

„Tem lepiej, ale jest to interes milionowy i traktujemy go poważnie” — 1000 razy po 1000 słów w 5 częściach

Joanna Bilińska, Magdalena Derwojedowa, Witold Kieraś,
Monika Kwiecień Danuta Skowrońska, Robert Wołosz

Uniwersytet Warszawski

Uniwersytet w Pécsu



Seminarium ZIL
Warszawa, 21 kwietnia 2016



Porządek wystąpienia

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

1 Wstęp

2 Makrostruktura

3 Mikrostruktura



Uwagi wstępne

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- 1 Wstęp
 - Projekt
 - Zestaw znaków
 - Cytaty, elementy obce
- 2 Makrostruktura
- 3 Mikrostruktura



Projekt

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Automatyczna analiza fleksyjna tekstów polskich z lat 1830-1918 z uwzględnieniem zmian w odmianie i pisowni (2013–2016; DEC-2012/07/B/HS2/00570)

- koncepcja leksykograficzna opisu zmian pisowni i fleksji w (elektronicznym) słowniku gramatycznym
- zrównoważony i oznakowany korpusu tekstów (1 mln segmentów)
- wzorce paradygmatyczne aktywne w badanym okresie
- systematyka kategorii fleksyjnych i ich wartości
- schemat opisu zjawisk ewolucyjnych w słowniku elektronicznym
- poszerzenie słownika analizatora morfologicznego o słownictwo dawniejsze



Zestaw znaków

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

174 różne znaki:

- cyfry i liczby (0...9, ½)
- znaki interpunkcyjne
- symbole walut, operatory i inne znaki (np. \$, £, , ×, ¬, §, &, †, °, ±, %, º)
- litery łacińskie (A-Z, a-z, Å, å, Æ, æ, Ć, ć, Ł, ł, Ń, ń, Ó, ó, Ś, ś, Ž, ž, Ž, ž, à, á, â, Ä, ä, ç, È, è, é, ê, ë, ò, ó, ô, Ö, ö, õ, ü, ů, ƚ)
- litery greckie (α, β, γ, Δ, δ, ζ, θ, π, ρ, ω)



Cytaty, elementy obce

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- litery greckie w formułach i wzorach, np.
prawdopodobieństwo trafienia celu ($x_2 - x_1$) będzie wynosić
 $2 - 1/2$ (1916)
 $MC^2 = mc^2 = 2a$ (1912)
- nazwy własne, np. *Göthe, Blücher, Molière*)
- tytuły, np. *der gute Fridolin und der böse Dietrich*
- wyrażenia, np. *C'est bête!*
„Es ist im Gegentheile ganz natürlich”



Makrostruktura

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

1 Wstęp

2 Makrostruktura

- Budowa korpusu
- Próbki w poszczególnych latach
- Style
- Zróżnicowanie stylistyczne korpusu
- Podstawowa charakterystyka słownictwa
- Nierozpoznane segmenty
- Nierozpoznane w latach
- Autorzy
- Biblioteki
- Miejsce wydania

3 Mikrostruktura



Budowa korpusu

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- 1000 próbek gronowych
- 1000 słów w próbce
- 5 stylów



Próbki w poszczególnych latach

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

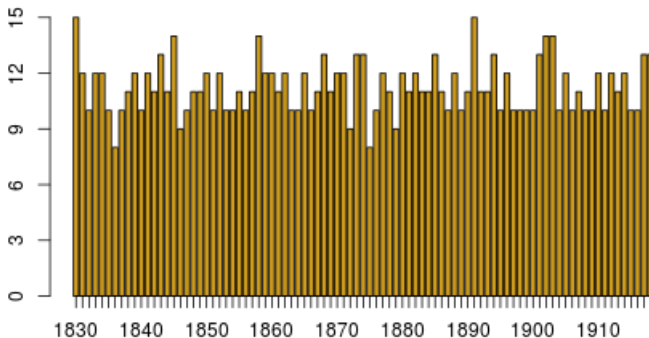
Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład





