

Przetwarzanie języka naturalnego

dr inż. Marcin Ciura

Wydział Informatyki i Telekomunikacji Politechniki Krakowskiej

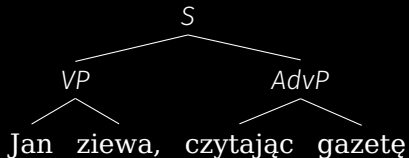
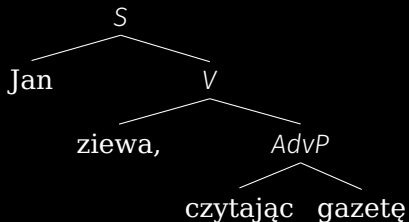
Plan na dziś: 12 slajdów, w tym półgodzinny film

- Analiza zależnościowa a analiza frazowa
- Universal Dependencies
- Współczesne podejście do analizy zależnościowej

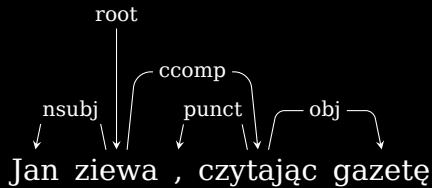
Analiza zależnościowa a analiza frazowa

Analiza frazowa

Fałszywe niejednoznaczności, bo hierarchia ma za dużo poziomów



Analiza zależnościowa (ang. *dependency analysis*)



Cechy analizy zależnościowej

- graf skierowany z jednym wierzchołkiem początkowym (*korzeniem*) na zdanie
- nie ma węzłów pośrednich; połączenia (*zależności*) występują tylko między segmentami tekstu
- role składniowe są etykietami krawędzi, a nie węzłów grafu

Universal Dependencies

Banki drzew (*treebanks*; 1)

Wzorcowe wyniki analizy wielu zdań danego języka,
zwanych korpusem (ang. *corpus*; l. mn. *corpora*)

Banki drzew (*treebanks*; 2)

Polski bank drzew z analizą frazową (wyszukiwarka):

- Składnica

Polskie banki drzew z analizą zależnościową (wyszukiwarka):

- UD Polish PUD
- UD Polish LFG
- UD Polish PDB

Universal Dependencies (1)

<https://universaldependencies.org/>

- spójne zasady opisu (ang. *annotation*) morfologicznego i składniowego
- w październiku 2020 r. obejmuje 90 języków i ponad 150 banków drzew

Universal Dependencies (2)

Każdemu wypowiedzeniu (zdaniu lub równoważnikowi zdania), nawet „niegramatycznemu”, odpowiada jeden wynik analizy: spójne drzewo, czyli taki graf, w którym każdy wierzchołek ma dokładnie jeden wierzchołek nadrzędny

Współczesne podejście do analizy zależnościowej

- analiza frazowa — reguły
- analiza zależnościowa — uczenie maszynowe

choć nic nie wymusza takiego doboru podejść

Uczenie maszynowe w analizie zależnościowej

- zbiór uczący — bank drzew
- algorytmy jednoprzebiegowe
- działanie „gorliwe/skwapliwe” (ang. *eager*):
dodawanie krawędzi do grafu w miarę
wczytywania kolejnych segmentów (wyrazów
lub znaków przestankowych)

Wykład o współczesnym podejściu (po angielsku)

Christopher D. Manning:
Greedy transition-based parsing

- Analiza zależnościowa a analiza frazowa
- Universal Dependencies
- Współczesne podejście do analizy zależnościowej

**Do zobaczenia
na następnym wykładzie**

o ???
