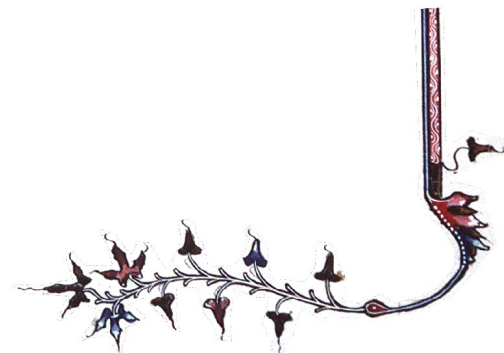
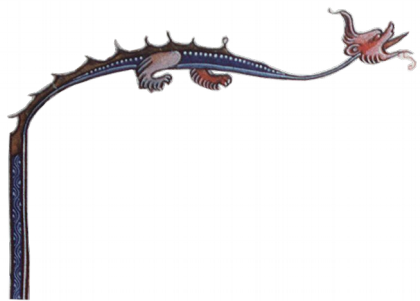


Norma i anomalia w leksykograficznym opisie łaciny średniowiecznej a metody korpusowe

Krzysztof Nowak

Pracownia Łaciny Średniowiecznej
Instytut Języka Polskiego
Polskiej Akademii Nauk





Plan

1) normatywizm wczesnych opisów łaciny średniowiecznej

2) słownik a norma

3) korpus a słownik

4) łączliwość w słowniku

5) frekwencja

6) naruszenie normy

7) okazjonalizmy?

8) poza normą

❖ materiał

🔪 słownik: (e)SŁŚ

<http://scriptores.pl/elexicon>

❖ źródła wiedzy


🔪 praca leksykograficzna

🔪 praca nad korpusem (*eFontes*: 2017)

🔪 studia leksykologiczne

leksykograf = leksykolog z deadlinem (*dedlajnem?* Fillmore 1988, za Kilgariff 1997)





Łacina średniowieczna jako przedmiot wypowiedzi normatywnych



Charles du Fresne
sieur Du Cange (1610-1688)

- ❖ „tradycja narzekania” (Milroy & Milroy 1999) w kręgu mitów językowych
- ❖ aprecjacja łaciny średniowiecznej
 - ‡ nowe spojrzenie na epokę
 - ‡ odrzucenie wartościowania zmiany językowej

‡ łacina klasyczna jako stała opisu i domyślna norma

- ‡ klasycyzm w badaniach literaturoznawczych (Mohrmann 1955)
- ‡ dydaktyka uniwersytecka (Stotz 2003)
- ‡ brak gramatyki łaciny średniowiecznej (do początków XXI w.)





Średniowiecze a „higiena językowa”

❖ teoria poprawności: „łacińskość” (Latinitas)

- § natura i analogia (reguły języka)
- § consuetudo (uzus)
- § auctoritas (wzorce literackie)
- § błąd (vitium) czy przymiot (virtus)?

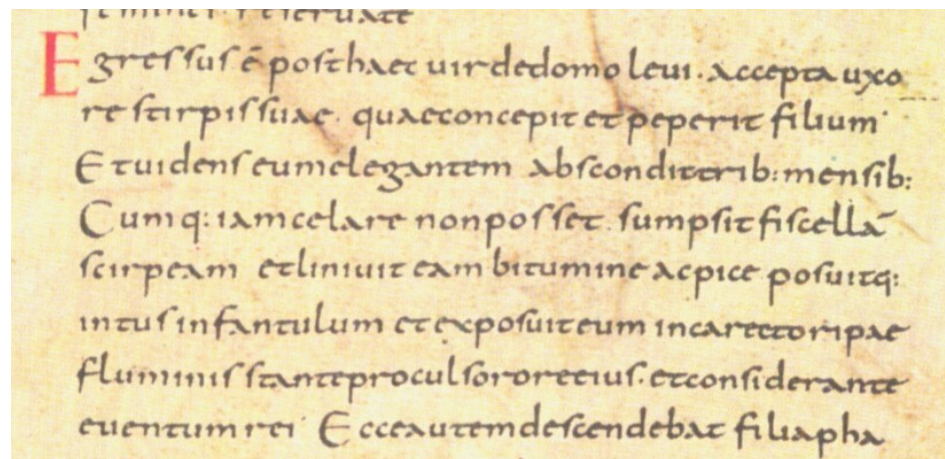
❖ kodyfikacja

- § *artes grammaticae*
- § traktaty ortograficzne
- § komentarze itp.