Style

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- 1 popularnonaukowy
- 2 drobne wiadomości prasowe
- 3 publicystyka
- 4 proza artystyczna
- 5 dramat artystyczny



Popularnonaukowy

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- podręczniki
- monografie
- artykuły w czasopismach (111):
 - „Biblioteka warszawska” (1841-1914; 17)
 - „Kosmos” (od 1876; 10)
 - „Ateneum” (1876-1901; 7)
 - „Wszechświat” (od 1882; 6)
- liczba nierozpoznanych:
 - min — 13
 - max – 253



PN: Z. Kramsztyk, *Cel jako wskazówka badania*, Krytyka Lekarska, 1/5, 1897 (13)

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Bo zasadą, podstawą, nauki jest to przekonanie, że świat z całą swoją różnorodnością stanowi jedność i całość. Tym węzłem, który służy do połączenia wszystkich zjawisk świata, jest dla nas zasada przyczyny: zjawisko każde staramy się odnieść do innych zjawisk, z których, jako z przyczyn, koniecznie wypływa; a tak postępując coraz dalej, dążymy do objęcia całości i wierzymy, że na tej, i jedynie na tej, drodze zadanie nauki spełnimy, osiągniemy poznanie zupełne. Ale **plódnem** dla nauki jest i zadanie odwrotne: zasada celu. Chcąc zgłębić jakieś zjawisko, złączyć je z **innemi**, możemy zapytać nie tylko: **z kąd** to zjawisko powstało? ale i: po co ono istnieje, dla jakich innych zjawisk i przedmiotów jest ono **korzystnem**, jakie inne zjawiska od niego zależą, jaki jest cel jego? W wytworach przemysłu ludzkiego to pytanie o cel zawsze wyraźnie stawiamy, a odpowiedź na nie również jest dla nas nauczającą, nieraz i bardziej nauczającą, niż pytanie o przyczynę układu czy zjawiska.



PN: F. Paszkowski, *O topografii gór*, 1830 (253)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

[...]Dawne te Lawy wulkanicznych roztopionych skał, stanowią osobną i obszerną dzielnicę niepodległą ({**terrains indépendans**}) skał **zagadliwych**, ({**problematicques**}), **iakimi** są Trachyty i **Basalty** ; **massy** lite z **feldspatu** i **Hornblendy**, (**rogomienia**) albo z feldspatu i Augitu{Augit, ({**Pyroxène**})} nazwany tak od blasku; więc spolszczyć go można na **blasczeń.**} złożone, które **Cordier** nazwał skałami **ogniorodnemi** ({**roches pyrogènes**}), a które przez niego i przez **Fleriau de Bellevue** **oryktognostycznie** rozkładane({**Journal de Physique**}) Tom LI. LX. i LXXIII.}, dotąd w sposobie powstania **swoiego** nie są **docieczone**. Liczą się one do utworów wulkanicznych, bo neptunicznych pochod **cale wyłączaia**, ale żadnego **niemaia** podobieństwa do Law **teraźnieyszei gorzelcami** wymiatanych. [...]



Drobne wiadomości prasowe

Interes milionowy

„Gazeta Lwowska”	16
„Gazeta Warszawska”	13
„Kurjer Warszawski”	8
„Tygodnik Petersburski”	7
„Czas”	5
„Gazeta Wielkiego Xięstwa Poznańskiego”, „Gazeta Polska”, „Dziennik Poznański”	4

- liczba nierozpoznanych:
min — 23
max — 253



DWP: „Przegląd Oświatowy”, 1918 (23)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Śmigiel. Wystawa książek otwartą została w pierwsze święto wielkanocne po sumie na salce Ducha św. na **wikarjacie**. Salka była pięknie przystrojona w biusty sławnych mężów polskich, kwiaty, zieleń, malowanki itd. Zainteresowanie się wystawą z początku niezbyt wielkie rośnie coraz więcej w miarę zapoznawania się z książkami. Najwięcej zajęta wystawa dzieci i młodzież ze Śmigla i bliższej okolicy. Wystawa pozostawiła korzystne i miłe wspomnienie u wszystkich zwiedzających. Największą w tym zastęgę mają pani Ł. i panienki ze Śmigla, które tak szczerze zajęły się i wystawą samą, przystrojeniem stołów i salki, **gustownem** i **pięknem** ułożeniem książek, za co im się szczerza należy podzięka.



