

Parsowanie semantyczne wypowiedzi w języku polskim z użyciem parsera ENIAM

Wojciech Jaworski

Instytut Podstaw Informatyki
Polskiej Akademii Nauk

3 lutego 2017

Co to jest parsowanie semantyczne?

- Zadanie znajdowania form logicznych.
- Wyrażanie treści zdania, akapitu, czy tekstu w sposób jawny, uporządkowany i zrozumiały dla komputera.
- Wymaga przyjęcia języka opisu zwanego językiem reprezentacji znaczenia
- oraz teorii (sposobu, koncepcji) opisu świata.
- Jako język reprezentacji znaczenia wykorzystuje się zwykle logikę
 - ▶ formalny język opisu
 - ▶ reguły wnioskowania
 - ▶ formalna semantyka oparta na teorii modeli
- Teoria opisu świata składa się z
 - ▶ ontologii definiującej typy bytów (pojęcia)
 - ▶ teorii poszczególnych dziedzin (teoria czasu, przestrzeni, zdarzeń itp.)

Demo oficjalne:

<http://eniam.nlp.ipipan.waw.pl>

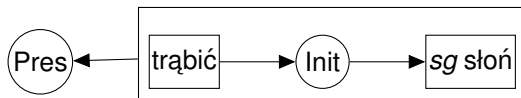
Demo nieoficjalne:

<http://students.mimuw.edu.pl/~wjaworski/eniam.html>

Język reprezentacji znaczenia

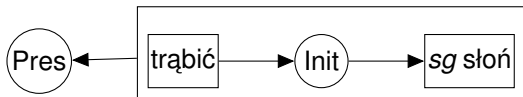
- Jako język reprezentacji znaczenia przyjmujemy rozszerzoną **logikę pierwszego rzędu**
- Formuły naszego języka reprezentacji znaczenia będziemy wyrażać graficznie w formie grafów semantycznych.

Słoń trąbi.



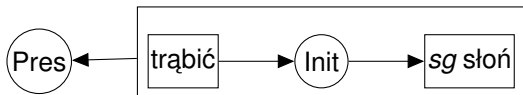
- W powyższym grafie pudełka reprezentują obiekty, o których jest mowa.
 - ▶ występuje zatem obiekt *słoń* i zdarzenie *trąbić*,
 - ▶ symbol *sg* określa licznosc obiektów jako dokładnie 1.
 - ▶ występujące w pudełkach napisy to typy bytów, są nimi pochodzące ze Słownosci sensy.

Słoń trąbi.



- Kółeczka reprezentują relacje między obiektami.
 - ▶ Init wskazuje na to, że *słoń* jest inicjatorem *trąbienia*,
 - ▶ a Pres na to, że zdarzenie jest równoczesne z czasem jego wypowiedzenia.
 - ▶ Strzałka wchodząca to pierwszy argument, wychodząca drugi.
 - ▶ Źródłem informacji o relacjach łączących czasowniki z ich argumentami jest słownik walencyjny Walenty.
- Zewnętrzna ramka to kontekst.
 - ▶ Reprezentuje on sytuację, czyli podzbiór czasoprzestrzeni,
 - ▶ w którym istnieją byty wskazane przez pojęcia w pudełkach
 - ▶ i zachodzą wymienione w kółeczkach relacje pomiędzy nimi.

Słóń trąbi.



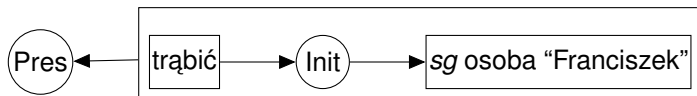
- Powyższy graf jest równoważny formule logicznej

$$\exists(x, \text{DSCR}(x, \exists(s, \text{TYPE}(s, \text{słóń}) \wedge |s| = 1, \\ \exists(t, \text{TYPE}(t, \text{trąbić}) \wedge \text{INIT}(t, s))))), \text{PRES}(x)),$$

w której każdy z obiektów jest identyfikowany przez zmienną.

- Predykat $\text{TYPE}(x, t)$ przypisuje typ t do zmiennej x ,
 - ▶ stwierdza, że zbiór bytów wskazywanych przez x należy do sensu t .
- Ze zmiennymi zawsze wiążemy zbiory bytów.
 - ▶ Pozwala to w prosty sposób zdefiniować licznosc.
- Predykat metajęzykowy DSCR wyraża kontekst.

Franciszek trąbi.



