

Metody dokonywania rozbioru składniowego zdań w języku naturalnym przypominają te stosowane do języków programowania, chociaż istnieją też nietypowe metody. Najczęściej stosuje się gramatyki bezkontekstowe. Czynnikiem dodatkowym w przetwarzaniu języka naturalnego bywa zastosowanie unifikacji.

$$\begin{bmatrix} \text{cecha}_1 & \text{wartość}_1 \\ \text{cecha}_2 & \text{wartość}_2 \\ \vdots & \vdots \\ \text{cecha}_n & \text{wartość}_n \end{bmatrix}$$

Podstawową strukturą wykorzystywaną w unifikacji jest macierz wartości cech.

Macierze wartości cech pozwalają na znaczące zmniejszenie liczby reguł składni, ponieważ pozwalają na umieszczenie w nich ogólnych związków, a następnie np. na wyrażenie związku zgody lub rządu. Macierze wartości cech mogą być zagnieżdżone, np.:

$$\left[ \begin{array}{cc} \text{kategoria} & \text{NP} \\ \text{zgoda} & \left[ \begin{array}{cc} \text{liczba} & \text{lp} \\ \text{osoba} & 3 \end{array} \right] \end{array} \right]$$

Ścieżka cechy jest listą nazw cech prowadzących do wartości danej cechy, np. wartość *lp* jest osiągalna za pomocą ścieżki *zgoda/liczba*.

# Analiza składniowa z wykorzystaniem unifikacji

Pewne struktury cech mogą być współdzielone z innymi cechami. Należy zauważyć, że nie oznacza to, że dwie cechy mają równe sobie wartości, ale że jest to ta sama, współdzielona wartość:

$$\left[ \begin{array}{cc} \text{kategoria} & \text{zdanie} \\ \text{element\_główny} & \left[ \begin{array}{cc} \text{zgoda} & \boxed{1} \left[ \begin{array}{cc} \text{liczba} & \text{lp} \\ \text{osoba} & 3 \end{array} \right] \\ \text{podmiot} & \left[ \begin{array}{cc} \text{zgoda} & \boxed{1} \end{array} \right] \end{array} \right] \end{array} \right]$$

Cecha *element\_główny/zgoda* jest jednocześnie cechą *element\_główny/podmiot/zgoda*.

# Analiza składniowa z wykorzystaniem unifikacji

Unifikacja dwóch takich samych cech daje tę samą cechę:

$$[ \text{liczba } l_p ] \sqcup [ \text{liczba } l_p ] = [ \text{liczba } l_p ]$$

natomiast unifikacja tej samej cechy o niezgodnych wartościach zawodzi:

$$[ \text{liczba } l_p ] \sqcup [ \text{liczba } l_m ] \text{ Błąd!!!}$$

Brak wartości danej cechy podczas unifikacji oznacza to samo, co zbiór wszystkich możliwych jej wartości:

$$[ \text{liczba } l_p ] \sqcup [ \text{liczba } \square ] = [ \text{liczba } l_p ]$$

Brak wartości danej cechy najczęściej występuje jako nieobecność cechy:

$$[ \textit{liczba} \quad \textit{lp} ] \sqcup [ \textit{osoba} \quad 3 ] = \begin{bmatrix} \textit{liczba} & \textit{lp} \\ \textit{osoba} & 3 \end{bmatrix}$$

Jeżeli wartości cech są podstrukturami zawierającymi kolejne cechy, to do tych wartości stosuje się rekurencyjnie unifikację. Unifikacja dotyczy też cech współdzielonych – współdzielenie musi zachodzić także po unifikacji.

# Analiza składniowa z wykorzystaniem unifikacji

Cechy poddane unifikacji mogą mieć podstruktury, możliwe jest też współdzielenie cech:

$$\begin{aligned} & \sqcup \left[ \begin{array}{l} \text{zgoda} \quad \boxed{1} \left[ \begin{array}{l} \text{liczba} \quad \text{lp} \\ \text{osoba} \quad 3 \end{array} \right] \\ \text{podmiot} \quad \left[ \begin{array}{l} \text{zgoda} \quad \boxed{1} \end{array} \right] \\ \text{podmiot} \quad \left[ \begin{array}{l} \text{zgoda} \quad \left[ \begin{array}{l} \text{osoba} \quad 3 \\ \text{liczba} \quad \text{lp} \end{array} \right] \end{array} \right] \end{array} \right] \\ & = \left[ \begin{array}{l} \text{zgoda} \quad \boxed{1} \left[ \begin{array}{l} \text{liczba} \quad \text{lp} \\ \text{osoba} \quad 3 \end{array} \right] \\ \text{podmiot} \quad \left[ \begin{array}{l} \text{zgoda} \quad \boxed{1} \end{array} \right] \end{array} \right] \end{aligned}$$

# Analiza składniowa z wykorzystaniem unifikacji

Poniższa próba unifikacji zawodzi:

$$\left[ \begin{array}{l} \text{zgoda} \\ \text{podmiot} \end{array} \begin{array}{l} \boxed{1} \\ \left[ \begin{array}{l} \text{zgoda} \\ \text{osoba} \end{array} \right] \end{array} \left[ \begin{array}{l} \left[ \begin{array}{l} \text{liczba} \\ \text{osoba} \end{array} \right] \\ 3 \end{array} \right] \begin{array}{l} \text{lp} \\ 3 \end{array} \end{array} \right]$$

$$\sqcup \left[ \begin{array}{l} \text{zgoda} \\ \text{podmiot} \end{array} \left[ \begin{array}{l} \left[ \begin{array}{l} \text{liczba} \\ \text{osoba} \end{array} \right] \\ \text{zgoda} \end{array} \right] \left[ \begin{array}{l} \text{liczba} \\ \text{osoba} \end{array} \right] \begin{array}{l} \text{lp} \\ 3 \end{array} \right] \left[ \begin{array}{l} \text{liczba} \\ \text{osoba} \end{array} \right] \begin{array}{l} \text{lm} \\ 3 \end{array} \end{array} \right]$$