DWP: „Gazeta Warszawska 324, 1 grudnia 1832” (253)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Hollandya. Z Hagi 18 Listopada. — Dnia 17 odczytano w **cytadelli Antwerpskiej** następujący rozkaz dzienny: „Waleczni towarzysze broni! Zbliża się chwila, w **ktoréy** synowie **dawnéj** Batawii nowe **maią** dać dowody **swéy** odwagi i wierności. Wkrótce armia **Francuzka** stanie pod murami **naszemi**, w celu zmuszenia nas siłą oręża do oddania **cytadelli**. **Polegaiąc** na słuszności sprawy **naszój**, na **waszój** **doświadczonéj** Odwadze, **waszém** przywiązaniu do Króla i **oyczyzny**, z niewzruszoną stałością oczekiwać będziemy nieprzyjacielskich zastępów. Towarzysze broni! Cała **Hollandya** a nawet cała Europa zwrócone ma na was oczy. Okażcie światu, że **iesteśmy** godni zaufania, **iakie** w nas ulubiony Monarcha pokłada, i postanówmy bić się do ostatniego. Niech **żyje** Król!” (podpisano) Baron **Chassé**.



Publicystyka

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- książki (113)
- prasa i czasopisma (87)
 - „Wiadomości polskich” (6)
 - „Rozwaga”, „Przeglądu Tygodniowego Życia Społecznego, Literatury i Sztuk Pięknych”, „Naród” (4)
- 70 tekstów anonimowych
- liczba nierozpoznanych:
min — 9
max – 182



Pub: „Dla Wszystkich” 6/18, 1918 (9)

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Pokuta jest zmianą umysłu. Czy do tego trzeba nam tylko myśli zmienić? **Istotnem** znamieniem pokuty nie jest tylko zmiana umysłu, ale i zmiana czynu. O Józafacie powiedziano: "On zniósł". Oczywiście zmienił się kierunek jego myśli, główną rzeczą jednakże było, że coś robił. Pewien kapitan brał często z sobą na okręt swojego chłopaka, który był zawsze bardzo nieszczęśliwy, ile razy nie pozwolono mu u steru zasiadać i kierować okrętem. Ojciec kazał mu w swojej kajucie sporządzić kółko, którem mógł dowoli obracać. To go bardzo uszczęśliwiło, bo wyobrażał sobie, że kieruje okrętem. Jeżeli zmiana umysłu nie zmieni życia, to jest to także zabawką, jak u tego chłopca. Ale jeżeli wszystkie nasze czyny i dążenia przybrały zupełnie inny kierunek, wtedy jest prawdziwą pokutą. Pokuta wymaga odwagi. Gdym się uczył konno jeździć, spotkałem pewnego razu podczas tej lekcji, doświadczonego człowieka, który mi radził z konia zsiść i siodło poprawić.



Pub: „Czasopism Naukowy...” 3/4 1830 (182)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Naypiérwsza iest nieufność zaszczepiona w nich przez nasze zamiary, nieufność zbyt usprawiedliwiona przez nasze postęпки. Przez trzy wieki czegożeśmy żądali od Afryki? Nic innego prócz niewolników i złota. **Cożeśmy iéy** dali **zato** w zamianę? Oto strzelbę i proch, a to dla tego żeby ułatwić **iéy dziecióm wzajemne** siebie wyplenianie; mocne **napoie** **zniżaiące** ich do zwierząt; kilka przedmiotów płocheho zbytku, bursztyn, ziarnka szklanne, galonowane kapelusze, suknie **szarletańskie**; toż to **iest** wszystko co opłaciło wolność i życie ludzkie! To **téż** oświata **nieiako** chroniąc się przed nami, skryła się w środek ich **kraiu**. To też gdy naród, który bądź przez miłość ludzkości, bądź przez politykę postawiwszy się na czele towarzyskiego poruszenia kuli **ziemnéy**, w ostatnich czasach zrobił kilka wspaniałych dążeń, skutek nie zupełnie odpowiedział **iego** staraniom; ukształcenie przezeń wprowadzone **nieprzestąpiło** granic osad **iego**; miasta przezeń zbudowane i zaludnione nieszczęśliwymi wyrwanymi z rąk **ludziokupców**, rozsiały trwogę po **kraiu**;