- Formuła odpowiadająca temu zdaniu ma postać

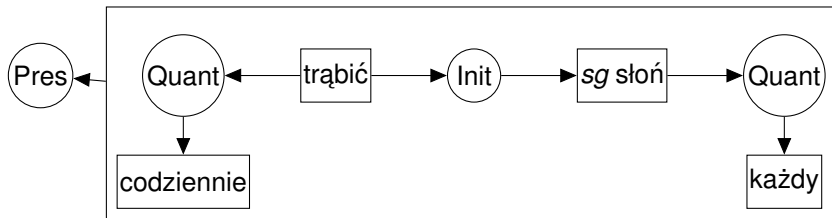
$$\exists(x, \text{DSCR}(x, \exists(o, \text{TYPE}(o, \text{osoba}) \wedge \text{HASNAME}(o, \text{'Franciszek'}) \wedge |o| = 1$$
$$\exists(t, \text{TYPE}(t, \text{trąbić}) \wedge \text{INIT}(t, o))), \text{PRES}(x)).$$

- Nazwy własne reprezentujemy za pomocą predykatu $\text{HASNAME}(x, \text{'name'})$,
 - ▶ który wiąże nazwę 'name' ze zbiorem bytów oznaczonym przez x .
- W grafach semantycznych nazwy własne umieszczamy w cudzysłowach.
- Nazwy własne nie definiują typu obiektu,
 - ▶ tylko identyfikują obiekt poprzez podanie przypisanej mu etykiety.

Kwantyfikatory

- Składnię rachunku predykatów pierwszego rzędu rozszerzamy o **kwantyfikatory** (np. *co dziesiąty*, czy *prawie każdy*).

Każdy słoń trąbi codziennie.

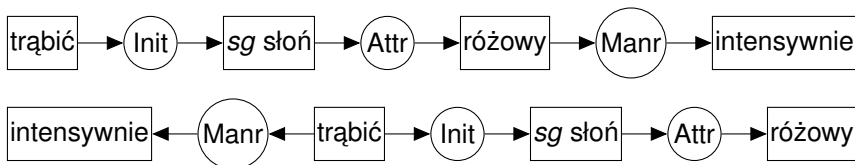


$$\exists(x, \text{DSCR}(x, \text{KAŻDY}(s, \text{TYPE}(s, \text{słoń}) \wedge |s| = 1, \text{CODZIENNIE}(t, \text{TYPE}(t, \text{trąbić}) \wedge \text{INIT}(t, s))), \text{PRES}(x)),$$

- Reprezentacja grafowa pozwala nie dospecyfikowywać zakresu kwantyfikatorów.
- Kwantyfikatory w formułach są wprowadzane zgodnie z kolejnością słów w zdaniu.

- Zazwyczaj wyrażane przez przymiotniki i przysłówki.
- Przyjmujemy, że cechy (atrybuty) to byty mające swoje instancje.
- Słowa odnoszą do instancji cech, a nie samych cech.
- Przykładowo instancjami koloru *różowy* mogą być różne odcienie tego koloru.

Intensywnie różowy słoń trąbi.

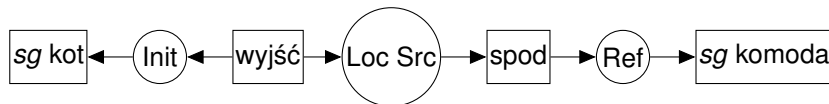


- Nazwę cechy łączymy z nazwą określonego przedmiotu za pomocą relacji Attr.
- Indywidua mogą mieć cechy i same cechy mogą mieć cechy.

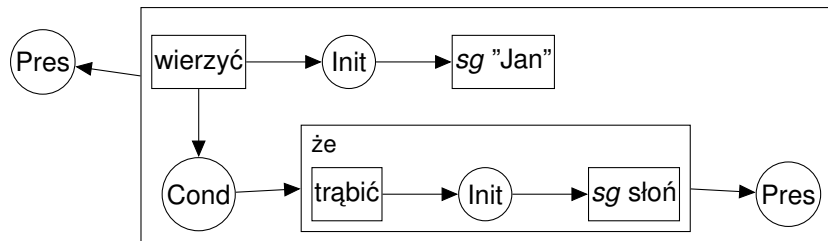
Określenia miejsca

- Zazwyczaj wyrażane przez wyrażenia przyimkowe i przysłówki.
- Przyimki lokatywne reprezentują relacje pomiędzy miejscami.
- Relacje te są zreifikowane.
- Relacje te mogą być modyfikowane, np. *dość głęboko w szafie*.
- Relacje Loc, Loc Src, Loc Goal, Path informują o obecności i kierunku ruchu.
- Relacja Ref łączy przyimek z jego podrzędnikiem.

