

Zastosowanie parsera zależnościowego do analizy składniowej tekstów średniopolskich

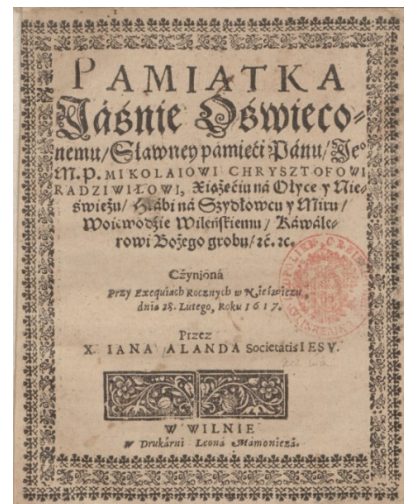
Aleksandra Wiczorek

Plan prezentacji

- materiał
- metodologia
- przebieg prac
- napotkane trudności
- plany na przyszłość

Źródło: Elektroniczny Korpus Tekstów Polskich z XVII i XVIII Wieku

- 13,5 → 25 mln segmentów
- zapis: warstwa oryginalna i uwspółcześniona
 - np. *woysko szweckie* → *wojsko szwedzkie*
- anotacja morfosyntaktyczna
 - (np. *stotem* → STÓŁ, subst, sg inst m)
- metadane (autor, tytuł, rok wydania...)



www.korba.edu.pl

- Kierownik: Włodzimierz Gruszczyński
- Finansowanie: Ministerstwo Nauki i Szkolnictwa Wyższego, Narodowy Program Rozwoju Humanistyki (projekty nr 0036/NPRH2/H11/81/2012 i 0413/NPRH7/H11/86/2018)
- Czas trwania: 2013-2018 i 2019-2023
- Instytucja prowadząca: Instytut Języka Polskiego PAN
- Współpraca: Instytut Podstaw Informatyki PAN

PAN **JP** Instytut Polskiej
Języka Akademii
Polskiego Nauk

Wybór wypowiedzeń do analizy składniowej

- z podkorpusu anotowanego ręcznie
 - podział na zdania
 - znaczniki gramatyczne
 - formy podstawowe
- od 10 do 50 segmentów ($\approx$ wyrazów)
- bez poezji
- bez zdań inkrustowanych łaciną

*Chcąc tedy podług obowiązku te tak szkodliwe Rzpltej rzeczy
avertere, to mi największą **difficultatem** czyniło, żebych był ostrzegł sławy
KJMci...*

Eksperymentalny korpus 1000 wypowiedzeń z XVII i XVIII wieku znakowanych składniowo

- wzorzec: Polski Bank Drzew Zależnościowych (Wróblewska, Rybak 2019*)
 - 22 tys. wypowiedzeń
 - anotacja ręczna
 - różnorodność konstrukcji składniowych (m.in. struktury porównawcze, elipsy, mowa zależna i niezależna)

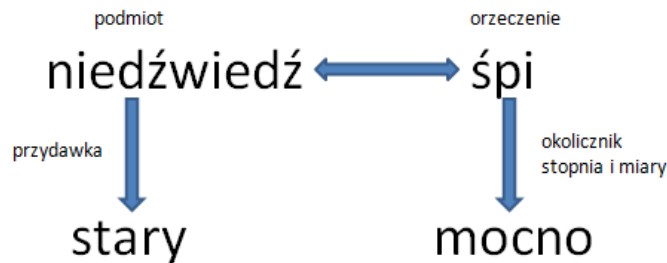
*A. Wróblewska, P. Rybak, *Dependency Parsing of Polish*, w: „Poznań Studies in Contemporary Linguistics” 55(2), 2019, s. 305-337.

