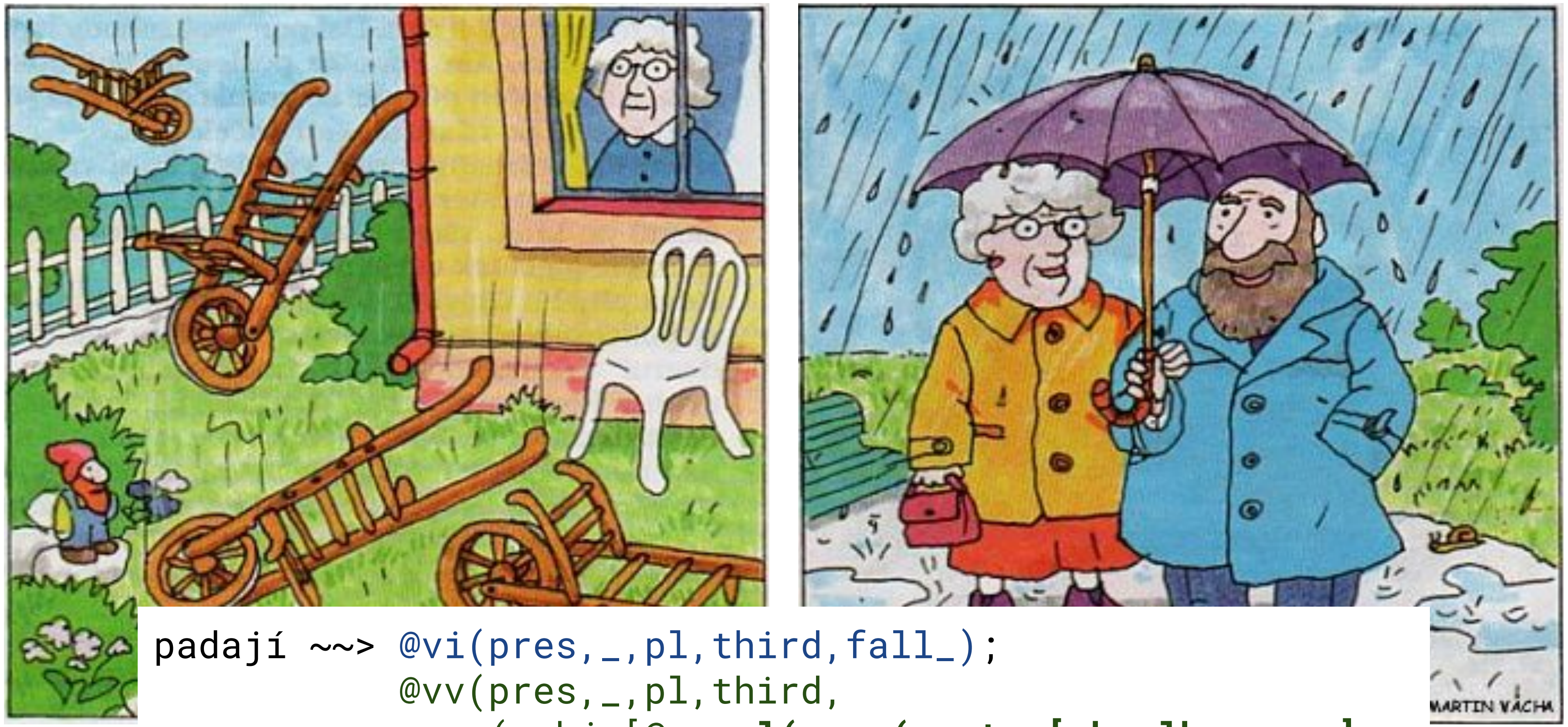
Joe & Molly meet Idioms

Ondřej Cífka

Idioms

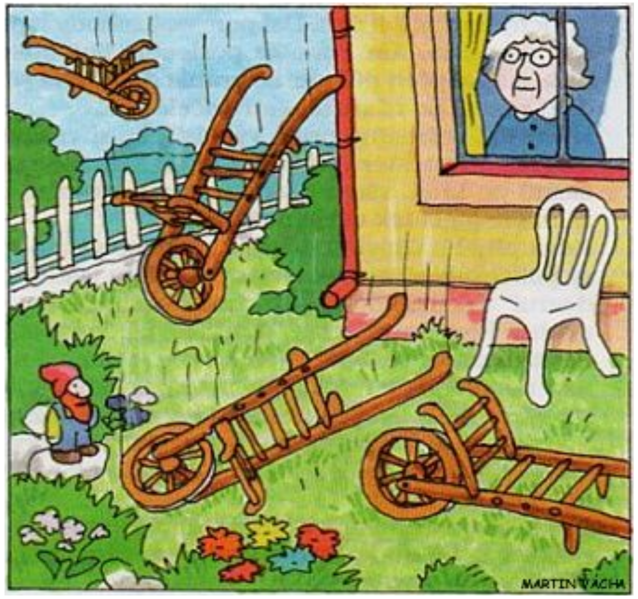
- expressions with meanings that cannot be derived from the meanings of their constituents
- some have a literal interpretation
    - *it's raining cats and dogs* (*padají trakaře*)
    - *kick the bucket* (*zaklepat bačkorama*)
- some don't
    - *dal mu* ***co proto***
    - *bylo to* ***raz dva***
    - *byl* ***ten tam***

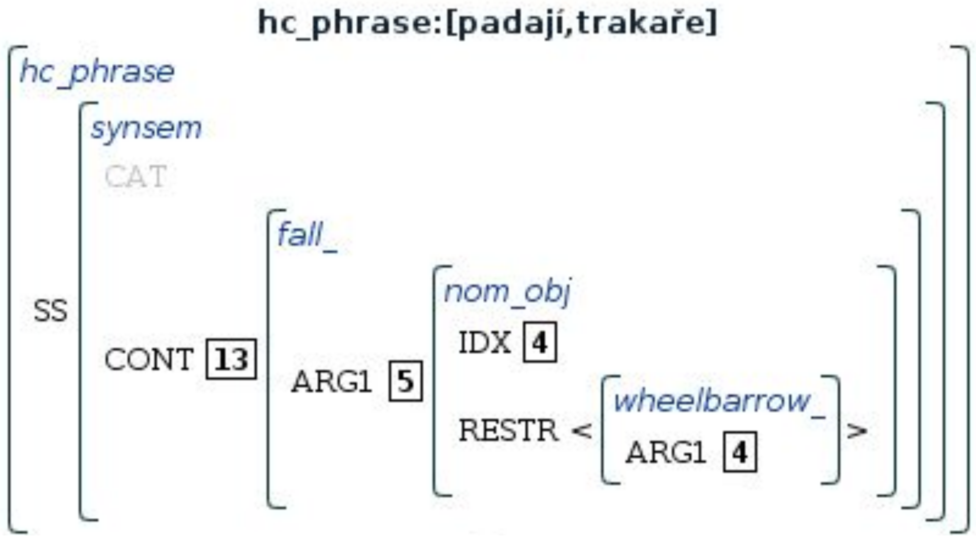
Padají trakaře



```
padají ~~> @vi(pres,_,pl,third,fall_);
           @vv(pres,_,pl,third,
               (subj:[@nnval(nom,(restr:[wheelbarrow_],
                                  idx:num:pl))],
                comps:[]),
               rain_).
```

