NPFL075 Practical Class 02

Jiří Mírovský

Charles University

Institute of Formal and Applied Linguistics

PDT

Prague **D**ependency **T**reebank

- **Czech journalistic texts** from 1990's

- **50 thousand** sentences annotated manually on **several layers**

    - morfological layer (part of speech, case, …)

    - analytical layer (surface syntax)

    - tektogrammatical layer (deep syntax)

Prague Dependency Treebank 2.0
(PDT 2.0)

- published in LDC in 2006 (registration required)

- (mostly) **manually annotated** texts from 1990's

  - 40% general newspaper articles
  - 20% economic news and analyses
  - 20% articles from a popular science magazine
  - 20% information technology texts

- the **annotation scenario** was developed between 1996–2000

- based on the **Functional Generative Description** (**FGD,** Sgall, Hajičová, Panevová, **1960's – 1980's**)

Prague Dependency Treebank 2.0
Three Layers of Annotation

- **morphological** layer (100% of data)

    – 7,110 documents, 115,844 sentences, 1,957,247 tokens

- **analytical** (surface syntax) layer (75%)

    – 5,330 documents, 87,913 sentences, 1,503,739 tokens

- **tectogrammatical** (deep syntax) layer (45%)

    – 3,165 documents, 49,431 sentences, 833,195 tokens

How does the annotation look like?

Máma mele maso.

[*Mom is_mincing meat.*]

Máma mele maso.
[*Mom is_mincing meat.*]

Máma [*Mom*] – noun, feminine, nominative, singular

mele [*is_mincing*] – verb, indicative, present, singular, 3rd person

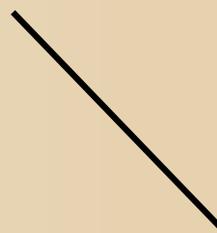**maso [*meat*]** – noun, neutral, accusative, singular

PDT – Analytical layer

Máma
[Mom] ———— mele
[is_mincing]
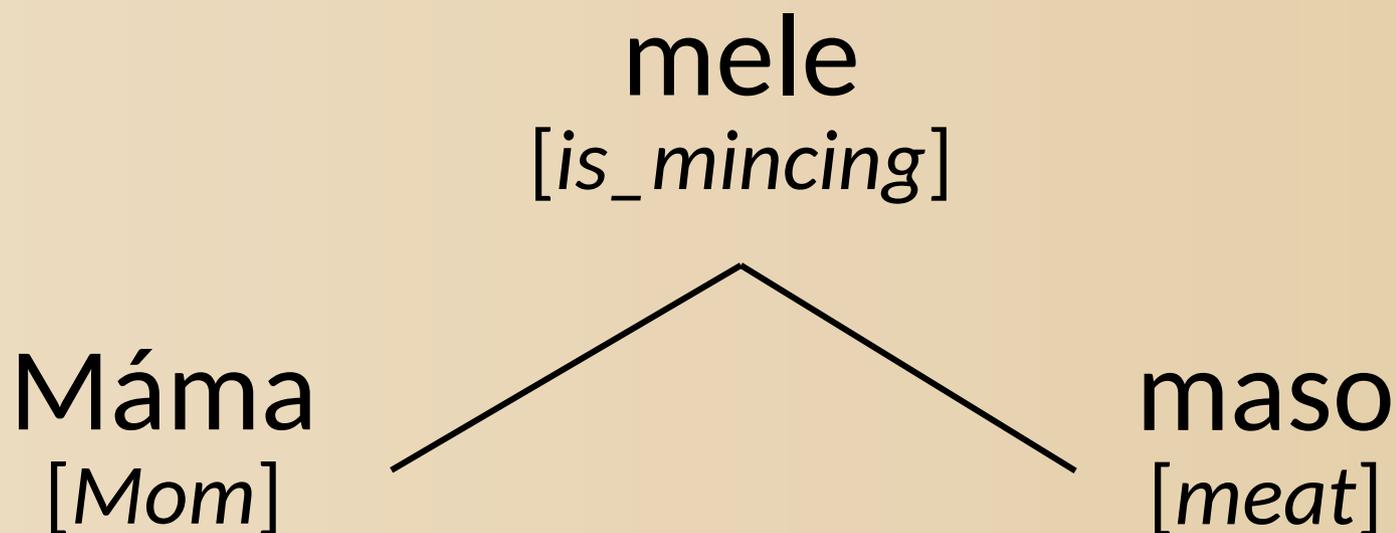
maso
[meat]

