Rudolf Rosa
rosa@ufal.mff.cuni.cz

Depfix:

Automatic post-editing of phrase-based machine translation outputs

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Monday seminar, 14th October 2013

Outline

- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- evaluation
- parsing of SMT outputs (MSTperl parser)

Motivation: Translation of negation

http://thedrivenclass.com/blog-detail/32876

Motivation: Errors in negation

- These **are not** actually errors.

Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.

Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.

Motivation: Errors in negation

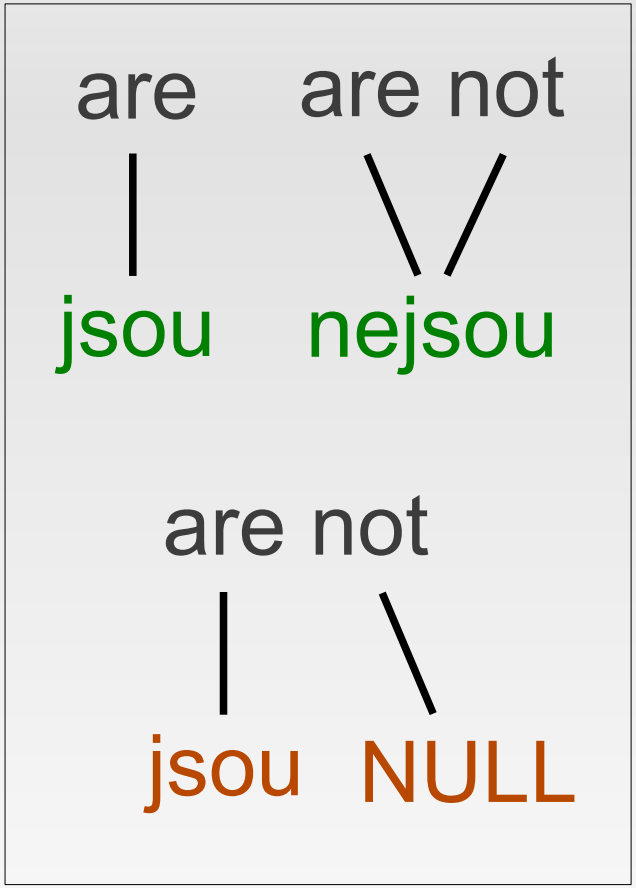
- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.
- I **would not cheat** on you.

Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.
- I **would not cheat** on you.
  - Moses: Já **bych** tě **podváděl**.
  - Gloss: I **would cheat** on you.
  - Ref.: Já **bych** tě **nepodváděl**.

Motivation: Errors in negation

- These **are not** actually errors.
  - Moses: **Jsou** to vlastně chyby.
  - Gloss: These **are** actually errors.
  - Ref.: **Nejsou** to vlastně chyby.
- I **would not cheat** on you.
  - Moses: Já **bych** tě **podváděl**.
  - Gloss: I **would cheat** on you.
  - Ref.: Já **bych** tě **nepodváděl**.



- some phenomena hard to get right with PBSMT

Simple to fix?

- there is a negation in the source
  - These are **not** actually errors

Simple to fix?

- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby

Simple to fix?

- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby

Simple to fix?

- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby
- add the negative prefix (ne-) into the target
  - **Ne**jsou to vlastně chyby

Simple to fix?

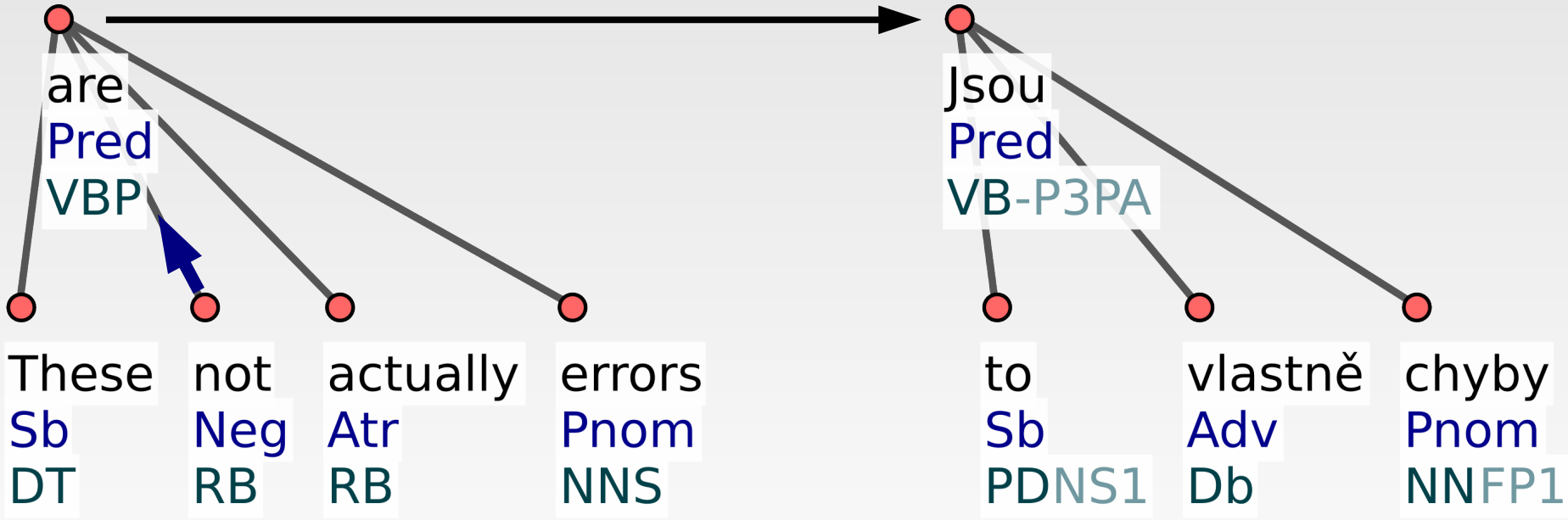
- there is a negation in the source
  - These are **not** actually errors
- there is no negation in the target
  - Jsou to vlastně chyby
- add the negative prefix (ne-) into the target
  - **Ne**jsou to vlastně chyby
- such a simple approach might be sufficient
- but usually useful to use some more NLP tools

Part-of-speech tagger

- run a POS tagger on the target sentence
  - Jsou to vlastně chyby
  - verb pronoun adverb noun
- a good heuristic: negate the (finite) verb!
  - **Ne**jsou to vlastně chyby
  - verb pronoun adverb noun
- fine-grained tags may even mark the negation
  - jsou VB-P---3P-**A**A---
  - nejsou VB-P---3P-**N**A---

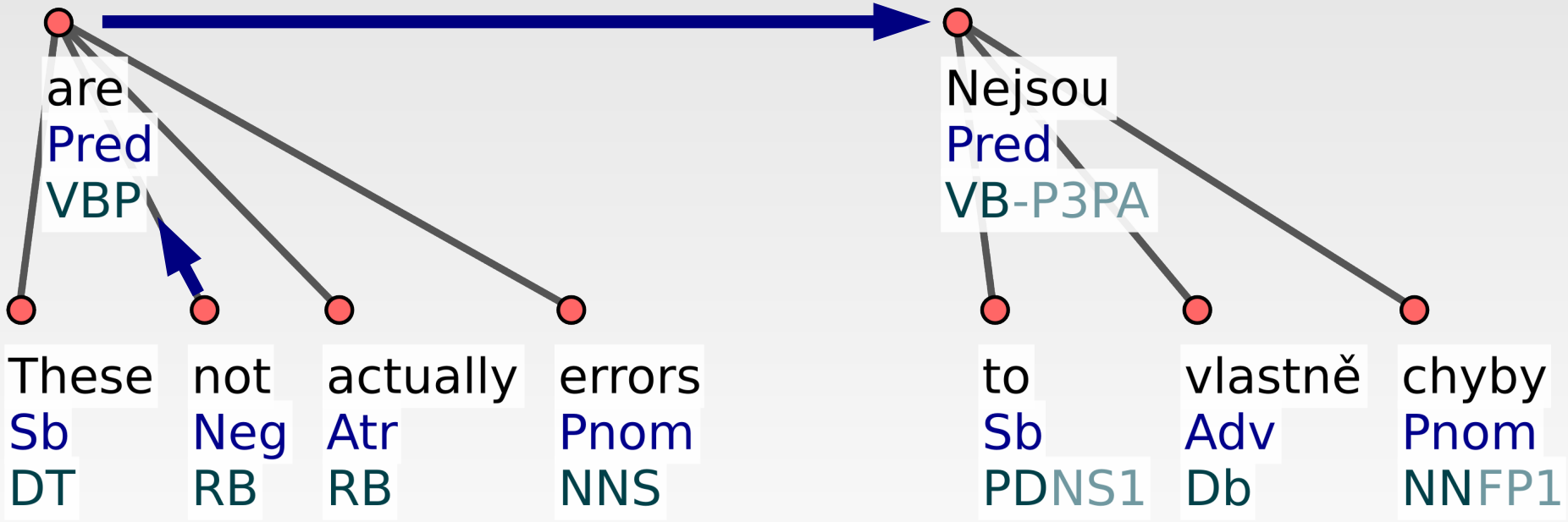
Dependency parser

- parse both source and target
- project negation through word alignment



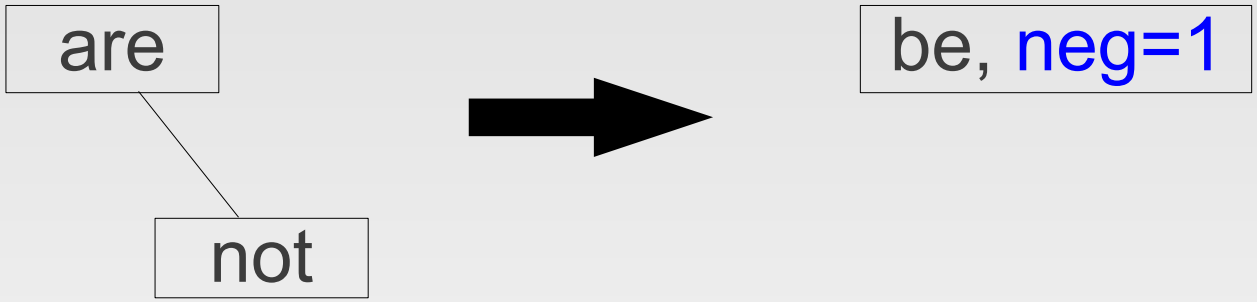
Dependency parser

- parse both source and target
- project negation through word alignment



Deep syntactic analysis

- auxiliary nodes collapsed into values of attributes on parent nodes



- abstract from various ways of expressing negation (not, no, un-, in-,...)
  - all marked by `neg=1` on the lexical node

Morphological generator

- `form = generate(word, morphological features)`

Morphological generator

- `form = generate(lemma, tag)`

Morphological generator

- `form = generate(lemma, tag)`
- instead of: `new_form = 'ne' + form`
  - `'nejsou' = 'ne' + 'jsou'`

Morphological generator

- `form = generate(lemma, tag)`
- instead of: `new_form = 'ne' + form`
  - `'nejsou' = 'ne' + 'jsou'`
- use the more sophisticated: `new_form = generate(lemma(form), negate(tag))`
  - `'nejsou' = generate(lemma('jsou'), negate('VB-P---3P-AA---'))`

Morphological generator

- `form = generate(lemma, tag)`
- instead of: `new_form = 'ne' + form`
  - `'nejsou' = 'ne' + 'jsou'`
- use the more sophisticated: `new_form = generate(lemma(form), negate(tag))`
  - `'nejsou' = generate(lemma('jsou'), negate('VB-P---3P-AA---'))`
  - `'nejsou' = generate( 'být' , 'VB-P---3P-NA---' )`

Depfix pipeline

- input
  - English sentence
  - its Czech translation, provided by an SMT system
- m-layer analysis and fixes (tagger, word-aligner)
- a-layer analysis and fixes (dependency parser)
- t-layer analysis and fixes (t-tree analyzer)
- output (morphological generator)
  - the Czech translation, corrected

Outline

- ✔ translation of negation (and its correction)
  - ✔ motivation
  - ✔ the fixing pipeline
- ➔ **analysis and corrections**
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- evaluation
- parsing of SMT outputs (MSTperl parser)

Outline

- ✔ translation of negation (and its correction)
  - ✔ motivation
  - ✔ the fixing pipeline
- ➔ analysis and corrections
  - ➔ **m-layer** (lemmas, tags, word-alignment)
  - ▪ a-layer (dependency trees, analytical functions)
  - ▪ t-layer (“tecto-trees”, formemes, grammatemes)
- ▪ evaluation
- ▪ parsing of SMT outputs (MSTperl parser)

M-layer

- analysis: lemmas, tags, word-alignment
  - +adding missing alignment links by string similarity
- corrections:
  - tokenization projection
  - morphological number projection
  - source-aware truecasing
  - vocalisation of prepositions

Tokenization projection

Source	Michèle **Alliot-Marie** had sent a communication
Moses	Michèle **Alliot - Marie** poslal sdělení
Gloss	Michèle Alliot - Marie sent$_{masc}$ a communication
Depfix	Michèle **Alliot-Marie** poslala sdělení
Gloss	Michèle Alliot-Marie sent$_{fem}$ a communication

Source-aware truecasing

Source	the director of the best hotel in **Pec**, Karel **Rada**.
Moses	ředitel nejlepší hotel v **peci**, Karel **rada**.
Gloss	the director of the best hotel in the **oven**, Karel **advice**.
Depfix	ředitel nejlepšího hotelu v **Peci**, Karel **Rada**.
Gloss	the director of the best hotel in **Pec**$_{\textbf{town}}$, Karel **Rada**$_{\textbf{surname}}$.

Vocalisation of prepositions

Source	The work being done by experts from three institutions
Moses	Práce odborníků **z** tří institucí
Gloss	Work by experts from three institutions
Depfix	Práce odborníků **ze** tří institucí
Gloss	Work by experts from three institutions

Outline

- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - **a-layer** (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- evaluation
- parsing of SMT outputs (MSTperl parser)

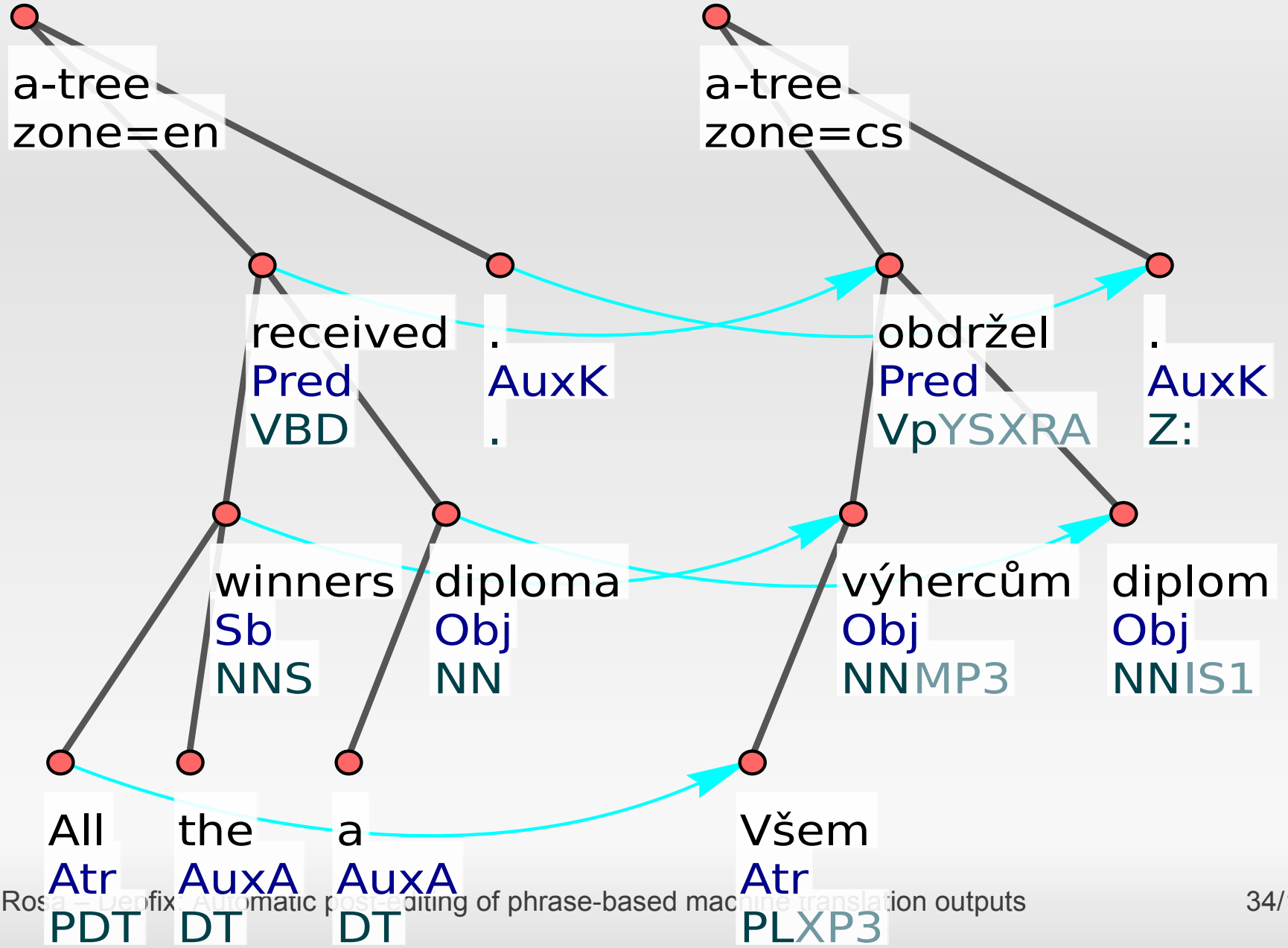
A-layer

- analysis: dependency trees, analytical functions
  - for Czech: MSTperl (adapted MST parser)
  - + fixing prepositions without children, auxiliary verbs with children, AuxT/AuxR...
- morphological agreement fixes:
  - preposition-noun, noun-adjective, subject-predicate...
- fixes of transfer of meaning to morphology
  - possessives, passives, subject...

