

Computational Linguistics — Text Analysis

★ Text

Keď dievča zaspalo, prišiel stavec, posvietil si naň, potriasol ho za plece a spustil do pivnice.

Na tretie ráno povedal drevorubač žene: „Dnes mi pošli s obedom najmladšiu. Bývala vždy dobrá a poslušná, iste pôjde po pravej ceste a nie kadetade, ako jej roztárané sestry.“

Matke sa to nevidelo: „Mám prísť aj o svoje najmilšie dieťa?“ — „O to sa neboj,“ vraví muž, „dievča nezablúdi, je rozvážne a múdre. Okrem toho vezmem hrachu a porozsýpam ho, je väčší ako šošovica, ukáže jej cestu.“

★ Grammar

- ▷ $S \rightarrow S \text{ CONJP } S$
- ▷ $S \rightarrow S \text{ ?}$
- ▷ $S \rightarrow S : (S \text{ DIR } +)$
- ▷ $(S \text{ DIR } +) \rightarrow (S \text{ DIR } +) (S \text{ DIR } -) (S \text{ DIR } +)$
- ▷ $(S \text{ GAP } + (NP \text{ GAP } + \text{AGR ?a})) \rightarrow (NP \text{ GAP } \text{AGR ?a}) (VP \text{ GAP } - \text{AGR ?a})$
- ▷ $S \rightarrow (S \text{ AUX } +) S \mid S (S \text{ AUX } +)$
- ▷ $(S \text{ AUX } + \text{AGR ?a VFORM ?v}) \rightarrow (S \text{ AUX } + \text{SUBCAT time AGR ?a VFORM ?v})$
- ▷ $(S \text{ AUX } + \text{SUBCAT time AGR ?a VFORM ?v}) \rightarrow$
 $(\text{PART SUBCAT time}) (S \text{ AGR ?a VFORM ?v})$
- ▷ $(S \text{ ?*}) \rightarrow \text{PREP (PRO SUBCAT dir)} (S \text{ ?*})$
- ▷ $(S \text{ ?*}) \rightarrow (\text{PART SUBCAT none}) (S \text{ ?*})$
- ▷ $(S \text{ INV } - \text{AGR ?a VFORM ?v}) \rightarrow (NP \text{ AGR ?a}) (VP \text{ AGR ?a VFORM ?v})$
- ▷ $(S \text{ INV } + \text{AGR ?a VFORM ?v}) \rightarrow (VP \text{ AGR ?a VFORM ?v}) (NP \text{ AGR ?a})$
- ▷ $(S \text{ GAP } + (NP \text{ GAP } + \text{AGR ?a})) \rightarrow$
 $(\text{ADVP SUBCAT time}) (NP \text{ CAS D}) (NP \text{ GAP } \text{AGR ?a})$
 $(V \text{ SUBCAT } _advp:time_np:objdir_advp_np:objdir \text{ AGR ?a VFORM ?v})$
 $(\text{ADVP SUBCAT form}) (NP \text{ CAS A})$
- ▷ $(S \text{ INV } - \text{AGR ?a VFORM ?v}) \rightarrow (\text{ADVP SUBCAT time})$
 $(V \text{ SUBCAT } _avdp:time_np:objdir \text{ AGR ?a VFORM ?v}) (NP \text{ AGR ?a}) (NP \text{ CAS A})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow (V \text{ AGR ?a VFORM ?v})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow$
 $(V \text{ SUBCAT } _vp:inf \text{ AGR ?a VFORM ?v}) (VP \text{ AGR ?a VFORM inf})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow (V \text{ SUBCAT } _np \text{ AGR ?a VFORM ?v}) (NP \text{ CAS N})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow$
 $(V \text{ SUBCAT } _np_adv:form \text{ AGR ?a VFORM ?v}) (NP \text{ CAS N}) (\text{ADVP SUBCAT form})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow (V \text{ SUBCAT } _np:objdir \text{ AGR ?a VFORM ?v}) (NP \text{ CAS A})$
- ▷ $(VP \text{ AGR ?a VFORM ?v}) \rightarrow$
 $(V \text{ SUBCAT } _avdp:time_np \text{ AGR ?a VFORM ?v}) (\text{ADVP SUBCAT time}) NP$

