Connective-Based Measuring of the Inter-Annotator Agreement in the Annotation of Discourse in PDT

Jiří Mírovský, Lucie Mladová, Šárka Zikánová

Coling 2010

Introduction

Discourse annotation

*We annotate both **intra-** and **inter-sentential** discourse relations marked by **explicit connectives**.*

*But in the Coling paper – only **inter-sentential***

An example

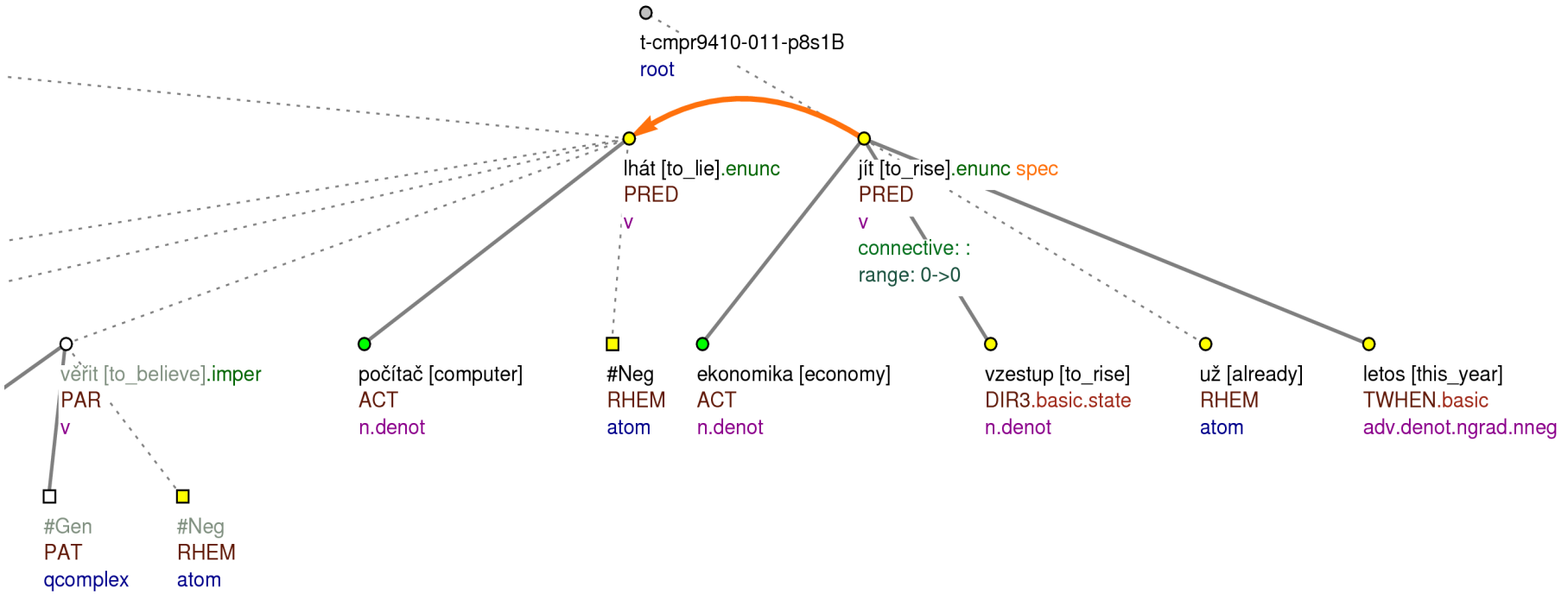
PANE KOLEGO, VĚŘTE NEVĚŘTE, POČÍTAČ ***NELŽE:***