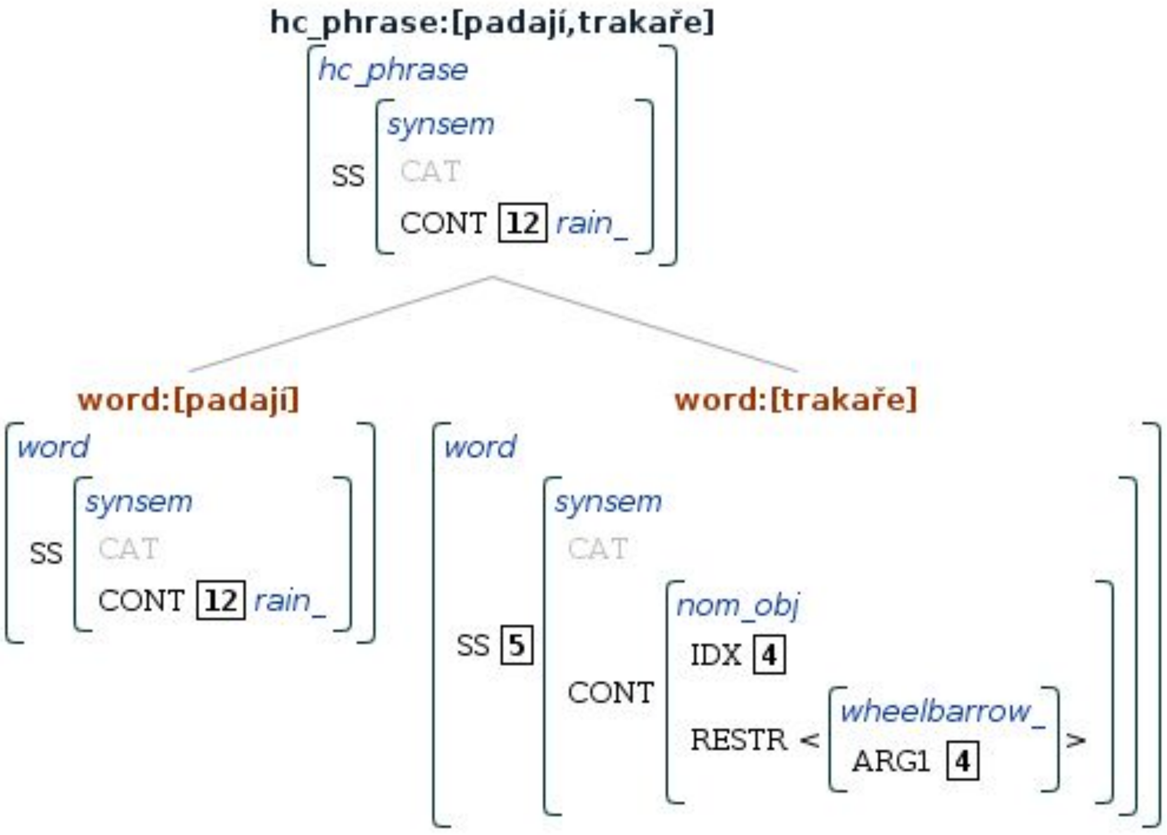
Padají trakaře



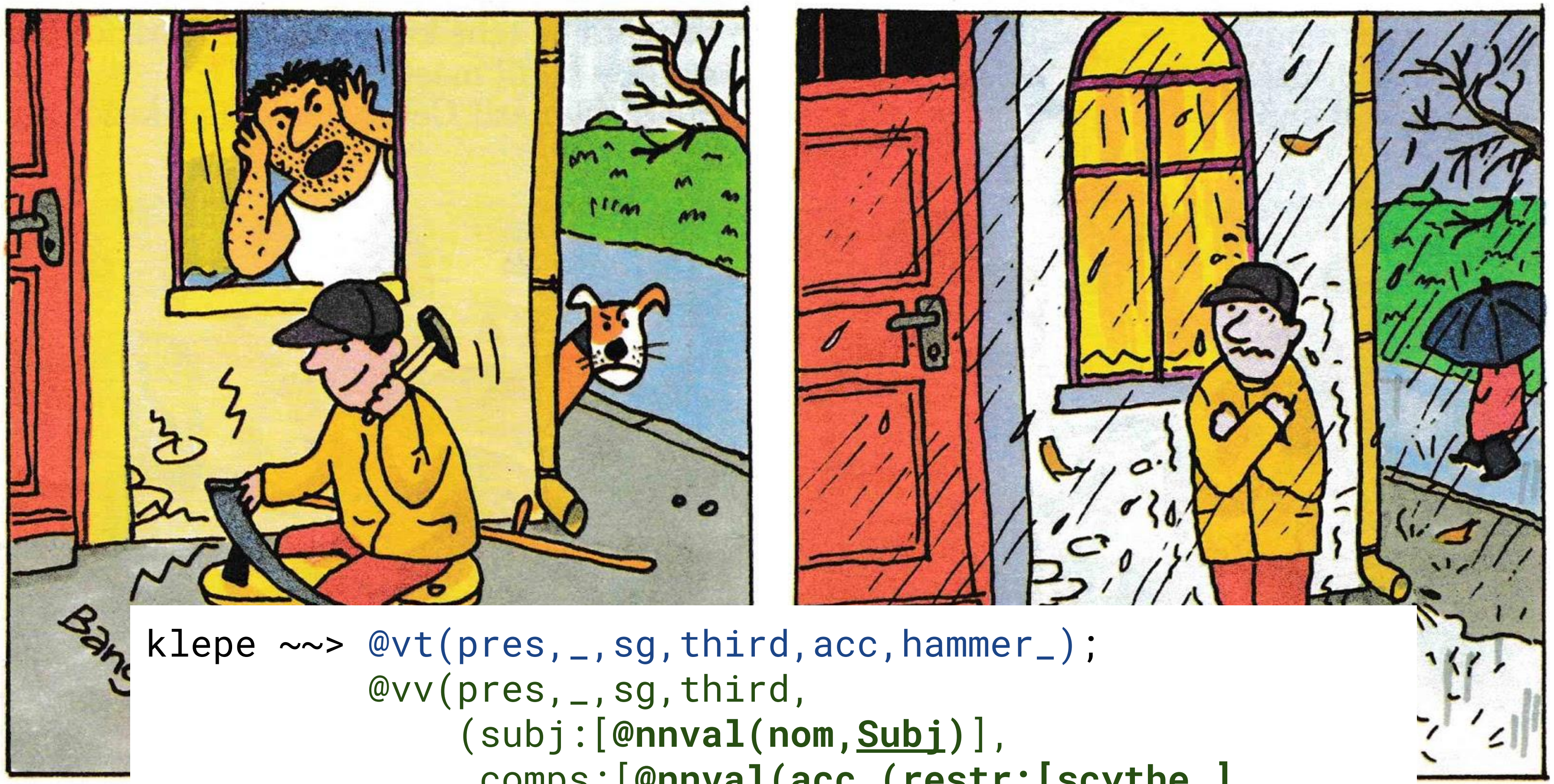


```
hc_phrase:[padají,trakaře]
[hc_phrase
 SS [synsem
     CAT
     CONT [13] [fall_
                ARG1 [5] [nom_obj
                          IDX [4]
                          RESTR < [wheelbarrow_
                                   ARG1 [4]      ] > ] ] ] ]
```