❖ instytucjonalizacja: reforma karolińska

(Bartsch 1982)

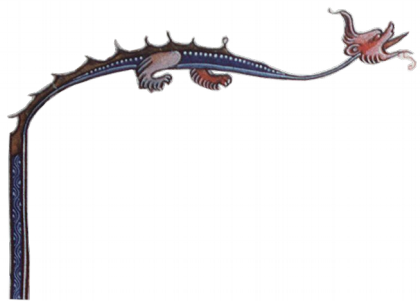
- § **kodyfikacja** traktaty
- § **promulgacja** okólniki
- § **promocja** sieć szkół
- § **standaryzacja** puryfikacja tekstów



Przykład minuskuły karolińskiej

(tzw. Biblia z Moutier-Grandval, ok. 835, Tours)





Słownik a norma

❖ norma „wewnątrzjęzykowa” a frekwencja

§ Zgusta (1971)

- standard-descriptive dictionaries: regularne zjawiska
- overall-descriptive dictionaries: szersza selekcja

§ Atkins & Rundell (2008): słownik opisuje normę

- normalne daje się poznać przez swą powtarzalność (*recurrence*)

§ Bańko (2001): fakty

- frekwencja daje pojęcie o uzusie
- język to domena przede wszystkim powtarzalności

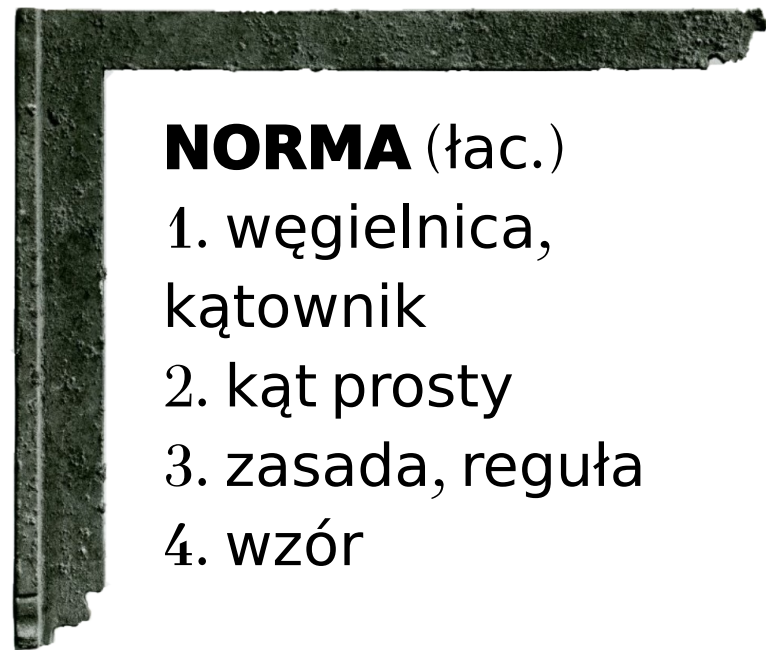
§ kognitywistyka

- Evans (2013): zdolność przeprowadzania operacji „statystycznych” w nauce języka (J. Bybee)
- Taylor (2016): wiedza o frekwencji stanowi część znajomości języka

❖ norma „zewnątrzęzykowa”

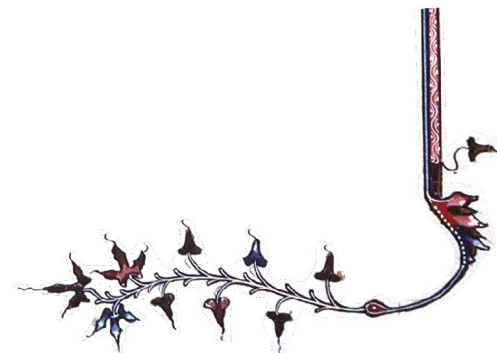
§ Bańko (2001): metafakty

Kątownik, I-III w., Efez (British Museum)



NORMA (łac.)