Kot wyszedł spod komody.



Jan wierzy, że słoń trąbi.

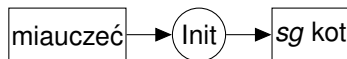


- Zdanie, które jest przedmiotem przekonań nie musi być obiektywnie prawdziwe.
- Umieszczamy je w pudełku oznaczającym, że jego prawdziwość należy określać ze względu na subiektywny model świata posiadany przez *Jana*.
- Aby reprezentować modele w języku przedmiotowym, rozszerzamy notację logiczną o metajęzykowy predykat DSCR.

Kot aranżuje.

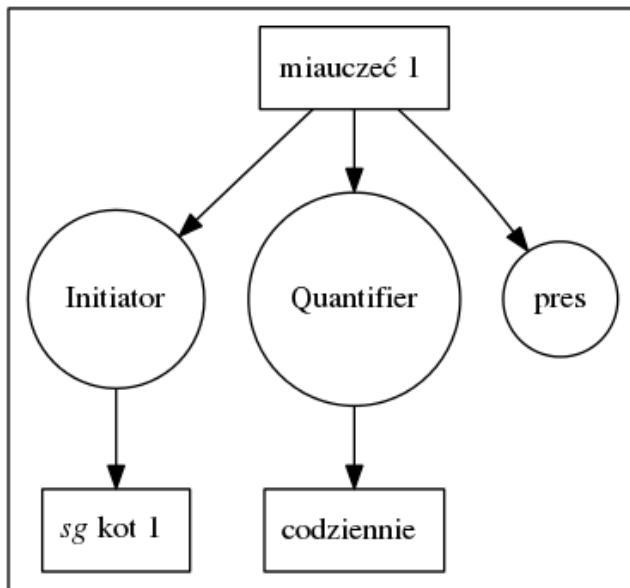


Kot miauczy.

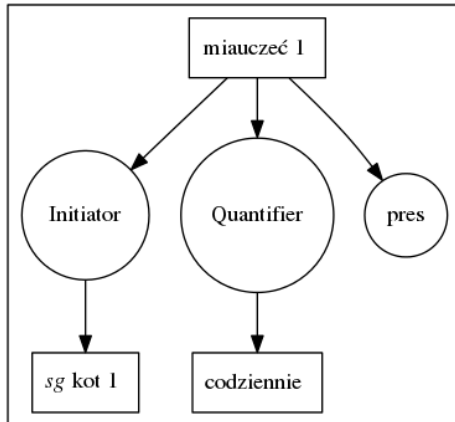


- Zasoby semantyczne pozwalają wykonać częściową dezambiguację na podstawie cech semantycznych
 - ▶ *Walenty* wyznacza preferencje selekcyjne podrzędników czasownika.
 - ▶ *Słowość* określa sensy słów uzgadniane z preferencjami selekcyjnymi.
 - ▶ *SGJP* wskazuje, że *Kot* jest nazwą własną i wskazuje jej typ.

Kot miauczy codziennie.



Kot miauczy codziennie.