Ekonomika ***jde*** *do vzestupu už letos.*

MR. COLLEAGUE, BELIEVE IT OR NOT, COMPUTER ***DOES NOT LIE:***

The economy ***rises*** *already this year.*

The example in trees



Another example

*Život ale **není** jen ekonomika.*

***Proto má** i socialismus budoucnost.*

*But life **is not** only the economics.*

***Therefore** also the socialism **has** a future.*

Another example

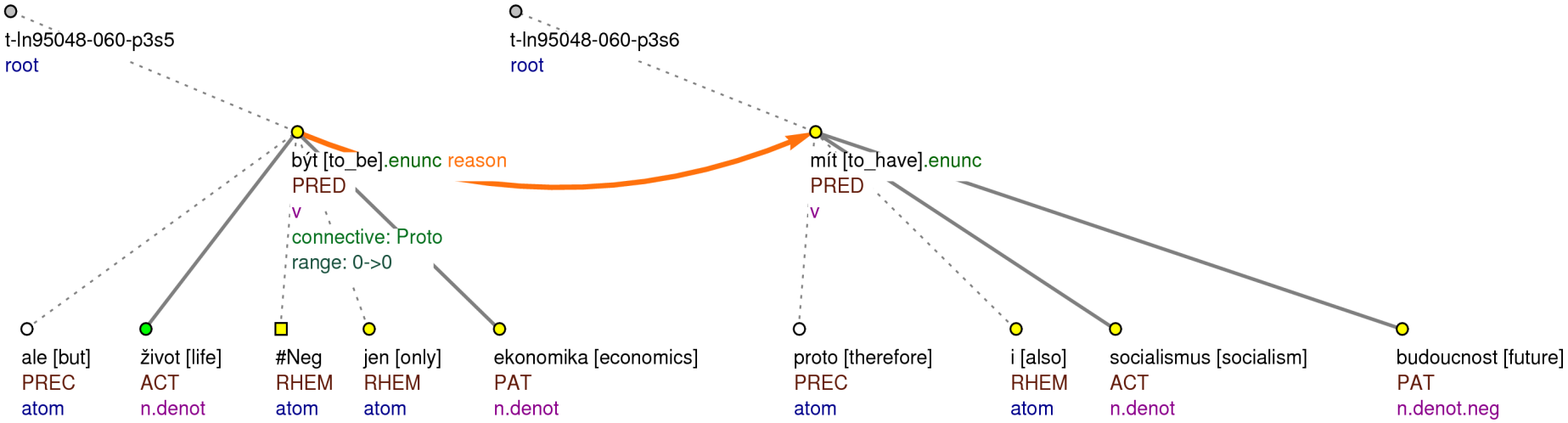
*Život ale **není** jen ekonomika.*
***Proto má** i socialismus budoucnost.*

(Miloš Zeman, Lidové noviny, 24.2.1990)

*But life **is not** only the economics.*
***Therefore** also the socialism **has** a future.*

(Milos Zeman, People's newspaper, 02/24/1990)

The other example in trees



Discourse types (23)

conc
cond
confr
conj
conjalt
corr
disjalt
equiv
exempl
explicat
f_cond
f_opp
f_reason
gener
grad
opp
preced
purp
reason
restr
spec
synchr
other

Discourse super types (4)

contrast
contingency
expansion
temporal

Data annotated

3,165 documents

49,431 sentences

833,195 tokens

6,192 discourse relations

Data annotated in parallel

44 documents

2,084 sentences

33,987 tokens

315 vs. **385** discourse relations

A question

How to measure the inter-annotator agreement?

An answer

like the agreement in the annotation of coreference

A strict measure

If the annotators **mark the same start and target nodes**, we take it as agreement on the recognition of a discourse relation.

The inter-annotator agreement (an example)

***Býval** šéfem tajné služby.*

***A** to **znamená**, že na své soky leccos ví.*

***He used to be** the head of the secret service.*

***And** that **means** that he knows a lot on his rivals.*

The inter-annotator agreement (the example with context)

***Býval** šéfem tajné služby.*

***A** to **znamená**, že na své soky leccos ví.*

Dnes je předsedou parlamentu.

***He used to be** the head of the secret service.*

***And** that **means** that he knows a lot on his rivals.*

Now he is the head of the parliament.

The inter-annotator agreement (the example with more context)

Čchiao se Tengovi podobá tím, že má rád koutky, do nichž se sbíhají všechny důležité nitky.

Býval šéfem tajné služby.

A to znamená, že na své soky leccos ví.

Dnes je předsedou parlamentu.

