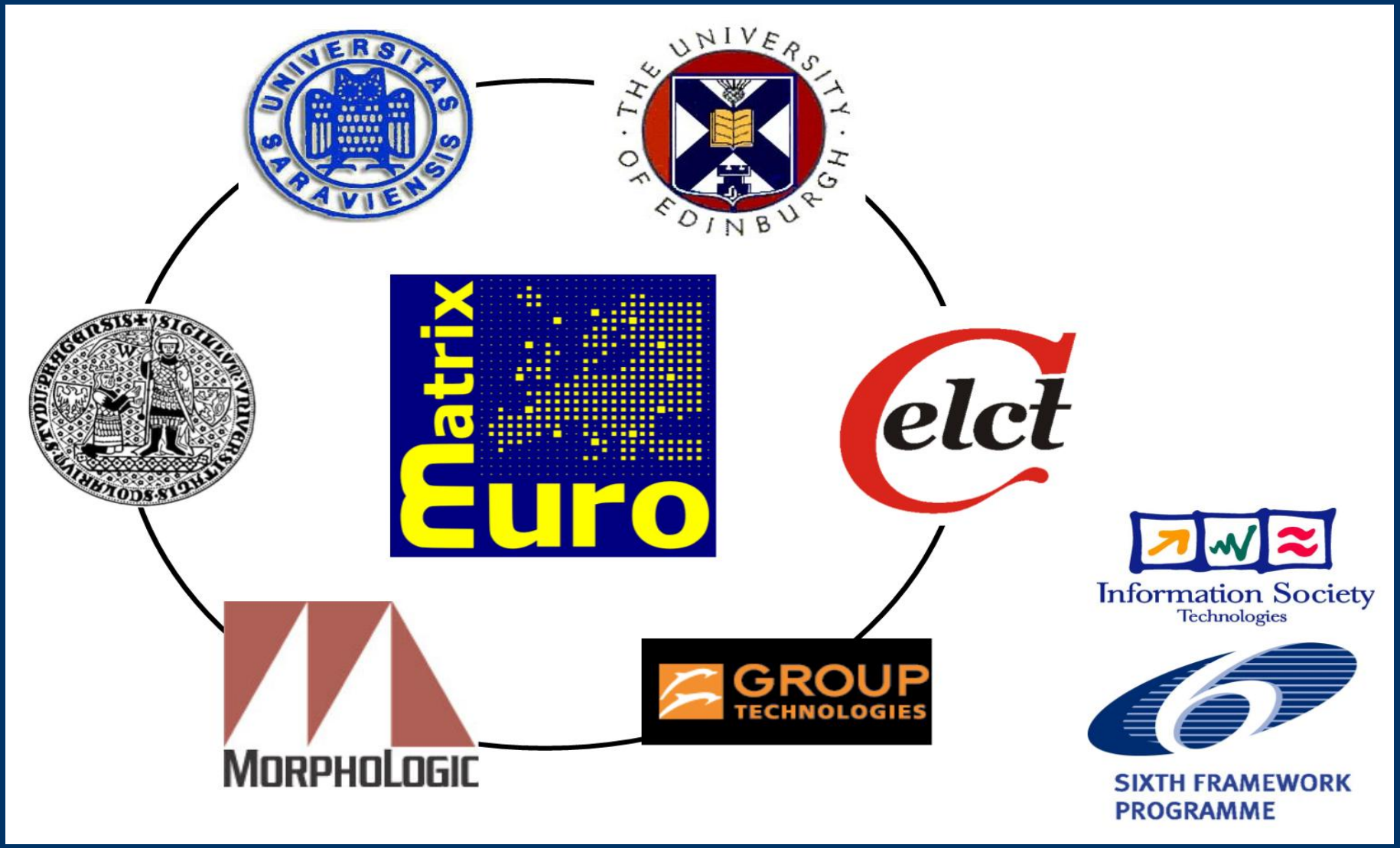
Automatic Acquisition of Semantic Transfer Rules for MT