- ▷ (VP **AGR** ?a **VFORM** ?v) →
(V **SUBCAT** _avdp:form **AGR** ?a **VFORM** ?v) (**ADVP SUBCAT** form)
- ▷ (VP **AGR** ?a **VFORM** ?v) → (V **SUBCAT** _pp:loc_avdp:form **AGR** ?a **VFORM** ?v)
(PP **SUBCAT** loc **AGR** ?a) (**ADVP SUBCAT** form)
- ▷ (VP **AGR** ?a **VFORM** ?v) → (V **SUBCAT** _pp **AGR** ?a **VFORM** ?v) (PP **AGR** ?a)
- ▷ (VP **AGR** ?a **VFORM** ?v) → (PP **AGR** ?a) (V **SUBCAT** _pp **AGR** ?a **VFORM** ?v)
- ▷ (VP **AGR** ?a **VFORM** ?v) →
(V **SUBCAT** _pp:loc **AGR** ?a **VFORM** ?v) (PP **SUBCAT** loc **AGR** ?a)
- ▷ (VP **AGR** ?a **VFORM** ?v) →
(V **SUBCAT** _np_pp:loc **AGR** ?a **VFORM** ?v) (NP **CAS** A) (PP **SUBCAT** loc)
- ▷ (VP **AGR** ?a **VFORM** ?v) →
(V _np:objindir_np:objdir **AGR** ?a **VFORM** ?v) (NP **CAS** D) (NP **CAS** A)
- ▷ (VP **AGR** ?a **VFORM** ?v zvrat1+) → (NP **CAS** D) (PRO **SUBCAT** zvrat1)
(NP **CAS** A) (V _np:objindir_np:objdir **AGR** ?a **VFORM** ?v zvrat1—)
- ▷ (V **VFORM** ?v zvrat1+) → (V **VFORM** ?v zvrat1—) (PRO **SUBCAT** zvrat1)
- ▷ (V **VFORM** ?v zvrat1+) → (PRO **SUBCAT** zvrat1) (V **VFORM** ?v zvrat1—)
- ▷ (V **VFORM** ?v zvrat2+) → (V **VFORM** ?v zvrat2—) (PRO **SUBCAT** zvrat2)
- ▷ (V **VFORM** ?v zvrat2+) → (PRO **SUBCAT** zvrat2) (V **VFORM** ?v zvrat2—)
- ▷ (**ADVP** ?*) → **PART** (**ADVP** ?*)
- ▷ (**ADVP SUBCAT** time) → (**ADV SUBCAT** time)
- ▷ (**ADVP SUBCAT** time) → (**PREP CAS** A **SUBCAT** time) (NP **CAS** A)
- ▷ (**ADVP SUBCAT** form) → (**PREP AGR** ? a **SUBCAT** form) (NP **AGR** ? a)
- ▷ (PP ?*) → (**PART SUBCAT** none) (PP ?*)
- ▷ (PP **CAS** ?p) → (**PREP CAS** ?p) (NP **CAS** ?p)
- ▷ (PP ?*) → (PP ?*) **CONJP** (PP ?*)
- ▷ (PP **SUBCAT** loc) → (PRO **SUBCAT** refer:loc)
- ▷ (PP **SUBCAT** loc) → (**ADV SUBCAT** loc)
- ▷ (NP **AGR** ? a) → (**ADJP AGR** ?a) (N **AGR** ?a)
- ▷ (NP **GAP** + (N **AGR** ?a)) → (**ADJP AGR** ?a) (N **GAP AGR** ?a)
- ▷ (NP ?*) → (NP **AGR** ? a) **CONJP** (NP ?*)
- ▷ (NP **AGR** ? a) → (N **AGR** ? a)
- ▷ (NP **AGR** ? a) → (PRO **AGR** ? a)
- ▷ (**ADJP** ?*) → (PRO **SUBCAT** poss **AGR** ?a) (**ADJP** ?*)
- ▷ (**ADJP** ?*) → (**ADJP** ?*) (**ADJP** ?*)
- ▷ (**ADJP AGR** ?a) → (**ADJ AGR** ?a)
- ▷ (**ADJP AGR** ?a) → (NUM **AGR** ?a)
- ▷ **CONJP** → (**CONJ SUBCAT** agr) | , | — |