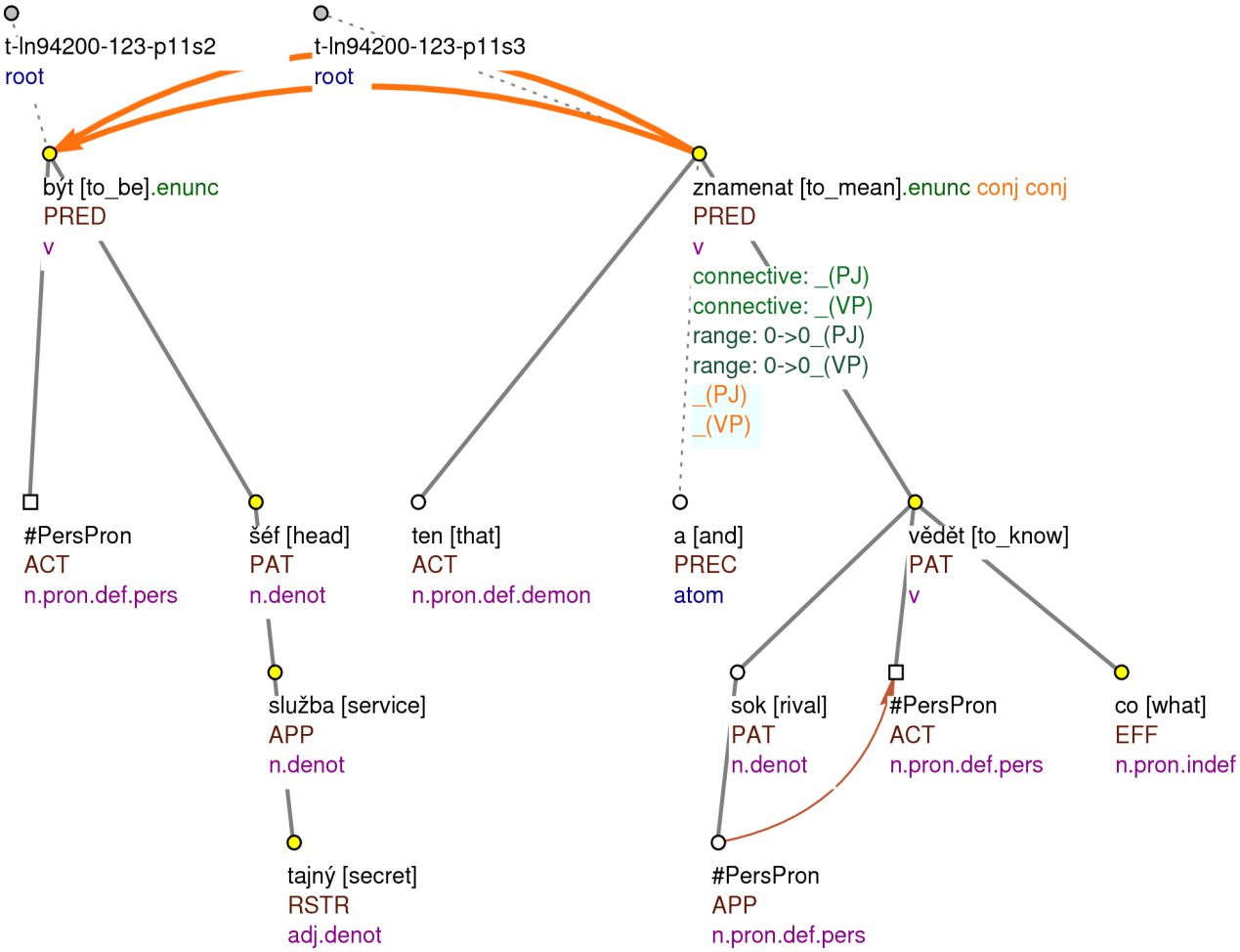
Cchiao resembles Teng in liking corners to which all important threads lead.

He used to be the head of the secret service.

And that means that he knows a lot on his rivals.

Now he is the head of the parliament.

