	2	3	4	5
Ranga $y_E$	1	2	3	5	4

Wartość krytyczna dla jednostronnego testu na poziomie 5% dla pięciu danych to 0,9. Obliczyłam, że współczynnik korelacji rangowej Spearmana dla tych danych to 0,9 (rys. 1.).



**Rys. 1.** Wykres dla poziomu korelacji  $\rho$

Przyjmuję, że hipoteza alternatywna  $H_1$  jest prawdziwa, co oznacza, iż prawo MA działa.

Dopiero po weryfikacji prawa MA, można wyliczać współczynniki we wzorze (3). Dla danych empirycznych podanych w tabeli 2 obliczyłam współczynniki  $y(1)$  i  $b$ , posługując się metodą przedstawioną w artykule Kułacka i Maćutek (2007: 28). Zastosowałam je do wzoru (3). Dane teoretyczne podane są w tabeli 2.

$$y(1) = 5,8664$$

$$b = -0,0760$$

Wykażę teraz znaczenie zastosowania formalnego testu opisanego powyżej, by potwierdzić działanie prawa MA p r z e d znalezieniem współczynników występujących we wzorze (3).

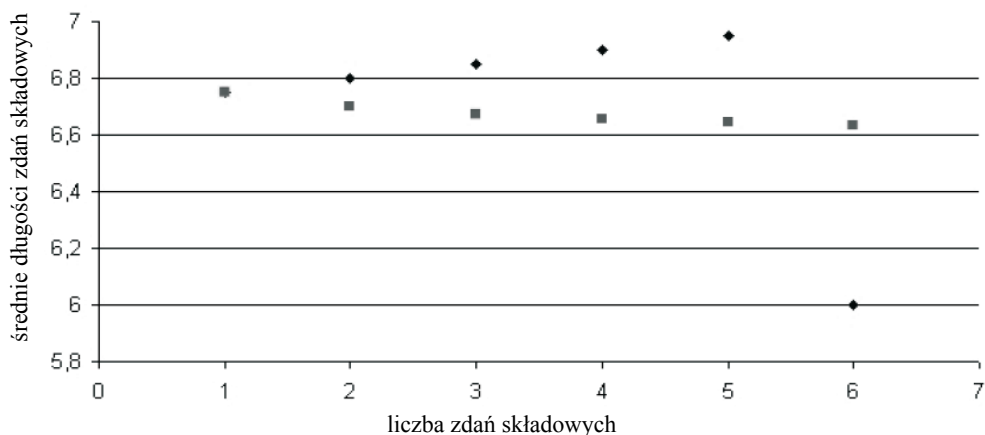
W tab. 4. podano dane fikcyjne, w tabeli 5. rangi dla danych fikcyjnych. Nie obserwuje się działania prawa MA dla tych danych (rys. 2.).

Tab. 4. Dane fikcyjne

x	$y_E$	$y_T$
1	6,7500	6,7500
2	6,8000	6,6994
3	6,8500	6,6743
4	6,9000	6,6576
5	6,9500	6,6450
6	6,0000	6,6351

Tab. 5. Rangi dla danych fikcyjnych

Ranga x	1	2	3	4	5
Ranga $y_E$	5	4	3	1	6



Rys. 2. Wykres danych fikcyjnych empirycznych (♦) i teoretycznych (■)

Wartość krytyczna dla testu jednostronnego na poziomie 5% dla 6 danych wynosi 0,8286. Współczynnik korelacji rangowej Spearmana wynosi dla tych danych  $-0,1428$ , co oznacza, że przyjmujemy hipotezę zerową  $H_0$ , oraz że prawo MA nie zachodzi. Jednakże możemy znaleźć współczynniki występujące we wzorze (3) i obliczyć dane teoretyczne z  $b = -0,00750 < 0$ . Obserwowany paradoks wynika z dobranego przedziału danych empirycznych i ostatniej liczby.

## 2. Prawo Shermmana i jego zastosowanie

Prawo Shermmana dotyczy rozkładu statystycznego długości zdań złożonych podrzędnie i współrzędnie, a także długości zdań składowych.