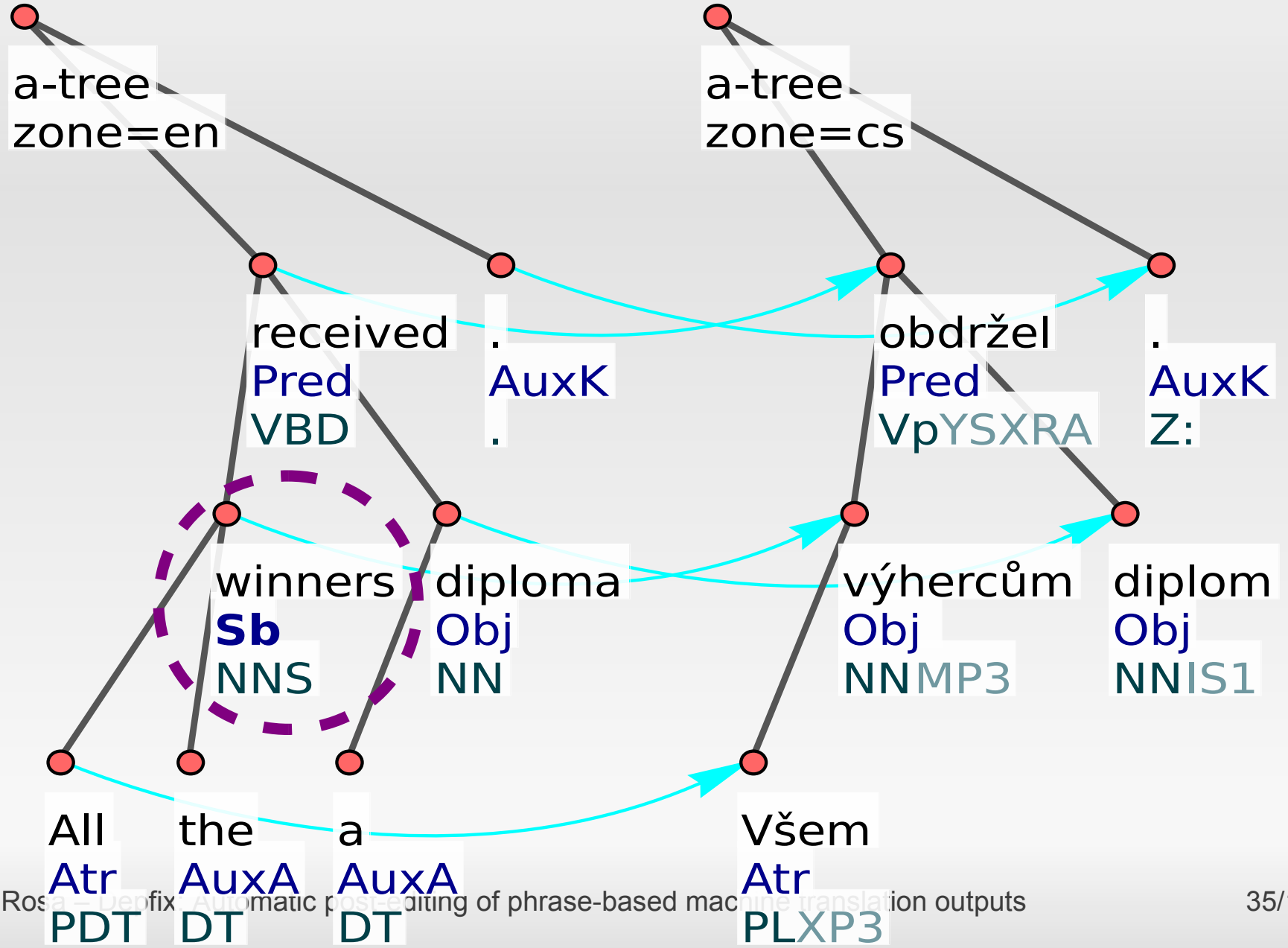
A-layer motivation

- Source:
  - *All the winners received a diploma.*
- Moses:
  - *Všem výhercům obdržel diplom.*
  - *To all the winners he received a diploma.*
- Depfix:
  - *Všichni výherci obdrželi diplom.*
  - *All the winners received a diploma.*

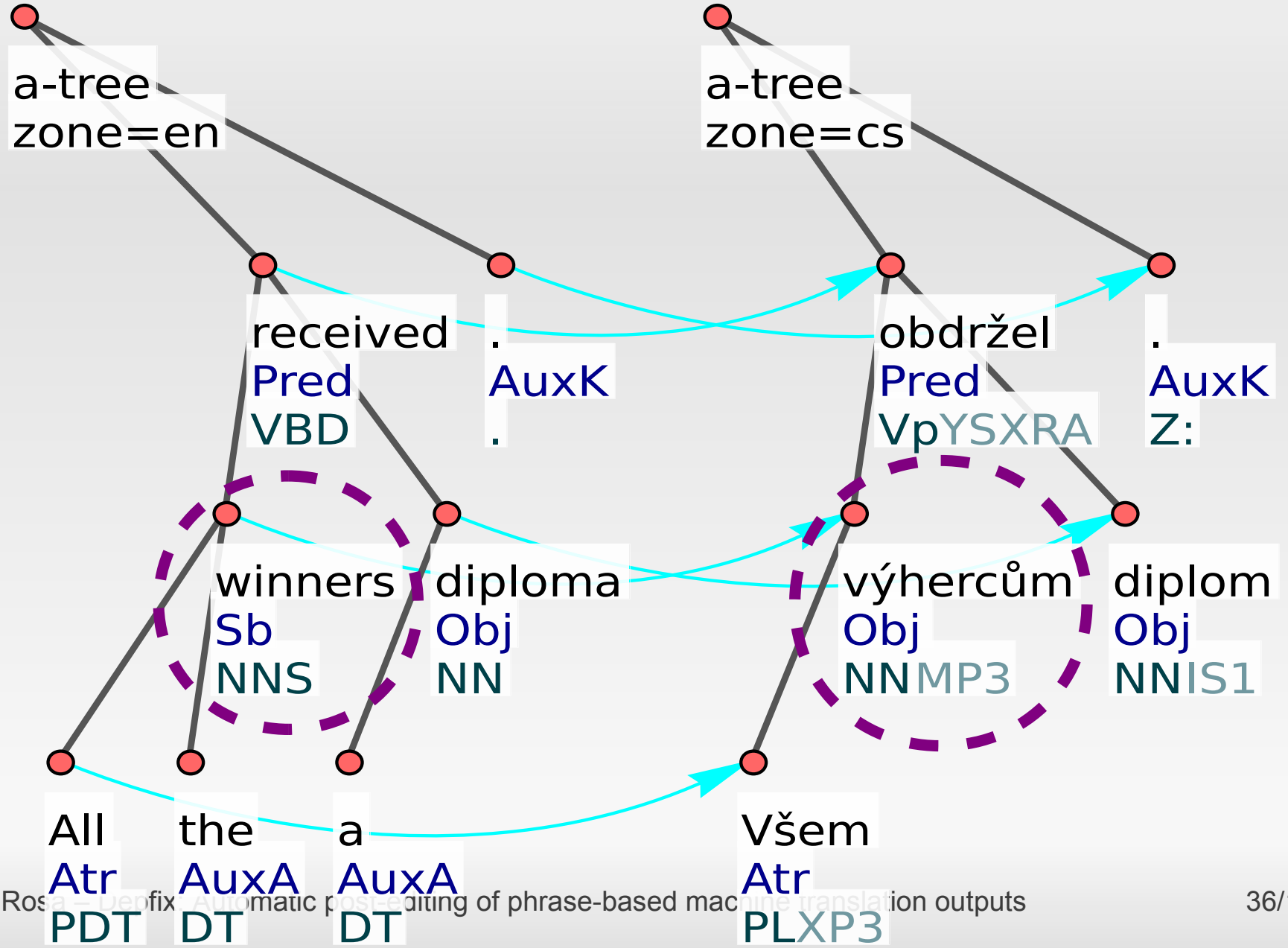
Všem výhercům obdržel diplom.