Proza

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- powieści
- opowiadania
- powieści poetyckie, poematy epickie („proza wierszowana”)
- liczba nierozpoznanych
min — 7
max — 208



Pr: M. Bałucki, *Grześ niepiśmienny*. *Powiastka*, 1889 (7)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Są tam i rodzice Grzesia. Piją, jedzą, bo stary wójt pamięta o tem, iżby im na **niczem** nie brakowało; ale markotni, że to Marcin nie Grześ bierze Marynkę. Czują w duszy urazę do syna, że zawiódł ich nadzieje i dał sobie zabrać taką bogatą dziewczynę, jak Marynka. — Pożałuje on tego — mówił Wojciech — ale po niewczasie. Wtem pomiędzy tym tłumem ktoś, jak cień przesunął się nagle i wpadł szybko do izby, gdzie tańczono. Za chwilę w izbie dał się słyszeć krzyk — muzyka grać przestała naraz — a ciekawi zaczęli się tłoczyć w tę stronę, aby się dowiedzieć, co się stało. Jakiś głos donośny słychać było.

— To tak? Toś ty taka? — Toś mnie przyrzekała a za innego poszłaś? Głos brzmiał donośnie wśród ciszy, jaka zapanowała między słuchaczami.

— Kto to taki? — pytano się między sobą.

— Jakiś wojak.

— To Grześ.



Pr: A. Nakwaska, *Aniela czyli Słubna obręczka: powieść narodowa*, 1831 (208)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Nieco zwiędła piękność Laury, wiele sobie **jeszcze** w **swoim** mniemaniu o własnych wdziękach i ich uroku obiecywała. Zdawało **ięy** się zawsze, że **jeszcze** **iest** piękną bo nią kiedyś była. Ale co **gorzēy**, zdawało **ięy** się że **iest** dowcipną i zabawną, a **tēm** nigdy **ięy** być nie dozwoliło niebo. Laura **domyślaiąc** się, że **ięy** wdzięki niestety troszkę **mniēy** **iak** **dawniēy** czynią **iuż** wrażenia; a że zawsze potrzebowała się bawić, tańcować i stroić, wzięła się na inny sposób: **ulegaiąc** nieubłaganym skutkom czasu, stała się ogólną doradczynią matek, młodych mężatek i panienek na wydaniu. Wszystkie modniarki podciągnęła pod swą krytykę. Towarzystwo żadne nie było eleganckie, kiedy ona listy Dam na **nię** proszonych nie dyktowała. To wszystko Pani D. umiała tak zgrabnie o sobie powiedzieć, i tak to zręcznie Pani **Różewskiēy** wyperswadować, że ta ostatnia nic bez rady pani Laury **przedsięwziąć** nie śmiała; a pani Laura, w **dobranēm** gronie swoich **przyjaciótek**, {{de la société par excellence}}, **iak** **iā** nazywała, mówiła o pani **Rożewskiēy**: {{Pauvre femme, elle est si bornée! Sa filie est riche et assez bien; il faut que j'en fasse quelque chose... Je me dévoue en femme d'esprit, et mon assiduité dans cette maison,}}



Dramat

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- dramaty
- wodewile
- libretta operetkowe
- utwory wierszowane (78)
- liczba nierozpoznanych
min — 10
max — 280



Dr: A. Świętochowski, *Poddanka. Dramat w I akcie*, 1879 (10)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

[[[Onufry ukazuje się]]]

Cóż?

[[ONUFRY.]]

Kazałem dać jej inną odzież.