<http://zil.ipipan.waw.pl/PDB>

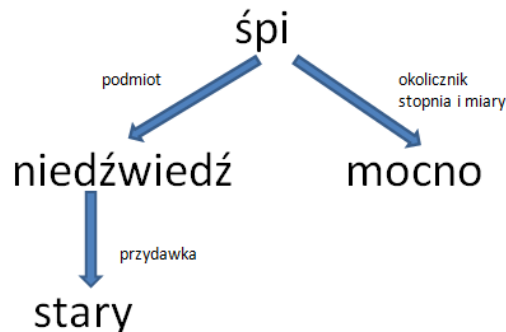
Drzewo zależności składniowych

- graf przedstawiający zależności składniowe pomiędzy składnikami zdania
- każdy wyraz w zdaniu ma dokładnie jeden nadrzędnik
- przykład: *Stary niedźwiedź mocno śpi.*

analiza „szkolna”:

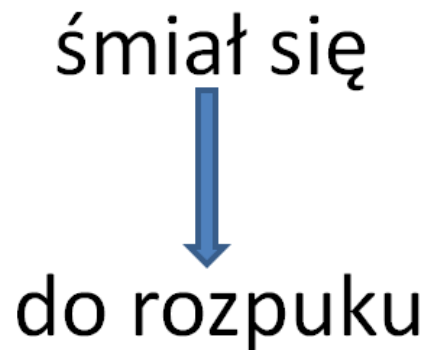


analiza w PBDZ:

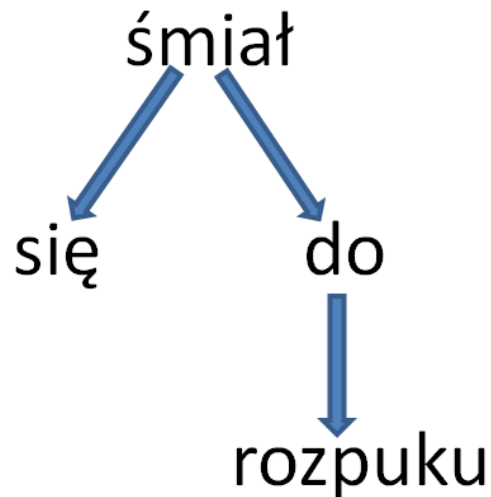


Drzewo zależności składniowych

analiza „szkolna”:

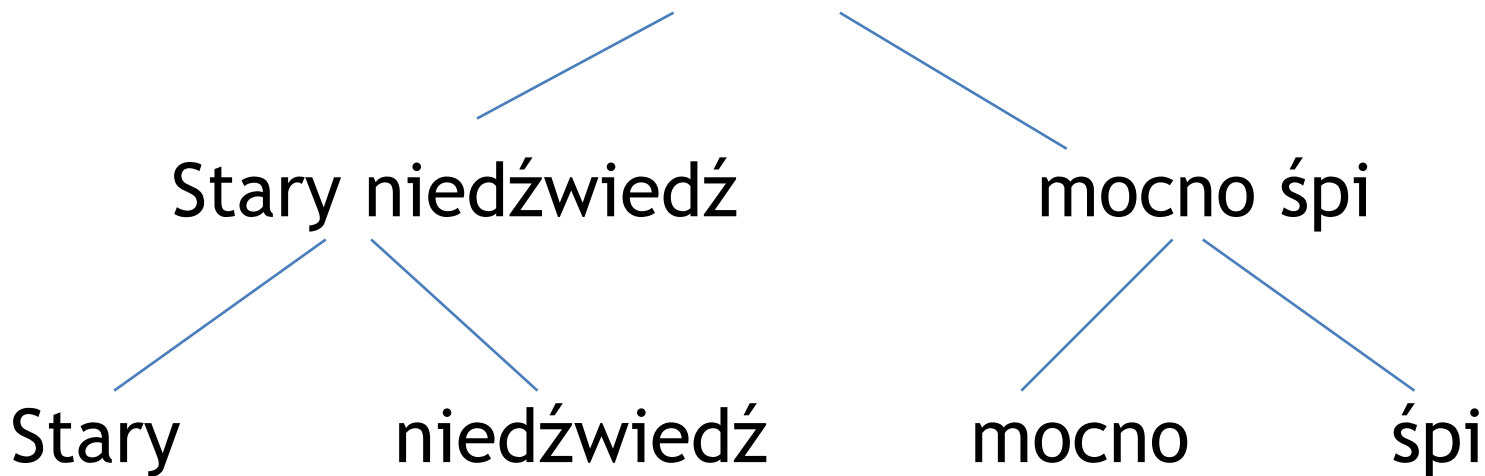


analiza w PBDZ:



Drzewo frazowe (składnikowe)

