

# Związki między wyrażeniami wydźwięku a atrybutami opinii

Aleksander Wawer

2 grudnia 2013

`axw@ipipan.waw.pl`

# Plan referatu

- 1 Zasób S: Słownik wydźwięku
- 2 Zasób A: Atrybuty opinii
- 3 Kolekcja tekstów
- 4 Związki składniowe pomiędzy S i A: metoda zależnościowa
- 5 Konkluzje

# Podstawowe zasoby i problem

## Leksykon wydźwięku (sentymetu), dalej S

Zawiera on nacechowane słowa, przenoszące opinie. Na przykład, *dobry*, *wspaniały* i *niesamowity* to słowa pozytywne, a *zły* i *straszny* wyrażają negatywne nastawienie.

## Atrybuty opinii, dalej A

Wszystko, co jest poddawane ocenie, obiekty i atrybuty obiektów. Przykładowo, *aparat fotograficzny ma świetną optykę, ale słabą baterię i często się zawiesza.*

Celem jest zbadanie związków między A i S na poziomie zdaniowym.

## Leksykon Wydźwięku (S)

- Pierwszy fragment leksykonu to 1200 przymiotników i przysłówków dzięki ręcznie stworzonym wzorcom leksykalnym i *web as corpus*.
- Lista ta została zweryfikowana przez dwójkę lingwistów.
- Natężenie wydźwięku na skalach 2-, 3-, 5- oraz 9-punktowych.
- Ręcznie przyznane oceny posłużyły do rozwoju algorytmów:
  - nienadzorowanych: orientacji semantycznej punktowej informacji wzajemnej (SO-PMI),
  - z nadzorem: modele klasyfikacyjne trenowane na kontekstach wystąpień w korpusie NKJP.
- Słownik poszerzony został automatycznie o kolejne 2600 słów, w większości z wysokimi bezwzględными wartościami SO-PMI.
- Słownik dostępny jest na stronie <http://zil.ipipan.waw.pl/SloownikWydzwieku>.

## Lista słów – atrybutów opinii (A)

- Punktem wyjścia jest zbiór recenzji dwóch typów produktów (uroda i odzież).
- Niewielki fragment tego zbioru uzyskał ręczne oznaczenia wystąpień atrybutów opinii.
- Posłużył on do pół-automatycznego wygenerowania dwóch list słów, potencjalnie oznaczających atrybuty opinii produktów każdego z typów.
- Obie listy słów zostały ręcznie zweryfikowane pod kątem ich poprawności.

# Informacje frekwencyjne

	uroda	odzież
słowa A	311	222
recenzje	5499	7197
zdania	30274	39691
zdania z A	18493	23628
zdania z S	18299	24113
zdania z A i S	14495	18519
średnio zdań w recenzji	5.5	5.5

# Wybrane informacje frekwencyjne: uroda

	Słowa A			
Słowa S	1	2	3	4
1	3443	2152	821	249
2	1768	1558	747	298
3	566	635	450	206
4	142	192	179	95

# Wybrane informacje frekwencyjne: odzież

	Słowa A			
Słowa S	1	2	3	4
1	1165	601	206	74
2	586	370	170	51
3	181	143	81	38
4	50	48	42	28



## Przetwarzanie zależnościowe: pytania

- W jakich sytuacjach słowa z wydźwiękiem odnoszą się do atrybutów opinii?
- Czy są to składniowe, czy semantyczne uwarunkowania?
- Odległość: czy ma wpływ i jaki na związek A-S?
- Jakie relacje zmieniają polaryzację wydźwięku, powodują przerwanie (brak) relacji A-S?
- Jaki wpływ na związek A-S ma koordynacja?
- Czy modyfikacja jest w każdym przypadku relacją odpowiadającą związkowi A-S?

# Przetwarzanie zależnościowe: odległość A-S w drzewie

Dla zdań z pojedynczym wystąpieniem A i pojedynczym S.

	uroda	odzież
0	938	213
1	1242	391
2	528	175
3	222	132
4	101	82
5	50	33
6	20	7
7	14	6
8	1	3
9	1	2

# Przetwarzanie zależnościowe: anotacja zdań

Dla podanego drzewa zależnościowego z zaznaczonym A i S:

- ❶ Błędna struktura składniowa pomiędzy A i S.
- ❷ Wystąpienie A nie jest poprawne w danym kontekście.
- ❸ Wystąpienie S nie jest poprawne w danym kontekście.
- ❹ W przypadku braku zajścia warunków 1-3:
  - S jest związane z A (S opisuje, odnosi się do, modyfikuje A),
  - S nie jest związane z A (jw.).