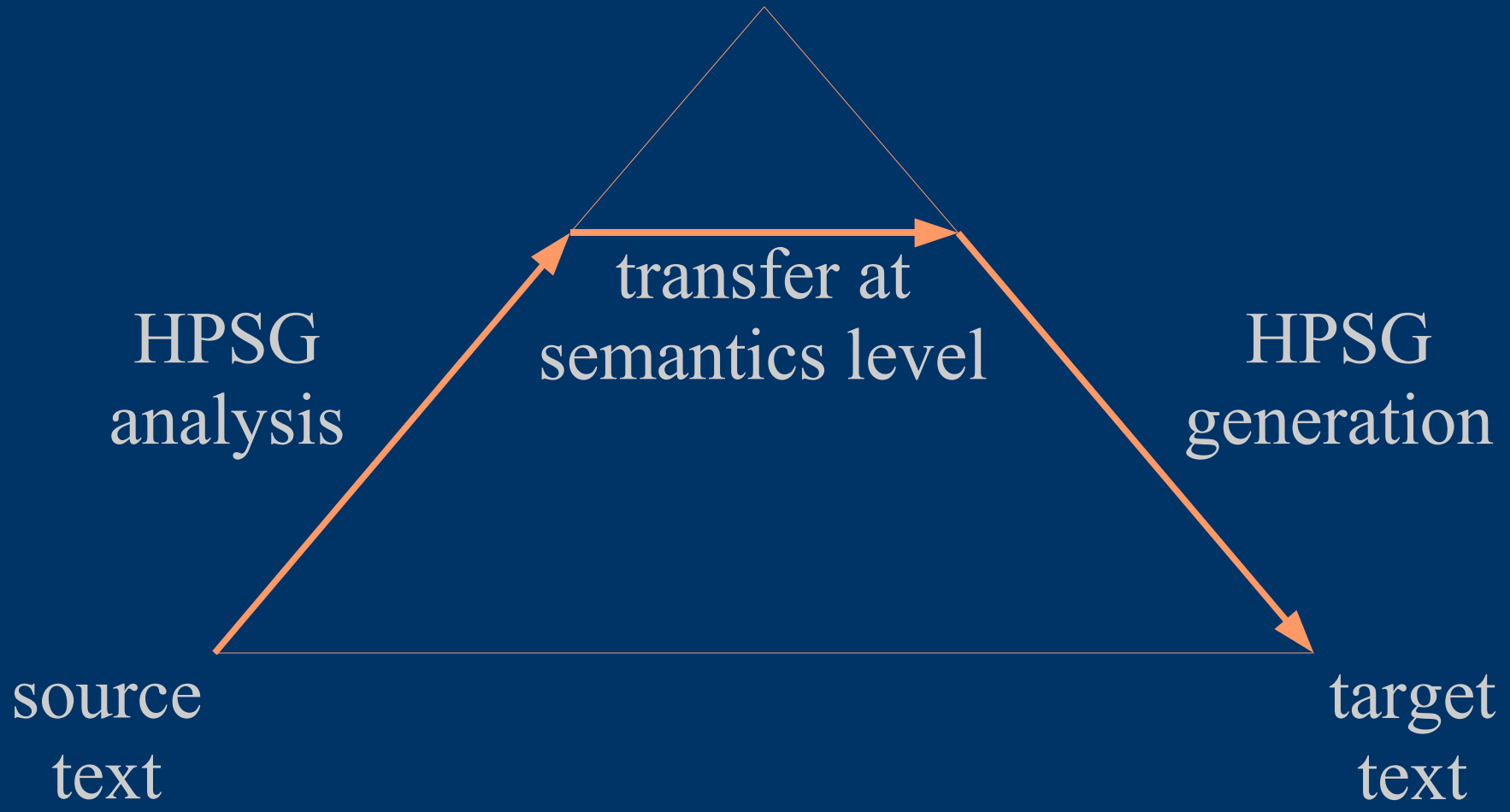
Michael Jellinghaus

Saarland University
micha@coli.uni-sb.de

Context: EuroMatrix Project



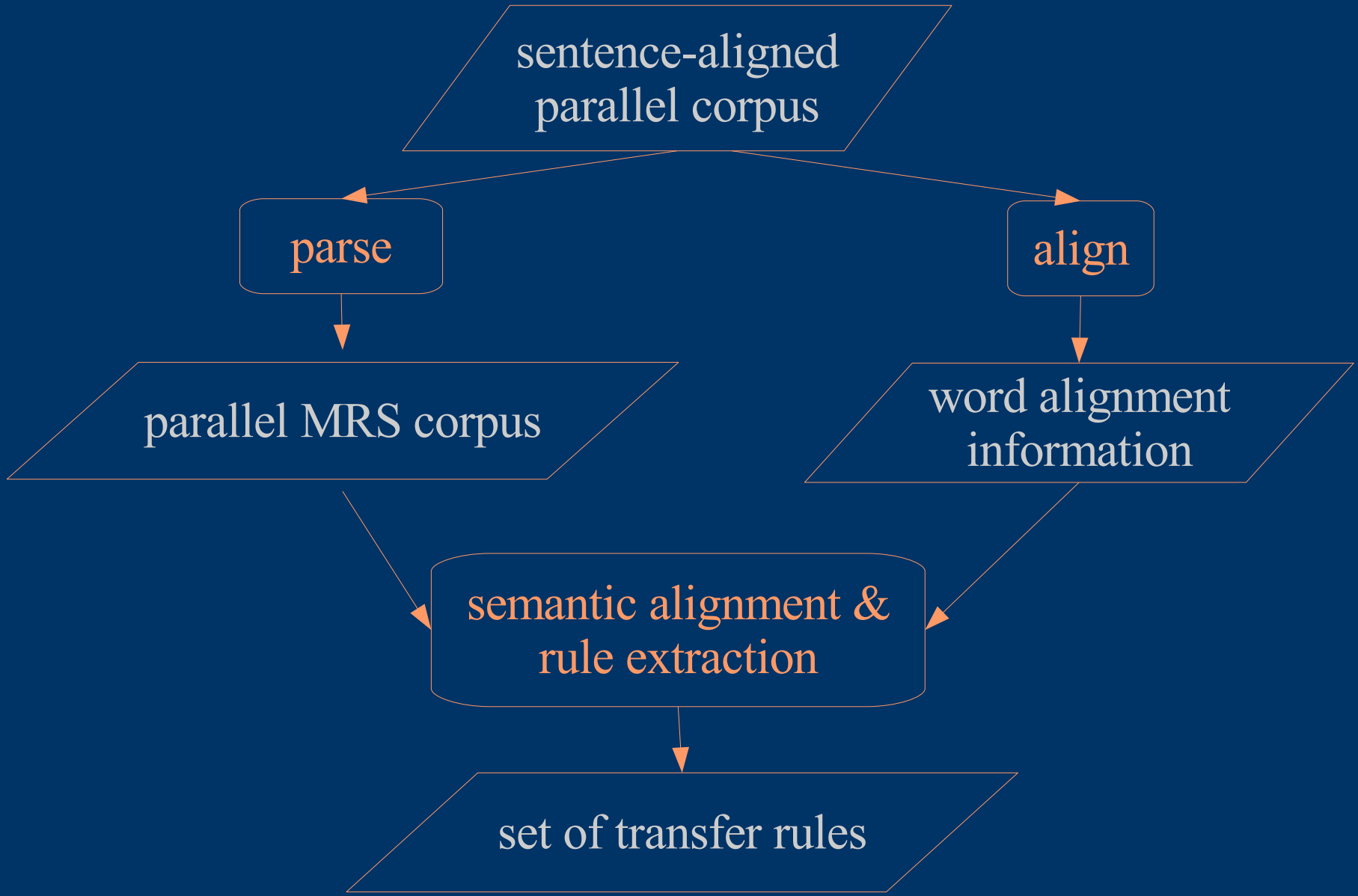
Deep Transfer-Based MT