1. węgielnica, kątownik
2. kąt prosty
3. zasada, reguła
4. wzór





Słownik historyczny a norma

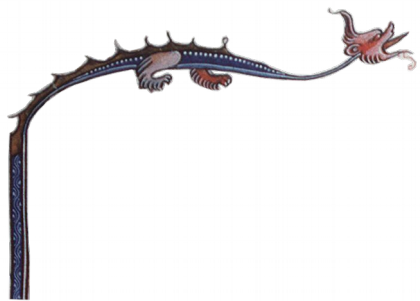
Bańko (2001)

- ❖ normatywizm selekcyjny: kompletność ujęcia
- ❖ normatywizm opisowy: norma dostępna, ale kłopotliwa (liczne źródła, „długie trwanie”, dialektyzacja)

Gruszczyński (2016)

- ❖ co rekonstruują słowniki historyczne?
 - ⌘ norma skodyfikowana w XVII-XVIII w. a uzus
 - ⌘ norma naturalna
- ❖ silna wariantywność
 - ⌘ rozchwianie ortografii zwł. rękopisów
 - ⌘ oboczności morfologiczne
 - ⌘ różne stopnie adaptacji zapożyczeń
- ❖ słownik historyczny musi notować okazjonalizmy, indywidualizmy, gry językowe





Korpus: nadzieje i

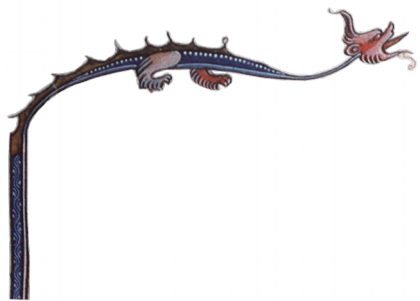
Ewolucja metody

- ❖ obiektywizacja opisu frekwencji
- ❖ „kontekstualizacja” znaczenia i wyczulenie na „efekty” kontekstowe (prozodia semantyczna, koligacja itd.)
- ❖ systematyzacja pracy z konkordancją (Sinclair 2003)

Nihil novi?

- ❖ Bańko (2001, 2012)
 - 📖 tradycyjny empiryzm leksykografii
- ❖ Geeraerts (2009)
 - 📖 leksykografia (CCELD) źródłem prac np. Sinclaira
 - 📖 analiza kontekstowa w leksykografii historycznej





Korpus: nadzieje 2

❖ techniki

🅒 *tickbox*

lexicography

🅒 szybkość i koszt

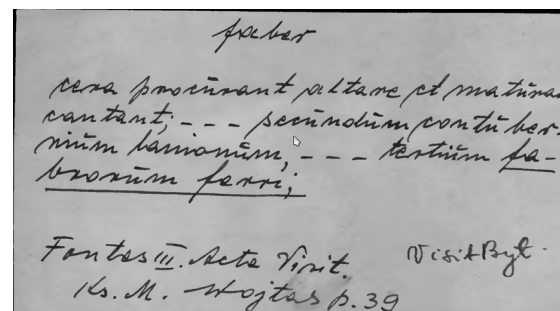
❖ mikrostruktura

🅒 inkorporacja

🅒 wizualizacja

❖ baza materiałowa

🅒 przewyciężenie ograniczeń ekscerpcji ręcznej (Tokarski 1969)



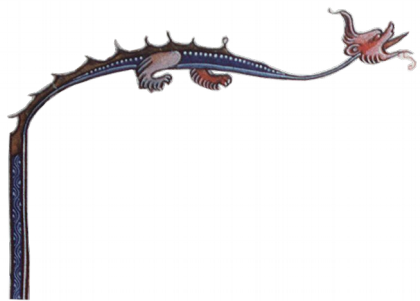
🅒 neutralizowanie „złych danych” (Labov)