Długość zdań jest cechą tekstu, która ani nie jest przypadkowa, ani zdeterminowana, ani też nie podlega ustalonym regułom. Podlega ona natomiast prawu, które bierze pod uwagę siły działające podczas wypowiedzi (Altmann 1992: 38).

Wyboru zdań nie należy dokonywać ani losowo, ani systematycznie, ale wykorzystać wszystkie zdania z zamkniętej całości, np. rozdziału książki (por. Altmann 1992: 38). Prawo to odgrywa znaczącą rolę podczas wyboru przez mnie metody badań nad prawem MA: analizuję całe rozdziały książek.

Aby wykazać, że zastosowanie wyżej opisanych wniosków wpływających z prawa Shermmana jest konieczne, przyjrzyjmy się kontrprzykładowi: prawo MA działa dla dostatecznie długiego tekstu. Dane empiryczne zebrane dla pierwszych ośmiu rozdziałów *Kodu Leonarda Da Vinci* wraz z prologiem zawarte są w tabeli 6. W tej samej tabeli podano dane zebrane z pierwszych dziesięciu rozdziałów książki wraz z prologiem. Zdania złożone z 5 lub mniej zdań składowych stanowią 98% populacji zdań w tych tekstach, dlatego też tylko one będą brane pod uwagę przy weryfikacji prawa MA. W kolumnach 2 i 4 znajdują się średnie długości zdań składowych zdań złożonych.

**Tab. 6.** Dane empiryczne dla rozdziałów P-R8 i P-R10

Liczba zdań składowych	Rozdziały P-R8	Rangi dla P-R8	Rozdziały P-R10	Rangi dla P-R10
1	6,1304	1	5,8815	1
2	5,7493	2	5,5483	2
3	5,7312	3	5,4748	3
4	5,0149	5	4,9534	5
5	5,1238	4	4,9929	4

Wartość krytyczna dla testu jednostronnego na poziomie 5% to 0,9. Obliczony współczynnik korelacji rangowej Spearmana dla obu zbioru danych wynosi  $r_s = 0,9$ . Odrzucamy hipotezę zerową, co oznacza, że prawo MA zachodzi.

W tabeli 7 podano dane zebrane dla pierwszych 8 rozdziałów książki wraz z prologiem oraz fragmentu rozdziału 9. Obliczony współczynnik korelacji rangowej Spearmana wynosi  $r_s = 0$ . Tym razem odrzucamy hipotezę alternatywną i uznajemy, że prawo MA nie działa na tych danych.

**Tab. 7.** Dane empiryczne dla P-R8 oraz fragmentu R9

Liczba zdań składowych	Rozdziały	Rangi
1	6,0764	2
2	5,8567	3
3	5,8470	4
4	5,1444	5
5	6,3909	1

Powyższe rozważania jasno pokazują, że wnioski z prawa Shermana muszą być stosowane do badań nad prawem MA. W przeciwnym wypadku prawo MA zachowuje się na tekście niekonsekwentnie.

### 3. Minimalna średnia długość zdań składowych

Przeprowadziłam badania na następujących tekstach polskich lub tłumaczeniach na język polski: *Hobbit* Johna R. R. Tolkiena, *Semantyka 2* Johna Lyonsa, *Zarys logiki matematycznej* Andrzeja Grzegorzcyka, *Zabić drozda* Harper Lee, *Demony dobrego* Dextera Jeffa Lindsaya. Kończyłam badania na tekście w momencie, kiedy jeden z następujących warunków zachodził: (1) długość tekstu wynosiła co najmniej 2000 zdań lub/i (2) prawo MA działało na danym tekście.

W tabeli 8 podano współczynniki korelacji rangowej Spearmana dla każdego z tekstów. Wartości krytyczne dla testu jednostronnego na poziomie 5% wynoszą: (1) dla 4 danych – 1, (2) dla 5 danych – 0,9, (3) dla 6 danych – 0,7714, (4) dla 7 danych – 0.6786. Prawo MA zachodzi na następujących tekstach: *Hobbicie*, *Semantyce 2* i *Zarysie logiki matematycznej*. Nie zachodzi natomiast dla *Zabić drozda* i *Demonów dobrego Dextera*.