Complementary Approaches to MT

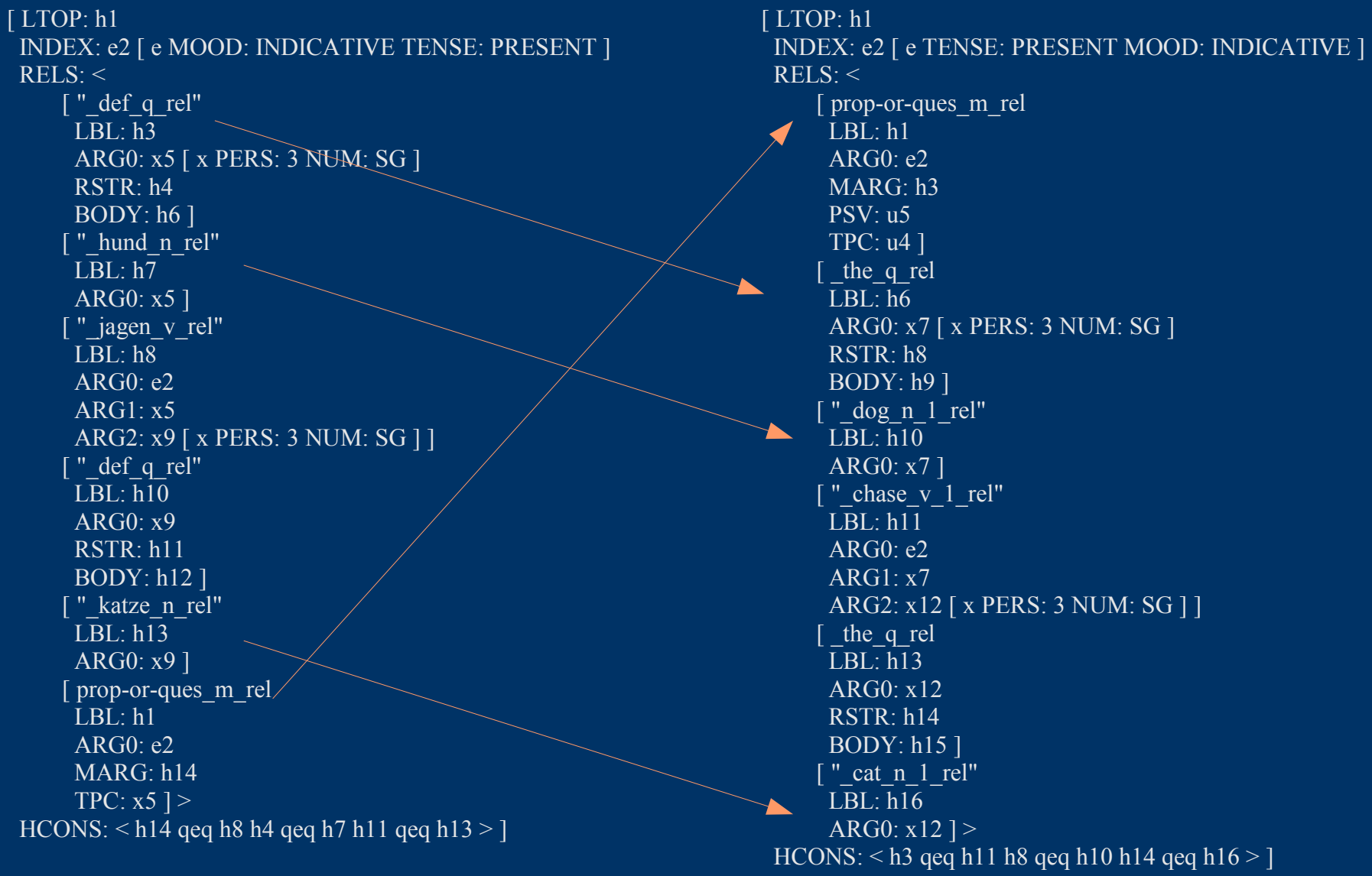
	statistical	deep transfer
development speed	+	-
grammaticality	-	+
lexical semantics	+	-
sentence semantics	-	+
coverage	+	-