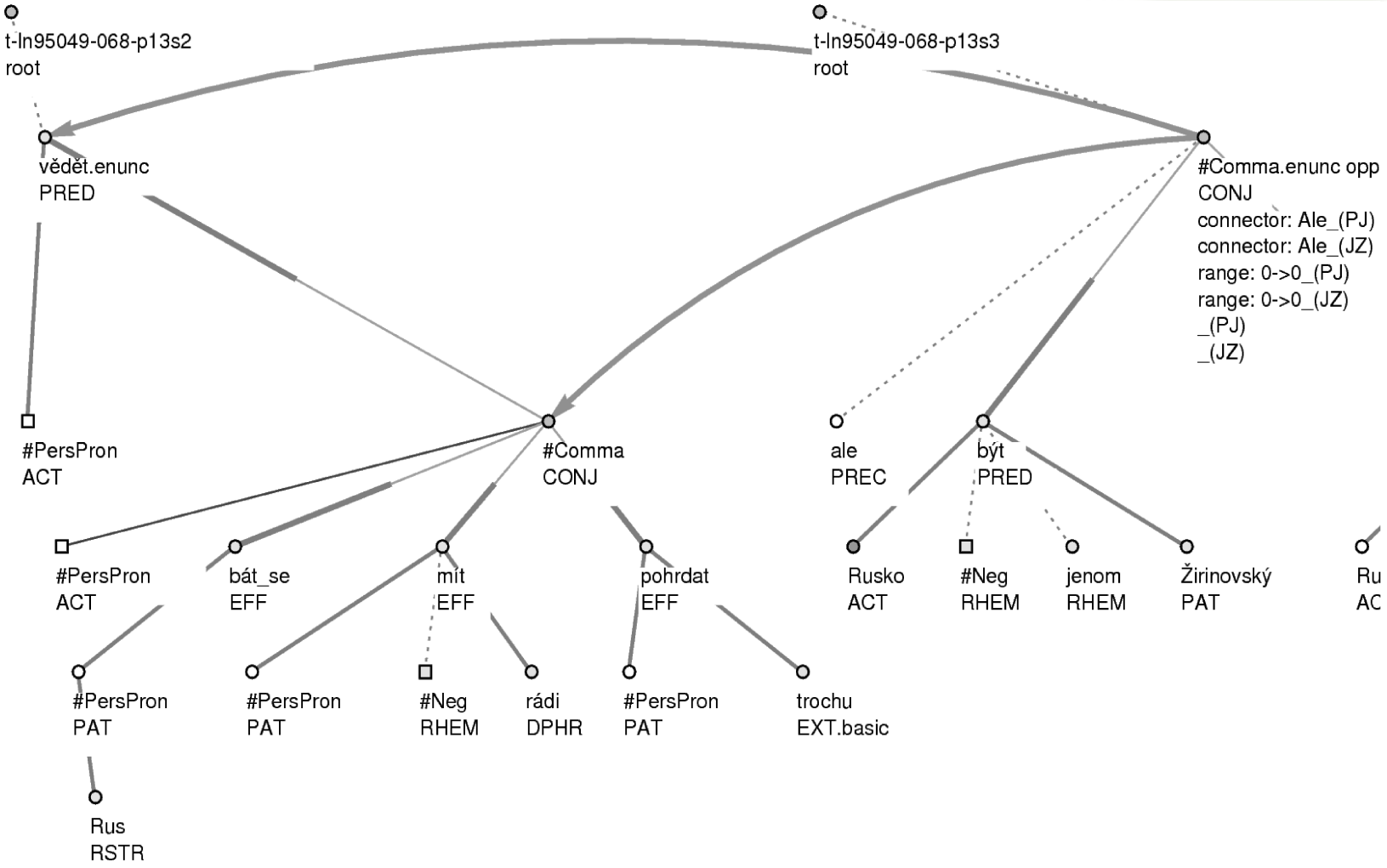
The inter-annotator agreement (the example in trees)



The inter-annotator agreement for the strict measure

measure	value
F_1-measure on relations	0.43
F_1-measure on relations + types	0.34
F_1-measure on relations + connectives	0.41
F_1-measure on rel. + types + connect.	0.32
agreement on types	0.8
agreement on connectives	0.95
Cohen's κ on types	0.74

Problems of the strict measure



Problems of the strict measure

*"**Vím, že se** nás Rusů **bojíte,** že nás **nemáte rádi,** že námi trochu **pohrdáte**. **Ale** Rusko **není** jenom Žirinovskij, Rusko **není** jenom vraždění v Čečensku."*

*"**I know that you are afraid** of us Russians**,** that **you dislike** us**,** that **you despise** us a little. **But** Russia **is not** only Zhirinovsky**,** Russia **is not** only murdering in Chechnya."*