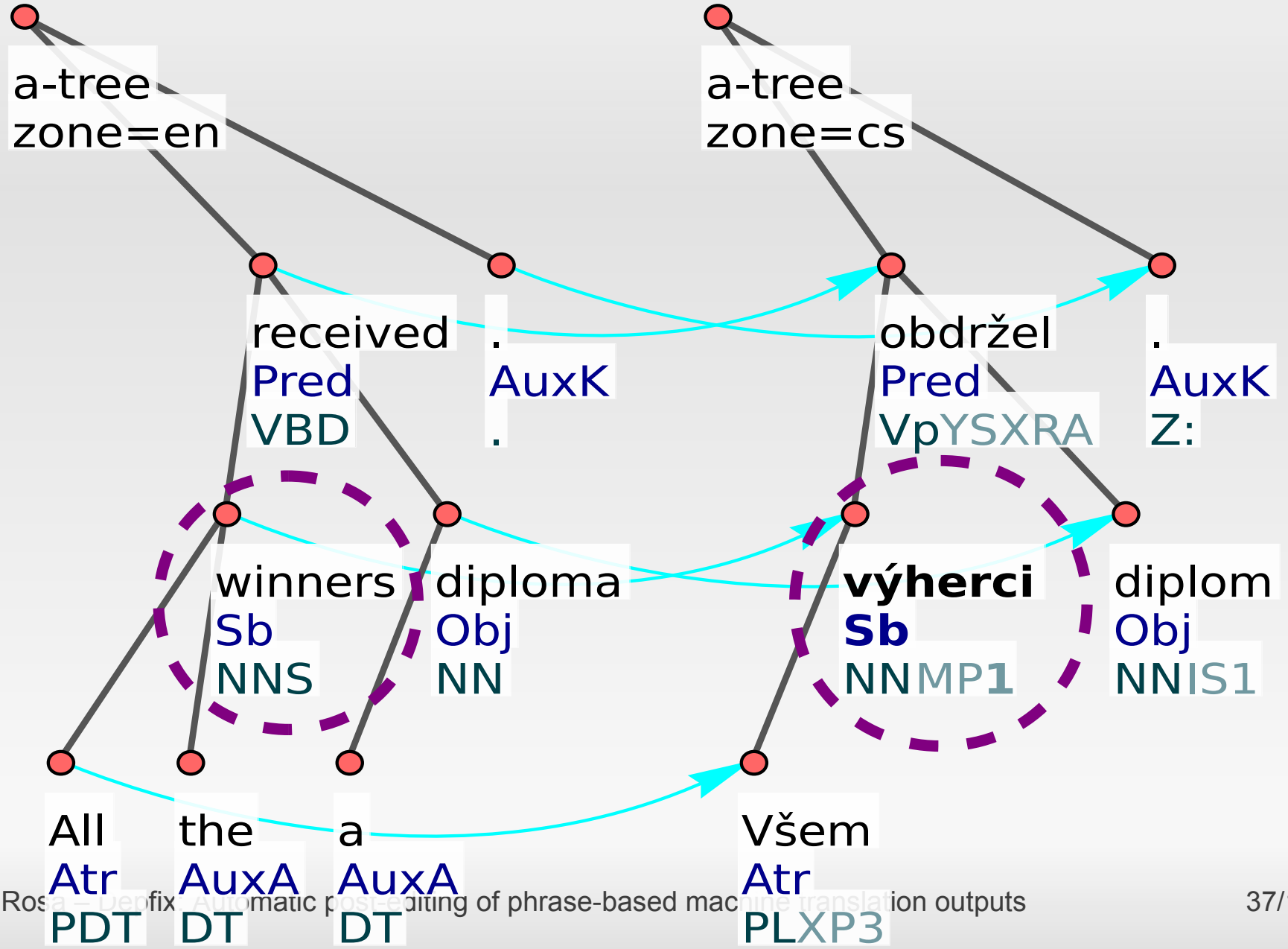
Transfer of meaning: subject



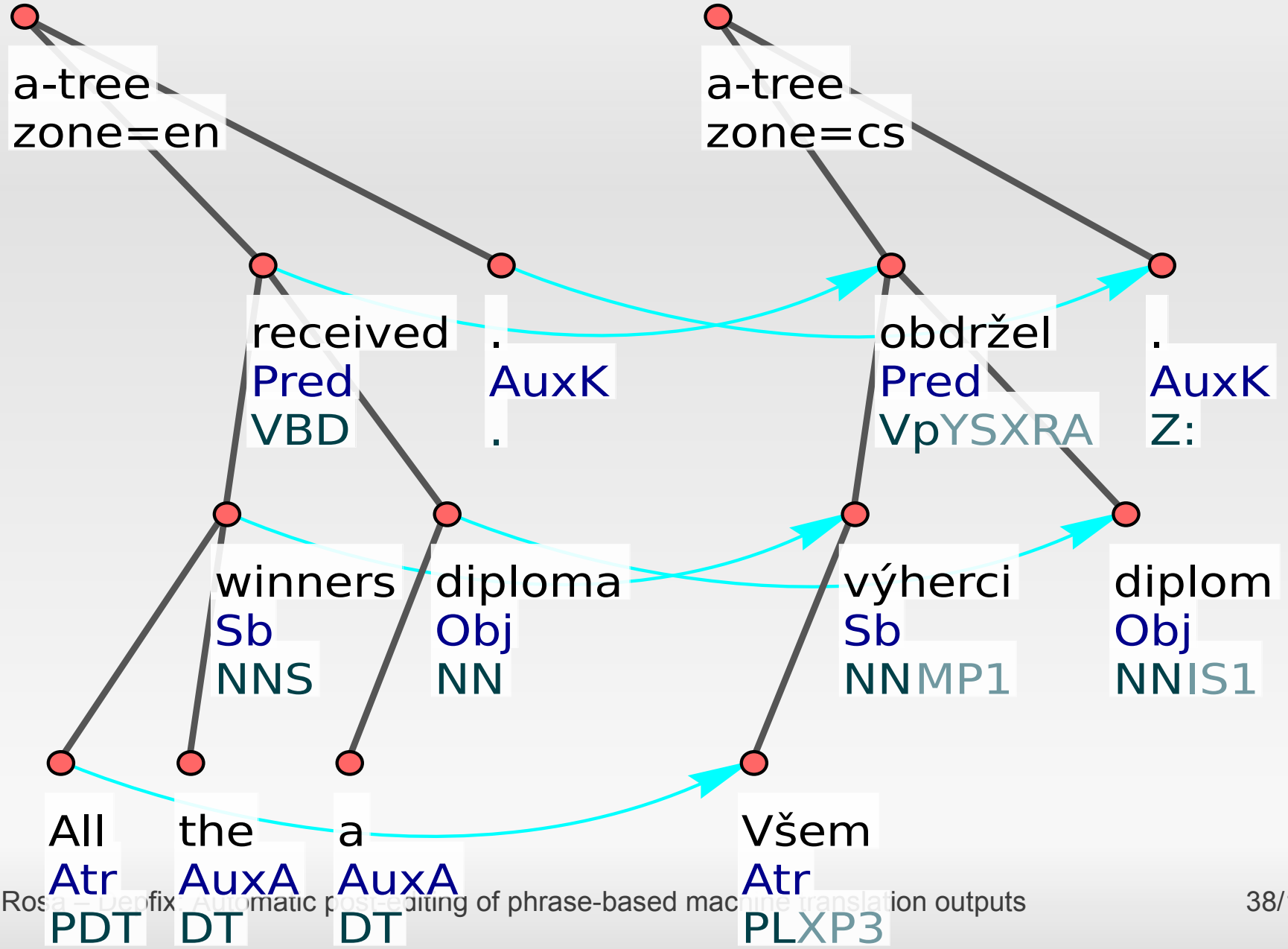
Transfer of meaning: subject



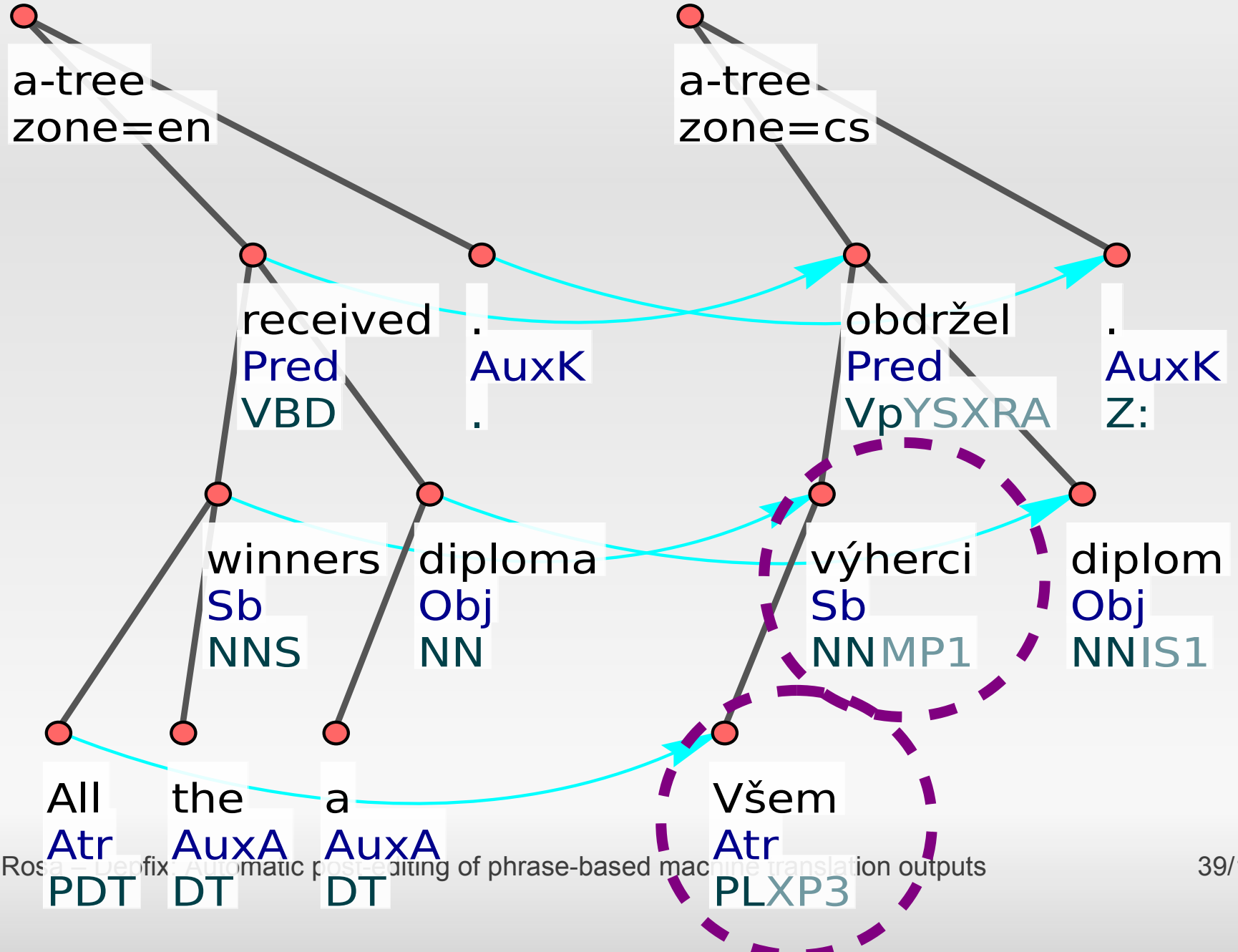
Subject → nominative



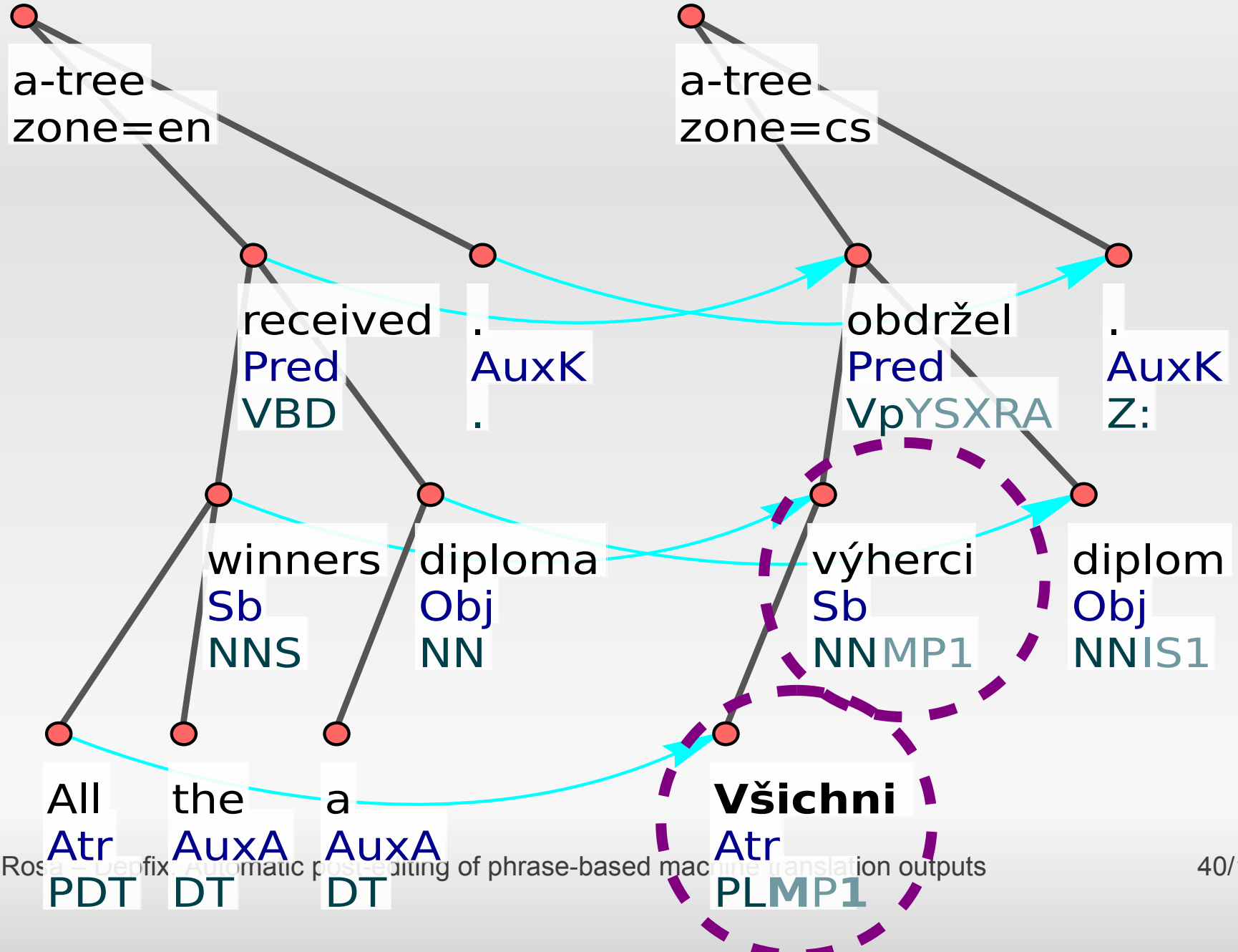
Všem výherci obdržel diplom.



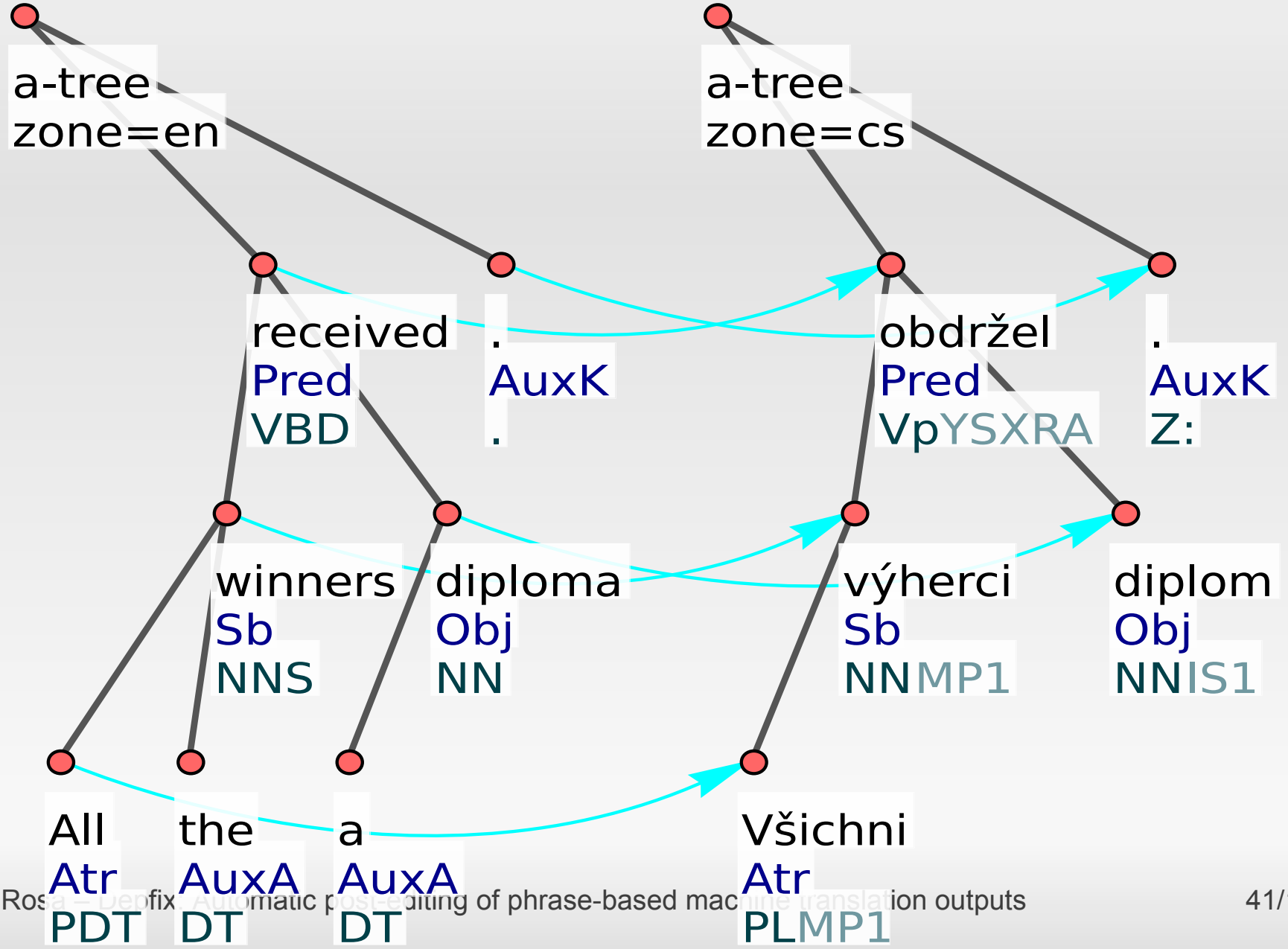
Noun-adjective agreement



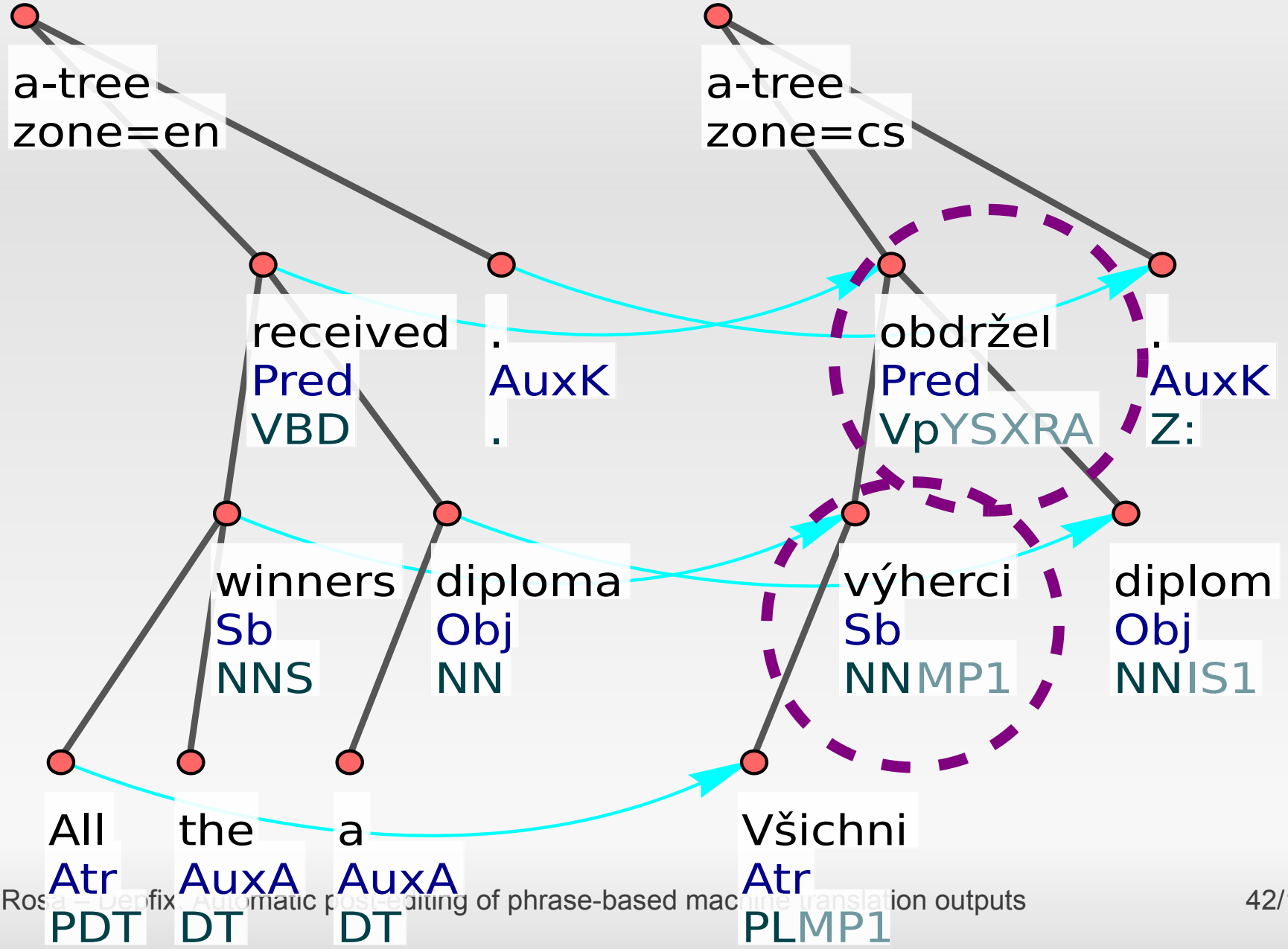
Agreement: gender, case (number)



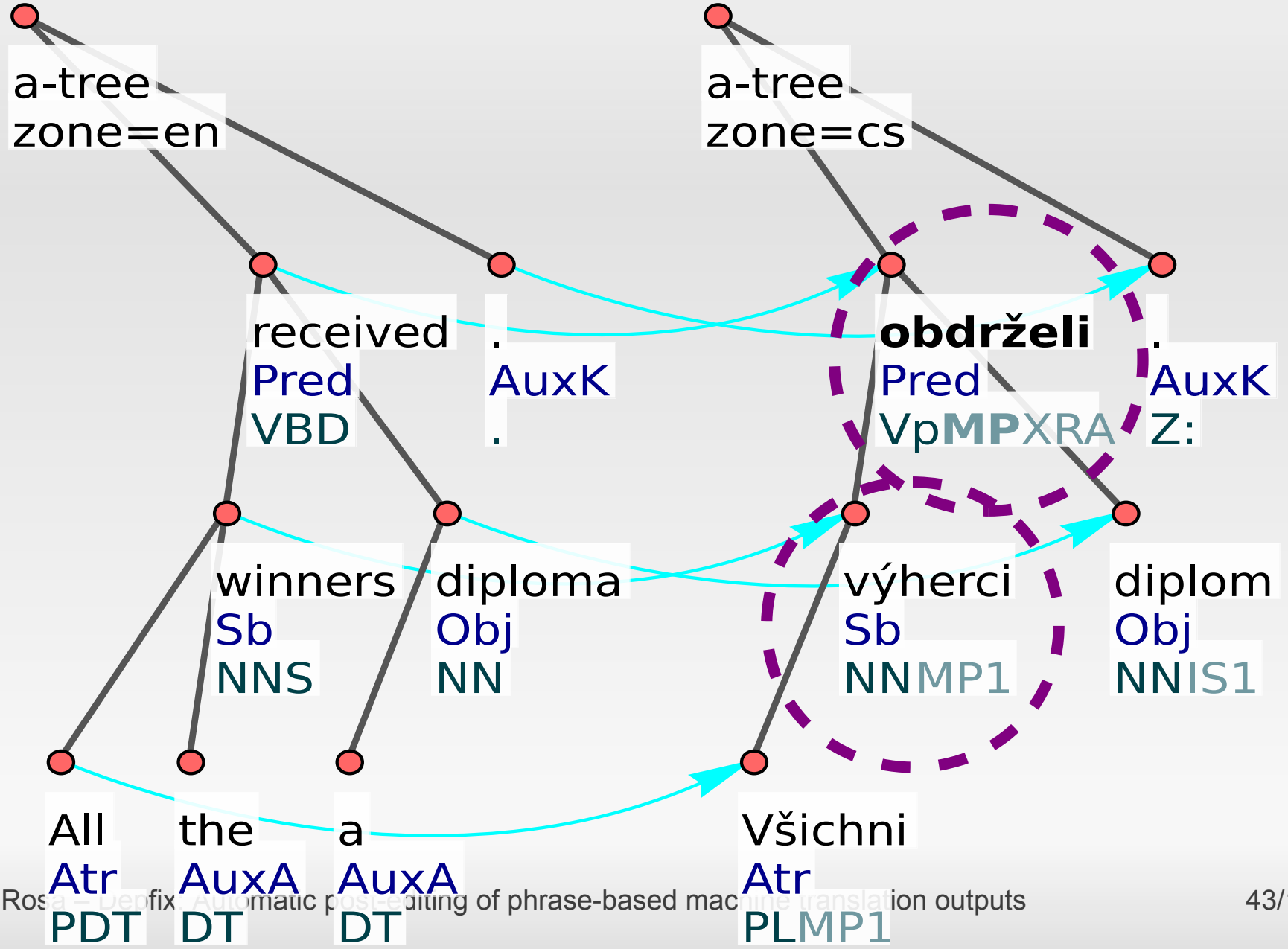
Všichni výherci obdržel diplom.



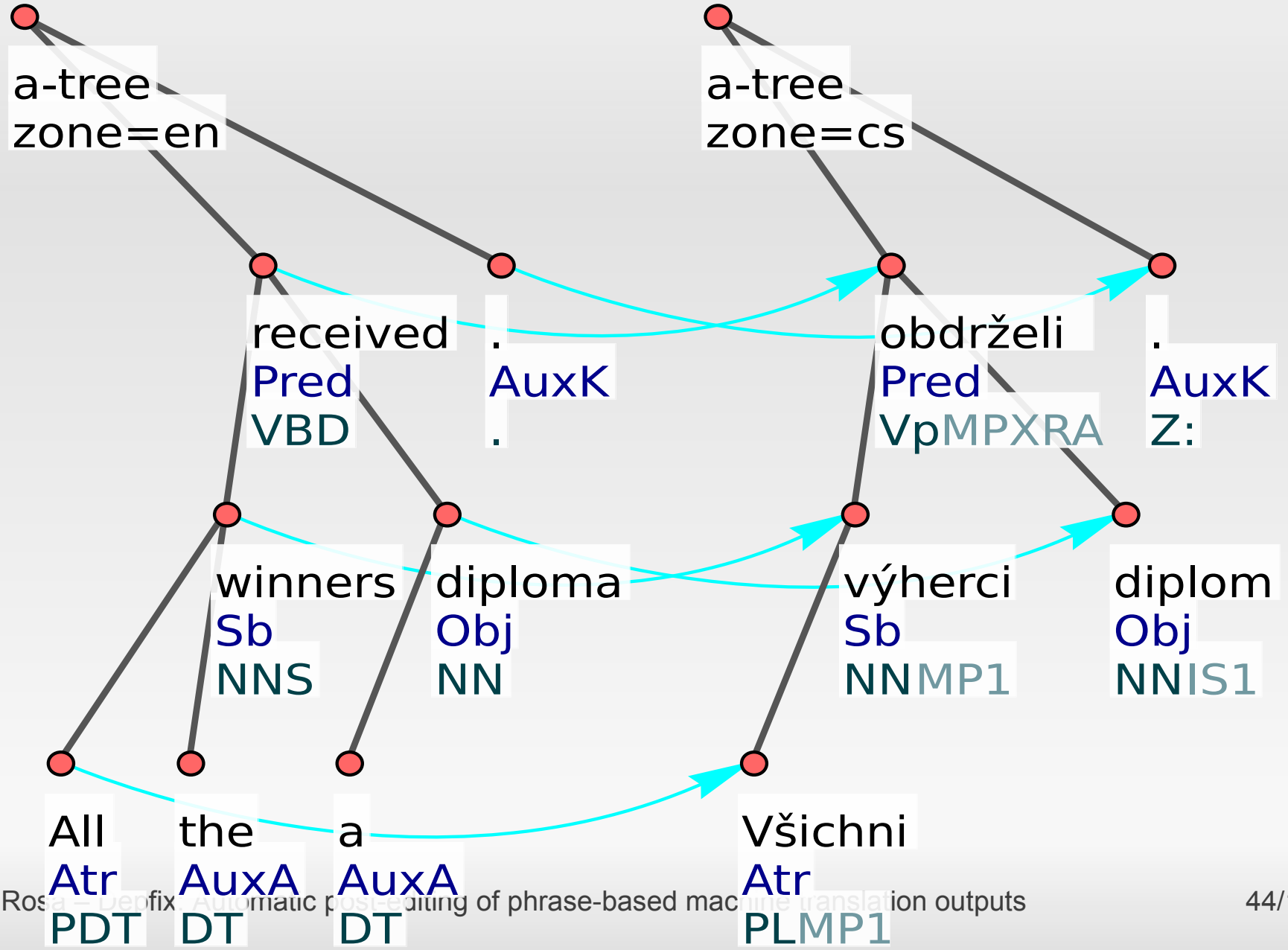
Subject-predicate agreement



Agreement: gender, num (person)



Všichni výherci obdrželi diplom.