Acquisition of Transfer Rules



MRS Example

“der Hund jagt die Katze” - “the dog chases the cat”



MRS Example

“der Hund jagt die Katze” - “the dog chases the cat”

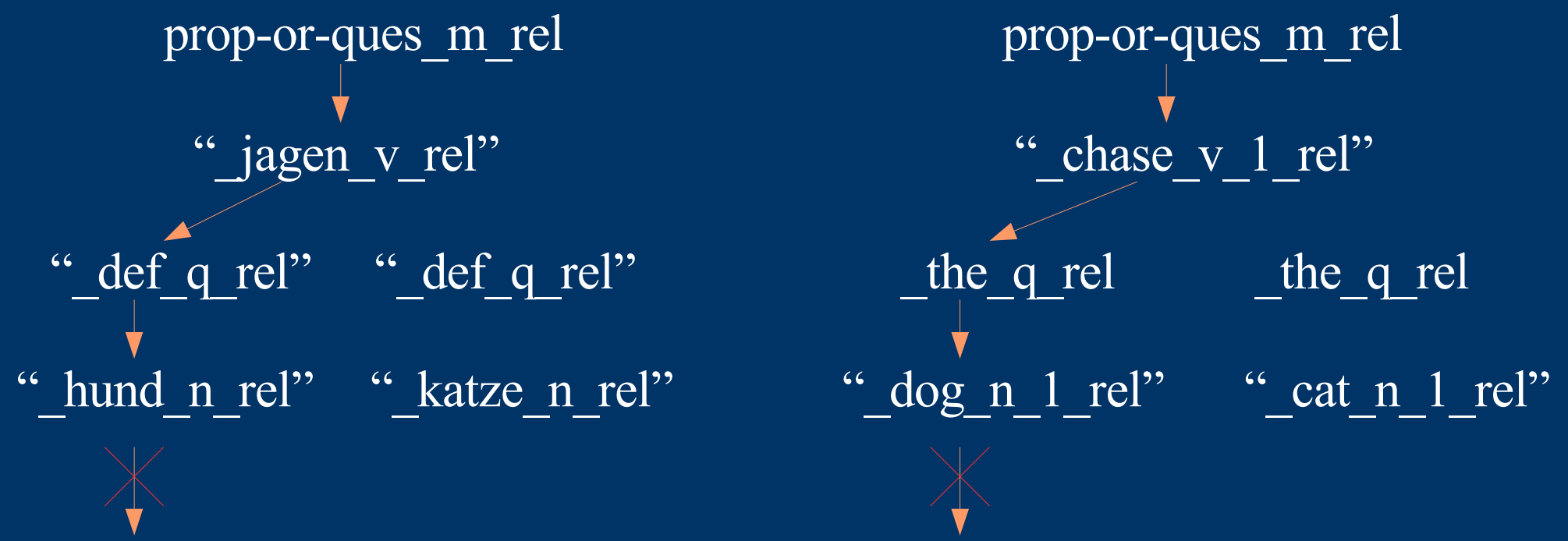
```
[ LTOP: h1
  INDEX: e2 [ e MOOD: INDICATIVE TENSE: PRESENT ]
  RELS: <
        [ "_def_q_rel"
          LBL: h3
          ARG0: x5 [ x PERS: 3 NUM: SG ]
          RSTR: h4
          BODY: h6 ]
        [ "_hund_n_rel"
          LBL: h7
          ARG0: x5 ]
        [ "_jagen_v_rel"
          LBL: h8
          ARG0: e2
          ARG1: x5
          ARG2: x9 [ x PERS: 3 NUM: SG ] ]
        [ "_def_q_rel"
          LBL: h10
          ARG0: x9
          RSTR: h11
          BODY: h12 ]
        [ "_katze_n_rel"
          LBL: h13
          ARG0: x9 ]
        [ prop-or-ques_m_rel
          LBL: h1
          ARG0: e2
          MARG: h14
          TPC: x5 ] >
  HCONS: < h14 qeq h8 h4 qeq h7 h11 qeq h13 > ]
```

