



Kraków, 19 kwietnia 2023

Wzorce interpunkcji w literaturze głównych języków świata są intrygująco matematyczne

Chwila zawahania... Tak, tu kropka – ale czy przecinek na pewno należy postawić tam? A może lepszy byłby myślnik? Interpunkcja potrafi przysparzać kłopotów; często się ją wręcz lekceważy. Błąd! Z najnowszych analiz statystycznych wyłania się inny jej obraz: interpunkcja zdaje się „wyrastać” z fundamentów współdzielonych przez wszystkie (poddane badaniu) języki, a jej cechy są dalekie od trywialnych.

Wielu osobom interpunkcja jawi się jako zło konieczne, gdy tylko można z chęcią ignorowane. Najnowsze analizy literatury napisanej we współcześnie wiodących językach świata każą zmienić ten osąd. W kilkuset dziełach powstałych w siedmiu językach, głównie zachodnich, zaobserwowano bowiem te same cechy statystyczne schematów używania znaków przestankowych! Interpunkcja, której wszystkich dziesięciu przedstawicieli można znaleźć we wprowadzeniu do niniejszego tekstu, okazuje się być uniwersalnym i niezbędnym dopełnieniem matematycznej perfekcji każdego badanego języka. Tak niezwykle wniosek o roli niepozornych przecinków, wykrzykników czy kropek płynie z artykułu naukowców z Instytutu Fizyki Jądrowej Polskiej Akademii Nauk (IFJ PAN) w Krakowie, opublikowanego na łamach czasopisma „Chaos, Solitons & Fractals”.

„Obecne analizy są rozwinięciem naszych wcześniejszych wyników nad multifraktalnymi cechami zmienności długości zdań dzieł literatury światowej. Czym bowiem jest długość zdania? To nic innego jak odległość do kolejnego, konkretnego znaku interpunkcyjnego: kropki. Teraz wzięliśmy więc pod statystyczną lupę wszystkie znaki interpunkcyjne, przyjrzelśmy się także, co się dzieje z interpunkcją podczas tłumaczeń”, mówi prof. dr hab. Stanisław Drożdż (IFJ PAN, Politechnika Krakowska).

Badaniom poddano dwa zbiory tekstów. Główne analizy dotyczące interpunkcji wewnątrz każdego z języków przeprowadzono na 240 cieszących się dużą popularnością dziełach literackich, napisanych w siedmiu językach, głównie zachodnich: angielskim (44), niemieckim (34), francuskim (32), włoskim (32), hiszpańskim (32), polskim (34) i rosyjskim (32). Taki a nie inny dobór języków wynikał z przyjętego kryterium: badacze założyli, że danym językiem powinno posługiwać się nie mniej niż 50 milionów ludzi, a napisane w nim dzieła powinny być nagrodzone nie mniej niż pięcioma literackimi Nagrodami Nobla. Ponadto dla statystycznej wiarygodności wyników badań każda książka musiała zawierać co najmniej 1500 sekwencji wyrazów rozdzielonych znakami interpunkcyjnymi. Osobny zbiór przygotowano pod kątem obserwacji stabilności interpunkcji w tłumaczeniach. Zawierał on 14 dzieł, z których każde było dostępne w każdym z badanych języków (dwie z 98 wersji językowych jednak pominięto z uwagi na ich niedostępność). Łącznie w obu zbiorach wśród autorów znaleźli się tacy twórcy jak J. Conrad, Ch. Dickens, A. C. Doyle, E. Hemingway, J. R. Kipling, G. Orwell, J. D. Salinger, A. V. Woolf, G. Grass, F. Kafka, T. Mann, F. Nietzsche, J. W. von Goethe, M. M. de La Fayette, A. Dumas, V. Hugo, M. Proust, J. Verne, U. Eco, M. de Cervantes, H. Sienkiewicz czy W. Reymont.

Uwagę krakowskich badaczy przykuł przede wszystkim rozkład statystyczny odległości między kolejnymi znakami interpunkcyjnymi. Szybko się okazało, że we wszystkich badanych językach najlepiej opisuje go jeden z precyzyjnie określonych wariantów rozkładu Weibulla. Krzywa tego typu ma charakterystyczny kształt: początkowo rośnie gwałtownie, by po osiągnięciu wartości maksymalnej nieco wolniej opaść do pewnej wartości krytycznej, poniżej której zdąża do zera już z małą i ciągle malejącą dynamiką. Rozkładem Weibulla zwykle opisuje się zjawiska związane z przetrwaniem (np. liczbę ludności w funkcji wieku), ale także różne procesy fizyczne, jak choćby rosnące zmęczenie materiałów.