Korpusy historyczne: problemy

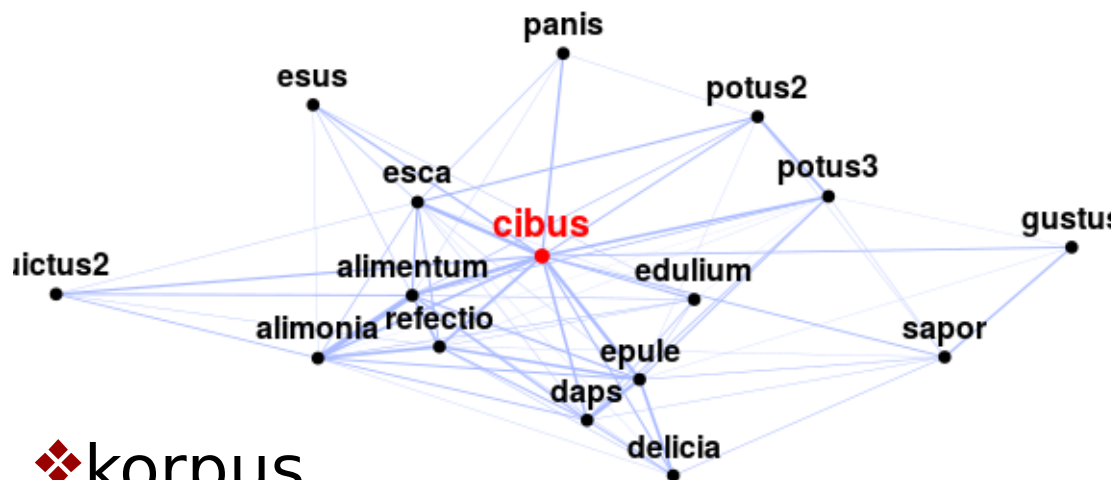
- ❖ przypadkowość zachowania
 - ❖ wątpliwa reprezentatywność
(Hernandez-Campoy & Schilling 2014; Gruszczyński 2016)
 - ❖ wątpliwe zrównoważenie
(*eFontes*)
 - ❖ wątpliwe metadane
 - ‡ czas
 - ‡ miejsce
 - ‡ gatunek
 - ❖ autentyczność?
 - ‡ konserwatyzm
 - ograniczona oralność
 - szkoła
 - skrypturalizm
 - ‡ stan zachowania źródeł
 - ‡ źródło cech językowych: skryba? autor? wydawca?
- 



Opis łączliwości

kookurencja jako epifenomen (Evert 2005)

- ❖ preferencja selekcyjna (SŁŚ: *additur, v.gr.*)
 - § PATERNUS 'ojcowski', np. uczucie, troska itd.
- ❖ kolokacje leksykalne (Cruse 2013; SŁŚ: *iuncturae*)
 - § PECUNIA 'pieniądze' *iuncturae*: przyjmować...
- ❖ idiomy, utrwalone metafory, klisze (SŁŚ: *locutiones, in imag., metaph.*)
 - § *tela araneorum* 'pajęczyna' (TELA 'sieć' + ARANEA 'pająk')



❖ korpus

- § ocena siły połączenia (+)
- § uspójnienie etykietowania
- § 1 *sense per collocation* (+)
- § pomijane efekty kontekstowe (+)
- § grupowanie kolokatów (+)





Frekwencja a dobór haseł

❖ w SŁŚ znajduje się ponad 8550 eksplicytnych *hapaksów*
(> 30% haseł, cf. 40-50% w badanych korpusach)

❖ dystrybucja *hapaksów*

‡ pojedynczy autorzy: **ok. 20 %**

(indywidualizmy? Długosz, Kadłubek,
Tomasz z Wrocławia...)

‡ słowniki: **ok. 8%** (inercja tradycji?)

(Bartłomiej z Bydgoszczy, Jan Tucholczyk...)