LTOP h1
prop-or-ques_m_rel LBL h1
MARG h14
h14 qeq h8
"_jagen_v_rel" LBL h8

"_def_q_rel" LBL h3
RSTR h4
h4 qeq h7
"_hund_n_rel" LBL h7

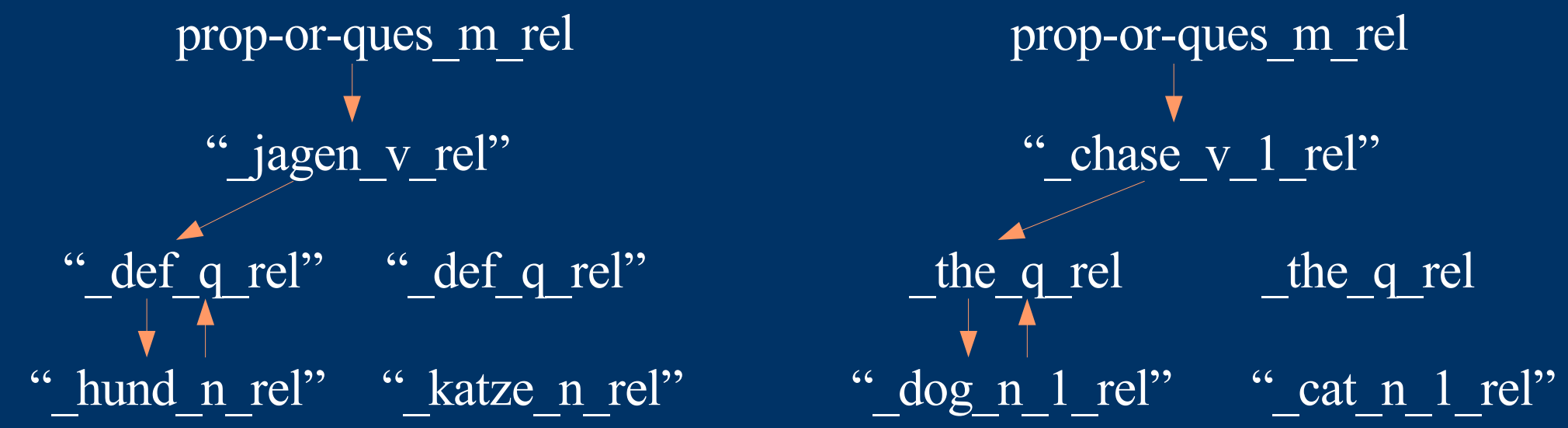
"_def_q_rel" LBL h10
RSTR h11
h11 qeq h13
"_katze_n_rel" LBL h13