★ Lexicon

- cesta: **CAT N**
- dieťa: **CAT N**
- dievča: **CAT N**
- drevorubač: **CAT N**
- hrach: **CAT N**
- matka: **CAT N**
- muž: **CAT N**
- obed: **CAT N**
- pivnica: **CAT N**
- plece: **CAT N**
- sestra: **CAT N**
- šošovica: **CAT N**
- ráno: **CAT N**
- starec: **CAT N**
- žena: **CAT N**
- dobrá: **CAT ADJ**
- múdre: **CAT ADJ**
- najmilšie: **CAT ADJ SUBCAT grad3**
- poslušná: **CAT ADJ**
- roztárané: **CAT ADJ**
- rozvážne: **CAT ADJ**
- väčší: **CAT ADJ SUBCAT grad2**
- ho: **CAT PRO SUBCAT dir CAS G**
- ja: **CAT PRO SUBCAT dir**
- jej: **CAT PRO SUBCAT poss**
- naň: **CAT PRO SUBCAT refer:loc CAS A**
- on: **CAT PRO SUBCAT dir**
- ona: **CAT PRO SUBCAT dir**
- sa: **CAT PRO SUBCAT zvrat1**
- si: **CAT PRO SUBCAT zvrat2**
- svoje: **CAT PRO SUBCAT poss**
- toho: **CAT PRO SUBCAT dir CAS G**

- tretie: **CAT NUM**
- byť: **CAT V SUBCAT** _np _advp:time_np _np_adv:form
- bývať: **CAT V SUBCAT** _advp:time_np _np _adv:form
- ísť: **CAT V SUBCAT** _pp:loc _advp:form _pp:loc_adv:form
- mať: **CAT V SUBCAT** _vp:inf
- nebáť: **CAT V VFORM** zvrat1+ **SUBCAT** _pp
- nevidieť: **CAT V VFORM** zvrat1+ **SUBCAT** _np:objindir_np:objdir
- nezablúdiť: **CAT V**
- porozsýpať: **CAT V SUBCAT** _np:objdir
- poslať: **CAT SUBCAT** _advp:time_np:objindir _advp_np:objdir
- posvietiť: **CAT V VFORM** zvrat2+ **SUBCAT** _pp:loc
- potriať: **CAT V SUBCAT** _np_pp:loc
- povedať: **CAT V SUBCAT** _advp:time_np:objdir
- prísť: **CAT V SUBCAT** _pp
- spustiť: **CAT V SUBCAT** _np_pp:loc _pp:loc
- ukázať: **CAT V SUBCAT** _np:objindir_np:objdir _np:objdir _np:objindir
- vravieť: **CAT V SUBCAT** _np:objindir_np:objdir _np:objdir _np:objindir
- vziať: **CAT V SUBCAT** _np:objdir
- zaspáť: **CAT V**
- dnes: **ADV SUBCAT** time
- kadetade: **ADV SUBCAT** loc
- vždy: **ADV SUBCAT** time
- ako: **CAT PREP SUBCAT** form
- do: **CAT PREP CAS** A
- na: **CAT PREP SUBCAT** time **CAS** A L
- o: **CAT PREP SUBCAT** form **CAS** A L
- okrem: **CAT PREP**
- po: **CAT PREP SUBCAT** form **CAS** A
- s: **CAT PREP SUBCAT** form **CAS** I
- za: **CAT PREP CAS** A I
- aj: **CAT PART SUBCAT** none
- iste: **CAT PART SUBCAT** none
- keď: **CAT PART SUBCAT** time
- nie: **CAT PART SUBCAT** none