# MaltEval jako narzędzie anotacji

MaltEval Tree Viewer 1.0

File Settings Navigate Help

Search in: [-:Choose data:-] Search by: [-:Choose search strategy:-] Search for: [ ] [ ] Negate Search Result: [ ] Search direction: << >>

Parsed 6: input1789\_2.conll

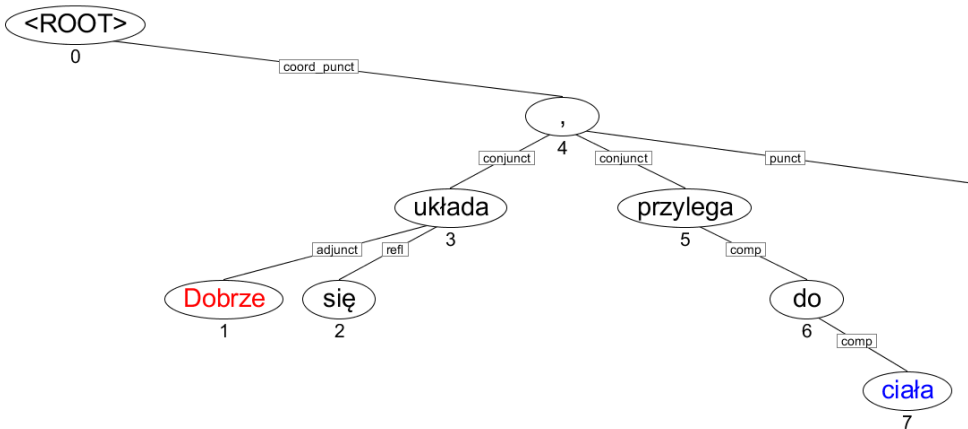
Legend:

- structure error between atr. and sent.
- attribute word incorrect
- sentiment word incorrect
- attribute related to sentiment

Prev sent.

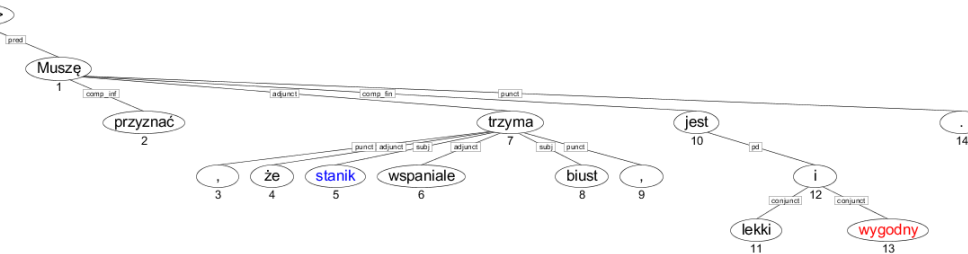
Next sent.