Preposition-noun agreement

Source	It is a story about sport, race relations, and Nelson Mandela.
Moses	Je to příběh **o sportu**, rasových **vztahů**, a Nelson **Mandela**.
Gloss	It is a story **about**$_{\textbf{loc}}$ **sport**$_{\textbf{loc}}$, race **relations**$_{\textbf{gen}}$, and Nelson$_{\text{nom}}$ **Mandela**$_{\textbf{nom}}$.
Depfix	Je to příběh **o sportu**, rasových **vztazích**, a Nelson **Mandelovi**.
Gloss	It is a story **about**$_{\textbf{loc}}$ **sport**$_{\textbf{loc}}$, race **relations**$_{\textbf{loc}}$, and Nelson$_{\text{nom}}$ **Mandela**$_{\textbf{loc}}$.

Noun-adjective agreement

Source	this half-hearted increase will bear the same fruit
Moses	**tato polovičatá nárůst** bude nést stejné ovoce
Gloss	**this**$_{\text{fem}}$ **half-hearted**$_{\text{fem}}$ **increase**$_{\text{masc}}$ will bear the same fruit
Depfix	**tento polovičatý nárůst** bude nést stejné ovoce
Gloss	**this**$_{\text{masc}}$ **half-hearted**$_{\text{masc}}$ **increase**$_{\text{masc}}$ will bear the same fruit

Translation of 'of'

Source	unsustainable deficit level **of public finances**.
Moses	neudržitelná úroveň schodku **veřejné finance**.
Gloss	unsustainable deficit level **public finances**.
Depfix	neudržitelná úroveň schodku **veřejných financí**.
Gloss	unsustainable deficit level **of public finances**.

Translation of 's

Source	**Janota's** possible **continuation** in office will be the topic of Friday's meeting.
Moses	**Janota je** možné **pokračování** ve funkci bude tématem páteční schůze.
Gloss	**Janota is** possible **continuation** in office will be the topic of Friday's meeting.
Depfix	možné **pokračování Janoty** ve funkci bude tématem páteční schůze.
Gloss	possible **continuation of Janota** in office will be the topic of Friday's meeting.

Translation of Subject

Source	At a time when **Swiss voters**$_{\text{subj}}$ have called for a ban on the construction of minarets
Moses	V době, kdy **švýcarské voliče**$_{\text{acc}}$ vyzvali k zákazu výstavby minaretů
Gloss	At a time, when **Swiss voters were called** for a ban on the construction of minarets
Depfix	V době, kdy **švýcarští voliči**$_{\text{nom}}$ vyzvali k zákazu výstavby minaretů
Gloss	At a time, when **Swiss voters called** for a ban on the construction of minarets

Outline

- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - **t-layer** (“tecto-trees”, formemes, grammatemes)
- evaluation
- parsing of SMT outputs (MSTperl parser)

T-layer

- analysis: t-trees, formemes, grammatemes
  - + systematic analysis of English verb tenses
- rule-based fixes:
  - ✔ negation
  - verb tenses
  - subject pronoun dropping
- statistical fixes:
  - valency

Translation of verb tenses

Source	This **will bring** problems for whoever is in office
Moses	To **přináší** problémy pro každého, kdo je v kanceláři
Gloss	This **brings** problems for anyone who is in office
Depfix	To **bude přinášet** problémy pro každého, kdo je v kanceláři
Gloss	This **will bring** problems for anyone who is in office

Subject pronoun dropping

Source	I don't blame them.
Moses	**Já** se jim nedivím.
Gloss	I myself them don't-blame$_{\text{1st person sg}}$.
Depfix	Nedivím se jim.
Gloss	Don't-blame$_{\text{1st person sg}}$ myself them.

Correction of valency errors

- Source text in English:

  EU criticizes not only the Greek government

- Google Translate to Czech (6th Aug 2013):

  *EU kritizuje nejen **řecká vláda*** **nominative (subject)**

  - *Not only **the Greek government** criticizes EU*

- Post-editation by Deepfix:

  *EU kritizuje nejen **řeckou vládu*** **accusative (object)**

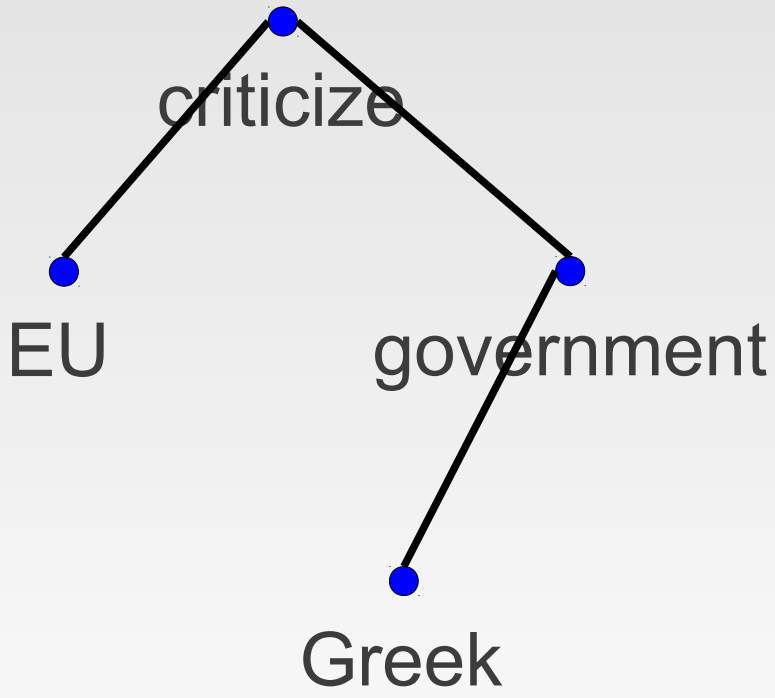
  - *EU criticizes not only **the Greek government***

Valency of *criticize* (*kritizovat*)

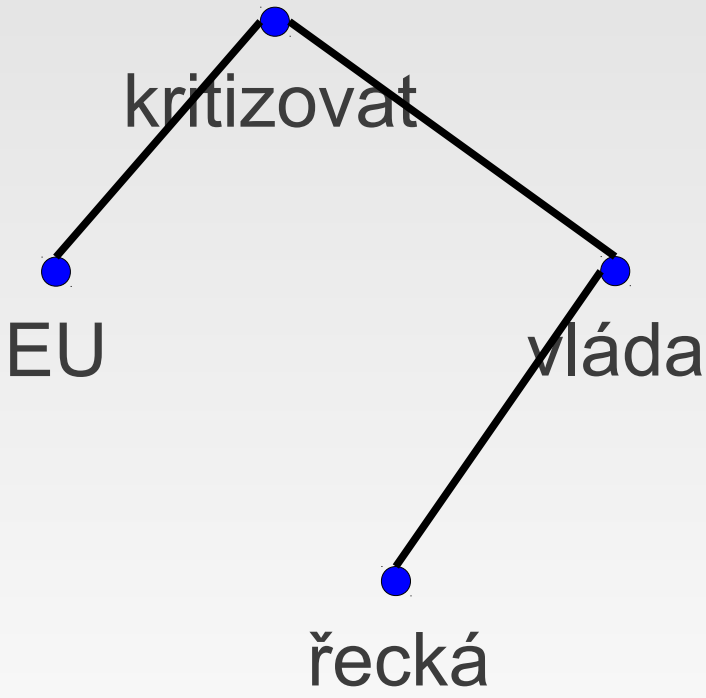
- *EU*$_{\text{subject}}$ *criticizes the Greek government*$_{\text{object}}$
- *EU*$_{\text{nominative}}$ *kritizuje řeckou vládu*$_{\text{accusative}}$
- a valency frame of a verb
  - subject *criticize* object (position)
  - nominative *kritizovat* accusative (cases)
- decomposition into head-argument pairs
  - (*to criticize, government*) ~ (*kritizovat, vládu*)
  - (*to criticize,* Object) ~ (*kritizovat,* accusative)