❖ korpus

‡ weryfikacja „hapaksowości” (+)

‡ pozyskanie (+)





Frekwencja względna w SŁŚ

- ❖ podstawa: liczba ekscerptów
- ❖ wykładniki
 - ‡ eksplicytne
 - *saepissime, saepe, et saepius*
 - ‡ implicytne
 - brak etykiety przy wyczerpaniu materiału
- ❖ zakres etykietowania
 - ‡ poświadczone znaczenia
 - ‡ składnia
 - ‡ kolokacje
- ❖ braki
 - ‡ brak cytowań, jeśli znaczenie poświadczone w łacinie klasycznej
 - ‡ warianty graficzne
- ❖ korpus
 - ‡ obiektywizacja frekwencji (Bańko 2001; Atkins & Rundell 2008)
 - ‡ włączenie informacji ilościowych
 - ‡ kwerendy *ad hoc*





Anomalia a frekwencja

❖ leksykograf a norma

§ powtarzalność (Kilgarif 1997)

§ przewidywalność (znaczenie: Kilgariff 1997)


§ „psychologizacja” nadawcy: Długosz czy prowincjonalny pisarz sądowy i in.

❖ leksykograf a przekroczenie normy

§ aberracja → odrzucenie, ingerencja (krytyka tekstu)

§ anomalia → kwalifikacja wystąpienia jako atypowego

§ norma → akceptacja



Anomalia: warianty graficzne

CIMOETERIUM 'cmentarz'



Eugène Atget, Grób Abelarda i Heloizy
(źródło: Wikimedia Commons)



caemeterium
cementerium
cemiterium
cimeterium
cimiterium
cimterium
cinisterium
cmiterium
coemeterium
cymmeterium

caemiterium
cemeterium
cimeterium
cimethirium
cimitherium
cinitarium
cminterium
coemetarium
cyminterium

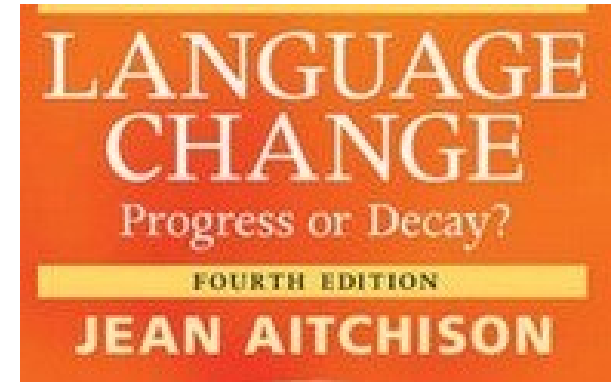
- ❖ notowane wszystkie, rekonstrukcja klasycznych (tylko dla hasłowania)
- ❖ brak frekwencji (+)





Anomalia: warianty morfologiczne

- ❖ fleksja
 - § nieregularność stwierdzona dla ok. 1000 haseł (3%)
 - § norma: łacina klasyczna
- ❖ fleksja rzeczownikowa (407) i czasownikowa (327)
 - § wahania paradygmatu (*patrones* zam. *patronos*; *poliunt* zam. *poluunt*)
 - § niestabilność rodzaju (*pecuda* r.n. zam. *pecudes* r.ż.)
 - § zmienność tematu (*abicis* zam. *abietis*; *pluuit* zam. *pluit*)
 - § pluralis abstraktów (*abstinentia* itp.), formy pasywne w miejsce aktywnych
 - § formy analityczne w miejsce syntetycznych



- ❖ korpus
 - § ocena kierunku zmiany językowej
 - § ile wystąpień tworzy normę?
 - § zjawiska obejmujące grupy leksemów (+)
 - § gramatyka ze słownika?





Znaczenia okazjonalne (I)

❖ ponad 4500 haseł (ok. 20% wszystkich) zawiera **znaczenia** opatrzone etykietą *occ.* ≠ okazjonalizm

❖ różne stopnie krystalizacji

‡ znaczenia kontekstowe

- OBERRO 'błądzić, mylić się': *occ.* 'o strunie: fałszować'
- OBSEQUOR 'służyć' *occ.* 'do mszy, na wojnie'

‡ gramatyka → semantyka

- OBNOXIUS *occ.* substantywizacja
- OBTENERO 'zaciemniać' → *occ. refl.* 'ściemniać się'





Znaczenia okazjonalne (2)

❖ korpus

- ¶ weryfikacja „okazjonalności” (+)
- ¶ niedookreśloność kontekstu (-)
- ¶ pogłębienie „kontekstualizacji” (?)