Padají trakaře



Klepat kosu

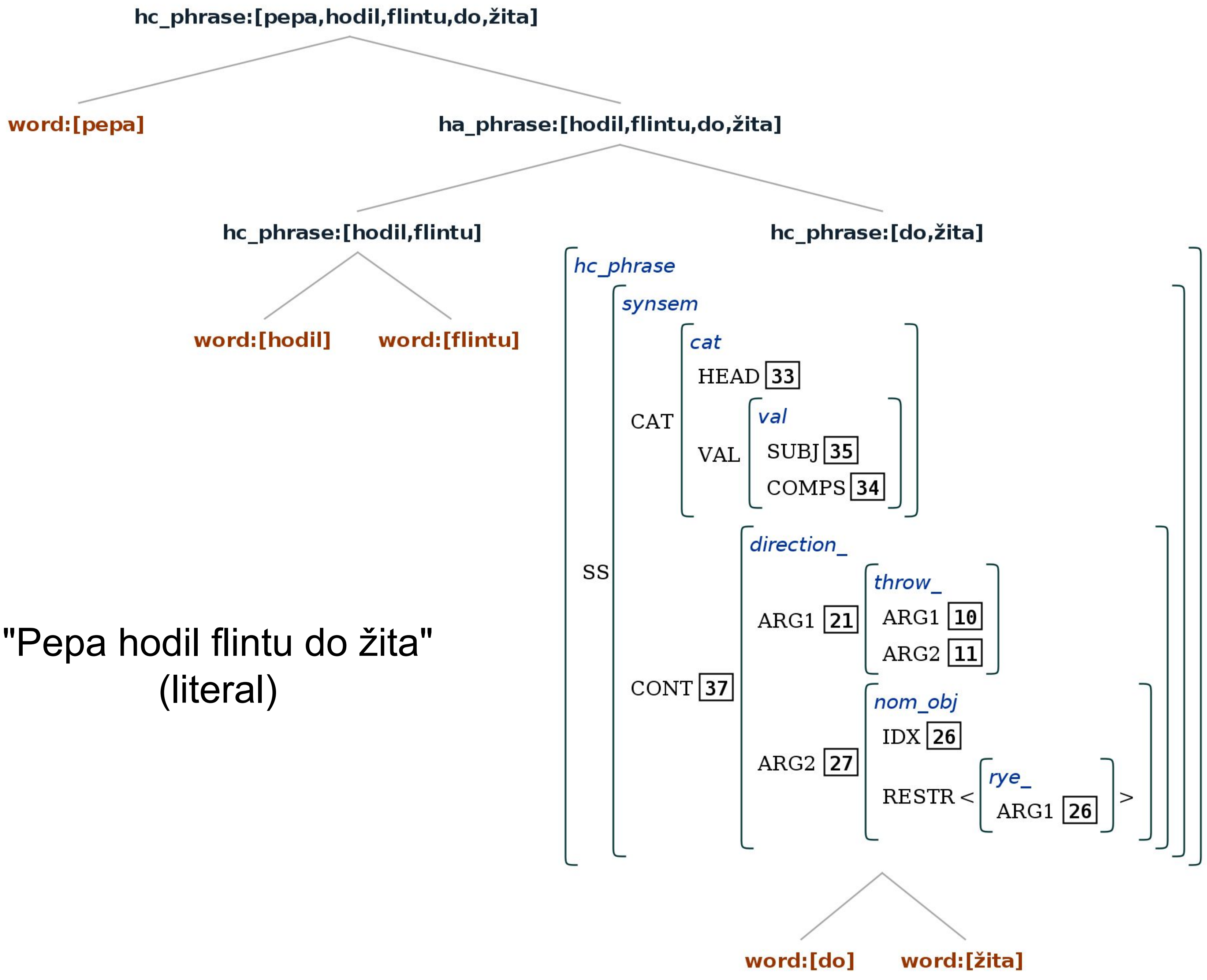


```
klepe ~~> @vt(pres,_,sg,third,acc,hammer_);
          @vv(pres,_,sg,third,
              (subj:[@nnval(nom,Subj)],
               comps:[@nnval(acc,(restr:[scythe_],
                                   idx:num:sg))]),
              (feel_cold_, arg1:Subj)).
```

Pepa klepe kosu (figurative)

word:[klepe]

$$
\left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{word}} \\
\text{SS} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{synsem}} \\
\text{CAT} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{cat}} \\
\text{HEAD} & \boxed{16} \\
\text{VAL} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{val}} \\
\text{SUBJ} & \boxed{18} \\
\text{COMPS} & \left\langle \boxed{11} \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{synsem}} \\
\multicolumn{2}{l}{\text{CAT}} \\
\text{CONT} & \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{nom\_obj}} \\
\text{IDX} & \boxed{8} \\
\text{RESTR} & \left\langle \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{scythe\_}} \\
\text{ARG1} & \boxed{8}
\end{array}\right] \right\rangle
\end{array}\right]
\end{array}\right] \mid \boxed{17} \langle\rangle \right\rangle
\end{array}\right]
\end{array}\right] \\
\text{CONT} & \boxed{20} \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{feel\_cold\_}} \\
\text{ARG1} & \boxed{10} \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{nom\_obj}} \\
\text{IDX} & \boxed{3} \\
\text{RESTR} & \left\langle \left[\begin{array}{ll}
\multicolumn{2}{l}{\textit{joe\_}} \\
\text{ARG1} & \boxed{3}
\end{array}\right] \right\rangle
\end{array}\right]
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$



"Pepa hodil flintu do žita"
(literal)

Solution (not very good)

```
do ~~> (word,
        ss:(cat:(head:(prep,
                       pred:minus,
                       pform:do_gen,
                       mod:[(cat:head:verb,
                             cont:VCont)]),
                 val:(subj:e_list,
                      comps:[@nnval(gen,PObjCont)])),
            cont:(direction_,
                  arg1:VCont,
                  arg2:PObjCont));
        ss:(cat:(head:(...,
                       mod:[(cat:head:verb,
                             cont:(throw_,
                                   arg1:SubjCont,
                                   arg2:(restr:[rifle_])))]),
                 val:(subj:e_list,
                      comps:[@nnval(gen,(restr:[rye_]))])),
            cont:(give_up_,
                  arg1:SubjCont))).
```