**Tab. 8.** Współczynniki korelacji rangowej Spearmana  $r_s$  dla badanych tekstów

Książka	Rozdziały	$r_s$	Liczba danych	Liczba zdań
<i>Hobbit</i>	1–5	0,7857	7	1712
<i>Semantyka 2</i>	10,1–10,4	1	4	537
<i>Zarys...</i>	1,0–1,5	0,9	5	541
<i>Zabić...</i>	1–8	–0,0857	6	2114
<i>Demony...</i>	1–7	0,4	4	2083

Czym różnią się te dwa teksty od pozostałych? W tabeli 9 podano średnie długości zdań pojedynczych dla każdego z tekstów. Średnie arytmetyczne zdań pojedynczych miały rangę 1 dla wszystkich tekstów, co oznacza, że średnio zdania pojedyncze były najdłuższe w tekstach.

**Tab. 9.** Średnie długości zdań pojedynczych badanych tekstów

Książka	<i>Hobbit</i>	<i>Semantyka 2</i>	<i>Zarys...</i>	<i>Zabić...</i>	<i>Demony...</i>
<b>Długość</b>	5,8004	14,8514	10,0218	5,5487	4,8042

Można zauważyć, że minimalna długość zdania pojedynczego dla polskich tekstów musi wynosić średnio powyżej 5,8004, by prawo MA zachodziło. Przyпуска się, że istnieje wartość graniczna średniej długości zdań pojedynczych w języku polskim, poniżej której prawo MA nie zachodzi. Reinhard Köhler podjął próbę wyjaśnienia zachodzenia prawa MA jako wyniku ograniczonej pojemności pamięci krótkotrwałej (patrz Köhler 1984). Jeśli zdania są wystarczająco krótkie, hierarchia opisywana przez prawo MA może nie istnieć dla danego tekstu.

W tym artykule podałam metodę, jaka powinna być stosowana do badań nad prawem MA. Przedstawiłam również warunki zachodzenia prawa MA. Podobne badania winny być przeprowadzone dla innych języków.

## Literatura

- ALTMANN G., 1980, *Prolegomena to Menzerath's Law*, vol. 2, s. 1–10.
- ALTMANN G., 1992, *Sherman's Laws of Sentence Length Distribution*, [w:] P. Saukkonen. *What is language synergetics?*, Oulu.
- CRAMER I. M., 2005, *The parameters of the Altmann-Menzerath law*, [w:] „Journal of Quantitative Linguistics”, vol. 12 (1), s. 41–52.
- HAMMERL R., SAMBOR J., 1993, *O statystycznych prawach językowych*, Warszawa.
- KÖHLER R., 1984, *Zur Interpretation Des Menzerathschen Gesetzes*, „Glottometrika”, vol. 6, s. 177–183.
- KUŁACKA A., MAĆUTEK J., 2006, *A Discrete Formula for the Menzerath-Altmann Law*, „Journal of Quantitative Linguistics”, vol. 13, no 2–3, s. 217–226.
- MENZERATH P., 1928, *Über einige phonetische Probleme*, [w:] *Actes du premier Congres international de linguistes*, Leiden, s. 104–105.
- STRAUSS U., ALTMANN G., 2007, *Hierarchic Relations*, [http://www.uni-trier.de/uni/fb2/ldv/lq1\\_wiki/index.php/Hierarchic\\_relations](http://www.uni-trier.de/uni/fb2/ldv/lq1_wiki/index.php/Hierarchic_relations) (dostęp 26 VII 2007).

## Research on Menzerath–Altmann Law Summary

The aim of the article is to present our most recent outcomes of investigation into the Menzerath-Altmann law: the conditions for it to hold. We will demonstrate (1) that the peculiarity of the law arises when the mathematical formula is applied before verification, (2) that the Sherman law must be applied to the methodology of the verification of the law, and finally (3) that the law does not hold for certain texts.