[[ANDRZEJ.]]

Co mówi?

[[ONUFRY.]]

Nic.

[[ANDRZEJ.]]

Powoli nawet śpiewać ją nauczymy. Ale ponieważ to wszystko dla ciebie, staraj się wytłomaczyć że tylko zupełna uległość mej woli zbawić ją może. Byłoby najlepiej, ażeby rozumiała tę konieczność szybko. Bo naprzód szkoda jej męczyć, a powtórnie nie trzeba robić we wsi i okolicy zbyt wielkiego zgorszenia. Jak wojewodzina z nią się obchodziła?

[[ONUFRY.]]

Jak z **własnem** dzieckiem.



Dr: K. Gaszyński, *Waryat z potrzeby. Krotofila ze śpiewkami...*, 1830 (280)

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

Tam huk chłopców, **iaki** taki
Pali do **swéy** koperczaki;
A Zagonek z **Wroniéy** Góry,
Do mnie wciąż **ieździ** w konkury,
Lecz to starzec **iuż** skrzywiony,
Do różańca nie do żony.
Więc na koszu go zostawię,
Ja **wotę tutaj** w Warszawie.

[[Hulakowski.]]

Wszystko to prawda, {{verum est}}, ale i to prawda że musisz iść za
Zagonka {{sine contradictione}}.

[[**Inocencia.**]]

Nie spodziewam się abyście mnie przymuszali Tateńku, bo **iak**
czytałam w Romansie Walter **Skota**...

[[Dydak.]]

Daymy temu na ten czas **pokóy**... **Oycze** i Siostro poświęćcie te
chwile mnie, którym was od tak dawna nie widział.

Zróżnicowanie stylistyczne korpusu

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

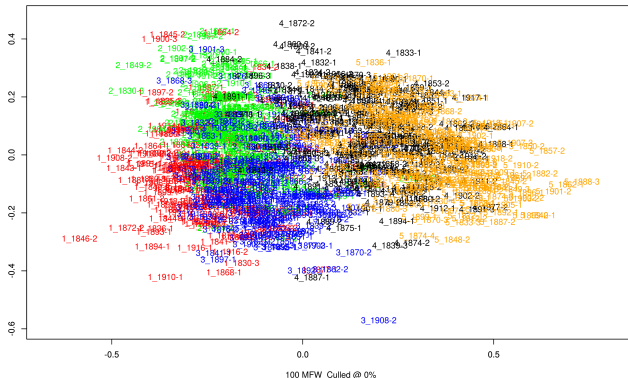
Mikrostruktura

Próbki

Konwencje

Przykład

korpus-f19-1.0-beta
Multidimensional Scaling





Podstawowa charakterystyka słownictwa

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

STYL	SEGM.	TYPY	HAPL	DISL	TRISL
PN	202342	40482	28378	6914	3022
DWP	200030	44354	32057	7310	3258
PUB	200995	42900	29907	6837	2976
PA	206176	44161	30245	6991	2934
DR	238013	41184	26383	5997	2651



Nierozpoznane segmenty

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

słowa nierozpoznane w stylach

drobne wiadomości prasowe	15 757
popularnonaukowe	18 260
publicystyka	12 743
proza artystyczna	11 164
dramat	29 015
suma	86 939