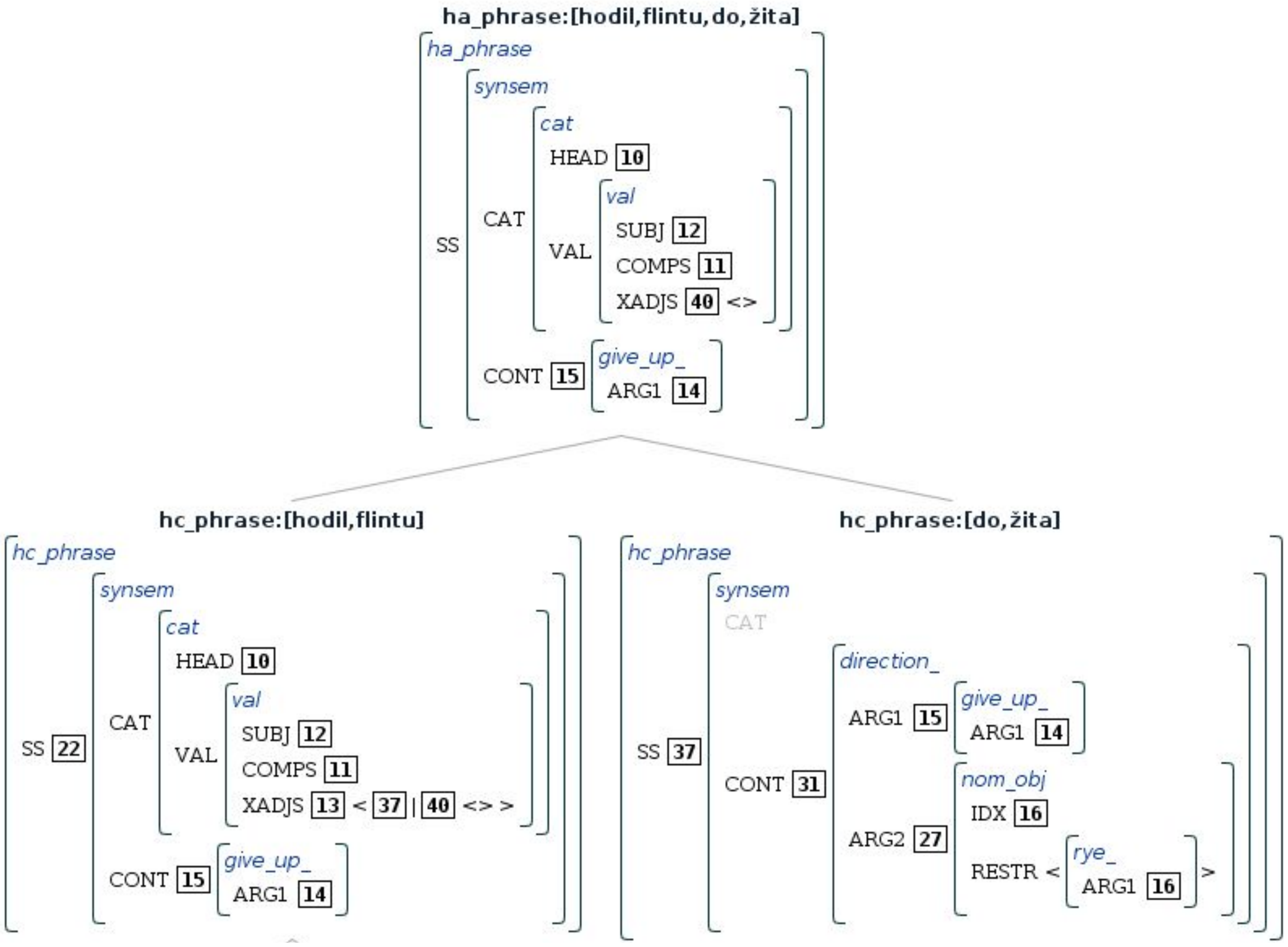
Solution (better)

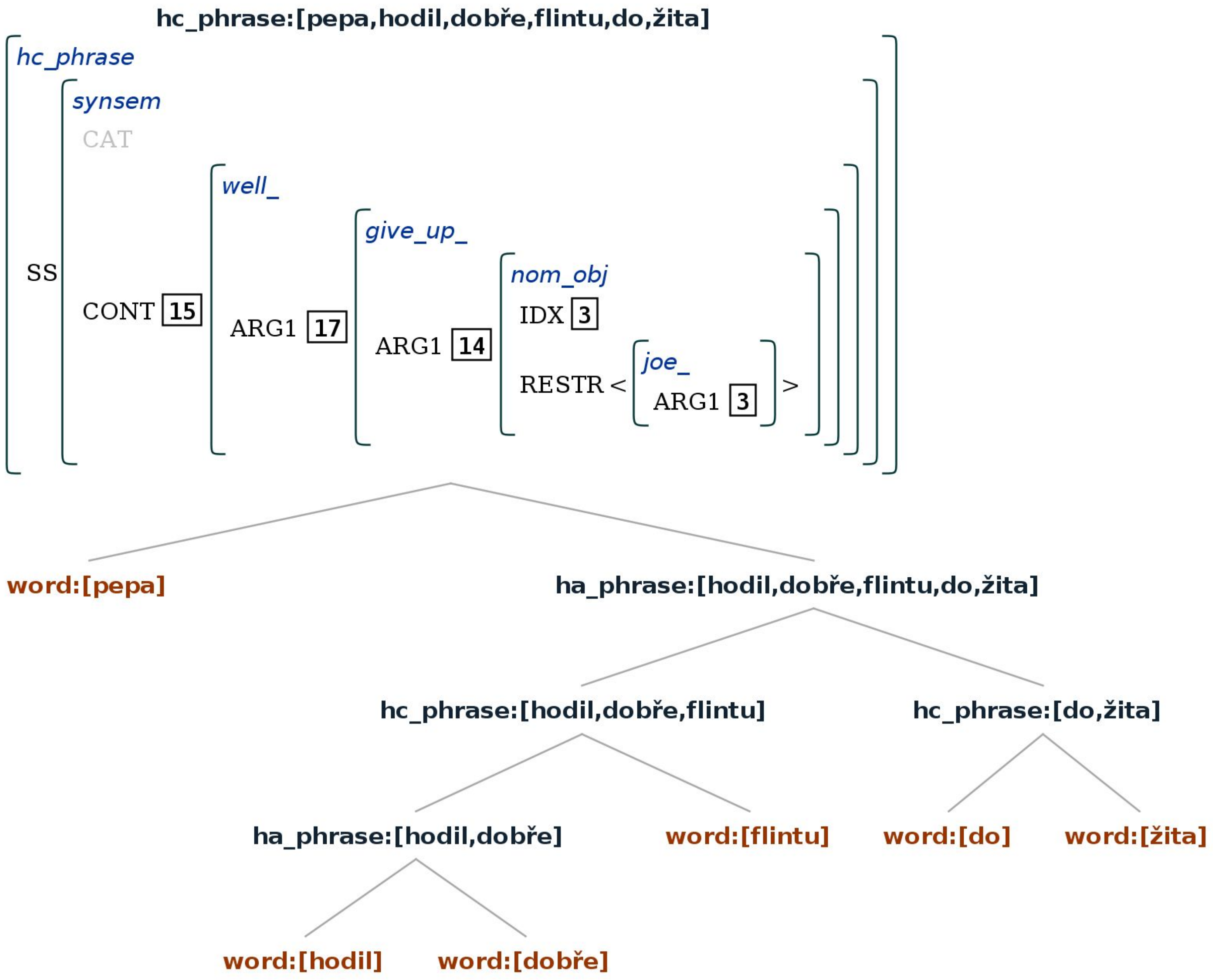
$$\begin{bmatrix} \textit{val} & \\ \ \text{COMPS} & \textit{list} \\ \ \text{SUBJ} & \textit{list} \\ \ \text{XADJS} & \textit{list} \end{bmatrix}$$

```
hodit ~~> @vinf((subj:[@nnval(nom,Subj)],
                 comps:[@nnval(acc,(restr:[rifle_],
                                    idx:num:sg))],
                 xadjs:[(cat:head:pform:do_gen,
                         cont:(direction_,
                               arg2:(restr:[rye_],
                                     idx:num:sg)))]),
                (give_up_, arg1:Subj)).

hodil ~~> @vv(past,ma,sg,third,
              (subj:[@nnval(nom,Subj)],
               comps:[@nnval(acc,(restr:[rifle_],
                                  idx:num:sg))],
               xadjs:[(cat:head:pform:do_gen,
                       cont:(direction_,
                             arg2:(restr:[rye_],
                                   idx:num:sg)))]),
              (give_up_, arg1:Subj)).
```