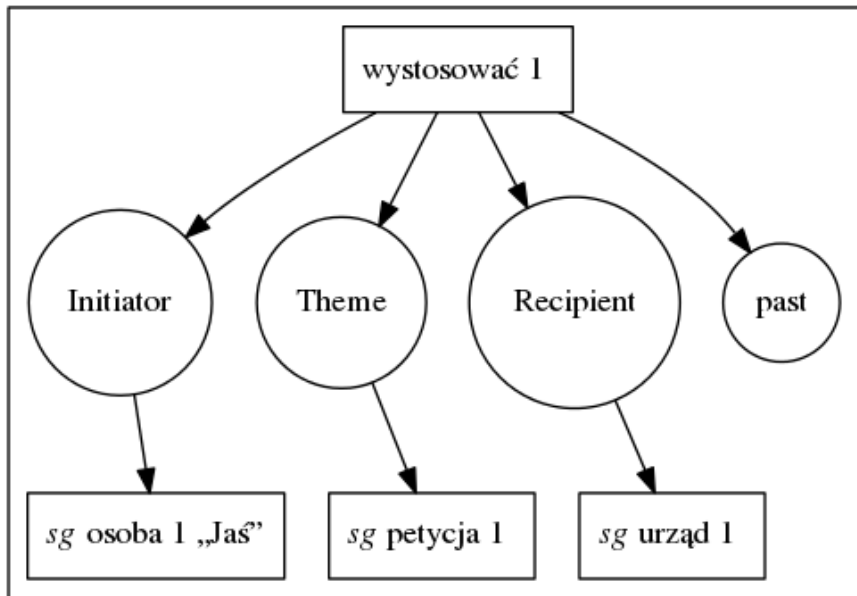
$\exists[x, \text{DSCR}([, x),$

$\exists[k, \text{TYPE}(k, \text{kot } 1) \wedge |k| = 1,$

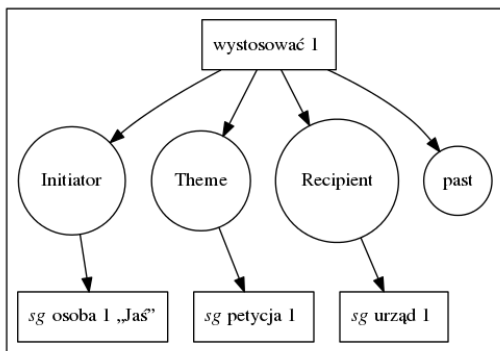
$\text{CODZIENNIE}[m, \text{TYPE}(m, \text{miauczeć } 1) \wedge \text{PRES}(m) \wedge$

$\text{INIT}(m, k)]]]]$

Jaś wystosował petycję do urzędu.



Jaś wystosował petycję do urzędu.



$\exists[x, \text{DSCR}([, x),$

$\exists[o, \text{TYPE}(o, \text{osoba } 1) \wedge \text{HASNAME}(o, \text{'Jaś'}) \wedge |o| = 1,$

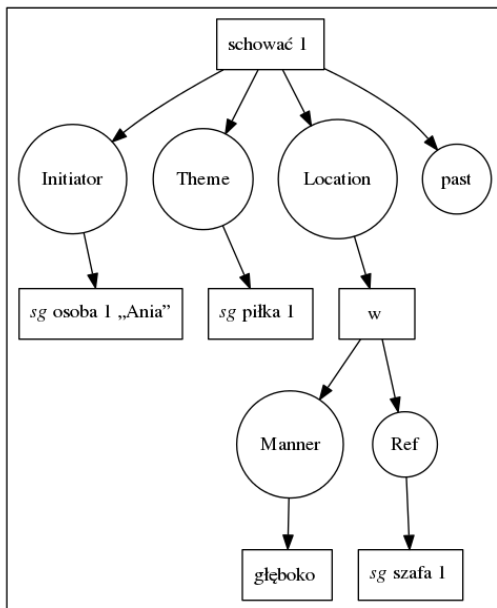
$\exists[w, \text{TYPE}(w, \text{wystosować } 1) \wedge \text{PAST}(w) \wedge$

$\exists[u, \text{TYPE}(u, \text{urząd } 1) \wedge |u| = 1, \text{RECIPIENT}(w, u)] \wedge$

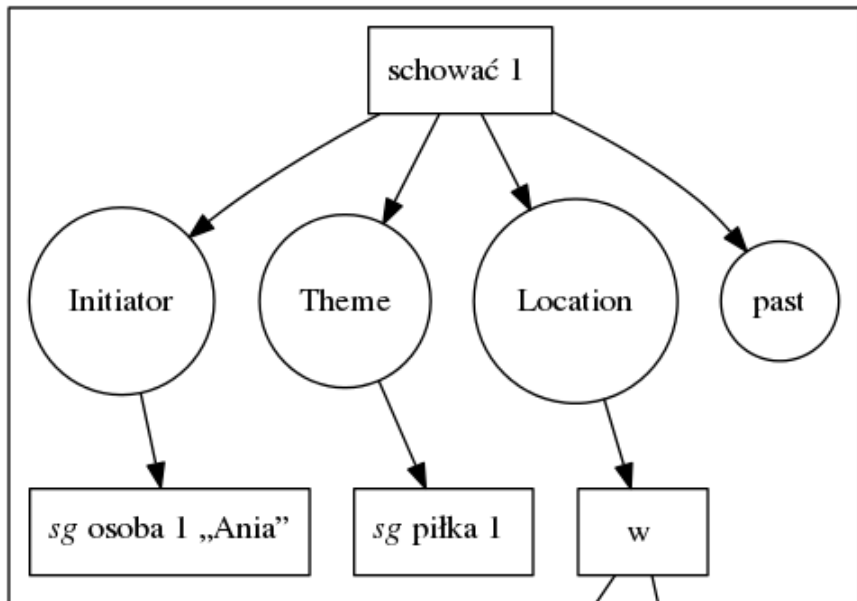
$\exists[p, \text{TYPE}(p, \text{petycja } 1) \wedge |p| = 1, \text{THEME}(w, p)] \wedge$