Nierozpoznane w latach

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

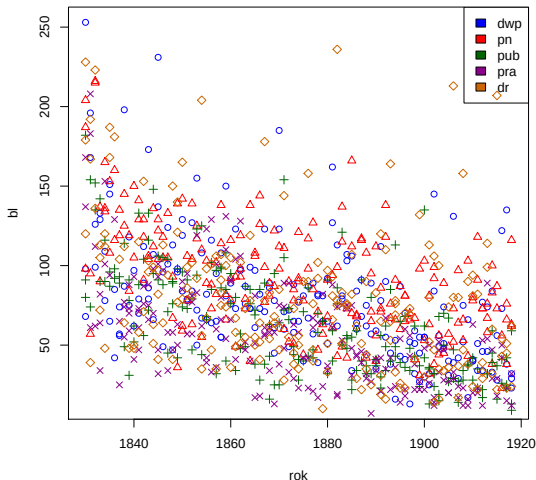
Mikrostruktura

Próbki

Konwencje

Przykład

Nierozpoznane ciągi



Autorzy

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

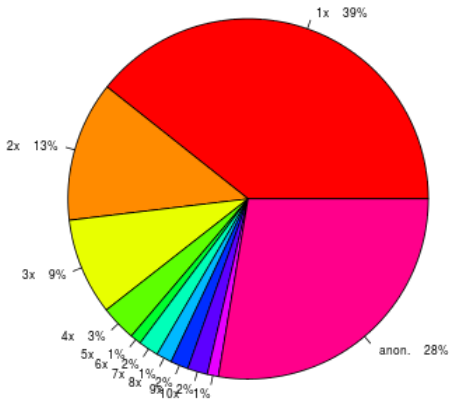
Mikrostruktura

Próbki

Konwencje

Przykład

liczba wystąpień/liczba autorów





111 autorów:

- 10 J. Słowacki
- 9 E. Orzeszkowa, J.I. Kraszewski
- 8 S. Przybyszewski, J. Korzeniowski
- 7 J. Dzierzkowski, M. Bałucki
- 6 E. Lubowski, Z. Krasiński, W. Koziembrodzki
- 5 S. Wyspiański, J. Blusiński
- 4 J. Żuławski, H. Sienkiewicz, W. Marrené, S. Kramsztyk, A. Gruszecki, W. Feldman, S. Bogusławski, W. Anczyk Ludwik
- 3 28, m.in. S. Żeromski, A. Świętochowski, W.S. Reymont, B. Prus, W. Łoziński, K. Libelt, M. Konopnicka, A. Fredro,
- 2 63, m.in. J.N. Baudouin de Courtenay, J. Szujski, W. Syrokomla, F. Skarbek, E. Romer, K. Przerwa-Tetmajer, W. Pol, A.E. Odyniec, C.K Norwid, T. Miciński, S. Leśniewski, J. Klaczko, W. Gomułicki, Z. Gloger, A. Dygasiński, K. Drzewiecki, F.S. Dmochowski, T. Chałubiński, S. Bełza,





Biblioteki

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

328	Polona
134	e-BUW
100	Wielkopolska BC
79	Śląska BC
67	Jagiellońska BC
64	Google Books
56	WikiSource
21	Kujawsko-Pomorska BC
14	Wojewódzka Biblioteka Publiczna im. H. Łopacińskiego w Lublinie
12	RCIN, Podkarpacka BC, Radomska BC
11	Podlaska BC
10	Dolnośląska BC

2-8	PBI, BC UŁ, BC UMCS, Małopolska BC, Pomorska BC, BC PW, BC UWM, Mazowiecka BC, Chełmska BC, BC Regionalia Ziemi Łódzkiej, BC PŁub, BC INiG - PIB, Archive.org, Akademicka BC AGH, Świętokrzyska BC, Biblioteka Literatury Polskiej w Internecie, BUW, Biblioteka UŁ, Biblioteka UG, Biblioteka ILP UW, BC UWM, Wolne Lektury
1	Zielonogórska BC, Tarnowska BC, RC IFiS PAN, Płocka BC, Łódzka Regionalna BC, Krośnieńska BC, Biblioteka UMK, Biblioteka Kórnicka PAN, Biblioteka Główna AGH, BC UWR, BC UE



Miejsce wydania

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- 166 różnych miejsc
- najczęstsze:

miejsce wyd.		pn	dwp	pub	pa	dr
Warszawa	390	116	75	53	90	56
Lwów	156	30	21	37	28	40
Kraków	114	23	12	29	14	36
Poznań	71	11	24	16	6	14
Paryż	60	6	–	32	10	12
Wilno	50	2	6	1	24	17