„Zgodność rozkładu długości sekwencji wyrazów między znakami interpunkcyjnymi z funkcjonalną postacią rozkładu Weibulla była tym lepsza, im więcej rodzajów znaków interpunkcyjnych uwzględnialiśmy w analizach; dla wszystkich znaków zgodność ta okazywała się niemal pełna. Jednocześnie między poszczególnymi językami są widoczne pewne różnice w rozkładach, ale sprowadzają się one tylko do doboru nieco innych wartości parametrów rozkładu, charakterystycznych dla danego języka. Interpunkcja wydaje się więc być integralną częścią wszystkich badanych języków”, zauważa prof. Drożdż, by po chwili z pewnym rozbawieniem dodać: *„...a ponieważ rozkład Weibulla dotyczy takich zjawisk jak przetrwanie, z wcale nie za dużym przybliżeniem oka można stwierdzić, że interpunkcja ma w swoją naturę literalnie wkomponowaną walkę o przeżycie”.*

Kolejny etap analiz polegał na wyznaczeniu funkcji hazardu. W przypadku interpunkcji opisuje ona, jak zmienia się warunkowe prawdopodobieństwo sukcesu – czyli prawdopodobieństwo wystąpienia kolejnego znaku interpunkcyjnego – jeśli w analizowanej sekwencji żaden taki znak jeszcze się nie pojawił. Wyniki są tu jednoznaczne: językiem charakteryzującym się najmniejszą skłonnością do używania interpunkcji jest angielski, z niewiele mu ustępującym hiszpańskim; najbardziej uzależnione od interpunkcji okazały się języki słowiańskie. Krzywe funkcji hazardu dla znaków interpunkcyjnych w sześciu badanych językach okazały się mieć podobny przebieg, różniły się bowiem głównie przesunięciem w pionie.

Wyjątkiem okazał się język niemiecki. Jego funkcja hazardu jako jedyna przecina większość krzywych skonstruowanych dla pozostałych języków. Niemiecka interpunkcja zdaje się więc łączyć cechy interpunkcyjne wielu języków, co czyni z niej swoiste interpunkcyjne esperanto. Powyższa obserwacja współgra z kolejną analizą, która polegała na sprawdzeniu, czy cechy interpunkcyjne oryginalnych dzieł literackich widać w ich tłumaczeniach. Zgodnie z oczekiwaniami, językiem najwierniej odwzorowującym interpunkcję od języka oryginału do języka docelowego okazał się niemiecki.

W przekazie mówionym przerwy można uzasadniać ludzką fizjologią, między innymi koniecznością zaczerpnięcia oddechu lub wygospodarowaniem chwili na ustrukturyzowanie w myślach dalszej wypowiedzi. A w przekazie pisany?

„Tworzenie zdania poprzez dokładanie do niego kolejnych wyrazów przy jednoczesnym dbaniu, by przekaz był zrozumiały i jednoznaczny, nieco przypomina napinanie cięciwy łuku: na początku jest to łatwe, ale z każdą chwilą staje się coraz bardziej wymagające. Jeśli w tekście nie ma żadnych elementów porządkujących (a przecież taką rolę pełni interpunkcja), trudność interpretacji rośnie wraz z wydłużaniem się ciągu wyrazów. Zbyt mocno napięta cięciwa może pęknąć, a zbyt długie zdanie może stać się niezrozumiałe. Dlatego autor staje przed koniecznością 'uwolnienia strzały', czyli zamknięcia fragmentu tekstu jakimś znakiem interpunkcyjnym. Obserwacja ta dotyczy wszystkich analizowanych języków, mamy więc do czynienia z czymś, co można byłoby nazwać prawem lingwistycznym”, stwierdza dr Tomasz Stanisław (IFJ PAN), pierwszy autor omawianego artykułu.