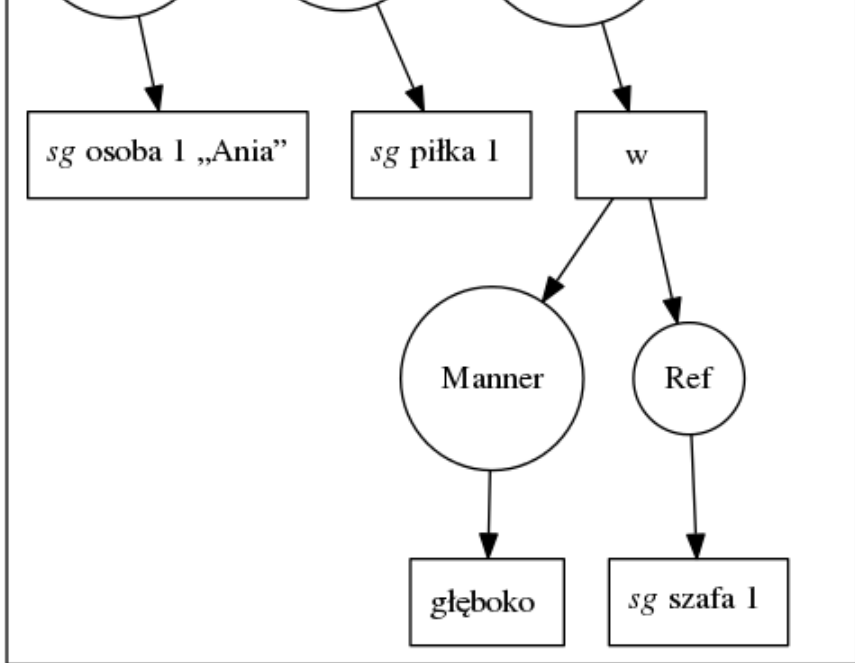
$\text{INIT}(w, o)]\text{]]]]]$

Ania schowała piłkę głęboko w szafie.

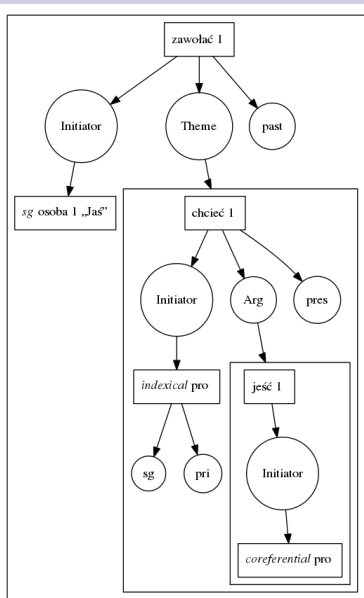


Ania schowała piłkę głęboko w szafie.

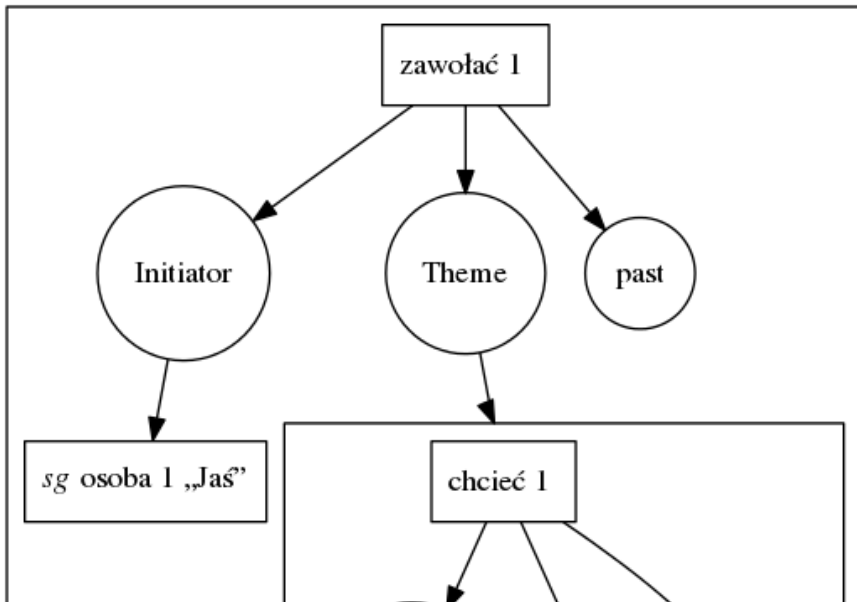




- Chcę jeść - zwołał Jaś.