Deep syntactic dependency trees

EU criticizes
the Greek government



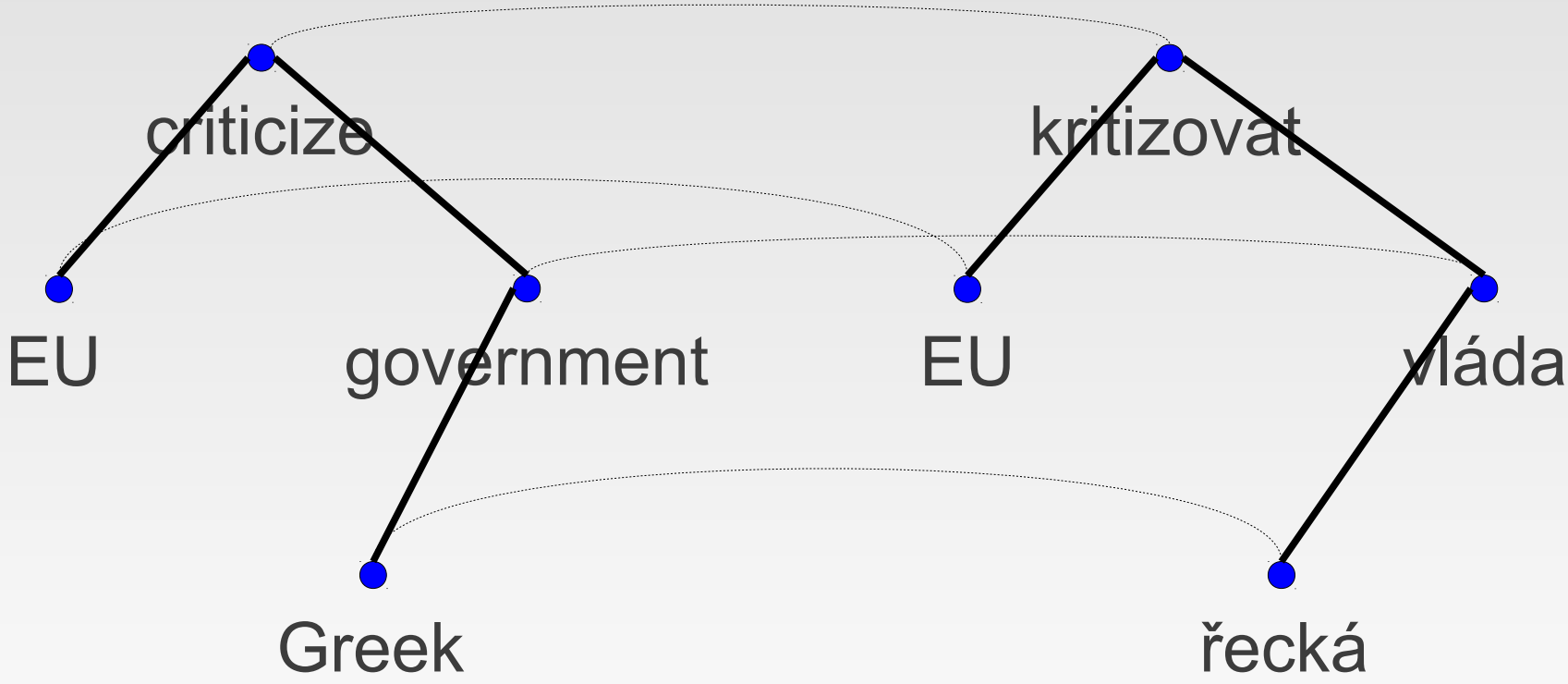
EU kritizuje
řecká vláda



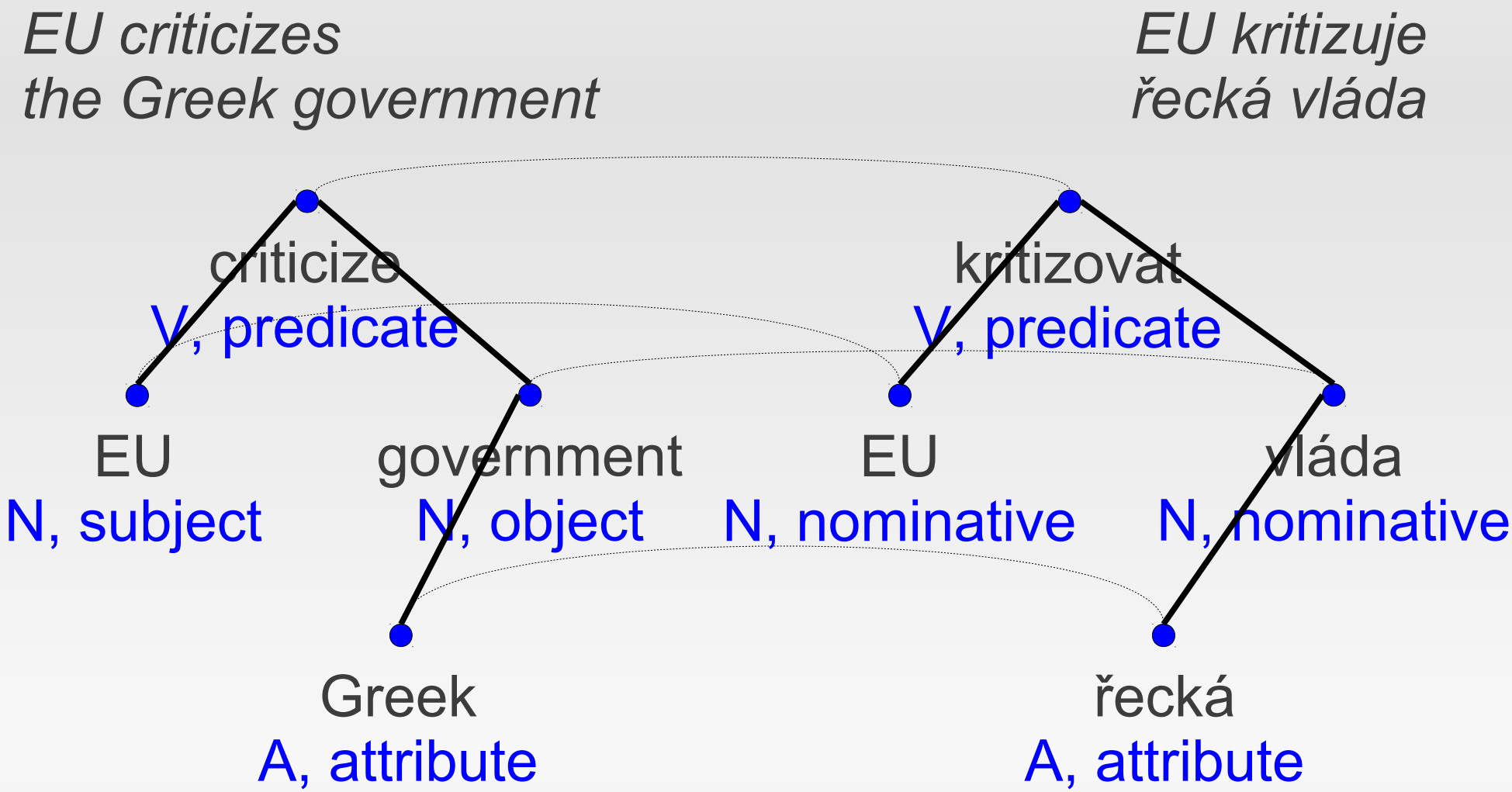
Deep syntactic dependency trees

EU criticizes
the Greek government

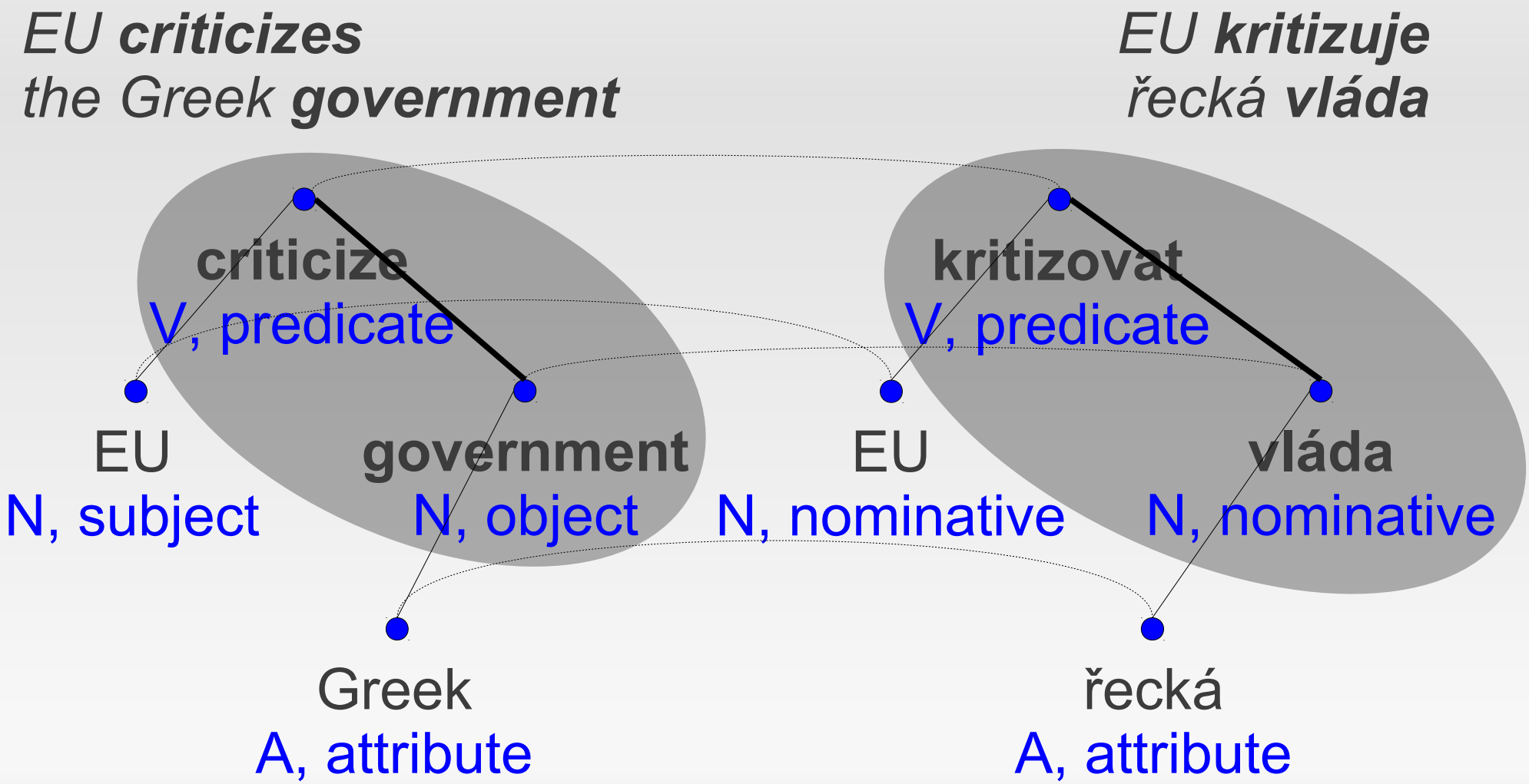
EU kritizuje
řecká vláda



Deep syntactic dependency trees

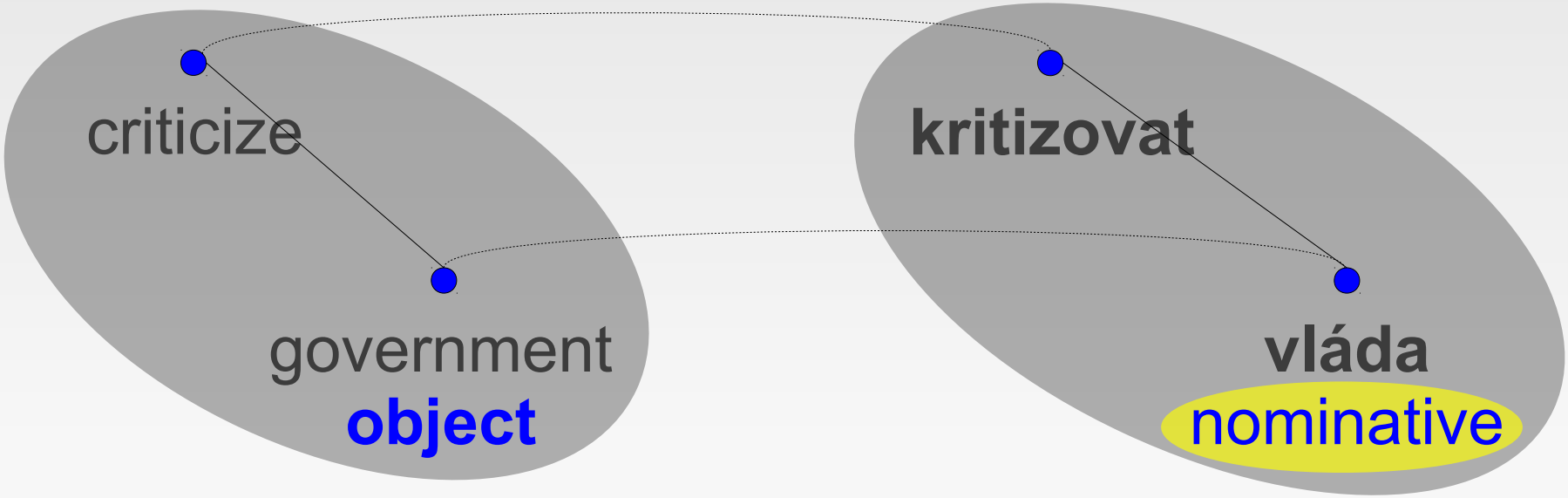


(head, arg) pair identification



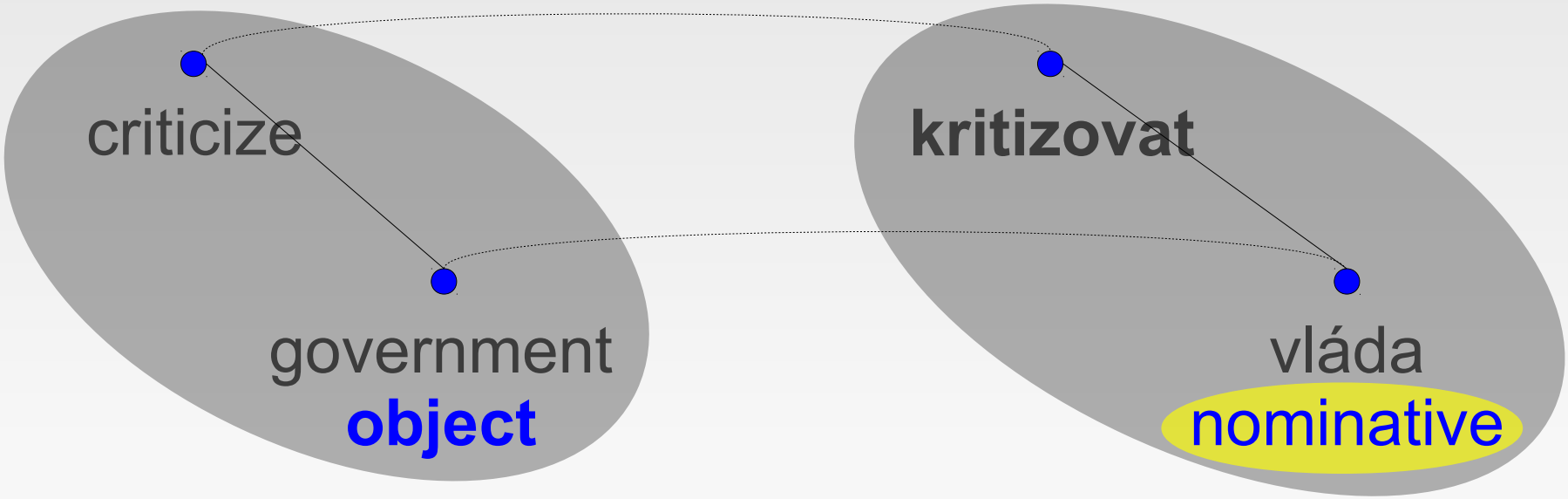
Valency models (FIX)

- $P(arg_{case} \mid head_{lemma}, English\_arg_{function})$
- $P(arg_{case} \mid head_{lemma}, English\_arg_{function}, arg_{lemma})$
- estimated from CzEng 1.0 (15M parallel stcs)



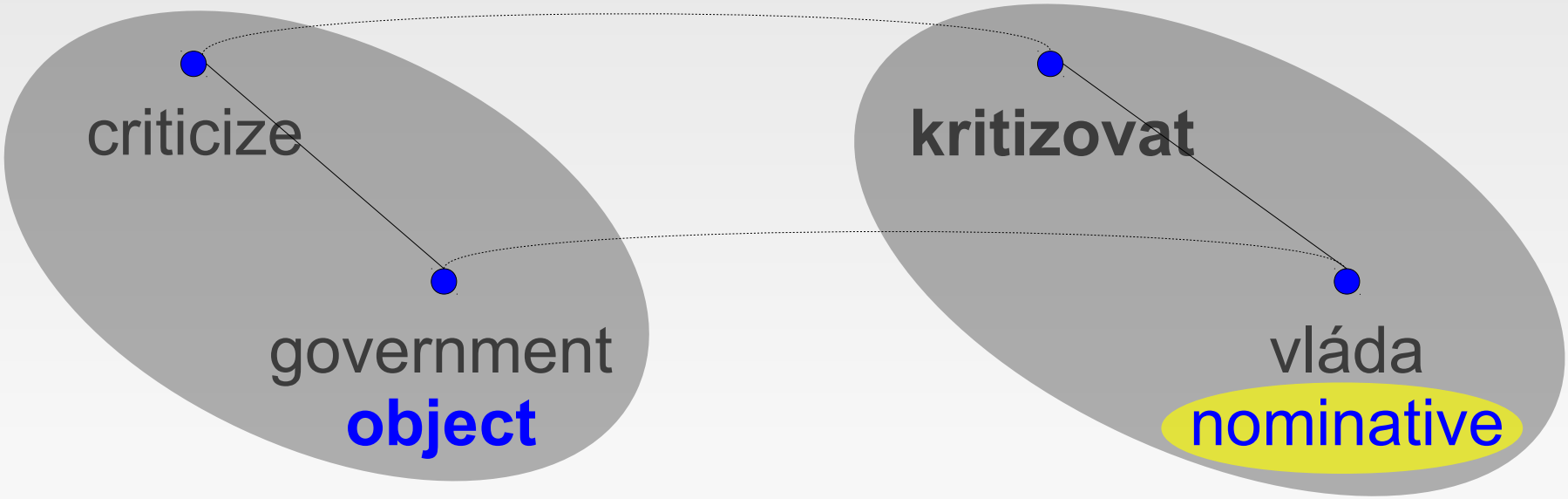
Argument case probabilities

- P(nominative | *kritizovat*, object) = 0.03
- P(accusative | *kritizovat*, object) = 0.80



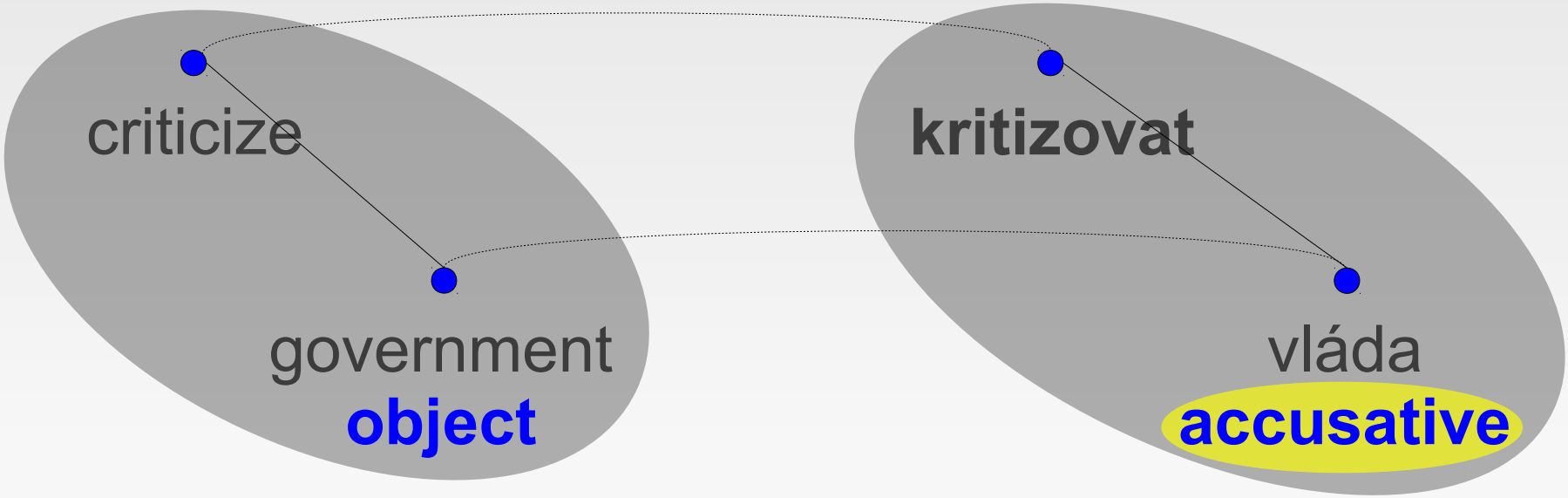
Argument case probabilities

- P(nominative | *kritizovat*, object) = 0.03
- P(accusative | *kritizovat*, object) = 0.80
- threshold: 0.55



Argument case correction

- P(nominative | *kritizovat*, object) = 0.03
- P(**accusative** | *kritizovat*, object) = **0.80**
- threshold: **0.55**



Sentence correction

- Statitical machine translation output:

  EU kritizuje nejen ***řecká***$_{\text{nominative}}$ ***vláda***$_{\text{nominative}}$

  - *Not only* ***the Greek government*** *criticizes EU*
- Valency model correction:

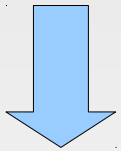
  EU kritizuje nejen ***řecká***$_{\text{nominative}}$ ***vládu***$_{\text{accusative}}$
- Agreement enforcement:

  EU kritizuje nejen ***řeckou***$_{\text{accusative}}$ ***vládu***$_{\text{accusative}}$

  - *EU criticizes not only* ***the Greek government***

Some interesting details

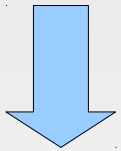
- the model actually works on formemes
  - functions (EN), cases (CS), **prepositions** (EN, CS)
  - in: *The government spends on the middle schools.*
  - SMT: *Vládá utrácí střední školy.*



    - (spend, on+X) → (utrácet, 4) P = 0.07
    - *The government destroys the middle schools.*
  - out: *Vládá utrácí za střední školy.*
    - (spend, on+X) → (utrácet, za+4) P = 0.89
    - *The government spends on the middle schools.*

Some interesting details

- the model actually works on formemes
  - functions (EN), cases (CS), **prepositions** (EN, CS)
  - in: *The government spends on the middle schools.*
  - SMT: *Vládá utrácí střední školy.*

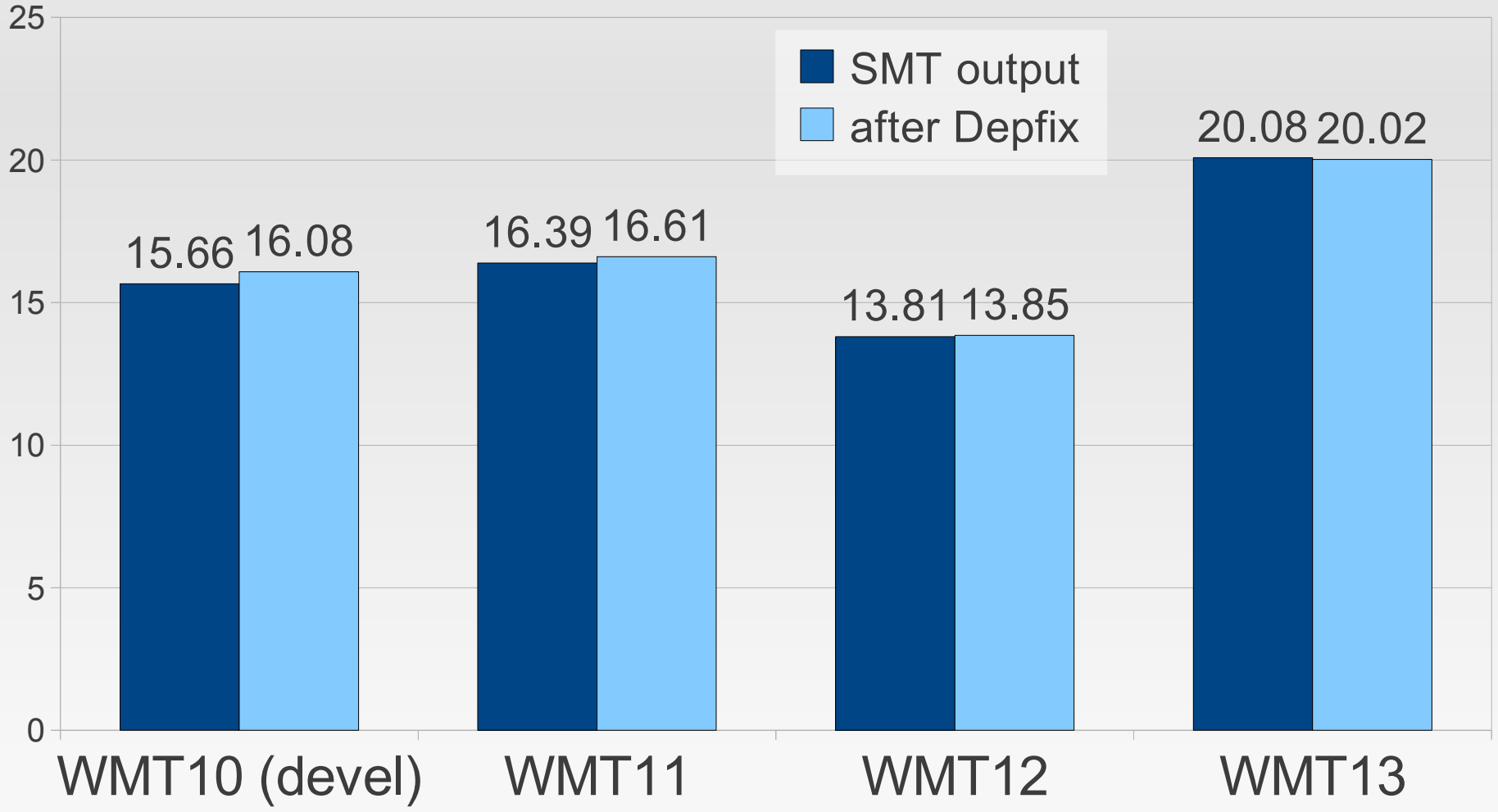


    - (spend, on+X) → (utrácet, 4) P = 0.07
    - *The government destroys the middle schools.*
  - out: *Vládá utrácí za střední školy.*
    - (spend, on+X) → (utrácet, za+4) P = 0.89
    - *The government spends on the middle schools.*
- we model both verb valency and noun valency