Semantic Alignment



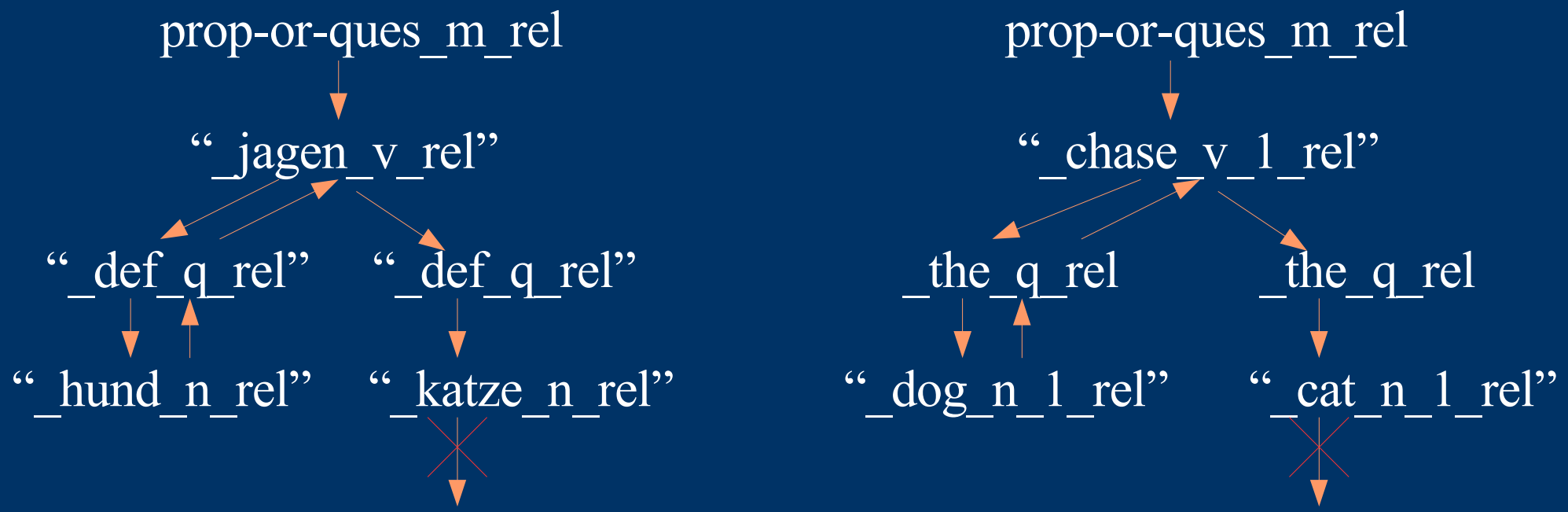

```
dog_rule_0 := monotonic_omtr &
 [ INPUT [ RELS < [ PRED "_dog_n_1_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4] ] > ],
 OUTPUT [ RELS < [ PRED "_hund_n_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4] ] > ] ].
```

Semantic Alignment



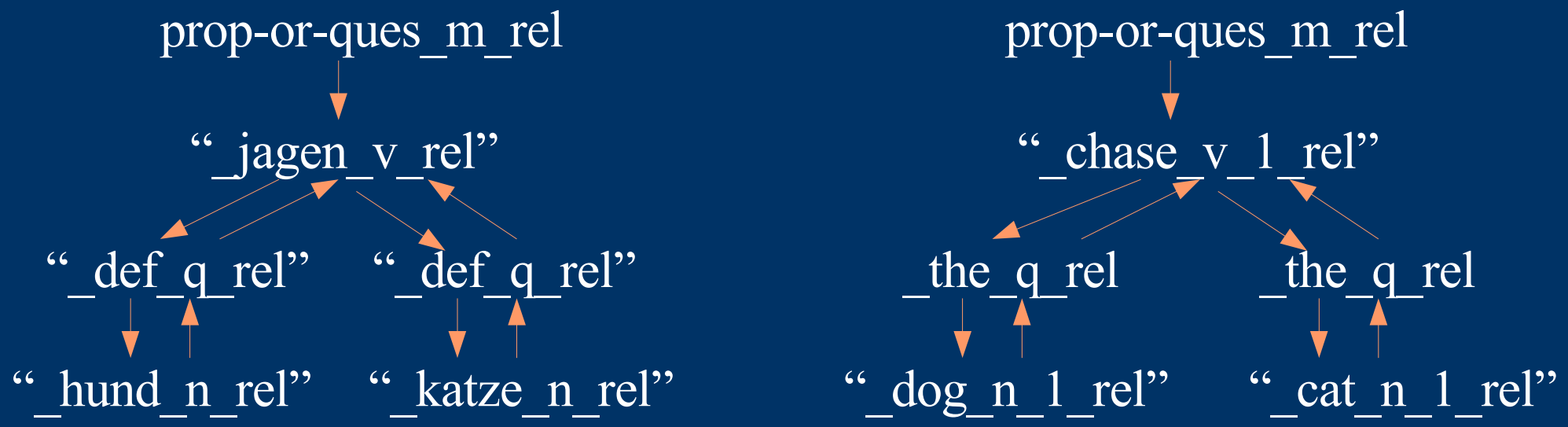

```
the_rule_0 :=monotonic_omtr &
  [ INPUT [ RELS < [ PRED _the_q_rel, LBL #1, RSTR #2, ARG0 #3 & [PERS #4, NUM #5], BODY #6 ] > ],
    OUTPUT [ RELS < [ PRED "_def_q_rel", LBL #1, RSTR #2, ARG0 #3 & [PERS #4, NUM #5], BODY #6 ] > ] ].

the_rule_1 := monotonic_omtr &
  [ INPUT [ RELS < [ PRED "_dog_n_1_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4] ],
                   [ PRED _the_q_rel, LBL #5, RSTR #6, ARG0 #2, BODY #7 ] >,
            HCONS < [qeq & HARG #6, LARG #1] > ],
    OUTPUT [ RELS < [ PRED "_def_q_rel", LBL #5, RSTR #6, ARG0 #2 & [PERS #3, NUM #4], BODY #7 ],
                    [ PRED "_hund_n_rel", LBL #1, ARG0 #2 ] >,
             HCONS < [qeq & HARG #6, LARG #1] >  ] ].
```