= *Błąd*

Unifikacja służy do modelowania zjawisk, które niewygodnie byłoby modelować za pomocą gramatyk bezkontekstowych. Należy do nich **związek zgody** (ang. *agreement*) występujący np. między grupą podmiotu (frazą rzeczownikową) i grupą orzeczenia (frazą czasownikową), między rzeczownikiem i określającym go przymiotnikiem itp.:

$$S \rightarrow NP VP$$
$$\langle NP \text{ zgoda} \rangle = \langle VP \text{ zgoda} \rangle$$



Także **związek rządu** (ang. *subcategorization*) jest modelowany za pomocą unifikacji. Określa on np. wymagania czasownika odnośnie dopełnień (np. **czekać** *na kogo, na co?*), wymagania niektórych rzeczowników (np. **proszek** *do prania, do pieczenia...*).

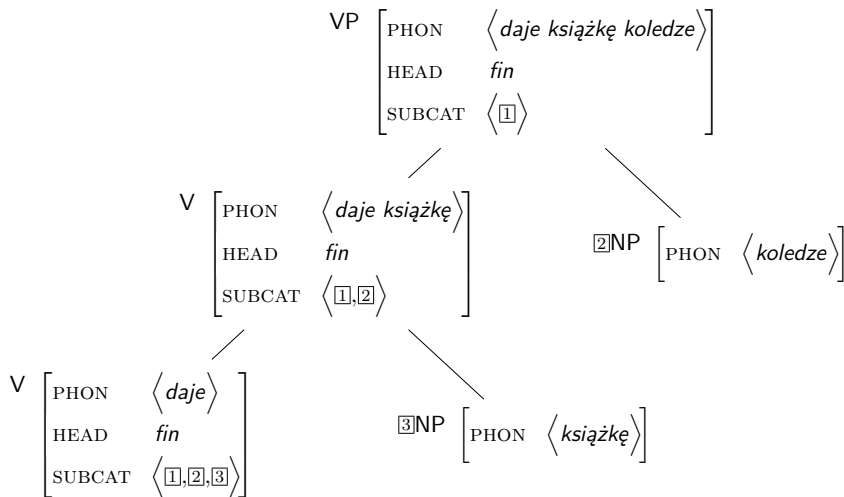
Związki czasowników w zdaniu są tak skomplikowane, że zapisywane są w **ramach** (ang. *subcategorization frames*). Dla języka angielskiego jest takich ram od 50 do 100. Np.:

$$\left[ \begin{array}{l} \text{ORTH} \quad \text{want} \\ \text{CAT} \quad \text{VERB} \\ \text{HEAD} \quad \left[ \text{SUBCAT} \quad \left\langle \left[ \text{CAT} \quad \text{NP} \right], \left[ \begin{array}{l} \text{CAT} \quad \text{VP} \\ \text{HEAD} \quad \left[ \text{VFORM} \quad \text{INFINITIVE} \right] \right] \right\rangle \right] \end{array} \right]$$

Unifikacja pozwala też na przekazywanie cech **elementu głównego** (ang. *head*) frazy wyżej, jako cech całej frazy. Np. cechy rzeczownika lub zaimka – podmiotu – są cechami całej grupy podmiotu. Cechy czasownika są cechami całej grupy orzeczenia.

Unifikacja jest jest głównym mechanizmem pozwalającym uzyskać analizę składniową zdania w gramatyce HPSG. Służy zarówno do przekazywania cech elementu głównego całej frazie, jak i do modelowania związków zgody i rządu.

# Zastosowanie unifikacji



Unifikacja pozwala także na modelowanie odległych zależności – takich, które przekraczają granice fraz. Np. w zdaniu:

*Którą książkę Marek pożyczył od Gosi?*

słowo *pożyczyć* wymaga dopełnienia. Rolę tego dopełnienia spełnia *którą książkę*, ale nie występuje w zwykłym miejscu – po czasowniku, tylko na początku zdania. „*Marek pożyczył □ od Gosi*” może być analizowane jako zdanie (*S*) z „dziurą” (ang. *gap*), w którą wstawiany jest ślad (ang. *trace*), natomiast pełne zdanie jest typu *S'* i wypełniacz musi zgodzić się z „*którą książkę*”.

# Realizacja unifikacji

```
1: function UNIFY(f1-orig,f2-orig)
2:   f1 ← dereference(f1-orig); f2 ← dereference(f2-orig)
3:   if f1 ≡ f2 then
4:     f1.pointer ← f2
5:     return f2
6:   else if f1 = null then
7:     f1.pointer ← f2
8:     return f2
9:   else if f2 = null then
10:    f2.pointer ← f1
11:    return f1
12:   else if complex(f1) ∧ complex(f2) then
13:     f2.pointer ← f1
14:     for all f2feature będących cechą f2 do
15:       f1feature ← znajdź lub utwórz odp. cechę w f1
16:       if Unify(f1feature.value, f2feature.value) = failure then
17:         return failure
18:       end if
19:     end for
20:     return f1
21:   else
22:     return failure
23:   end if
24: end function
```

- 1 Elżbieta Dobryjanowicz, *Podstawy przetwarzania języka naturalnego. Wybrane metody analizy składniowej*, Akademicka Oficyna Wydawnicza RM, Warszawa, 1992.
- 2 Christer Samuelsson, *Fast Natural-Language Parsing Using Explanation-Based Learning*, Swedish Institute of Computer Science, ISRN SICS/D-13-SE, 1994.