Outline

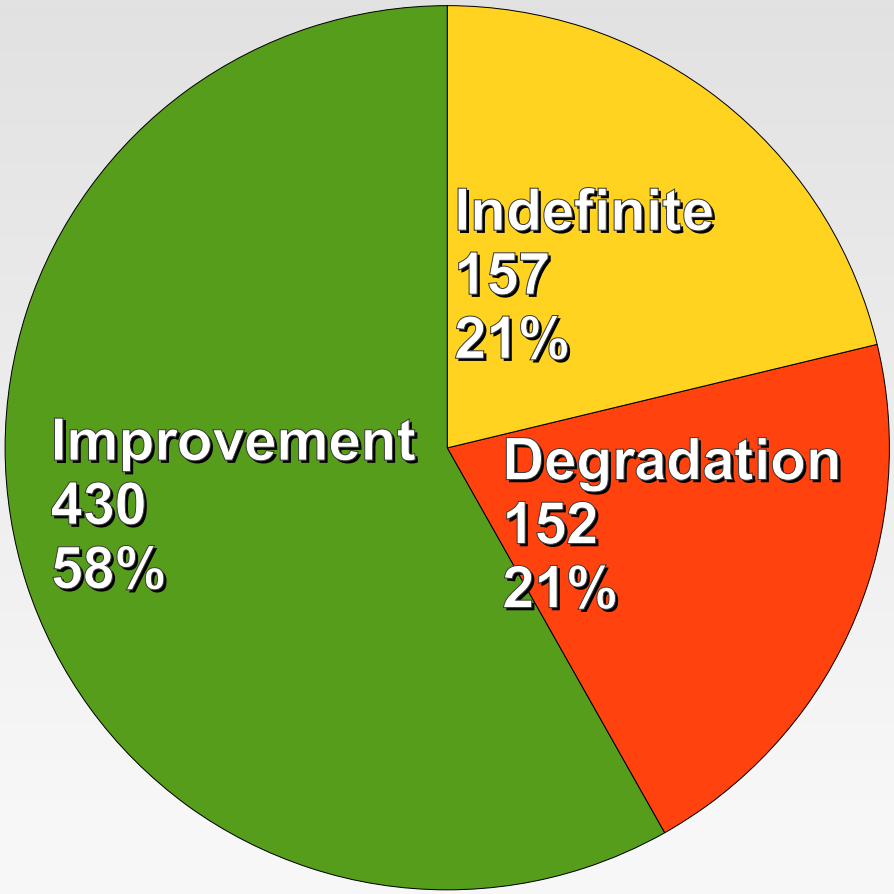
- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- **evaluation**
- parsing of SMT outputs (MSTperl parser)

Automatic evaluation (BLEU)

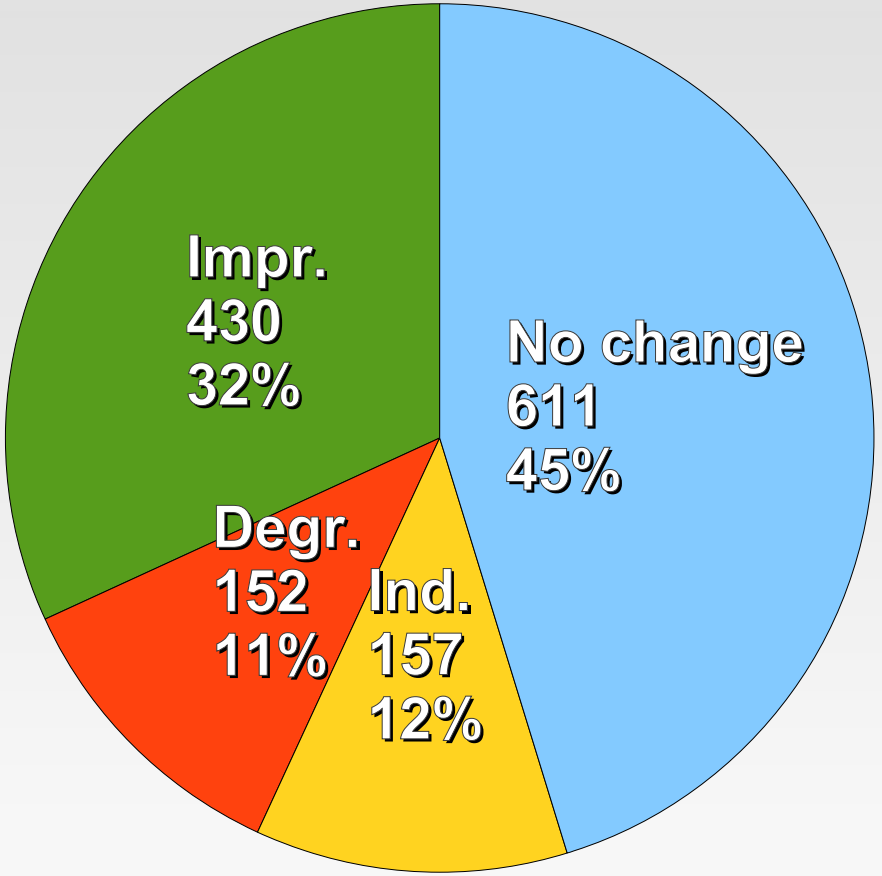


Manual evaluation

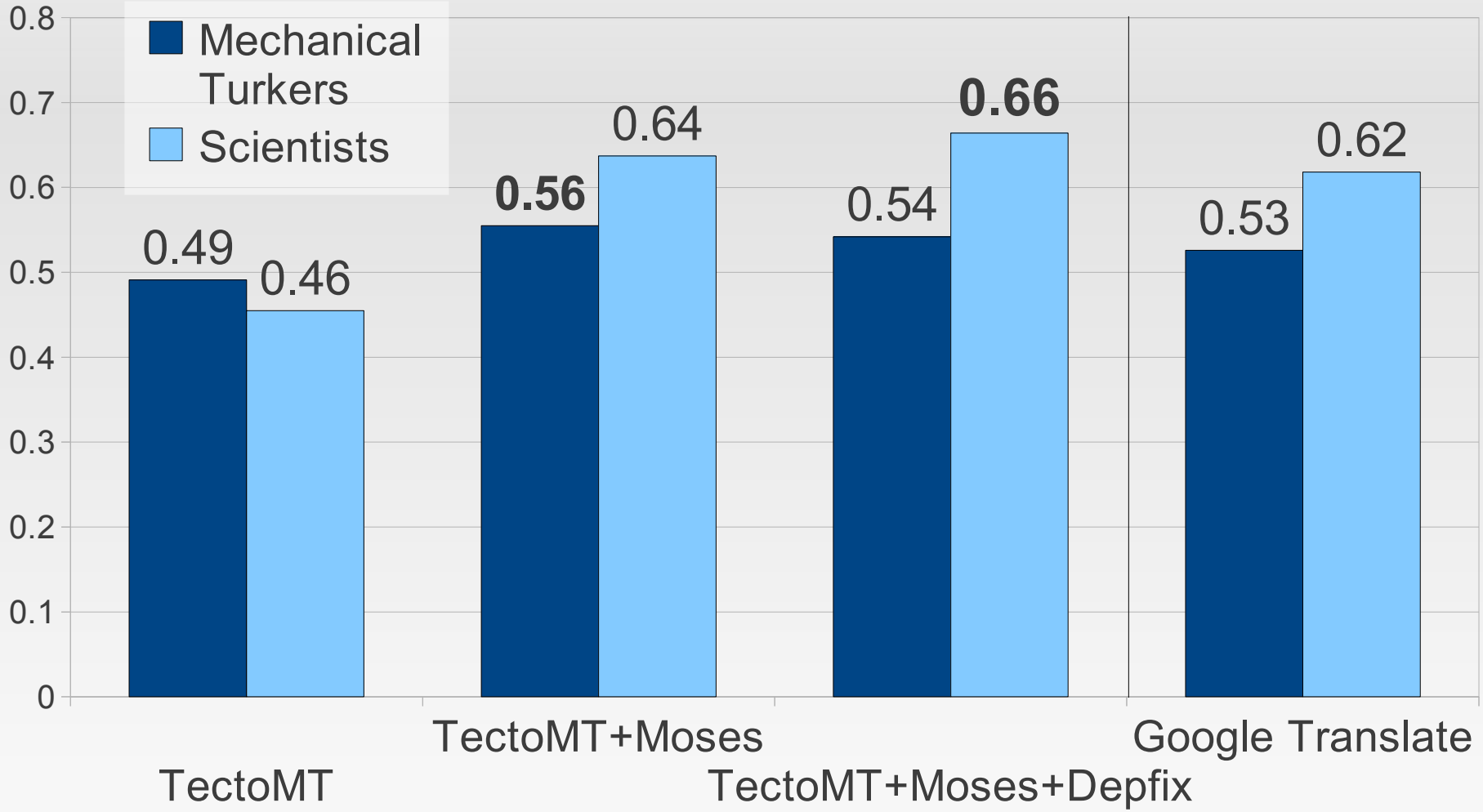
- changed sentences



- all sentences



Manual evaluation (WMT13)



iHNed.cz evaluation (M. Kalina)

Jak překladač z Matfyzu porazil Google

(...)

Pomohl odstraňovač chyb

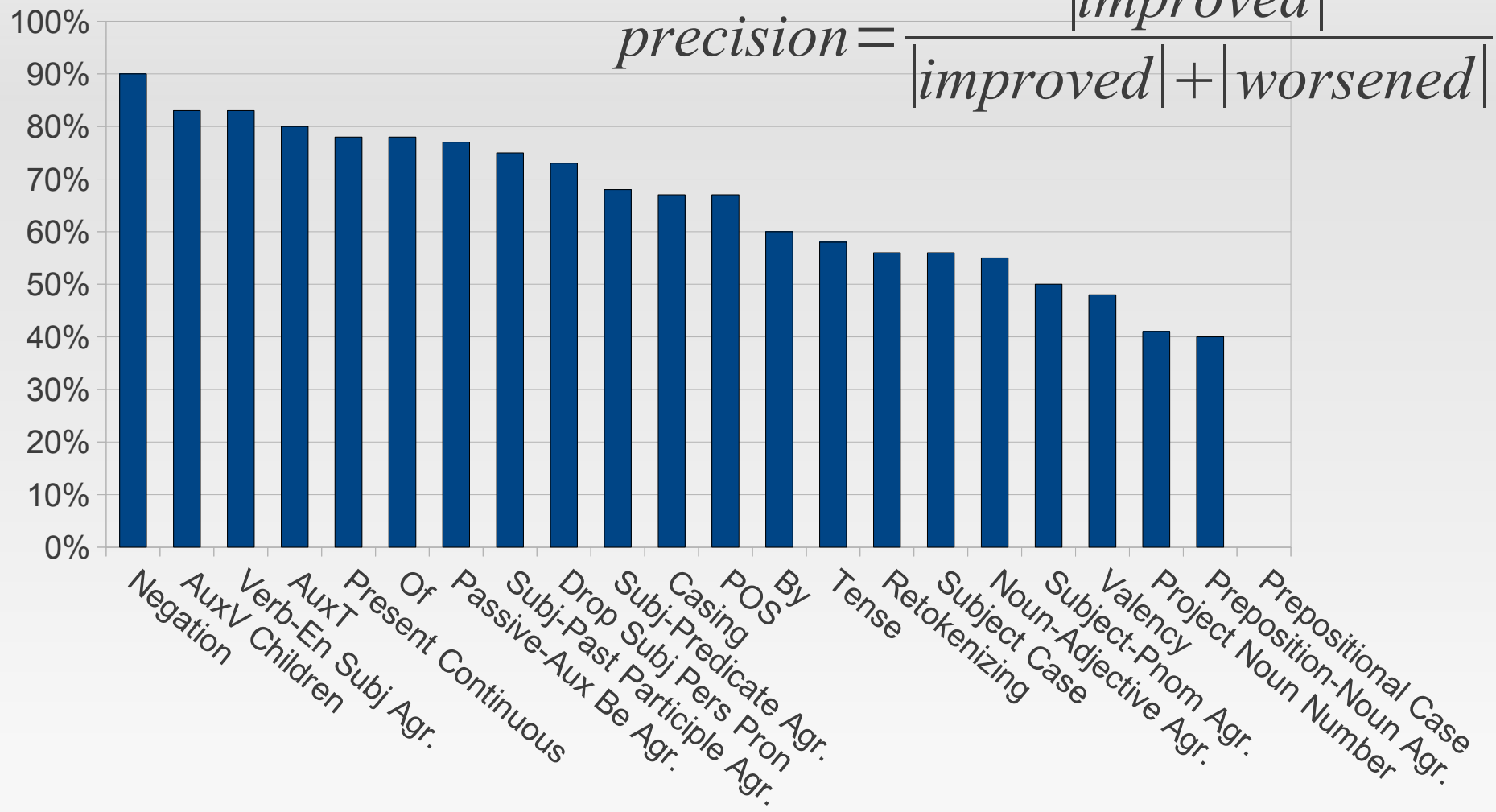
(...)

Aby byl výsledek ještě lepší, prošel výsledný text ještě automatickou korekturou pomocí českého systému **Depfix**, odstraňovače chyb, jenž opravil například špatně přeložené negativní věty a pády.

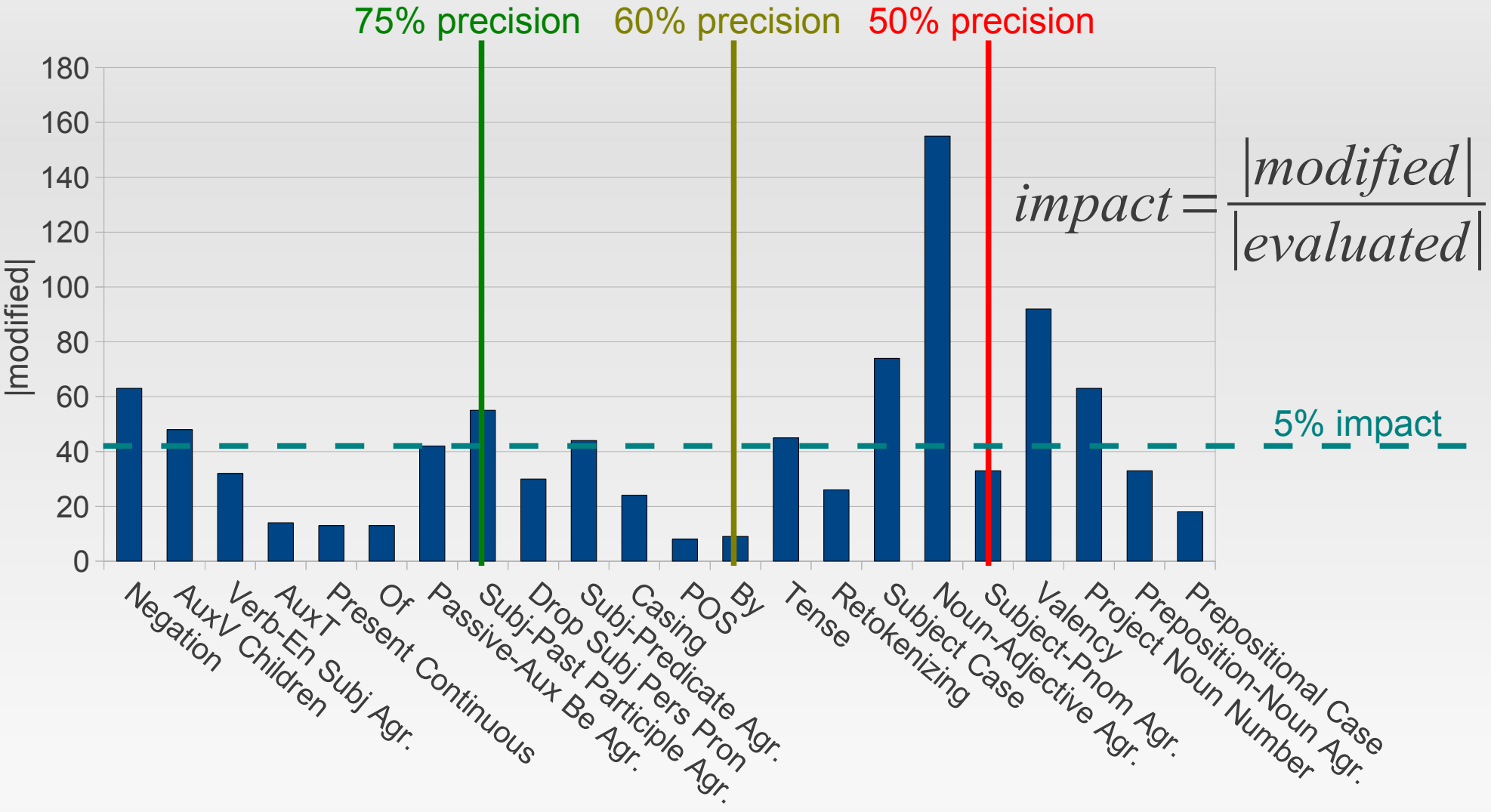
http://tech.ihned.cz/hnfuture/c1-60978500-prekladac-google-maffyz-system

Precision of rules

$$precision = \frac{|improved|}{|improved| + |worsened|}$$



Impact of rules



Outline

- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- evaluation
- → **parsing of SMT outputs** (MSTperl parser)