Solution (better)

```
% Semantics Principle

ha_phrase *> ((ss:(cont:NHCont,
                  cat:val:xadjs:XAdjs),
              nonh_dtr:ss:cont:NHCont,
              head_dtr:ss:cat:val:xadjs:XAdjs);
             (ss:(cont:HCont,
                  cat:val:xadjs:del(NH_ss, XAdjs)),
              nonh_dtr:ss:NH_ss,
              head_dtr:ss:(cont:HCont,
                           cat:val:xadjs:XAdjs))).
```


ha_phrase:[hodil,flintu,do,žita]
ha_phrase
SS
synsem
CAT
cat
HEAD 10
VAL
val
SUBJ 12
COMPS 11
XADJS 40 <>
CONT 15
give_up_
ARG1 14
hc_phrase:[hodil,flintu]
hc_phrase
SS 22
synsem
CAT
cat
HEAD 10
VAL
val
SUBJ 12
COMPS 11
XADJS 13 < 37 | 40 <> >
CONT 15
give_up_
ARG1 14
hc_phrase:[do,žita]
hc_phrase
SS 37
synsem
CAT
CONT 31
direction_
ARG1 15
give_up_
ARG1 14
ARG2 27
nom_obj
IDX 16
RESTR <
rye_
ARG1 16
>

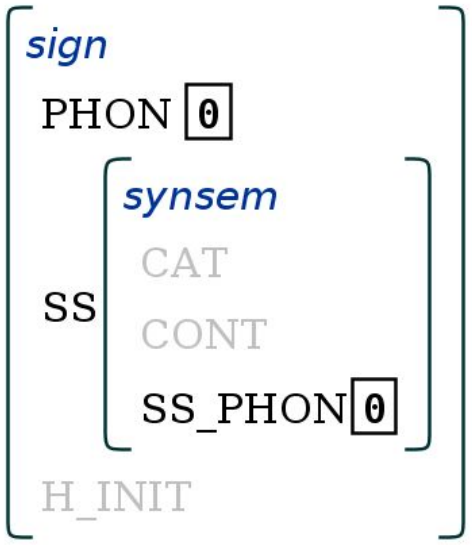

hc_phrase:[pepa,hodil,dobře,flintu,do,žita]
hc_phrase
SS
synsem
CAT
CONT 15
well_
ARG1 17
give_up_
ARG1 14
nom_obj
IDX 3
RESTR < joe_ ARG1 3 >
word:[pepa]
ha_phrase:[hodil,dobře,flintu,do,žita]
hc_phrase:[hodil,dobře,flintu]
hc_phrase:[do,žita]
ha_phrase:[hodil,dobře]
word:[flintu]
word:[do]
word:[žita]
word:[hodil]
word:[dobře]

hc_phrase:[děvčata,chtějí,hodit,flintu,do,žita]

$$\left[\begin{array}{l} \textit{hc\_phrase} \\ \text{SS}\ \left[\begin{array}{l} \textit{synsem} \\ \text{CAT}\ \left[\begin{array}{l} \textit{cat} \\ \text{HEAD}\ \boxed{46} \\ \text{VAL}\ \left[\begin{array}{l} \textit{val} \\ \text{SUBJ} <> \\ \text{COMPS}\ \boxed{47} \\ \text{XADJS}\ \boxed{49} \end{array}\right] \end{array}\right] \\ \text{CONT}\ \boxed{50}\ \left[\begin{array}{l} \textit{want\_} \\ \text{ARG1}\ \boxed{6}\ \left[\begin{array}{l} \textit{nom\_obj} \\ \text{IDX}\ \boxed{3} \\ \text{RESTR} < \left[\begin{array}{l} \textit{girl\_} \\ \text{ARG1}\ \boxed{3} \end{array}\right] > \end{array}\right] \\ \text{ARG2}\ \boxed{8}\ \left[\begin{array}{l} \textit{give\_up\_} \\ \text{ARG1}\ \boxed{6}\ \left[\begin{array}{l} \textit{nom\_obj} \\ \text{IDX}\ \boxed{3} \\ \text{RESTR} < \left[\begin{array}{l} \textit{girl\_} \\ \text{ARG1}\ \boxed{3} \end{array}\right] > \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

Fixed (multi-word) expressions

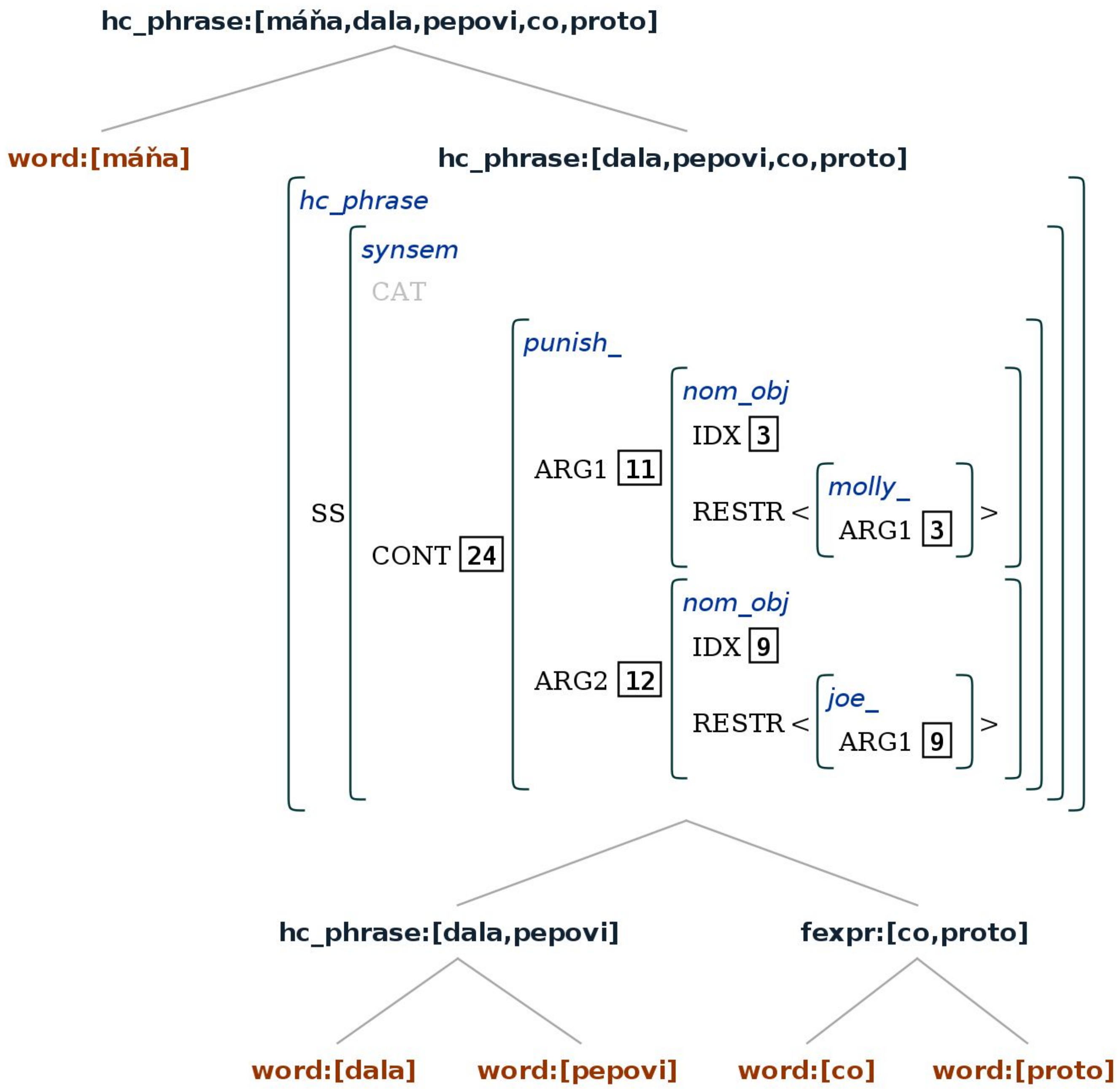
- e.g. *bylo to* ***raz dva***, *byl* ***ten tam***, *dal mu* ***co proto***



```
fixed_expr ##
    (fexpr,
     h_init:plus,
     head_dtr:Head,
     nonh_dtr:NonH,
     ss:cat:head:fixed)
===>
    cat> (Head,
          word,
          h_init:plus),
    cat> (NonH,
          (fexpr;word),
          ss:cat:head:fixed).
```