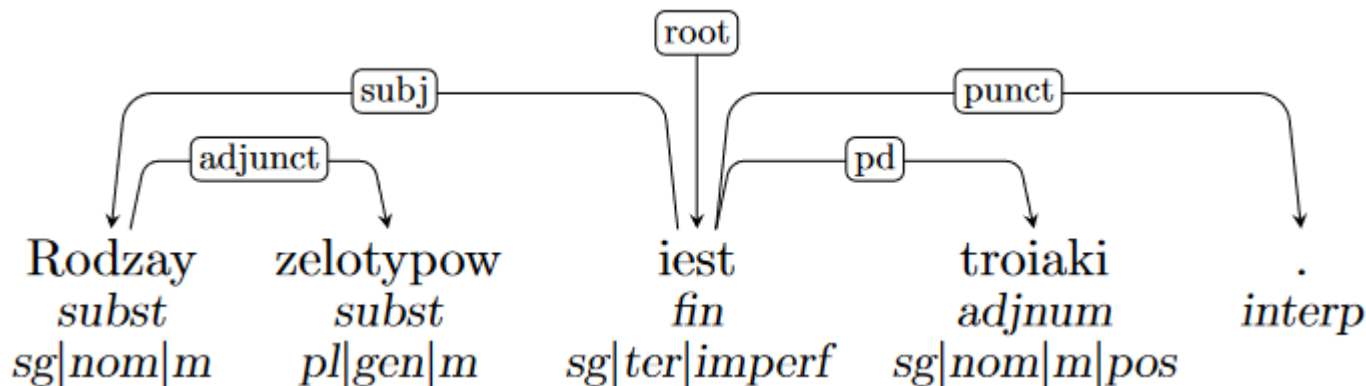
Stary niedźwiedź mocno śpi.



Parser zależnościowy

- program do automatycznej analizy składniowej
- trenowany na ręcznie przygotowanym korpusie wzorcowym
- zalety:
 - rezultatem jest jedno drzewo zależności składniowych
 - dobre wyniki dla języków o (stosunkowo) swobodnym szyku
 - istnienie bardzo wielu banków drzew zależnościowych dla innych języków, w tym historycznych

Przykład z korpusu składniowego zdań XVII-XVIII-wiecznych



- root = element główny
- subj = podmiot
- adjunct = przydawka
- pd = orzecznik
- punct = znak interpunkcyjny niepełniący funkcji składniowej
zelotyp 'zazdrośnik w miłości'

Przebieg pracy

automatyczna analiza
składniowa (parsowanie)



korekta dwóch
anotatorów



wybór dobrej
interpretacji

Napotkane trudności

- interpunkcja
- błędny podział na zdania
- szyk przestawny
- brak orzeczenia
- zdania trudne do zrozumienia
- ...i inne

Interpunkcja średniopolska

- nieuregulowana, niekonsekwentna
- często oddawała intonację języka mówionego, a nie strukturę zdania

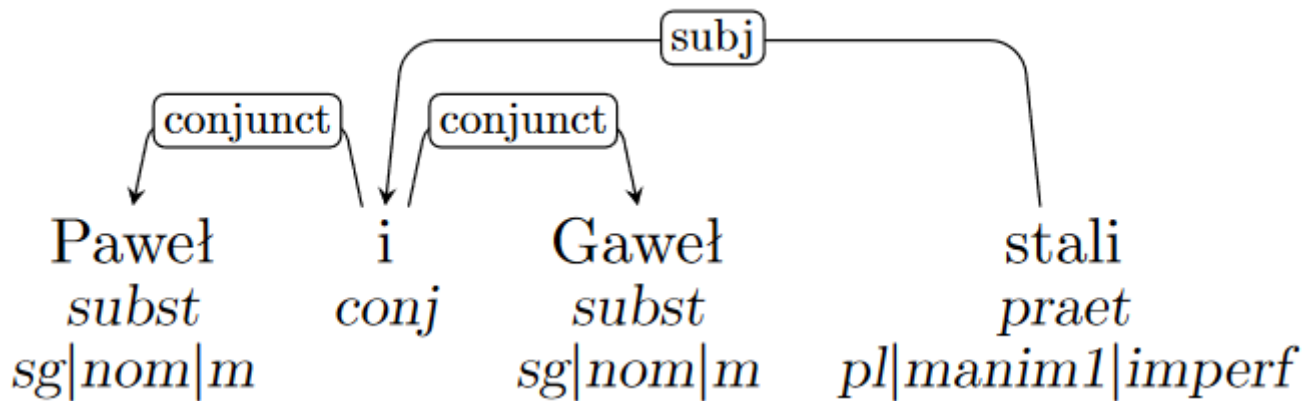
Zachciało mi się Morzēm lechać na Nabozęństwo Wielkanocne do Harrazon gdzie stał Woiewoda bo y blizey było Morzēm nizeli lądem y dlatego tez zeby koni nie turbować.