Parsing of SMT Outputs

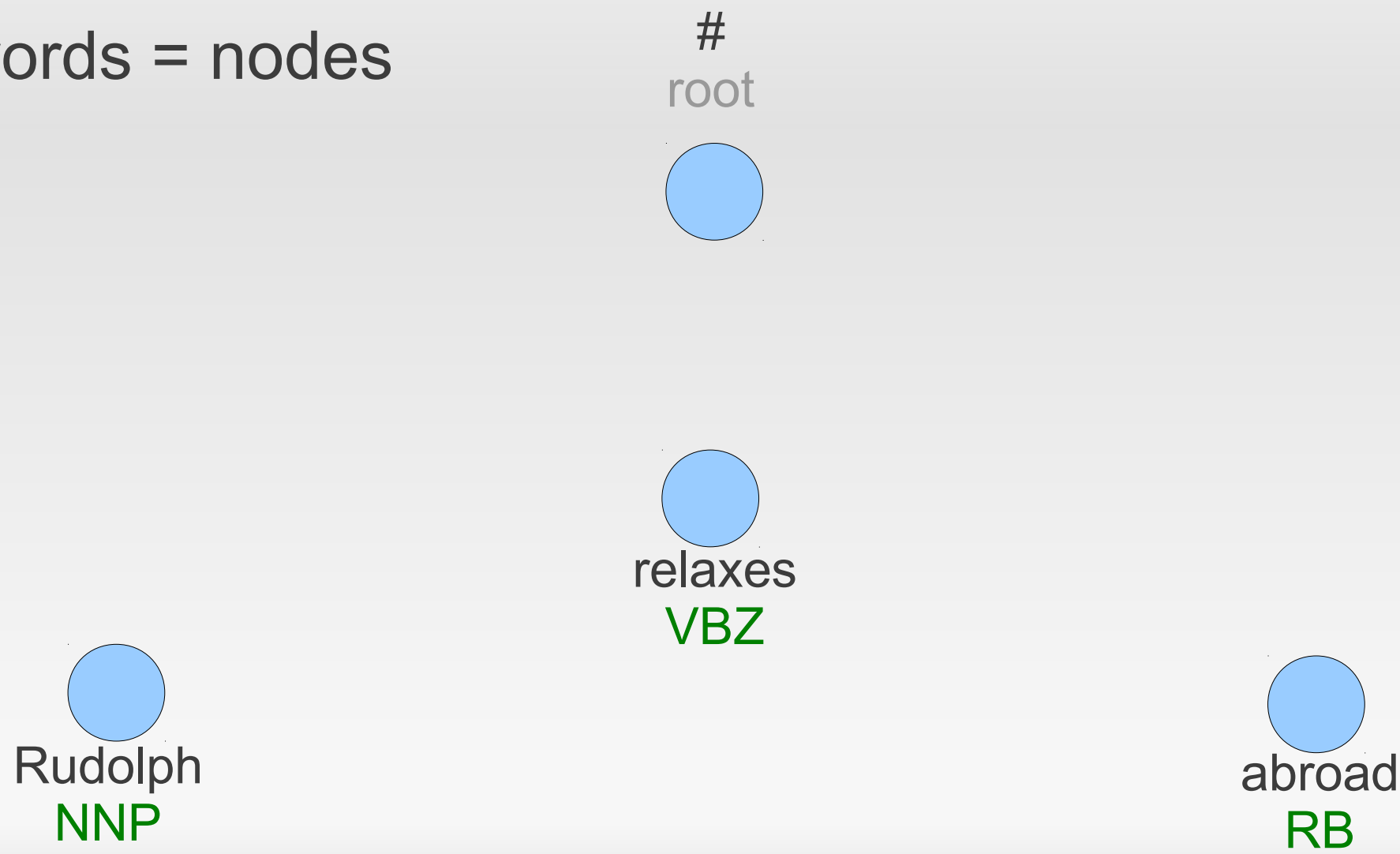
- can be useful in many applications
  - automatic classification of translation errors
  - **automatic correction of translation errors (Depfix)**
  - multilingual question answering...
- ✔ we have the source sentence available
  - Can we use it to help parsing?
- ✗ SMT outputs noisy (errors in fluency, grammar...)
  - parsers trained on gold standard treebanks
  - Can we adapt parser to noisy sentences?

MST parser

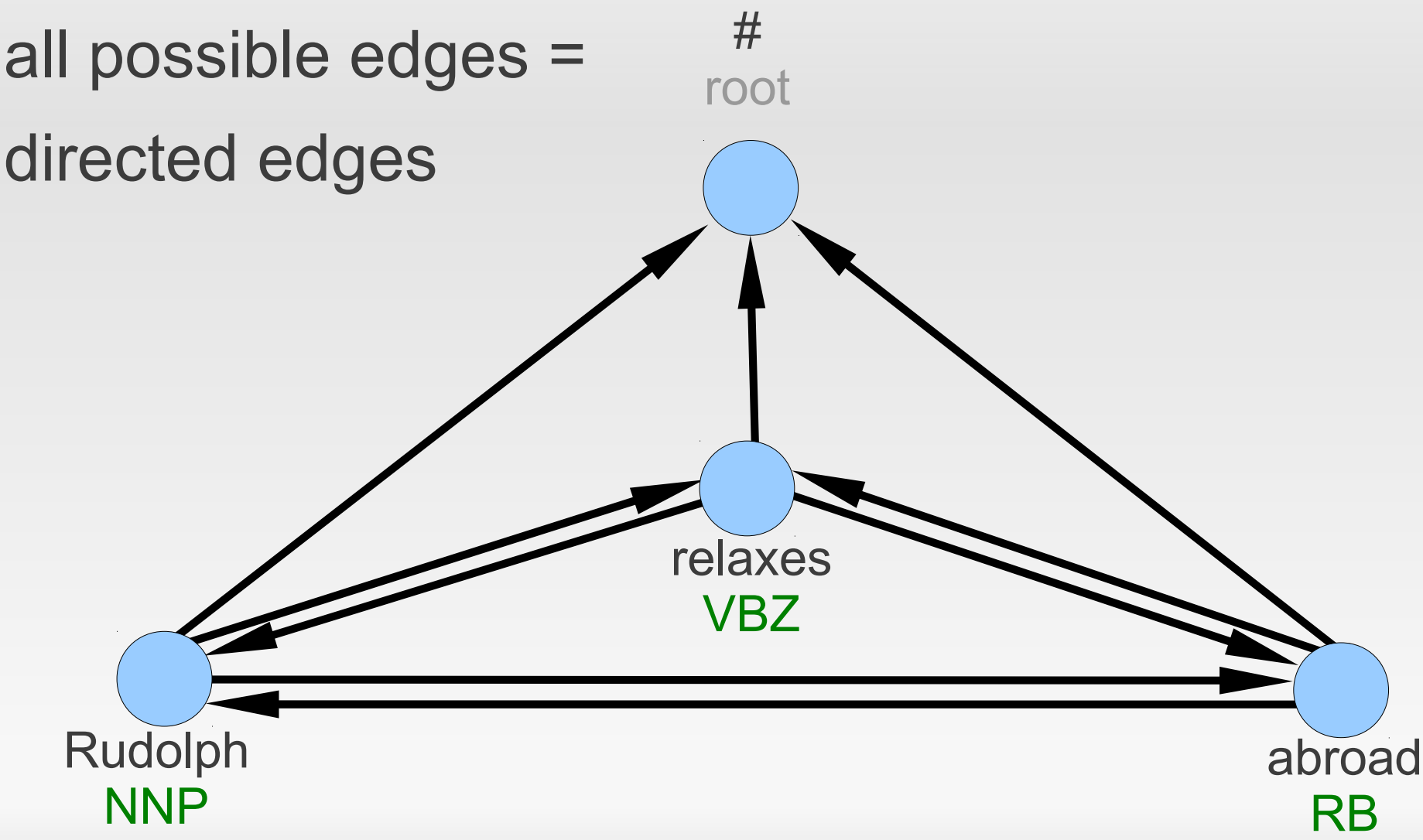
- Maximum Spanning Tree parser
- McDonald, Crammer, Pereira (2005)
  - Online large-margin training of dependency parsers
- McDonald, Pereira, Ribarov, Hajič (2006)
  - **Non-projective** dependency parsing using spanning tree algorithms

(1) Words and Tags

words = nodes



(2) (Nearly) Complete Graph

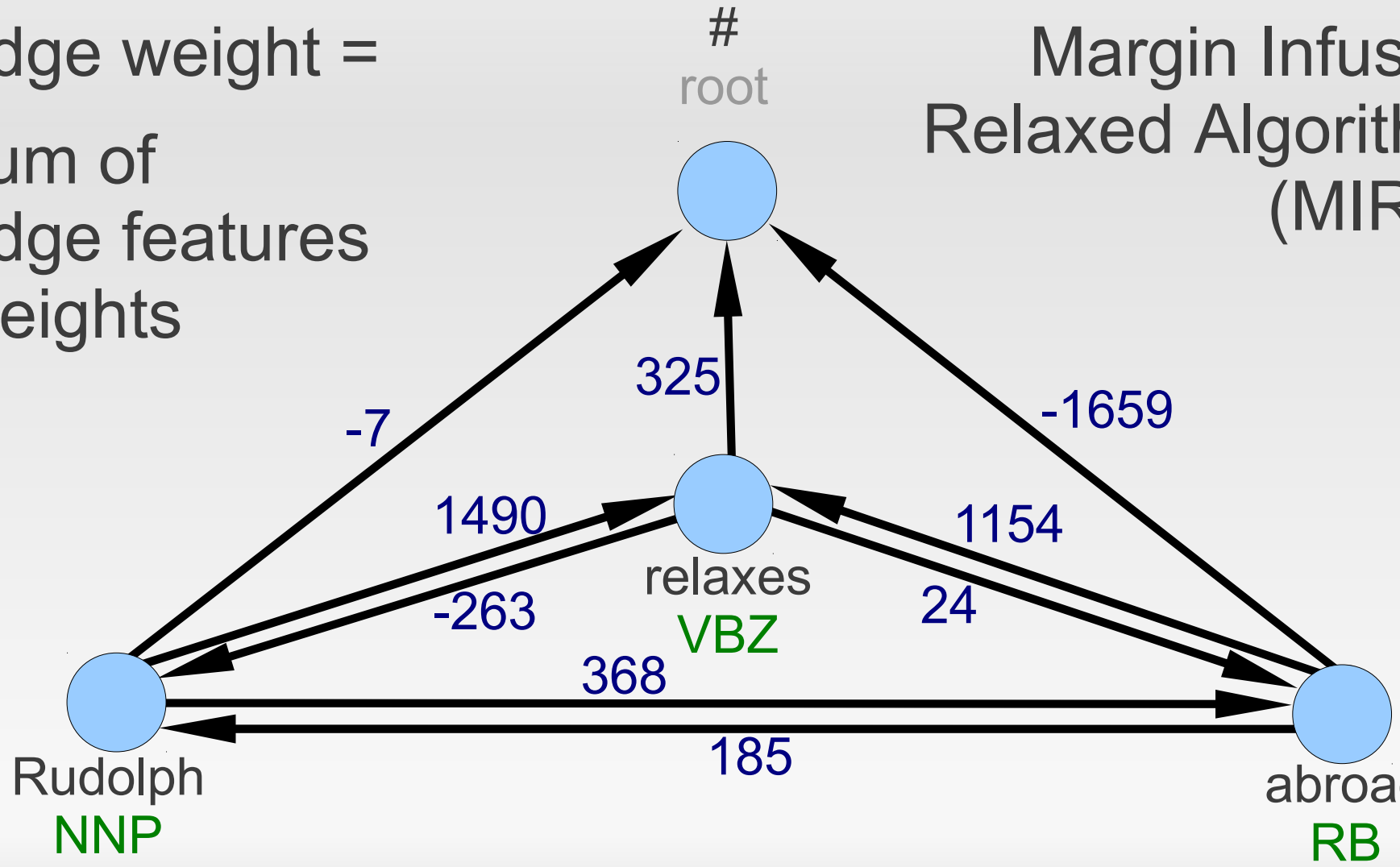


(3) Assign Edge Weights

edge weight =

sum of
edge features
weights

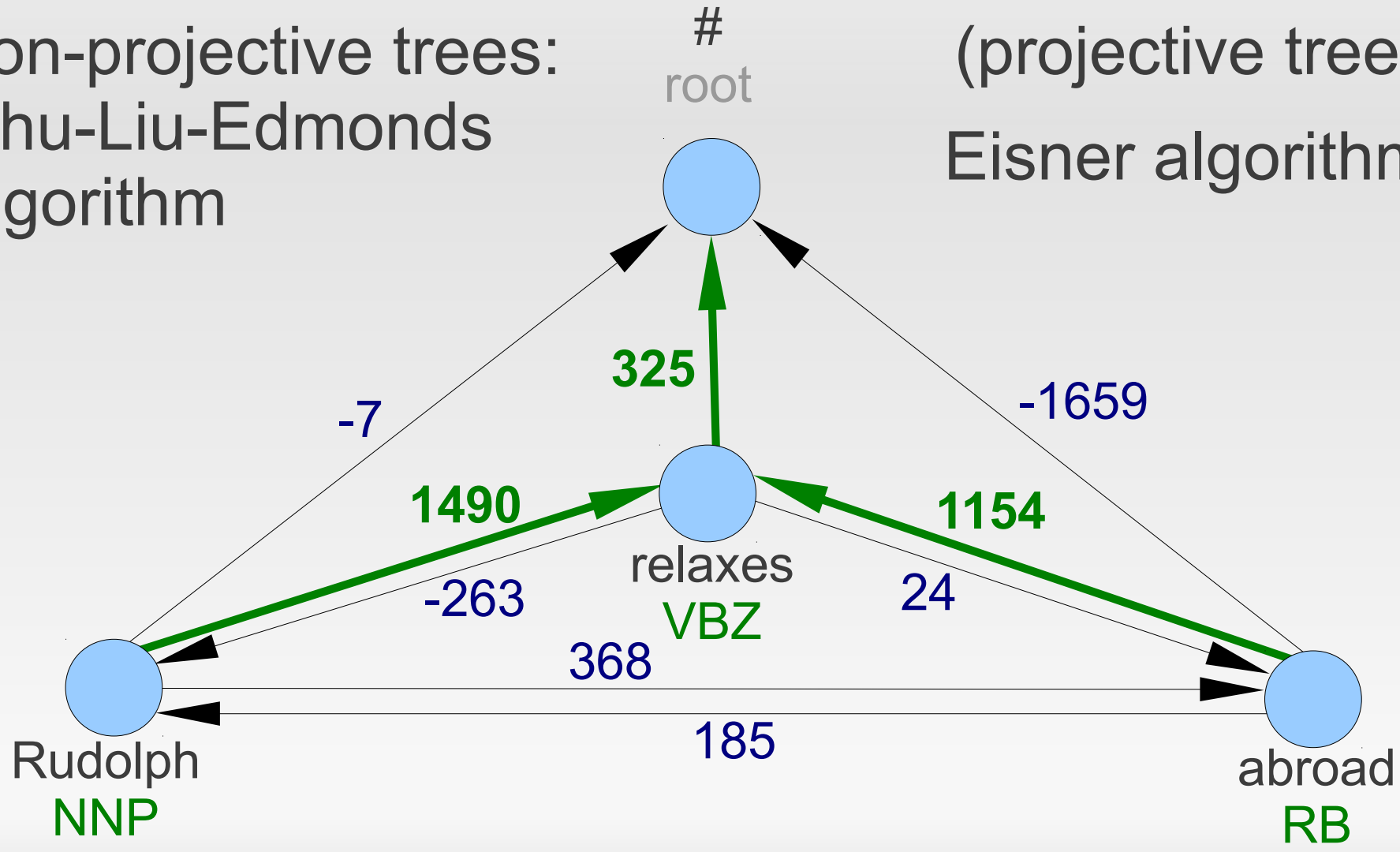
Margin Infused
Relaxed Algorithm
(MIRA)



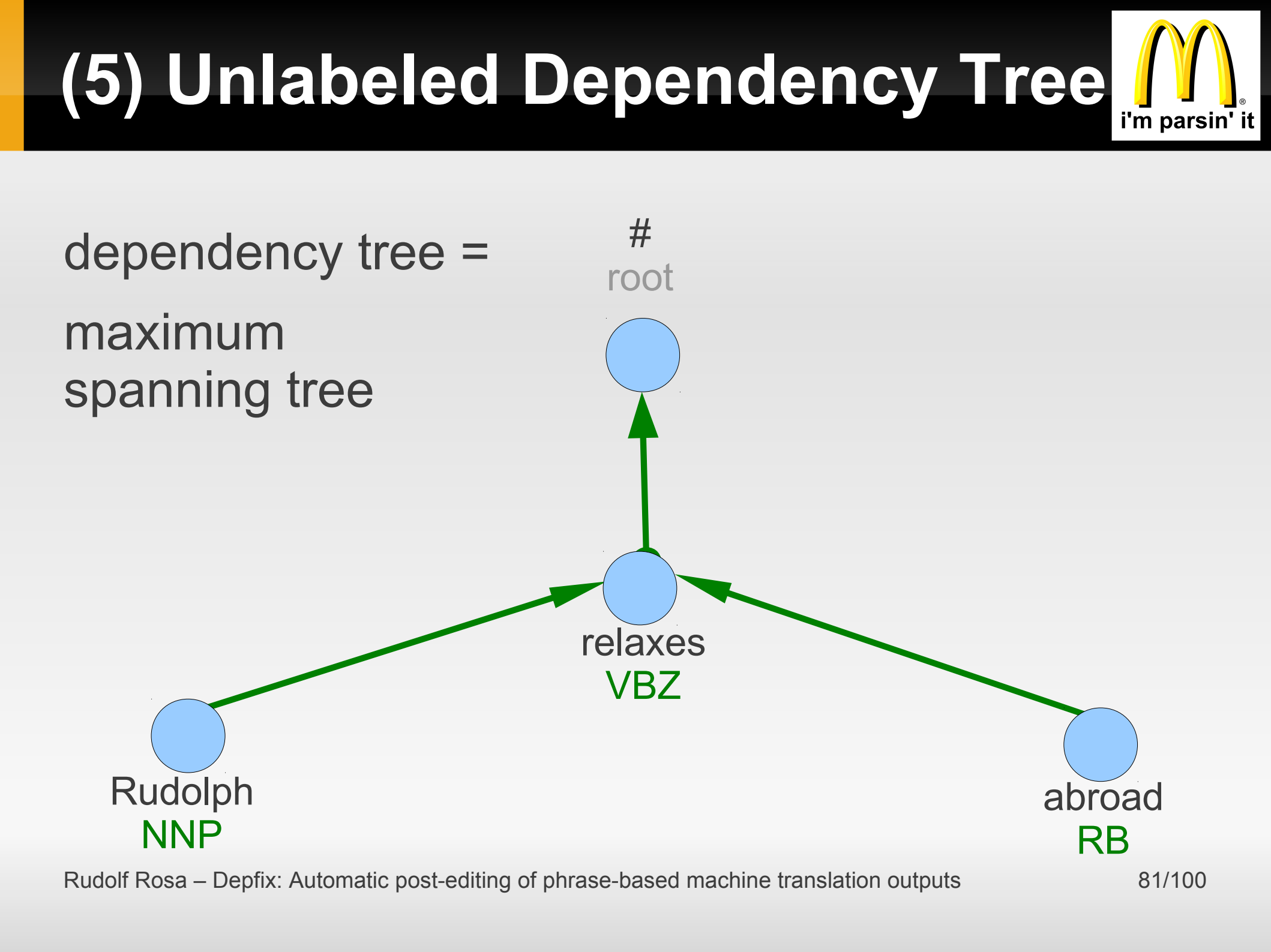
(4) Maximum Spanning Tree

non-projective trees:
Chu-Liu-Edmonds
algorithm

(projective trees:
Eisner algorithm)