Mikrostruktura

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

1 Wstęp

2 Makrostruktura

3 Mikrostruktura

- Próbki
- Konwencje notacyjne
- Przykład



Próbki

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

• tekst



Próbki

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- tekst
- metryczka



Próbki

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- tekst
- metryczka
- źródło (skan; , pdf, . jpg, DjVu)



Próbki

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- tekst
- metryczka
- źródło (skan; , pdf, . jpg, DjVu)
- w wersji roboczej również plik korekty



Konwencje notacyjne

Interes milionowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

- fragment nieczytelny lub obcy { { . . . } }
- przypis { . . . }
- *dramatis personae* [[. . .]]
- didaskalia # . . . #



Przykład

Interes miljonowy

Wstęp

Projekt

Zestaw znaków

Obce

Makrostruktura

Budowa

P/R

Style

PN

DWP

Pub

Pr

Dr

MDS

Słownictwo

Nierozpoznane

Ign/r

Autorzy

Biblioteki

Miejsce

Mikrostruktura

Próbki

Konwencje

Przykład

W. Rapacki, *Odbijanego. Komedia w pięciu aktach*, Warszawa
1888, 1888_5.1_meta.txt.

```
[[Starosta.]] #(do kozaczka)#; Nalej mu!  
[[Paweł.]] #(biorąc kielich)# Za przezacne drogie  
zdrowie wasze!  
#(pije)# Ha! po miodzie gładko i słodko interesik  
nasz pójdzie.  
Wczoraj w Glinianach miałem to szczęście widzieć  
dostojne oblicze  
pańskie i jego małżonki.  
[[Starosta.]] #(odkrząknąwszy mocno)# {{Ad rem}}!
```



Tekst

Interes miljonowy

Ostrzeżenie dla cyklistów. Podpisane grono cyklistów, które w ostatnich dniach przejeżdżało przez Delatyn ostrzega wszystkich cyklistów którzyby tę turę odbywali, by pomijali sklep Narodnej Torhowli zostający pod zarządem niejakiego p. Kwiatkowskiego. Pan ów w stosunku do zatrzymujących się u niego gości, okazał się już kilkakrotnie takim gburem i człowiekiem pozbawionym najprostszych zasad grzeczności. ie wszelki stosunek z nim jest nad wyraz przykry a nieraz wprost kompromitujący. Notatkę tę, którą pomieszczamy równocześnie w »Kole« organie galicyjskiej kolarzy, podajemy do wiadomości kolarzy i turystów z tem przeświadczeniem że oszczędzimy im możliwych nieprzyjemności, o jakie z wojowniczym czy chorym na nerwy p. Kwiatkowskim nie trudno. Edmund Olankowski, Eustachy Turzański i Michał Pietrustewicz, Karol Chrzanowski.





Metryczka

Interes miljonowy

autor: Rapacki, Wincenty

tytuł: Odbijanego. Komedia w pięciu aktach

data wydania: 1888

miejsce wydania: Warszawa

redaktor:

tytuł książki:

tytuł gazety, czasopisma, serii wydawniczej:

nr:

wydawnictwo: Gebethner i Wolff

numery stron: 22-26

styl: dramat

źródło: Podlaska Biblioteka Cyfrowa

link: <http://pbc.biaman.pl/dlibra/doccontent?id=32080>





Źródło

Interes milionowy

wielu innych występów ochrania? Ależ gdy z drugiej strony, sławimy chciwość wielu dzierżawców, i niektórych dziedziców, mianowicie spekulantów, kupujących dobra bez pieniędzy, którzy w folwarkach, gdzie w porównaniu pól dworskich, pańszczyzna jest zamała, nie chcą wydawać na najem, a pragnąc przez wyciągnięcie intrat wyjść z interesów, nieraz

Jako bezwarunkową tezę stawiam:

„Ropa w Galicyi, gdziekolwiek takowa występuje, znajduje się na powtórne złóżysku” (sekundäre Lagerstätte) i jest skutkiem suchej destylacji z wielkiej ilości nagromadzonych szczątków organicznych, zawartych w starszych warstwach, jak są dotychczas w Karpatach znane pokłady łupku bitumicznego utworu neokomskiego.