Kto zgromądzi/ w mych oczách/ obfitych też brody?

Interpunkcja

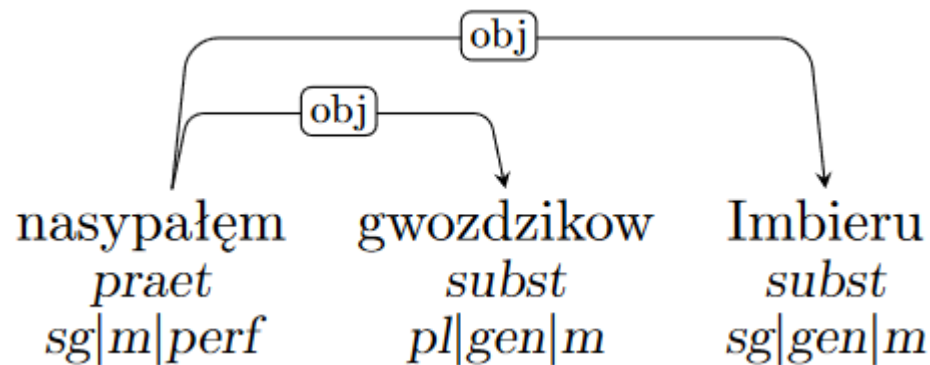
- obecność znaku interpunkcyjnego tam, gdzie dziś go nie postawimy
 - ...co pokázuie iásnie **nikczemność/ y ospáłość** więtszey części *Cesárzow Greckich y woysk ich.*
 - *Kray zasłoneczny **południowy, iest to tá** część ziemi...*
 - *Ta sfera ziemi ták wyrobiona do czegoby **służyłá? następuiące** opiszą reguły.*
 - ***Handel!** roźszerzył granice krolestwa.*
- dwukropek zamiast myślnika lub średnika
 - *Są w nim cztery chóry: owo opisać trudno takiej nowej inwentiej.*
- średnik zamiast przecinka
 - *Bywaią y melnicy w tym konszcie doskonali; ktorzy od maystrow lepszych niech przeymuią konszt ciesielski.*
- zapis wyrażen przyimkowych z dywizem
 - *w-koronie*
- brak oczekiwanego znaku interpunkcyjnego
 - ...nasypałęm gwoździkow *Imbieru...*
 - ...od synow mężow braci...

Koordinacja w schemacie zależnościowym



- conjunct = jeden z elementów koordynacji

Koordinacja bez znaku interpunkcyjnego



- obj = dopełnienie bliższe

Trudności z podziałem na wypowiedzenia

- *Gdzie Bog stworzył ten świat? Odpowiadam. Ze nigdzie.*
- *Tám łáskáwego Czytelniká odsyłam. bo mi sie o tym powtarzać nie zdáło.*
- *Młoda rádá/ Márcowa pogodá/ Pánieńska łáská. Iednáko státeczne bywáią.*
- *Co władzą powierzoną ten akt rugowego prawa od I. O. Xiążęcia etc. ręką własną podpisuię Działo się w Ptaszkowey d. 23 Marca 1736.*

Szyk przestawny

- *dziwna trafiła się awantura*
- *Opaci dobrą uprowadzili fundacyą*
- *na boku gdzie swego życzyłbym połączyć zdrowia*
- *u mego widział mnie stryia*
- *gdzie ieszcze pewną wielkiego y dawnego/ rozeznasz ruinę
budynku*
- *wiele królestw Pan Bóg karać raczył*