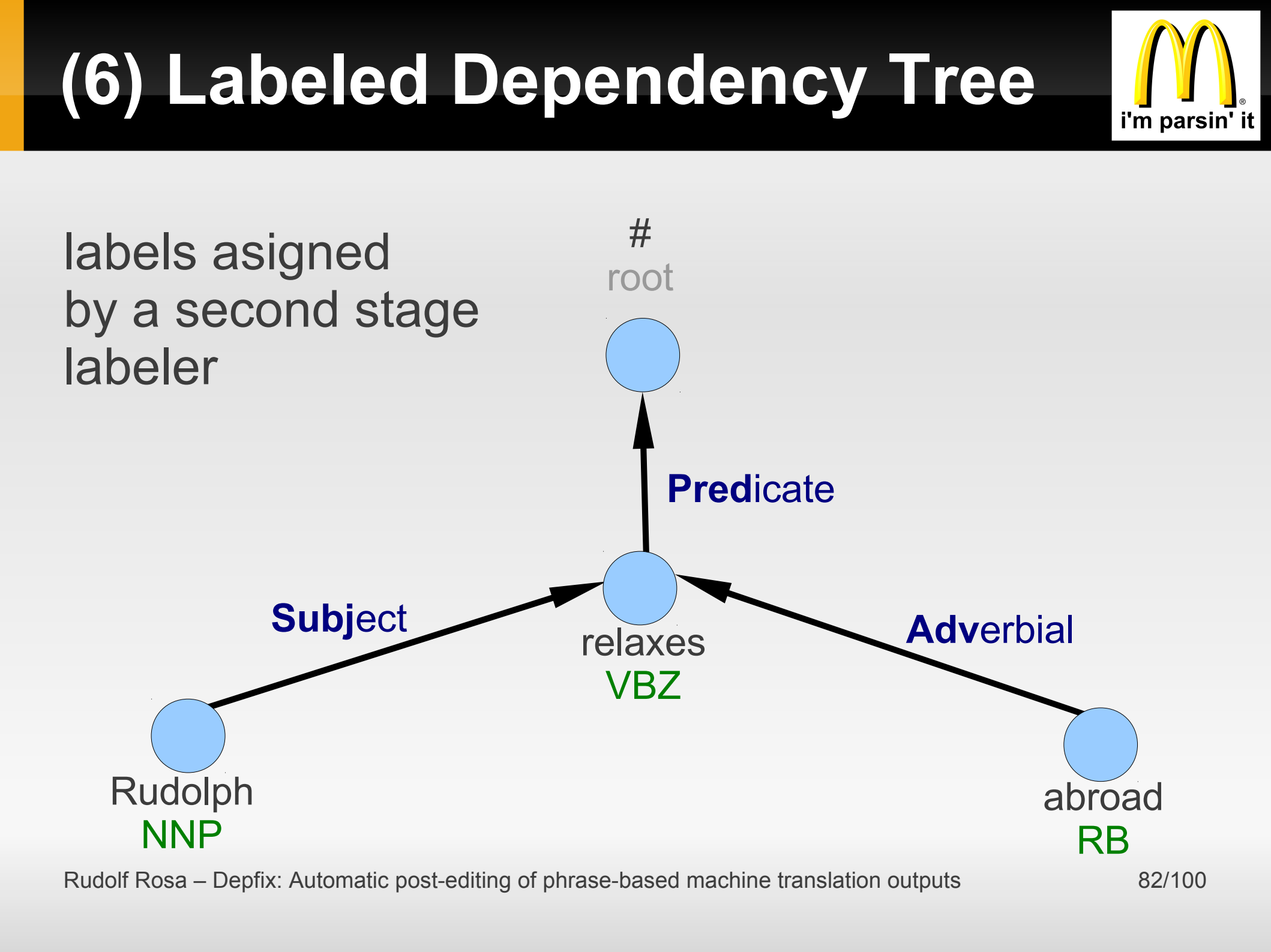
(5) Unlabeled Dependency Tree



dependency tree =

maximum spanning tree

(6) Labeled Dependency Tree



labels asigned
by a second stage
labeler

#
root

Predicate

Subject

relaxes
VBZ

Adverbial

Rudolph
NNP

abroad
RB

Tool::Parser::MSTperl

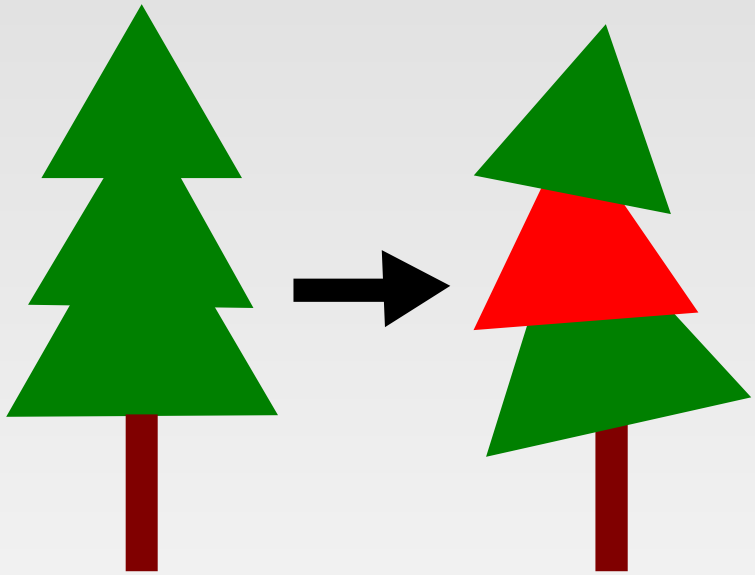
- reimplementation of MST Parser in Perl
  - Treex::Block::W2A::CS::ParseMSTperl
  - (so far only) first-order, non-projective
- adapted for SMT outputs parsing
  - worsening the training data (David Mareček)
  - adding parallel information
  - manually boosting feature weights
  - exploiting large-scale data

Parser Training Data

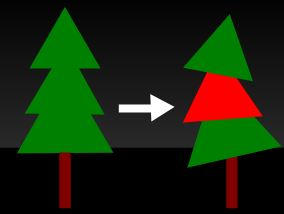
- Prague Czech-English Dependency Treebank
  - parallel treebank
  - 50k sentences, 1.2M words
  - morphological tags, surface syntax, deep syntax
  - word alignment

Worsening the Treebank

- treebank used for training contains correct sentences
- SMT output is noisy
  - grammatical errors
  - incorrect word order
  - missing/superfluous words
  - …
- let's introduce similar errors into the treebank!
  - so far, we have only tried inflection errors

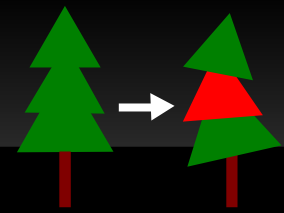


Worsen (1): Apply SMT



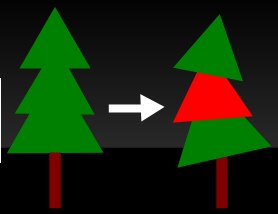
- translate English side of PCEDT to Czech
  - by an SMT system (we used Moses)
- now we have e.g.:
  - Gold English
    - Rudolph's car is black.
  - Gold Czech
    - Rudolfovo$_{NEUT}$ auto$_{NEUT}$ je černé$_{NEUT}$.
  - SMT Czech
    - Rudolfova$_{FEM}$ auto$_{NEUT}$ je černý$_{MASC}$.

Worsen (2): Align SMT to Gold



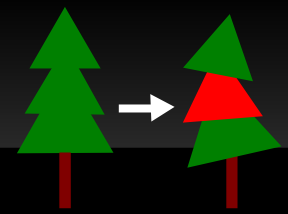
- align SMT Czech to Gold Czech
- Monolingual Greedy Aligner (Martin Popel)
  - alignment link score = linear combination of:
    - similarity of word forms (or lemmas)
    - similarity of morphological tags (fine-grained)
    - similarity of positions in the sentence
    - indication whether preceding/following words aligned
  - repeat: align best scoring pair until below threshold
  - no training: weights and threshold set manually

Worsen (3): Create Error Model



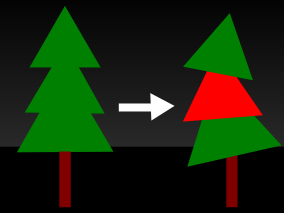
- for each tag:
  - estimate probabilities of SMT system using an incorrect tag instead of the correct tag (Maximum Likelihood Estimate)
- Czech tagset: fine-grained morphological tags
  - part-of-speech, gender, number, case, person, tense, voice...
  - 1500 different tags in training data

Worsen (3): Error Model



- Adjective, Masculine, Plural, Instrumental case (AAMP7), e.g. *lingvistickými* (linguistic)
  - **0.2** Adjective, Masculine, Singular, Nominative case
    - e.g. *lingvistický*
  - **0.1** Adjective, Masculine, Plural, Nominative case
    - e.g. *lingvističtí*
  - **0.1** Adjective, Neuter, Singular, Accusative case
    - e.g. *lingvistické*
- … altogether 2000 such change rules

Worsen (4): Apply Error Model

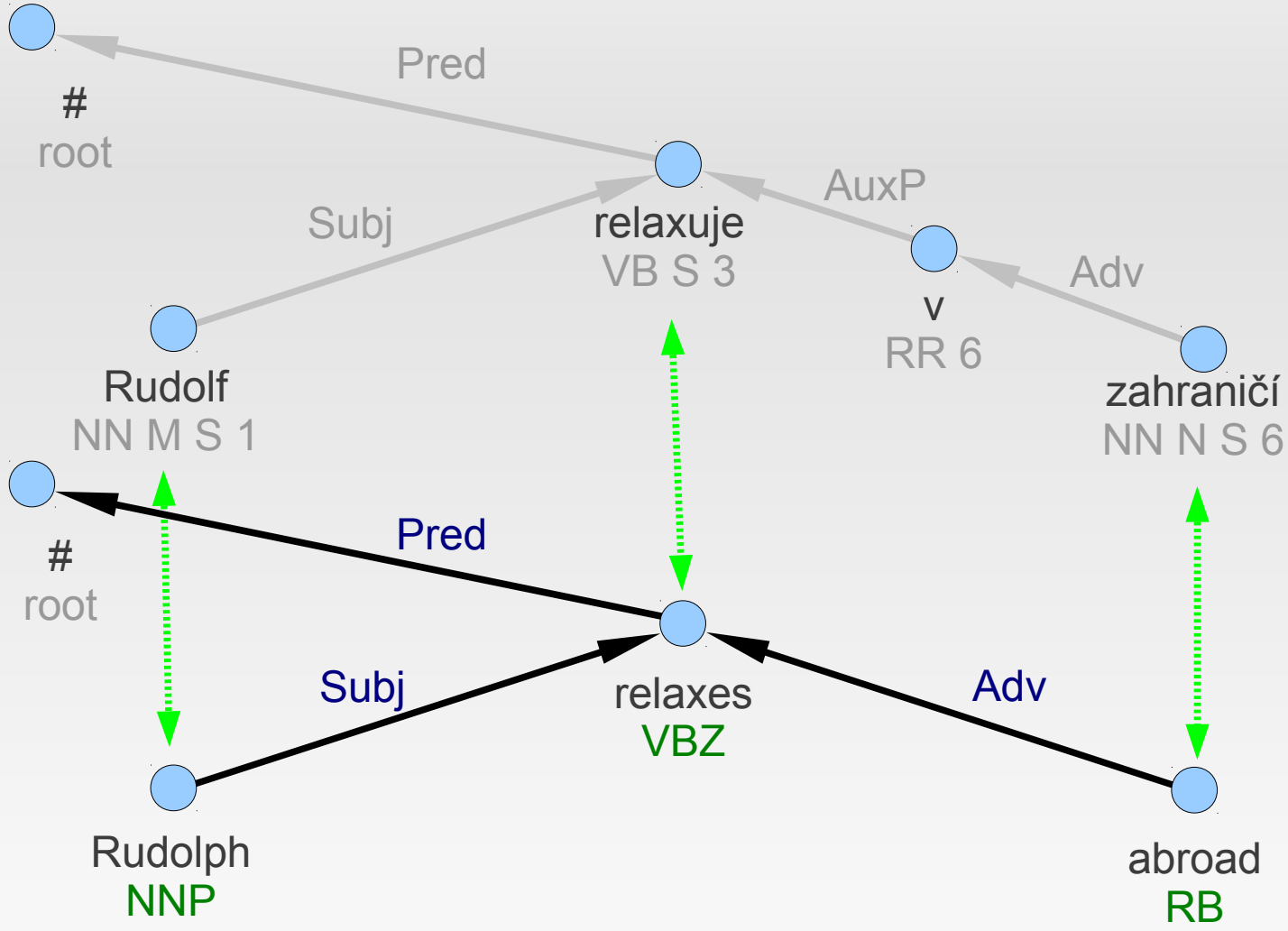


- take Gold Czech
- for each word:
  - assign a new tag randomly sampled according to Tag Error Model
  - generate a new word form
    - rule-based generator, generates even unseen forms
    - new_form = generate_form(lemma, tag) || old_form
- → get Worsened Czech
- use resulting *Gold English-Worsened Czech* parallel treebank to train the parser

Parallel Features

- word alignment (using GIZA++)
- additional features (if aligned node exists):
  - aligned tag (NNS, VBD...)
  - aligned dependency label (Subject, Attribute...)
  - aligned edge existence (0/1)

Parallel Features Example



Manually boosting feature weights

- **aligned edge existence** is the key feature here
- observation: since the worsening is probably too mild, its weight is too low
  - edge exists: -0.57
  - edge does not exist: -0.83
  - missing aligned node(s): -0.67

Manually boosting feature weights

- **aligned edge existence** is the key feature here
- observation: since the worsening is probably too mild, its weight is too low
  - edge exists: **-0.57**
  - edge does not exist: -0.83
  - missing aligned node(s): -0.67
- experiment: manually increase its weight
  - edge exists: **-0.25**

Manually boosting feature weights

- **aligned edge existence** is the key feature here
- observation: since the worsening is probably too mild, its weight is too low
  - edge exists: **-0.57**
  - edge does not exist: -0.83
  - missing aligned node(s): -0.67
- experiment: manually increase its weight
  - edge exists: **-0.25**
- success – manual changing of weights feasible

Exploiting large-scale data

- exploiting large-scale parsed data (CzEng) to provide additional lexical features
- lexical features are important for the parser
- CzEng has 10 times more word types (lemmas) than PCEDT (400k vs. 40k)
- training the parser on whole CzEng infeasible
- new feature: pointwise mutual information

$$PMI'(parent, child) = \log \frac{count([parent, child])}{count([parent, *]) \cdot count([*, child])}$$

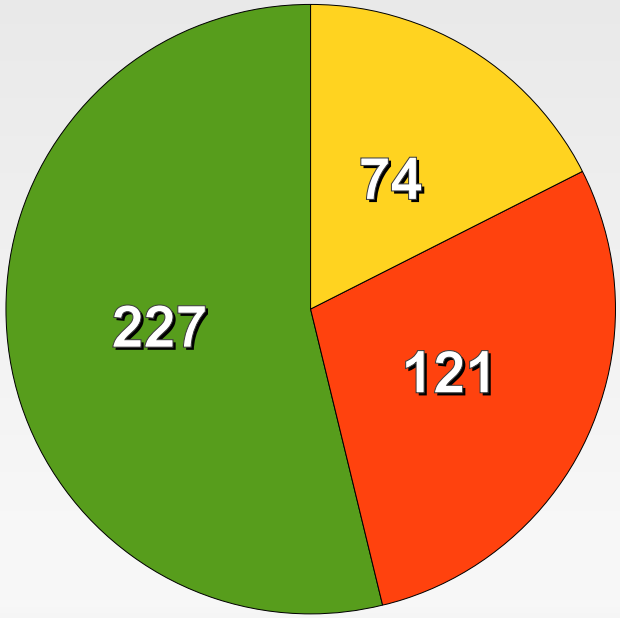
Direct Evaluation: by Inspection

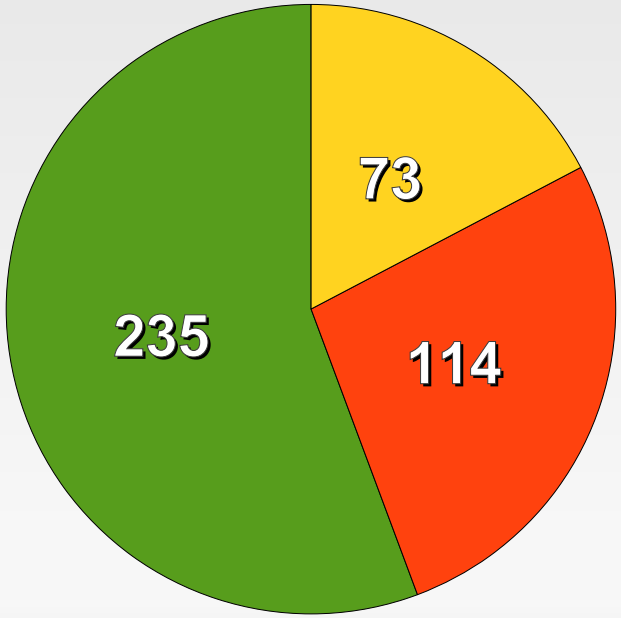
- manual inspection of several parse trees
  - comparing baseline and adapted parser ouputs
- examples of improvements:
  - subject identification even if not in nominative case
  - adjective-noun dependence identification even if agreement violated (gender, number, case)
- hard to do reliably
  - trying to find a correct parse tree for an (often) incorrect sentence – not well defined

Indirect Evaluation: in Depfix

- improvements and deteriorations in comparison to Depfix employing a baseline parser:
  - original McDonald's MST Parser 

  - our baseline setup, without adaptations





Outline – all done!

- translation of negation (and its correction)
  - motivation
  - the fixing pipeline
- analysis and corrections
  - m-layer (lemmas, tags, word-alignment)
  - a-layer (dependency trees, analytical functions)
  - t-layer (“tecto-trees”, formemes, grammatemes)
- evaluation
- parsing of SMT outputs (MSTperl parser)

Thank you for your attention

Rudolf Rosa
rosa@ufal.mff.cuni.cz

Depfix:

**Automatic post-editing
of phrase-based machine translation outputs**

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

For this presentation and other information, please visit:

http://ufal.mff.cuni.cz/~rosa/