Na zakończenie warto zauważyć, że wynalazek interpunkcji jest stosunkowo niedawny – w najstarszych tekstach znaki przestankowe nie występowały w ogóle. Pojawienie się optymalnych wzorców interpunkcji we współczesnych językach pisanych może być więc interpretowane jako efekt ich zaawansowania ewolucyjnego. Jednak nadmierna potrzeba użycia interpunkcji niekoniecznie musi być oznaką takiego zaawansowania. Języki angielski i hiszpański, wspólnie najbardziej uniwersalne, w świetle powyższych badań okazują się być mniej rygorystyczne w kwestii częstości stosowania interpunkcji. Prawdopodobnie języki te są tak sformalizowane pod względem konstrukcji zdania, że jest w nich mniej miejsca na niejednoznaczności, które musiałyby być rozwiązywane za pomocą znaków interpunkcyjnych.

Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego Polskiej Akademii Nauk (IFJ PAN) w Krakowie prowadzi badania podstawowe i aplikacyjne w obszarze fizyki oraz nauk pokrewnych. Główna część działalności naukowej Instytutu koncentruje się na badaniu struktury materii, w tym własności oddziaływań fundamentalnych od skali kosmicznej po cząstki elementarne. Częścią Instytutu jest nowoczesne Centrum Cyklotronowe Bronowice, unikalny w skali europejskiej ośrodek, obok badań naukowych zajmujący się terapią protonową nowotworów. IFJ PAN prowadzi też cztery akredytowane laboratoria badawcze i pomiarowe. Wyniki badań – obejmujących fizykę i astrofizykę cząstek, fizykę jądrową i oddziaływań silnych, fizykę fazy skondensowanej materii, fizykę medyczną, inżynierię nanomateriałów, geofizykę, biologię radiacyjną i środowiskową, radiochemię, dozymetrię oraz fizykę i ochronę środowiska – są każdego roku przedstawiane w ponad 600 artykułach publikowanych w recenzowanych wysoko punktowanych czasopismach naukowych. Corocznie Instytut jest organizatorem lub współorganizatorem wielu międzynarodowych i krajowych konferencji naukowych oraz szeregu seminariów i innych spotkań naukowych. IFJ PAN jest członkiem Krakowskiego Konsorcjum Naukowego „Materia-Energia-Przyszłość”, któremu, na lata 2012-2017, nadany został status Krajowego Naukowego Ośrodka Wiodącego (KNOW). Wiele projektów i przedsięwzięć realizowanych przez Instytut jest wpisanych na Polską Mapę Infrastruktury Badawczej (PMIB). Instytut zatrudnia ponad pół tysiąca pracowników. Komisja Europejska przyznała IFJ PAN prestiżowe wyróżnienie „HR Excellence in Research” jako instytucji stosującej zasady „Europejskiej Karty Naukowca” i „Kodeksu Postępowania przy rekrutacji pracowników naukowych”. W kategoryzacji MEIN Instytut został zaliczony do najwyższej kategorii naukowej A+ w obszarze nauk fizycznych.

KONTAKT:

prof. dr hab. **Stanisław Drożdż**
Instytut Fizyki Jądrowej im. Henryka Niewodniczańskiego Polskiej Akademii Nauk
tel.: +48 12 6628220
email: stanislaw.drozd@ifj.edu.pl

PUBLIKACJE NAUKOWE:

„Universal versus system-specific features of punctuation usage patterns in major Western languages”
T. Stanisław, S. Drożdż, J. Kwapien
Chaos, Solitons & Fractals, 168, 113183, 2023
DOI: <https://doi.org/10.1016/j.chaos.2023.113183>

POWIĄZANE STRONY WWW:

<http://www.ifj.edu.pl/>
Strona Instytutu Fizyki Jądrowej Polskiej Akademii Nauk.

<http://press.ifj.edu.pl/>
Serwis prasowy Instytutu Fizyki Jądrowej PAN.

MATERIAŁY GRAFICZNE:

IFJ230419b_fot01s.jpg HR: http://press.ifj.edu.pl/news/2023/04/19/IFJ230419b_fot01.jpg
Widoczne na ilustracji funkcje hazardu pokazują prawdopodobieństwo wystąpienia kolejnego znaku interpunkcyjnego w zależności od długości pozbawionej interpunkcji sekwencji wyrazów. Pod względem interpunkcyjnym najbardziej „międzyjęzykowy” jest niemiecki (wykres zielony). (Źródło: IFJ PAN)