Hanks (2000), *Do Word Meanings Exist?*: mnożyć czy uogólniać?

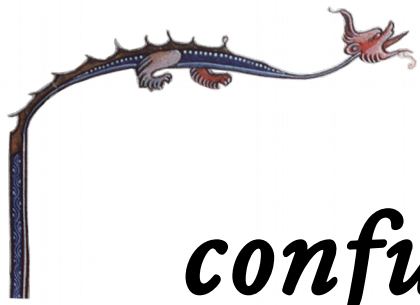
¶ definicje zbyt wąskie

- GUBERNO 'sprawować opiekę nad osobami' *occ.* causam 'nad sprawą'

¶ w kierunku ekwiwalentów tłumaczeniowych

- EDICO 'wypowiadać, oświadczać' *occ.* 'zeznać w sądzie'



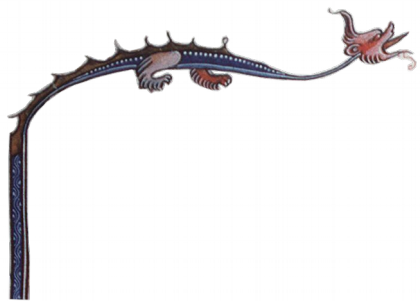


Poza normą?

confunditur cum X 'pomyłone z X'

- ❖ łatwe: mała różnica grafii, duża różnica znaczenia (krytyka tekstu) (-)
 - ⌘ PISCOR 'łowić ryby' → EXPISCO 'wypytywać'
 - ⌘ OBSERVO 'zachowywać' → OBVERSO 'wywrócić, zniszczyć'
- ❖ trudniejsze przypadki, np. złożenia (+-)
 - ⌘ oczywiste omyłki: OBLATIO 'ofiarowanie' → ABLATIO 'odebranie'
 - ⌘ zmiana znaczenia? PRAEHABEO → HABEO, PROHABEO itd..
 - ⌘ wątpliwe (puryzm leksykografa):
PRAECLUSIO 'zamknięcie; zakaz' → CONCLUSIO 'zakończenie'
- ❖ podtrzymywanie normy wbrew frekwencji (-)
 - ⌘ LIMES 'granica' → LIMEN 'próg domu' (regularnie w korpusie i słownikach łaciny europejskiej!)



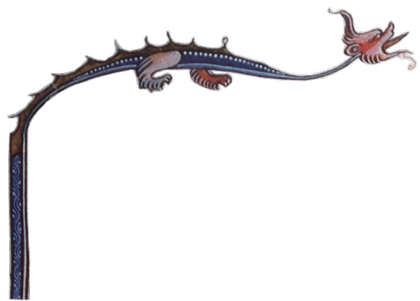


Podsumowanie

rola korpusu

- ‡ obiektywizacja i weryfikacja regularności
- ‡ uwrażliwienie na kontekst w opisie znaczenia
- ‡ porównanie (korpusy innych języków i dialektów)
- ‡ opis zróżnicowania
 - diatopicznego
 - diachronicznego
 - (Gruszczyński 2016: „pozorna synchronia słowników historycznych”)
 - gatunkowego
- ‡ rozbudowa mikrostruktury haseł (wizualizacja, informacja ilościowa itd.): słowniki historyczne są zaprzeczeniem *user friendliness*, dlaczego więc nie skomplikować ich jeszcze bardziej?





Pytania? Sugestie?

- ❖ w poszukiwaniu nowych form leksykografii historycznej
(K. Nowak, B. Bon, R. Alexandre)

<http://medialatinitas.eu>

<http://scriptores.pl/wiki>

- ❖ Elektroniczny słownik łaciny średniowiecznej (PŁŚ IJP PAN)

<http://scriptores.pl/elexicon>

- ❖ Elektroniczny korpus łaciny średniowiecznej

(początek 2017, PŁŚ IJP PAN) <http://scriptores.pl/efontes>

- ❖ email: krzysztof.nowak@ijp-pan.krakow.pl