Máma [*Mom*] – Subject

mele [*is_mincing*] – Predicate

maso [*meat*] – Object

PDT – Analytical layer

mele
[*is_mincing*]

Máma
[*Mom*]

maso
[*meat*]

Máma [*Mom*] – Subject

mele [*is_mincing*] – Predicate

maso [*meat*] – Object

PDT – Tectogrammatical layer

mele
[*is_mincing*]
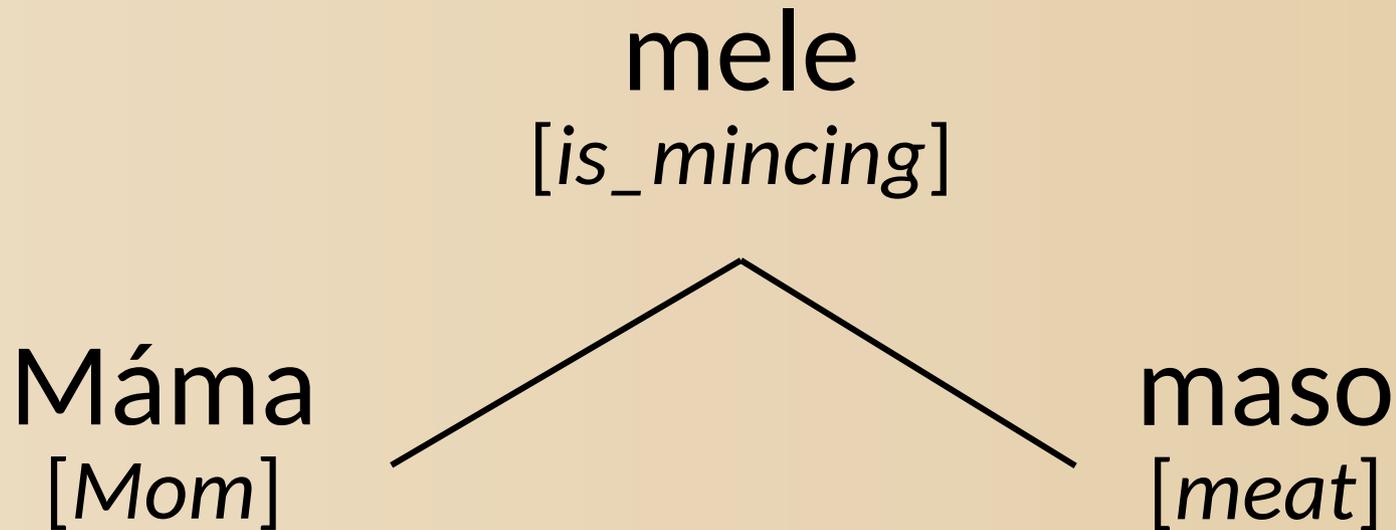
Máma
[*Mom*]

maso
[*meat*]

Máma [*Mom*] – Actor

mele [*is_mincing*] – Predicate

maso [*meat*] – Patiens

a-layer

AuxS

Můžete [Can_you] Pred   ? AuxK

vysvětlit [explain] Obj

to [it] Obj   na [on] AuxP

příkladu [an_example] Adv

t-layer

root

vysvětlit.inter [to_explain] PRED

ten [it] PAT   #PersPron ACT   #Gen ADDR   příklad [example] MEANS

m-layer

Můžete	to	vysvětlit	na	příkladu	?
[Can_you]	[it]	[explain]	[on]	[an_example]	
moci	ten	vysvětlit	na	příklad	?
VB-P---2P-AA---	PDNS4----------	Vf-------A----	RR--6----------	NNIS6-----A---1	Z:-------------

w-layer

Můžete	to	vysvětlit	napříkladu	?
[Can_you]	[it]	[explain]	[onan_example]	