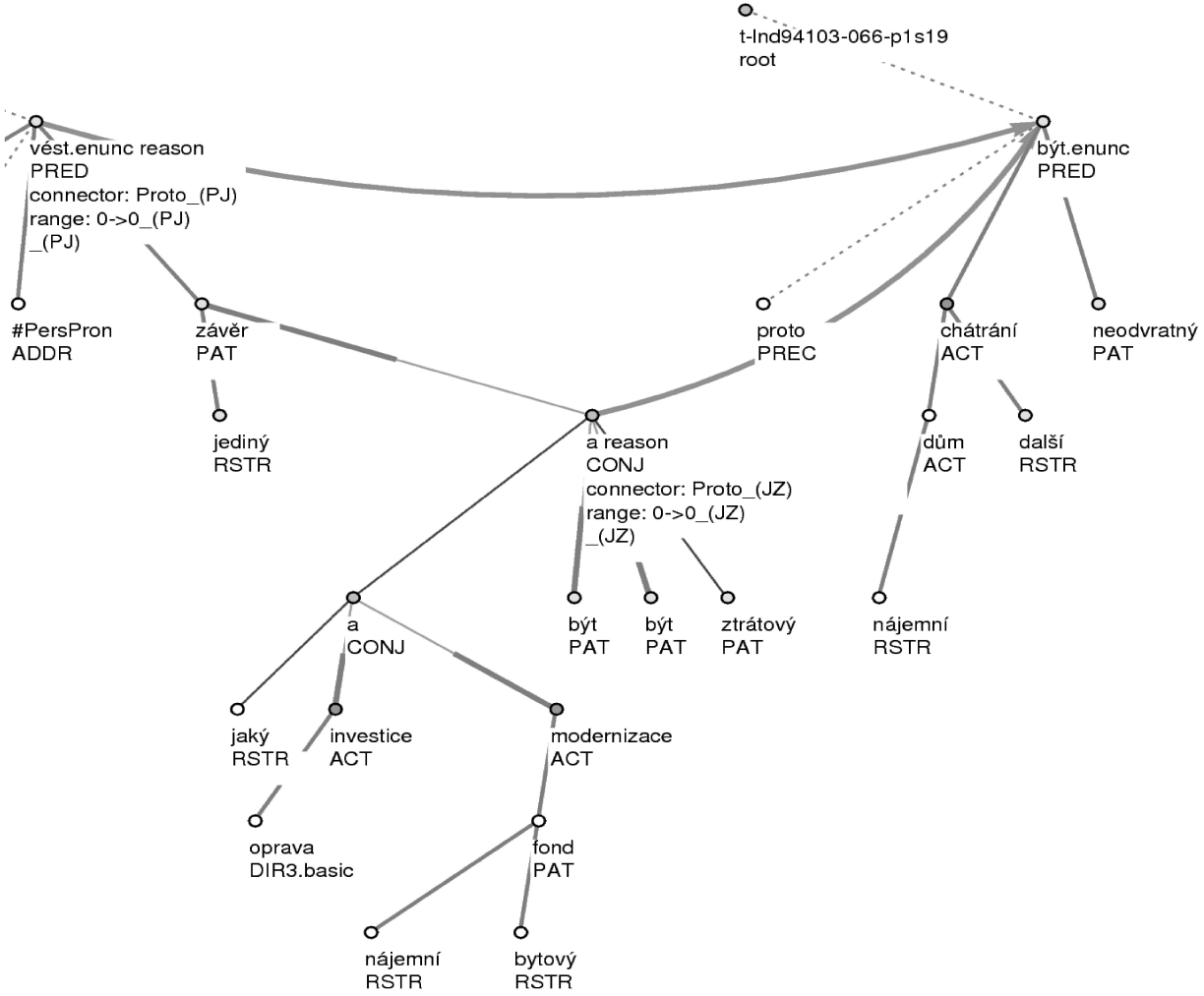
A 1-level skipping measure

If the annotators **mark the same start and target nodes, or if they differ by one level either on the start or target nodes (and agree on the other ones)**, we take it as agreement on the recognition of a discourse relation.