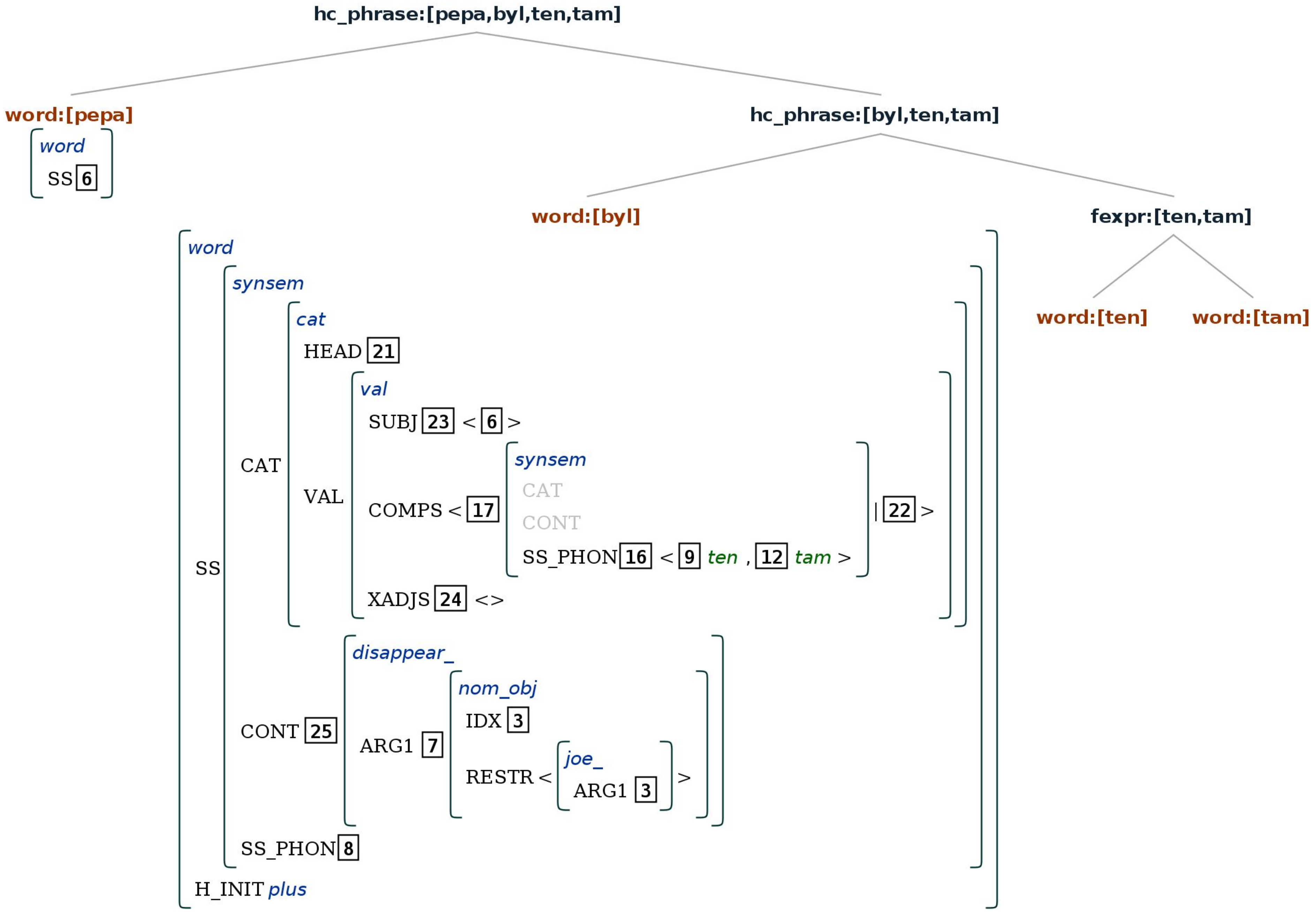
Dát co proto (give what for)

```
dala ~~> @vd(past,f,sg,third,acc,dat,give_);
         @vv(past,f,sg,third,
             (subj:[@nnval(nom,Arg1)],
              comps:[@nnval(dat,Arg2),
                     (ss_phon:[(a_ co),(a_ proto)])],
              xadjs:[]),
             (punish_,arg1:Arg1,arg2:Arg2)).
```


hc_phrase:[máňa,dala,pepovi,co,proto]
word:[máňa]
hc_phrase:[dala,pepovi,co,proto]
hc_phrase
synsem
CAT
SS
CONT 24
punish_
ARG1 11
nom_obj
IDX 3
RESTR < molly_ ARG1 3 >
ARG2 12
nom_obj
IDX 9
RESTR < joe_ ARG1 9 >
hc_phrase:[dala,pepovi]
fexpr:[co,proto]
word:[dala]
word:[pepovi]
word:[co]
word:[proto]

Být ten tam (disappear)

```
byl ~~> ...;
        @vv(past,m,sg,third,
            (subj:[@nnval(nom,Subj)],
             comps:[(ss_phon:[(a_ ten),(a_ tam)])],
             xadjs:[]),
            (disappear_,
             arg1:Subj)).

byla ~~> ...;
         @vv(past,f,sg,third,
             (subj:[@nnval(nom,Subj)],
              comps:[(ss_phon:[(a_ ta),(a_ tam)])],
              xadjs:[]),
             (disappear_,
              arg1:Subj)).

...
```


hc_phrase:[pepa,byl,ten,tam]
word:[pepa]
word
SS 6
hc_phrase:[byl,ten,tam]
word:[byl]
fexpr:[ten,tam]
word:[ten]
word:[tam]
word
synsem
cat
HEAD 21
val
SUBJ 23 < 6 >
CAT
VAL
synsem
CAT
CONT
COMPS < 17
SS_PHON 16 < 9 ten , 12 tam >
| 22 >
SS
XADJS 24 <>
disappear_
nom_obj
IDX 3
CONT 25
ARG1 7
joe_
RESTR <
ARG1 3
>
SS_PHON 8
H_INIT plus