Brak orzeczenia

- *W tejże izbie, gdzie Kuchenka Jej Mści Dobrodziejki, stół jeden, ława jedna, szafka do schowania jedna.*
- *SYriyczyk/ álbo Pietruszká syriyska/ źiele nam nieznáíome/ bo tylko w Syryey roście.*
- *Miasto dość wielkie w swej osiadłości i obronę ma swą dobrą.*

Brak orzeczenia

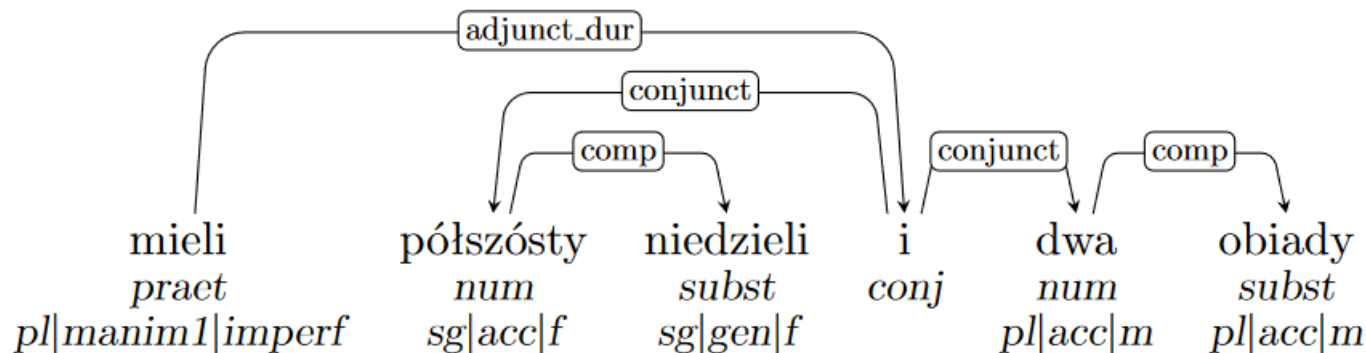
- W tejże izbie, gdzie *Kuchenka* Jej Mści Dobrodziejki, *stół* jeden, *ława* jedna, *szafka* do schowania jedna.
- *SYriyczyk/ álbo Pietruszká* *syriyska/ źiele nam nieznáíome/ bo tylko w Syryey roście.*
- *Miasto dość wielkie* w swej osiadłości i obronę ma swą dobrą.

Zdania trudne do zrozumienia

- *Mieli go (słyszę) półszosty niedzieli z dziećmi i dwa obiady.*
- *Ma chałupę nową, stada pilnuje i ordynaryją bierze, myto i przysiewek.*

Zdania trudne do zrozumienia

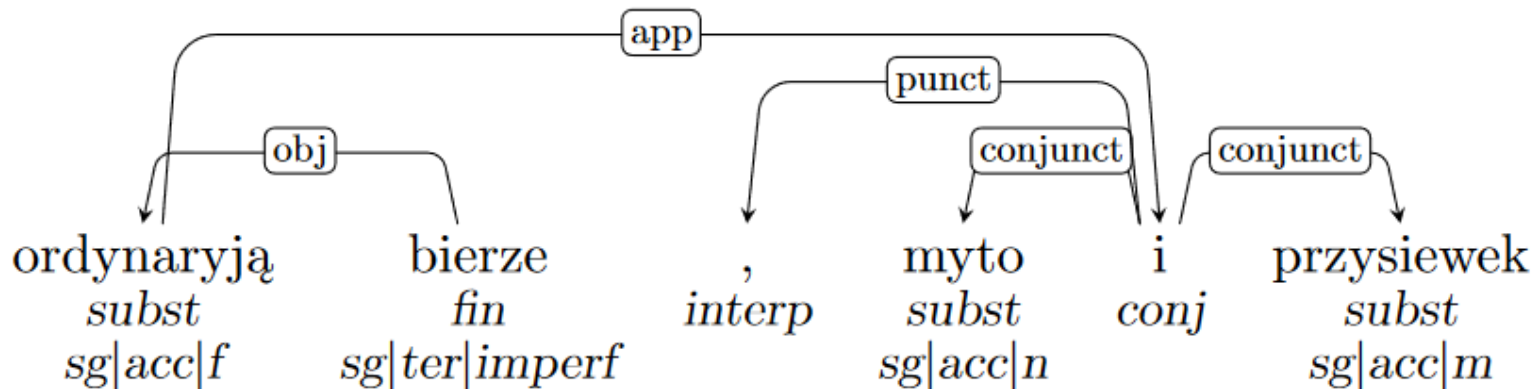
- Mieli go (słyszę) [przez] półszosty niedzieli z dziećmi i [przez] dwa obiady.