The inter-annotator agreement with 1-level skipping

measure	strict	1-level skipping
F_1-measure on relations	0.43	0.54
F_1-measure on relations + types	0.34	0.43
F_1-measure on relations + connectives	0.41	0.49
F_1-measure on rel. + types + connect.	0.32	0.39
agreement on types	0.8	0.8
agreement on connectives	0.95	0.91
Cohen's κ on types	0.74	0.73

Problems of the 1-level skipping measure

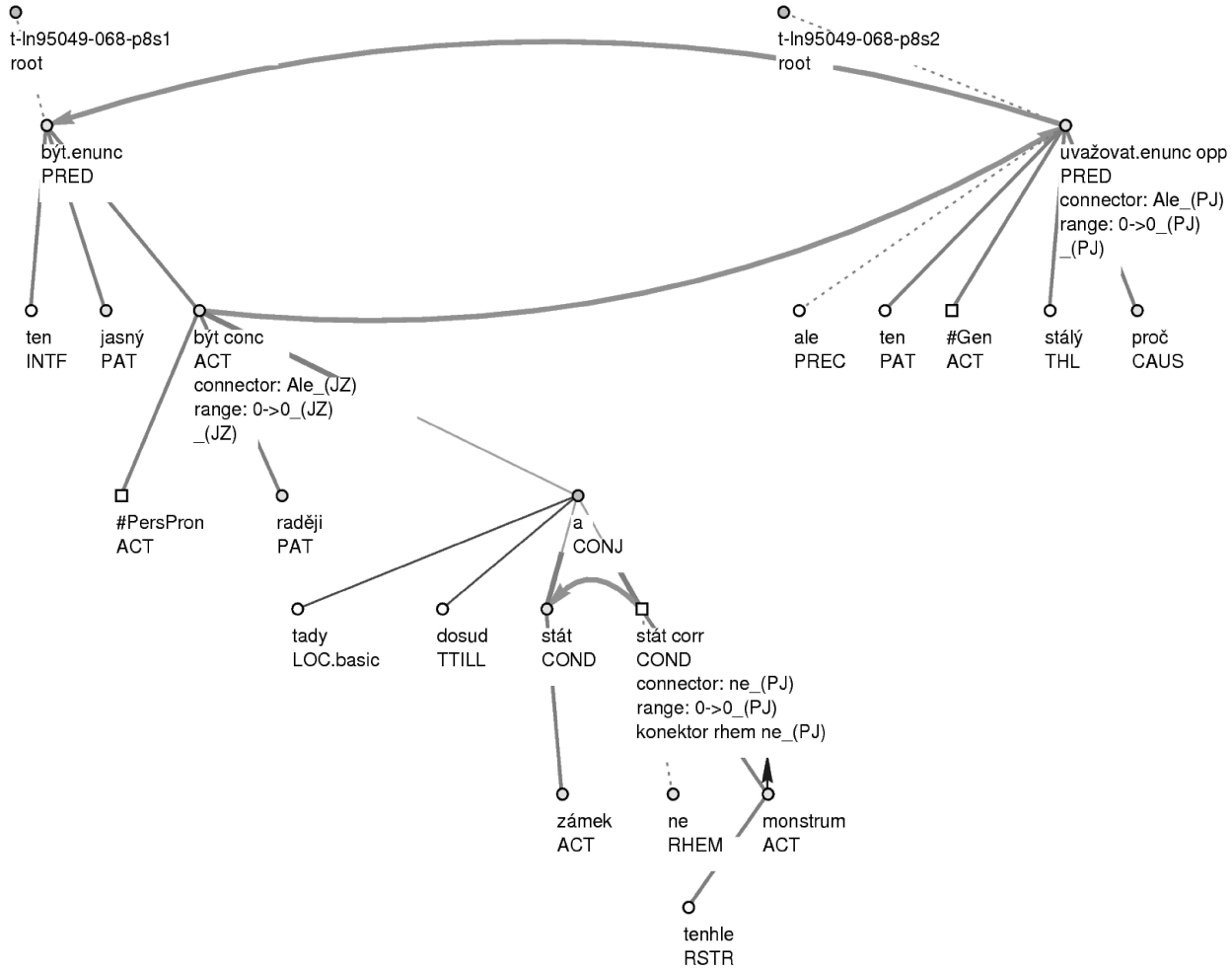


Problems of the 1-level skipping measure

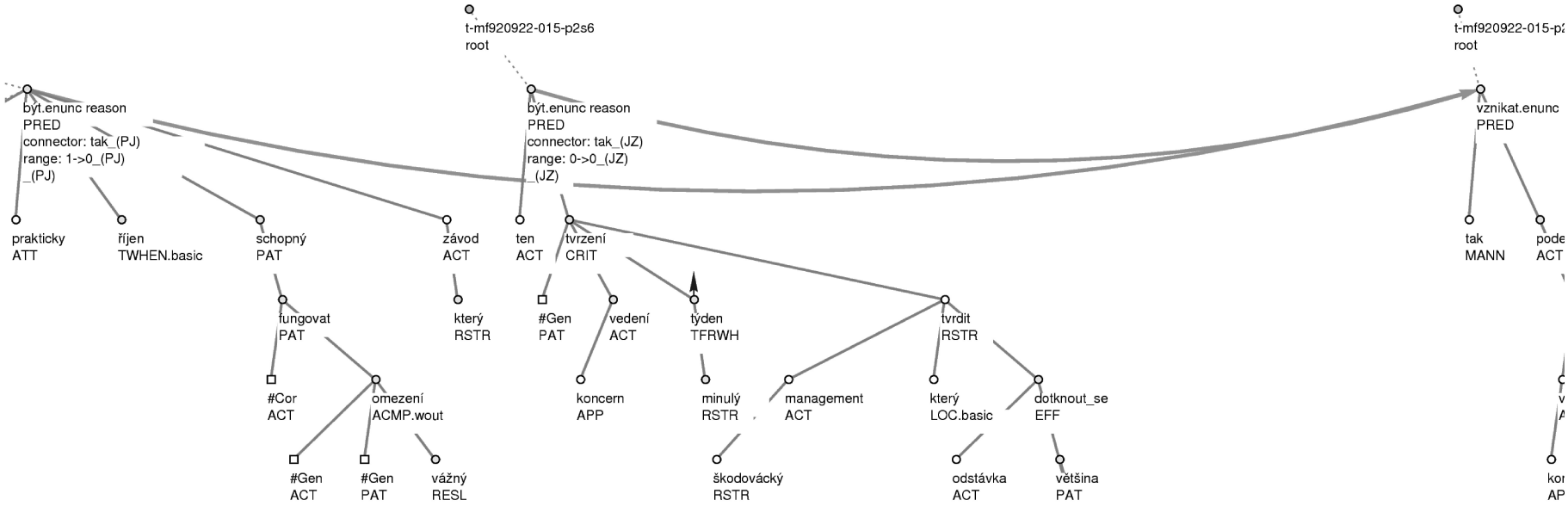
*Racionální kalkulace ... **povede k** jedinému **závěru**: jakékoliv investice do oprav ... **budou** ztrátové. **Proto je** další chátrání nájemních domů neodvratné.*

*A rational calculation ... **will lead to a** single **conclusion**: any investment to repairs ... **will be** loss-making. **Therefore**, further dilapidation of the apartment buildings **is** inevitable.*

Problems of the 1-level skipping measure



Problems of the 1-level skipping measure



The inter-annotator agreement with 1-level skipping

measure	strict	1-level skipping
F_1-measure on relations	**0.43**	**0.54**
F_1-measure on relations + types	0.34	0.43
F_1-measure on relations + connectives	**0.41**	**0.49**
F_1-measure on rel. + types + connect.	0.32	0.39
agreement on types	0.8	0.8
agreement on connectives	**0.95**	**0.91**
Cohen's κ on types	0.74	0.73