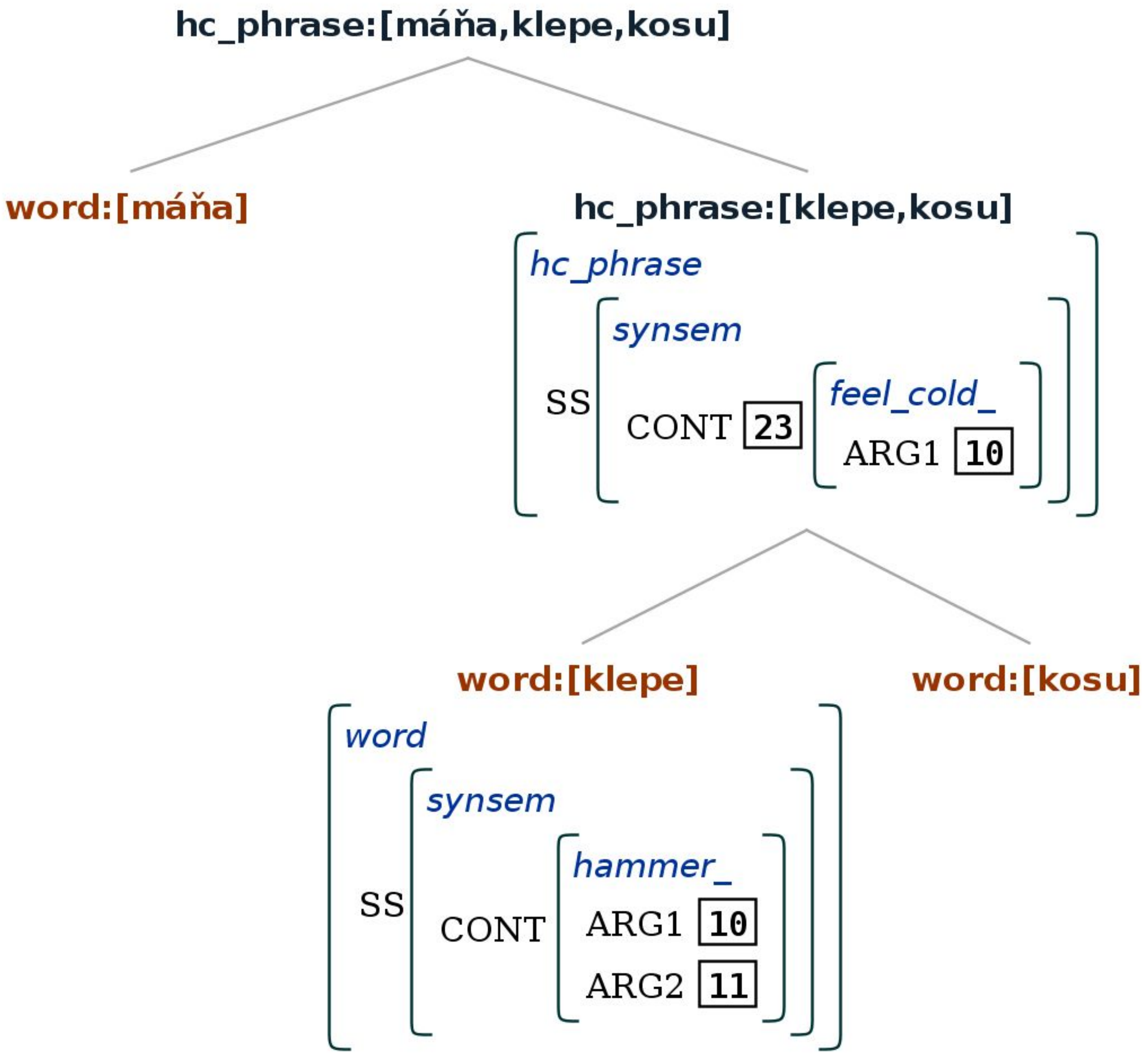
Fully semantics-based approach

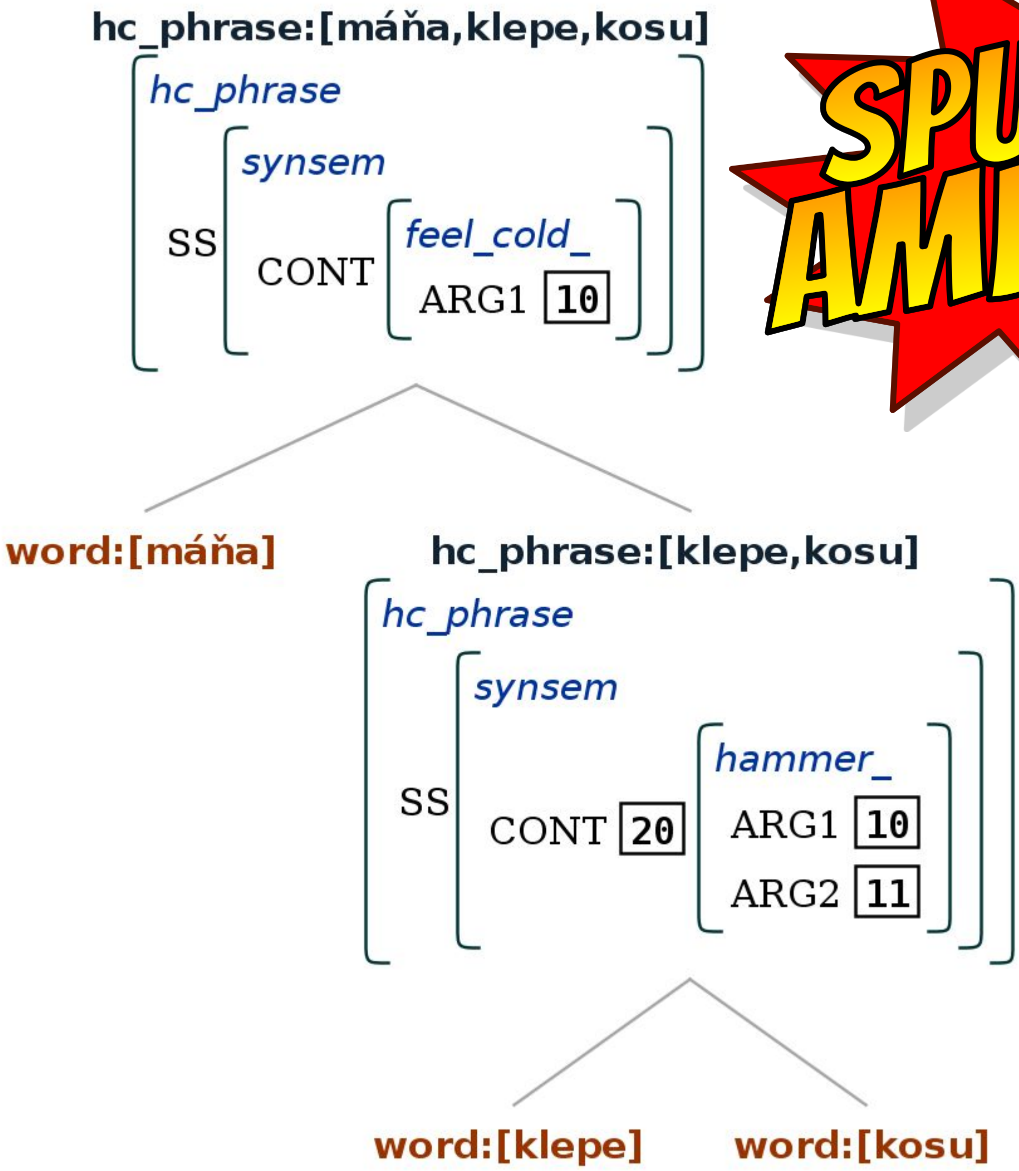
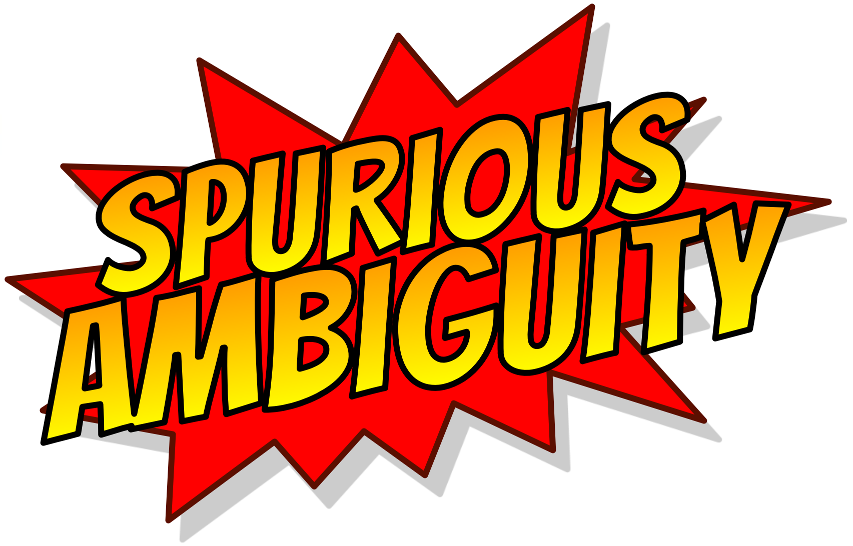
- `idiom` predicate pairing literal and figurative meanings:

```
idiom((fall_, arg1:(restr:[wheelbarrow_], idx:num:pl)),
      rain_) if true.
idiom((hammer_, arg1:Arg1, arg2:(restr:[scythe_], idx:num:sg)),
      (feel_cold_, arg1:Arg1)) if true.
idiom((direction_,
       arg1:(throw_,
             arg1:Arg1,
             arg2:(restr:[rifle_], idx:num:sg)),
       arg2:(restr:[rye_], idx:num:sg)),
      (give_up_,
       arg1:Arg1)) if true.
```