# Brak związku A-S

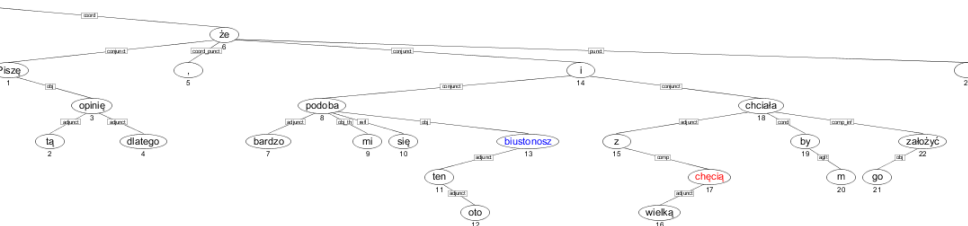




# Jest związek A-S



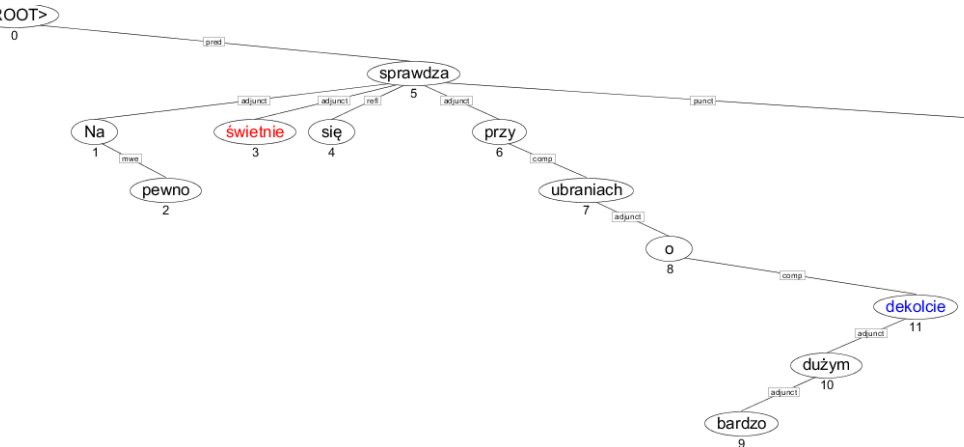
# Jest związek A-S: koreferencja



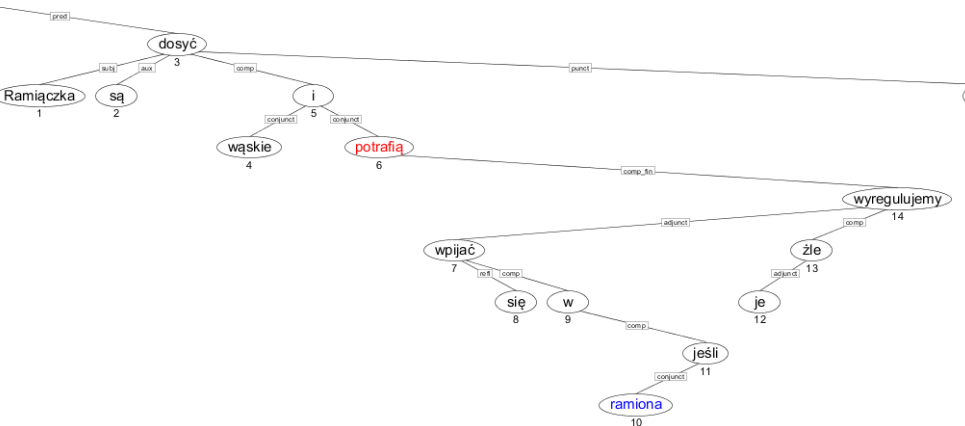




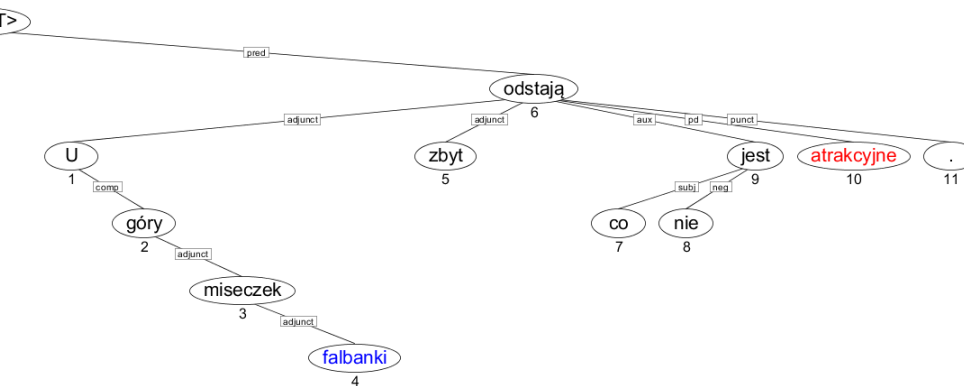
# Brak związku A-S: zdanie trudne



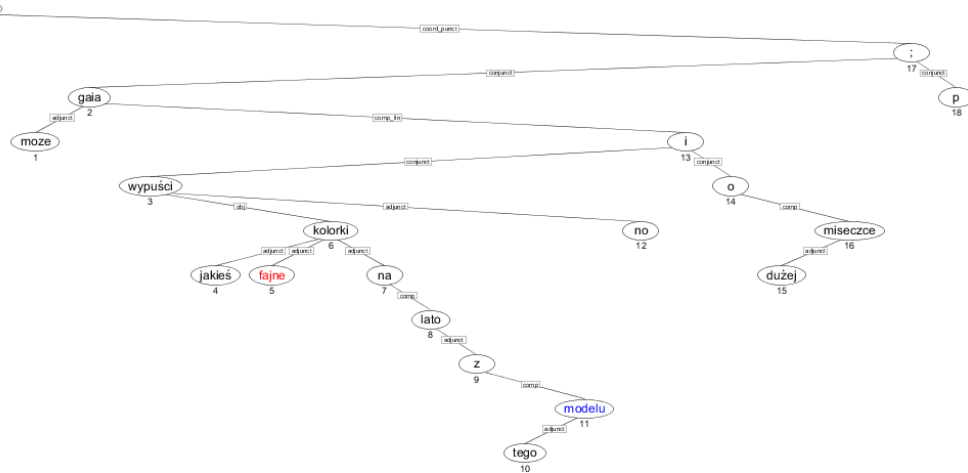
# Błędy parsowania



# Błędy parsowania

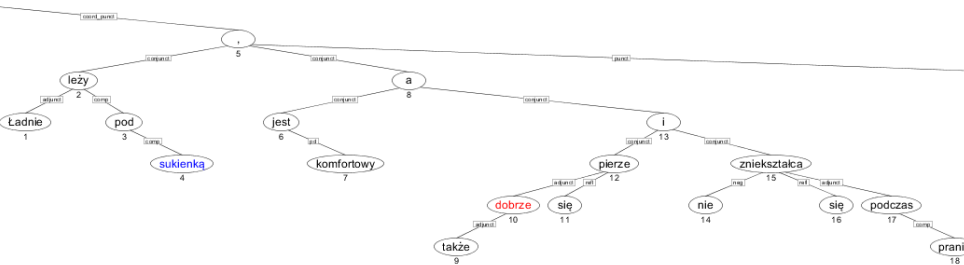


# Błędy parsowania

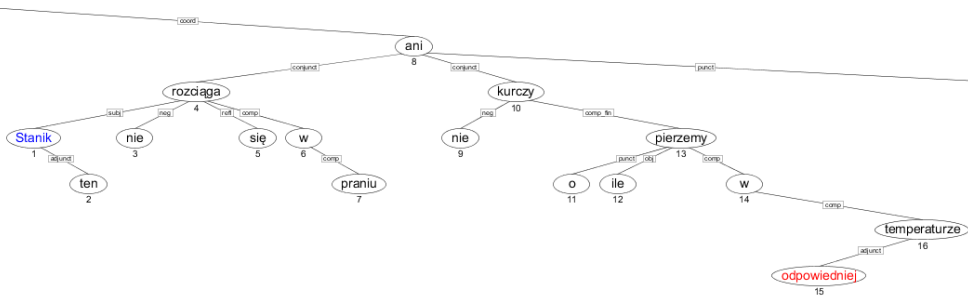




# Odległość A-S: 6



# Odległość A-S: 6





## Anotowane zdania: dane frekwencyjne

Podzbiór zdań został oznaczony podwójnie.

	N	N zgodnych	% zgodność
poprawność struktury	82	75	91%
poprawność A	75	64	85%
poprawność S	75	70	93%
A związany z S	54	42	77%

Zadanie jest relatywnie dobrze określone i niezbyt trudne.

# Automatyczne rozpoznawanie relacji A-S: przestrzeń cech

Dla każdego węzła na ścieżce między A i S generujemy wektor cech zapamiętując:

- Lemat (LEM)
- Klasę gramatyczną i część mowy (POS)
- Typy relacji zależnościowych (REL)

Skonstruowana przestrzeń cech nie uwzględnia informacji o pozycji (odpowiednik BOW).

# Automatyczne rozpoznawanie relacji A-S: wyniki

Ze względu na niskie liczebności zastosowana metoda to LOO (Leave One Out).

	Dokładność	Precyzja	Pełność