- adjunct_dur = okolicznik czasu trwania
- comp = podrzędnik wymagany

Zdania trudne do zrozumienia

- Ma chałupę nową, stada pilnuje i ordynaryją bierze, myto i przysiewek.*



- app = przydawka apozycyjna

Dalsze cele i plany na przyszłość

- rozbudowa Banku Drzew XVII-XVIII
- uzupełnienie pełnej KorBy o warstwę składniową
- analiza wybranych zjawisk składniowych

Rozbudowa Banku Drzew XVII-XVIII

- więcej wypowiedzeń tego samego typu
lub
- wypowiedzenia zawierające wtręty łacińskie
lub
- wypowiedzenia zawierające określone konstrukcje składniowe

Rozbudowa pełnego korpusu o warstwę składniową – przykład: nowa edycja NKJP

NKJP **DRZEWA ZALEŻNOŚCIOWE** **OPIS JĘZYKA ZAPYTAŃ**

Klasyczne **Kompaktowe** Pudełkowe

Takie ADJ rozmowienia SUBST bywają FIN zaraźliwe ADJ . INTERP

KONSTRUKTOR ZAPYTAŃ METADANE

Liczba wyników na stronę
10

Wyszukaj

Znaleziono 1 wyników.

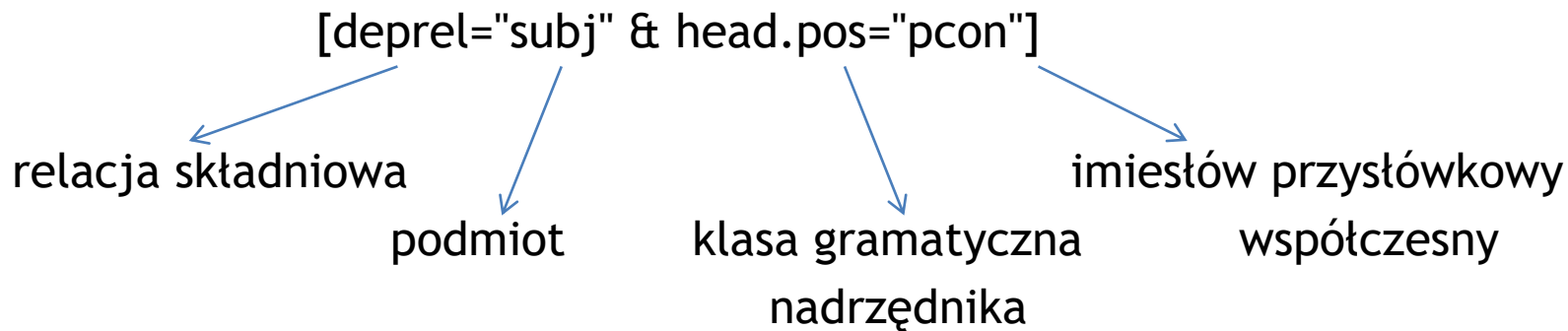
Lp	Lewy kontekst	Rezultat	Prawy kontekst	ID
1	je oddać w kolportaż, więc do roboty".	Takie [taki:adj:pl:nom:n:pos] rozmowienia [rozmowienie:subst:pl:nom:n:ncol] bywają [bywać:fin:pl:ter:imperf]	zaraźliwe. Gdy przed laty przygotowywałem do druku pierwszą	PWN_1202900001120_text-27

nkjp.nlp.ipipan.waw.pl

Możliwości badań składniowych w KorBie

- wypowiedzenia z imiesłowami przysłówkowymi i podmiotem
- szyk przestawny przydawki
- rekcja czasowników

Imiesłów przysłówkowy współczesny z podmiotem?



→ ...czytając strona po stronie. (NKJP)

→ ...mając wewnątrz Klasztor w sobie/ zbyt wesołe cele... (KorBa)

Dziękuję za uwagę

aleksandra.wieczorek@ijp.pan.pl

www.korba.edu.pl