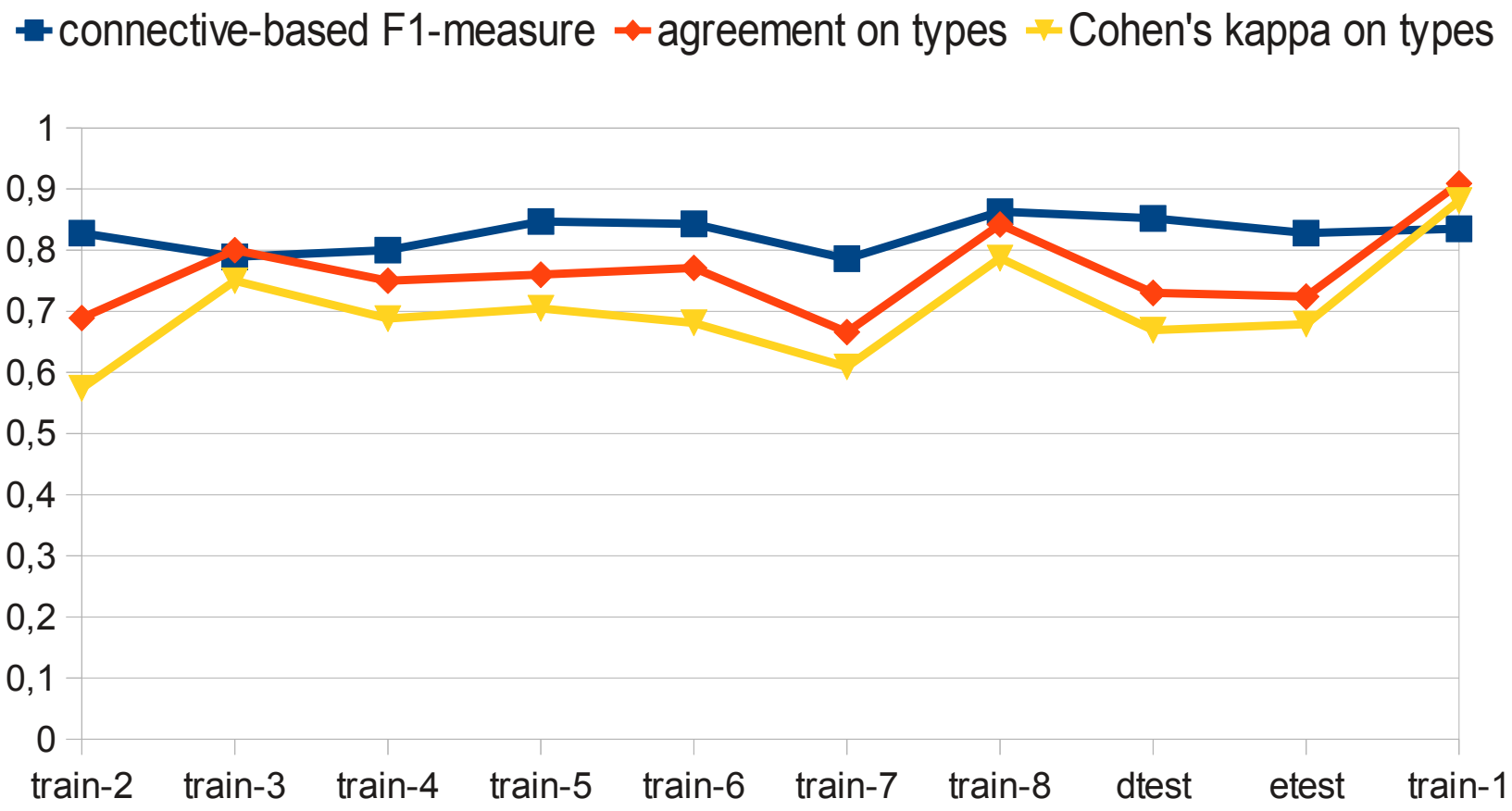
A connective-based measure

If the annotators **mark the same connective**, we take it as agreement on the recognition of a discourse relation.

The inter-annotator agreement, the connective-based measure

measure	strict	1-level skipping	connective -based
F_1-measure on relations	0.43	0.54	0.86
F_1-measure on relations + types	0.34	0.43	0.55
F_1-measure on rel. + start/end nodes	-	-	0.53
F_1-measure on rel. + types + nodes	0.32	0.39	0.33
agreement on types	0.8	0.8	0.64
agreement on start/end nodes	-	-	0.62
Cohen's κ on types	0.74	0.73	0.54

Agreement in the course of the annotation



Agreement in the course...

measurement	F1	agreement on types	kappa on types
train-2	0.83	0.69	0.57
train-3	0.79	0.8	0.75
train-4	0.8	0.75	0.69
train-5	0.85	0.76	0.71
train-6	0.84	0.77	0.68
train-7	0.79	0.67	0.61
train-8	0.86	0.84	0.79
dtest	0.85	0.73	0.67
etest	0.83	0.72	0.68
train-1	0.84	0.91	0.88
all par. data	**0.83**	**0.77**	**0.71**

Agreement in comparison

measurement	F1	**agreement on types**	kappa on types
all par. data	0.83	**0.77**	0.71
Penn Discourse Treebank	-	**0.8**	-

Conclusion


Thank you!
© TemplatesWise.com