Fully semantics-based approach

```
fun idiom(+,-).


% Semantics Principle

hc_phrase *> (ss:cont:(Cont;idiom(Cont)),
              head_dtr:ss:cont:Cont).

ha_phrase *> (ss:cont:(Cont;idiom(Cont)),
              nonh_dtr:ss:cont:Cont).
```

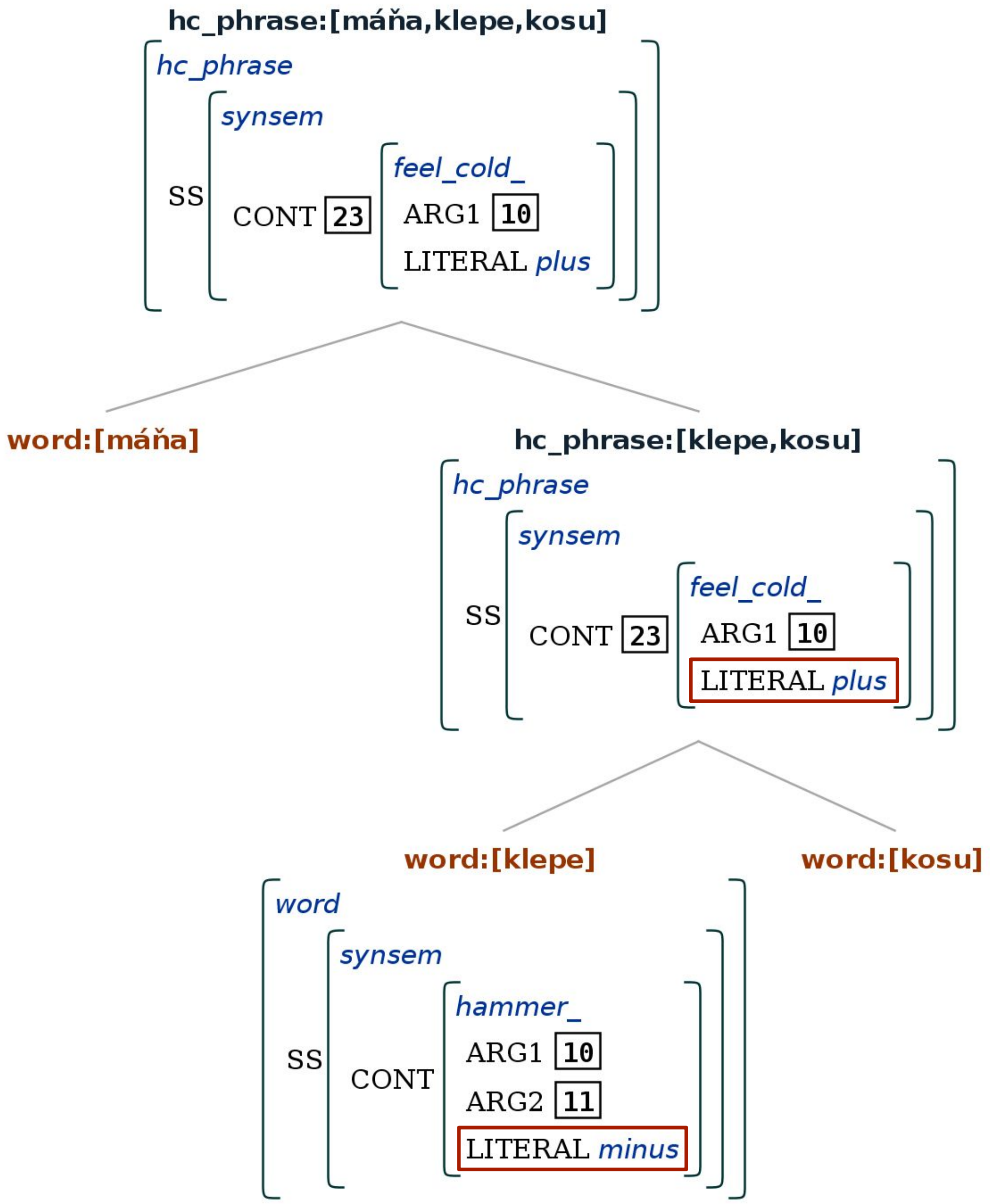

hc_phrase:[máňa,klepe,kosu]
word:[máňa]
hc_phrase:[klepe,kosu]
hc_phrase
SS
synsem
CONT 23
feel_cold_
ARG1 10
word:[klepe]
word:[kosu]
word
SS
synsem
CONT
hammer_
ARG1 10
ARG2 11


hc_phrase:[máňa,klepe,kosu]
hc_phrase
SS
synsem
CONT
feel_cold_
ARG1 10
word:[máňa]
hc_phrase:[klepe,kosu]
hc_phrase
SS
synsem
CONT 20
hammer_
ARG1 10
ARG2 11
word:[klepe]
word:[kosu]

SPURIOUS
AMBIGUITY

Solution

- Mark every `cont` as `literal:plus` or `literal:minus`
- Use `idiom` if and only if marked as `literal:minus`

```
fun make_cont(+,-).
make_cont(ss:cont:(Cont, literal:plus), Cont) if true.
make_cont(ss:cont:(LCont, literal:minus), ICont)
   if idiom(LCont, ICont).


% Semantics Principle

hc_phrase *> (ss:cont:make_cont(Dtr),
              head_dtr:Dtr).

ha_phrase *> (ss:cont:make_cont(Dtr),
              nonh_dtr:Dtr).
```


hc_phrase:[máňa,klepe,kosu]
hc_phrase
SS
synsem
CONT 23
feel_cold_
ARG1 10
LITERAL plus
word:[máňa]
hc_phrase:[klepe,kosu]
hc_phrase
SS
synsem
CONT 23
feel_cold_
ARG1 10
LITERAL plus
word:[klepe]
word:[kosu]
word
SS
synsem
CONT
hammer_
ARG1 10
ARG2 11
LITERAL minus

Links

- **https://github.com/cifkao/ltgf-project**/tree/val_approach
- **https://github.com/cifkao/ltgf-project**/tree/master