- Chcę jeść - zawałał Jaś.



Initiator

Theme

past

sg osoba 1 „Jaś”

chcieć 1

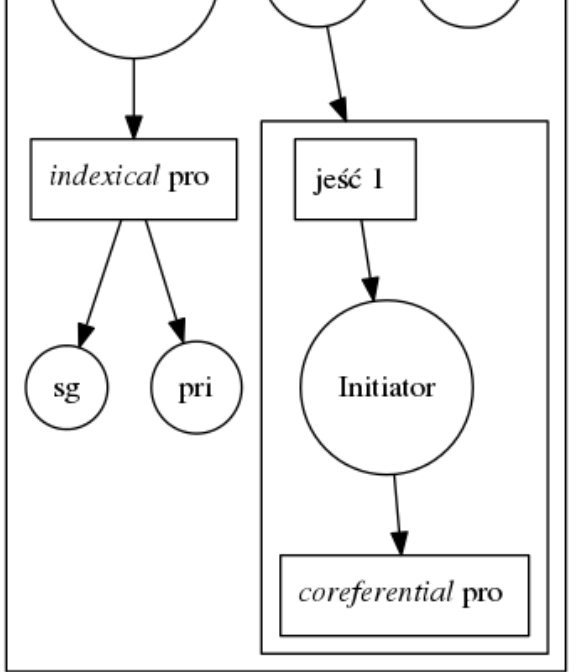
Initiator

Arg

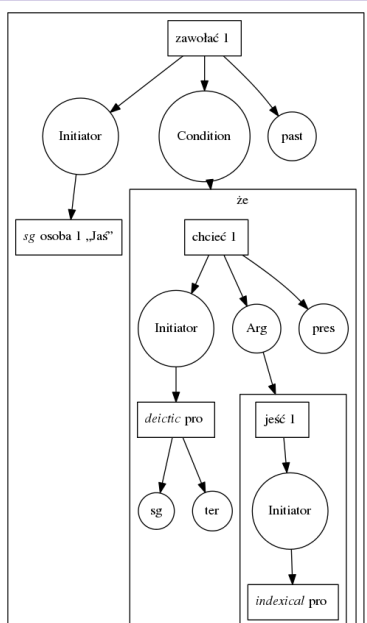
pres

indexical pro

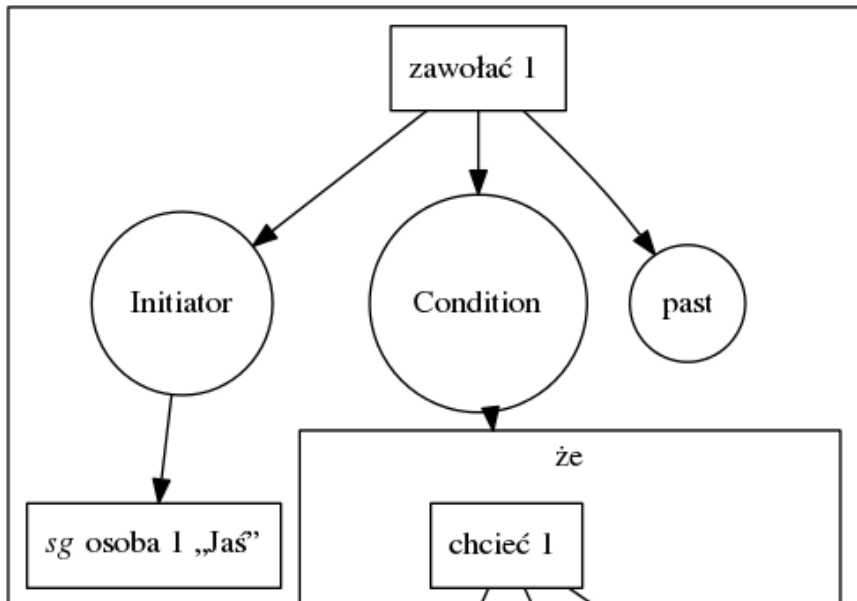
jeść 1



Jaś zwołał, że chce jeść.