Semantic Alignment



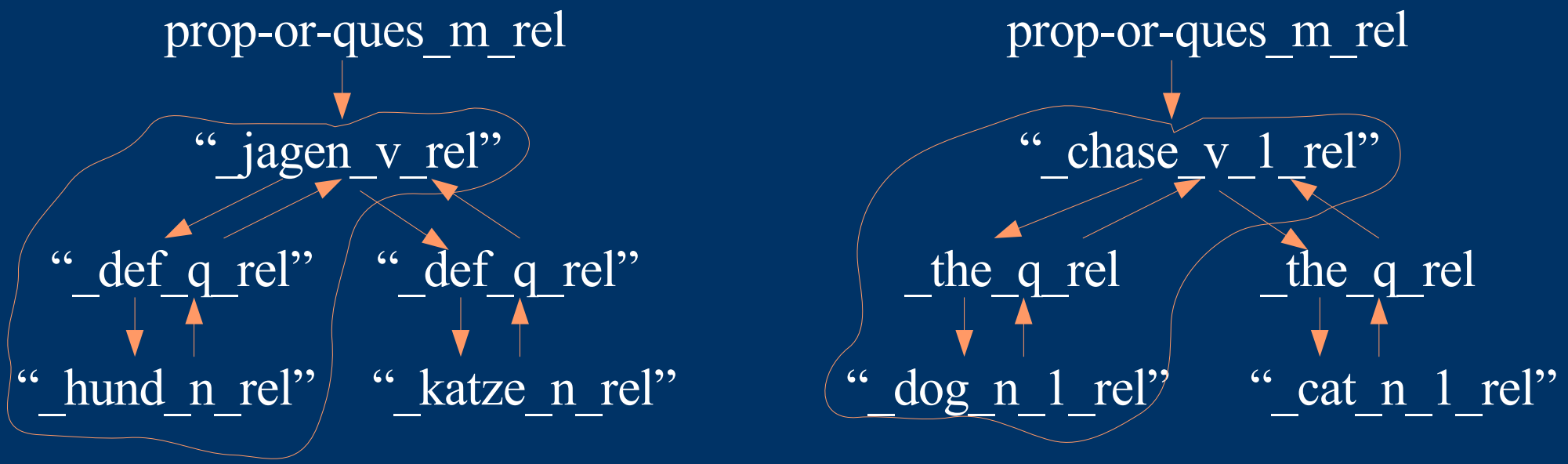
cat_rule_0 := monotonic_omtr &
[INPUT [RELS < [PRED "_cat_n_1_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4]] >],
OUTPUT [RELS < [PRED "_katze_n_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4]] >]].