Regresja Logistyczna	0.886	0.8	0.061
liniowy SVM	0.876	0.412	0.106

Dokładność =  $TP+FN / TP+FP+TN+FN$

Precyzja =  $TP / TP+FP$

Pełność =  $TP / TP+FN$

# Podsumowanie

- Powstał zbiór zdań, opisujący potencjalne związki między atrybutami opinii (A) a wydźwiękiem (S).
- Podzbiór 570 zdań, o odległości między A i S większej lub równej 3 węzły, został oznaczony ręcznie.
  - 66 z tych zdań zawiera w sobie związek między A i S
- Wniosek: istnieją zdania, posiadające automatycznie wygenerowany, poprawny opis zależnościowy, w których związek A-S wykracza poza zakres frazy.
- Wykonane zostały wstępne prace nad automatycznym rozpoznawaniem związków tego typu.

## Planowane prace

- Poszerzenie zbioru zdań o składnicę zależnościową.
  - Ze względu na brak oznaczeń A, potencjalne słowa tego typu można wskazać automatycznie na podstawie informacji składniowych oraz wystąpień słów S (CRF).
  - Zdania takie, zawierające słowa S i potencjalne A, poddane zostaną ręcznej weryfikacji.
- Implementacja i przetestowanie algorytmów uczenia (np. metody jądrowe, korzystające z funkcji splotu / convolution kernels)
- Problem przypisania słowa A do poprawnego słowa S, w przypadku kilku słów S w zdaniu.
- Problem przypisania słowa S do poprawnego słowa A, w przypadku kilku słów A w zdaniu.