Jaś zwołał, że chce jeść.



Initiator

Condition

past

sg osoba 1 „Jaś”

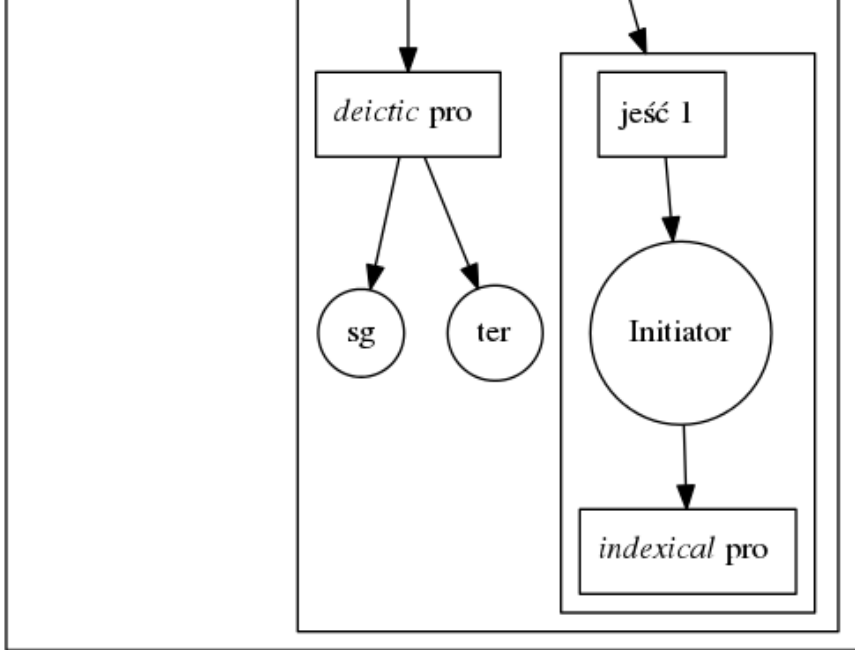
że

chcieć 1

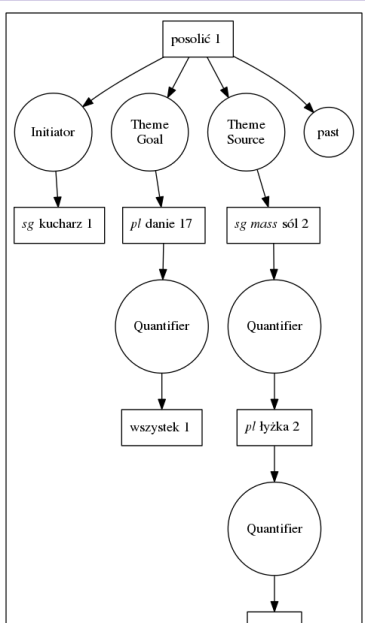
Initiator

Arg

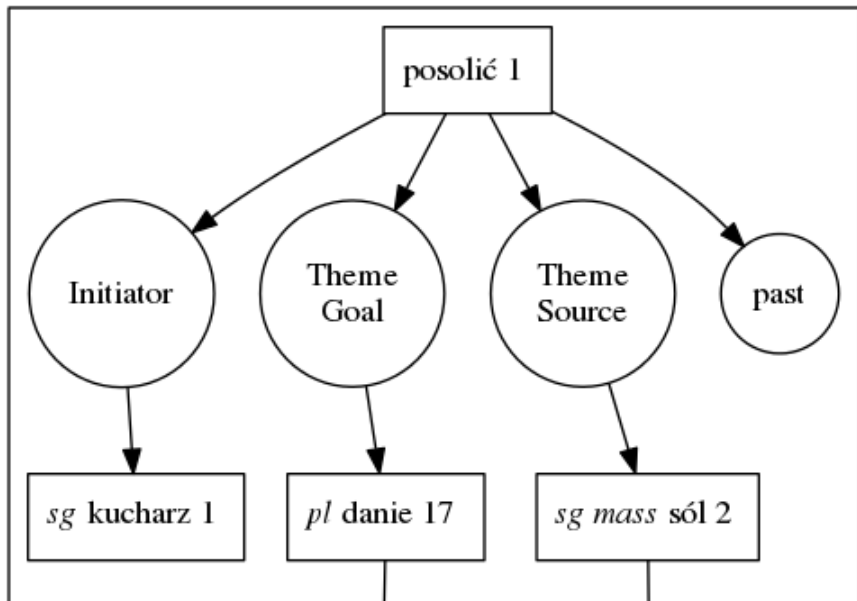
pres



Kucharz posolił wszystkie dania dwoma łyżkami soli.