Semantic Alignment



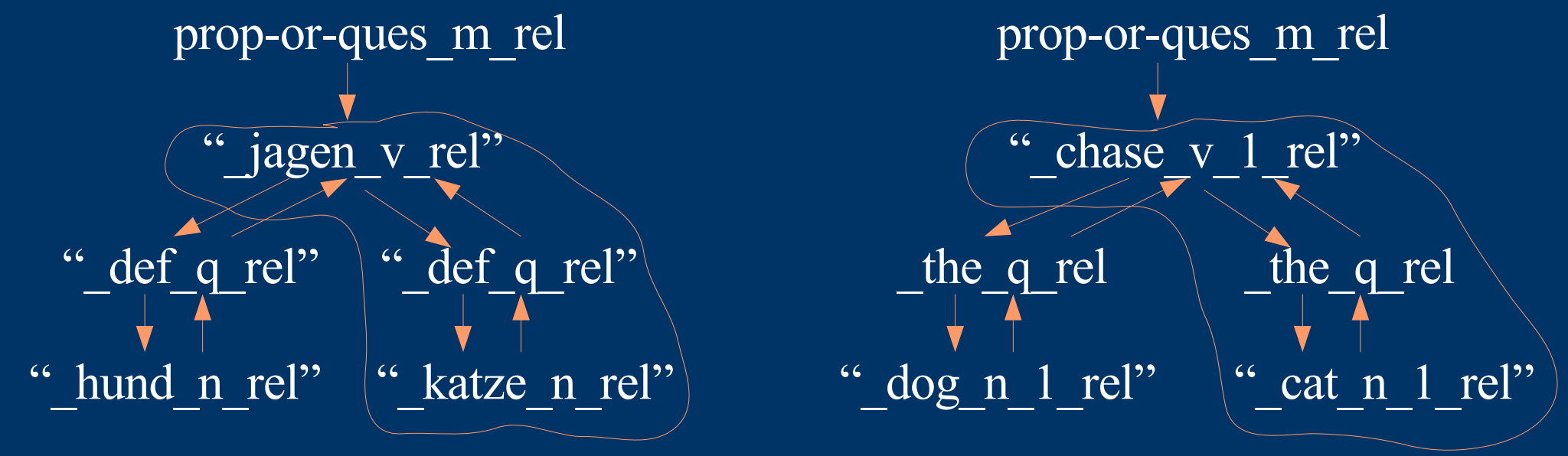

```
chase_rule_0 := monotonic_omtr &
[ INPUT [ RELS < [ PRED "_chase_v_1_rel", LBL #1, ARG0 #2 & [MOOD #3, TENSE #4], ARG2 #5, ARG1 #6 ] > ],
OUTPUT [ RELS < [ PRED "_jagen_v_rel", LBL #1, ARG0 #2 & [TENSE #4, MOOD #3], ARG2 #5, ARG1 #6 ] > ] ].
```

Semantic Alignment




```
chase_rule_1 := monotonic_omtr &
[ INPUT [ RELS < [ PRED "_dog_n_1_rel", LBL #1, ARG0 #2 & [PERS #3, NUM #4] ],
                 [ PRED _the_q_rel, LBL #5, RSTR #6, ARG0 #2, BODY #7 ],
                 [ PRED "_chase_v_1_rel", LBL #8, ARG0 #9 & [MOOD #10, TENSE #11], ARG2 #8, ARG1 #2 ] >,
         HCONS < [qeq & HARG #6, LARG #1] > ],
  OUTPUT [ RELS < [ PRED "_jagen_v_rel", LBL #8, ARG0 #9 & [TENSE #11, MOOD #10],
                    ARG2 #8, ARG1 #2 & [PERS #3, NUM #4] ],
                  [ PRED "_def_q_rel", LBL #5, RSTR #6, ARG0 #2, BODY #7 ],
                  [ PRED "_hund_n_rel", LBL #1, ARG0 #2 ] >,
           HCONS < [qeq & HARG # 6, LARG #1] > ] ].
```

Semantic Alignment




```
chase_rule_2 := monotonic_omtr &
  [ INPUT [ RELS < [ PRED _the_q_rel, LBL #1, RSTR #2, ARG0 #3 & [PERS #4, NUM #5], BODY #6 ],
                   [ PRED "_chase_v_1_rel", LBL #7, ARG0 #8 & [MOOD #9, TENSE #10], ARG2 #3, ARG1 #11 ],
                   [ PRED "_cat_n_1_rel", LBL #12, ARG0 #3 ] >,
            HCONS < [qeq & HARG #2, LARG #12] > ],
    OUTPUT [ RELS < [ PRED "_jagen_v_rel", LBL #7, ARG0 #8 & [TENSE #10, MOOD #9],
                      ARG2 #3 & [PERS #4, NUM #5], ARG1 #11 ],
                    [ PRED "_def_q_rel", LBL #1, RSTR #2, ARG0 #3, BODY #6 ],
                    [ PRED "_katze_n_rel", LBL #12, ARG0 #3 ] >,
             HCONS < [qeq & HARG #2, LARG #12] > ] ].
```

Complementary Approaches to MT

	statistical	deep transfer
development speed	+	-→+
grammaticality	-	+
lexical semantics	+	-→+
sentence semantics	-	+
coverage	+	-

State of Affairs

Functional prototype

- complete pipeline:
  - data preparation
  - rule acquisition
  - translation: parsing, transfer, generation
- MRS Test Suite

Currently expanding to bigger training sets

- CLEF data (Question Answering)
- EuroParl

State of Affairs

Current challenge: noisy data

- loose translations
  “Wer...?” -- “What group...?”
- parse selection errors
  “Das Auto jagt die Katze”
  -- “The car chases the cat” or “The cat chases the car”
- alignment errors

⇒ stochastification (assign weights to rules) or
⇒ stochastification light (sort by #EPs, frequency)

State of Affairs

Rule size

- Complete sentences?
- Scope of interesting collocations?
- Generalizations?
- Size of rule set?

Extensions

- other language pairs
- phrases
- ...

Thank you!