Kucharz posolił wszystkie dania dwoma łyżkami soli.



sg kucharz 1

pl danie 17

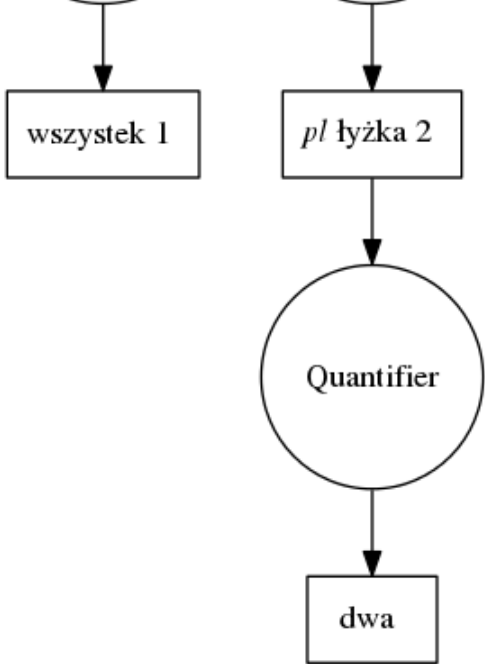
sg mass sól 2

Quantifier

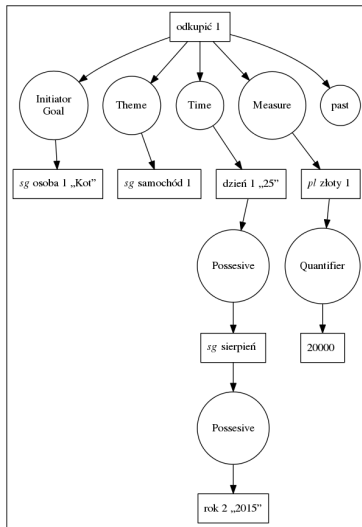
Quantifier

wszystek 1

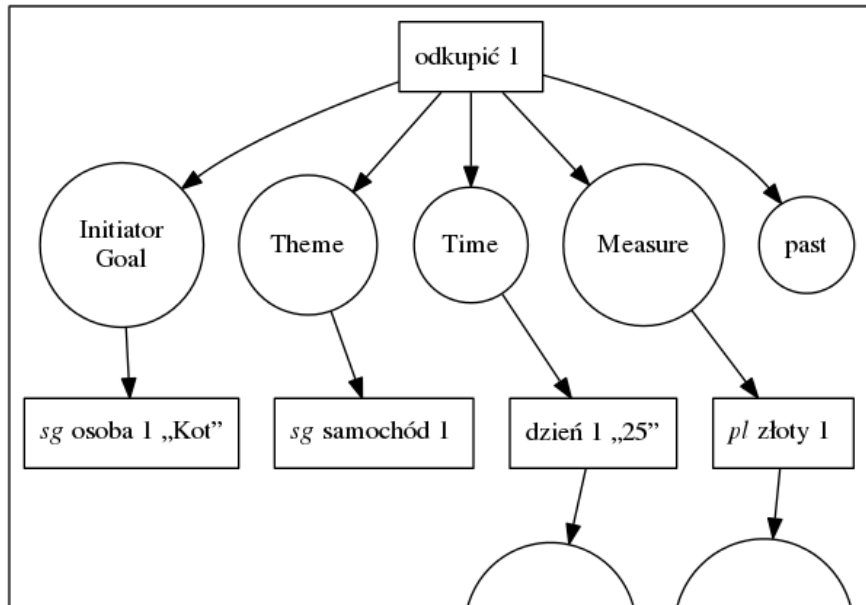
pl łyżka 2



Kot odkupił 25 sierpnia 2015 samochód za 20000zł.



Kot odkupił 25 sierpnia 2015 samochód za 20000zł.

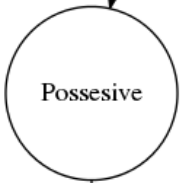


sg osoba 1 „Kot”

sg samochód 1

dzień 1 „25”

pl złoty 1



sg sierpień

20000



Possesive

Quantifier

sg sierpień

20000

Possesive

rok 2 „2015”