Class 07: Semantic roles and PropBank

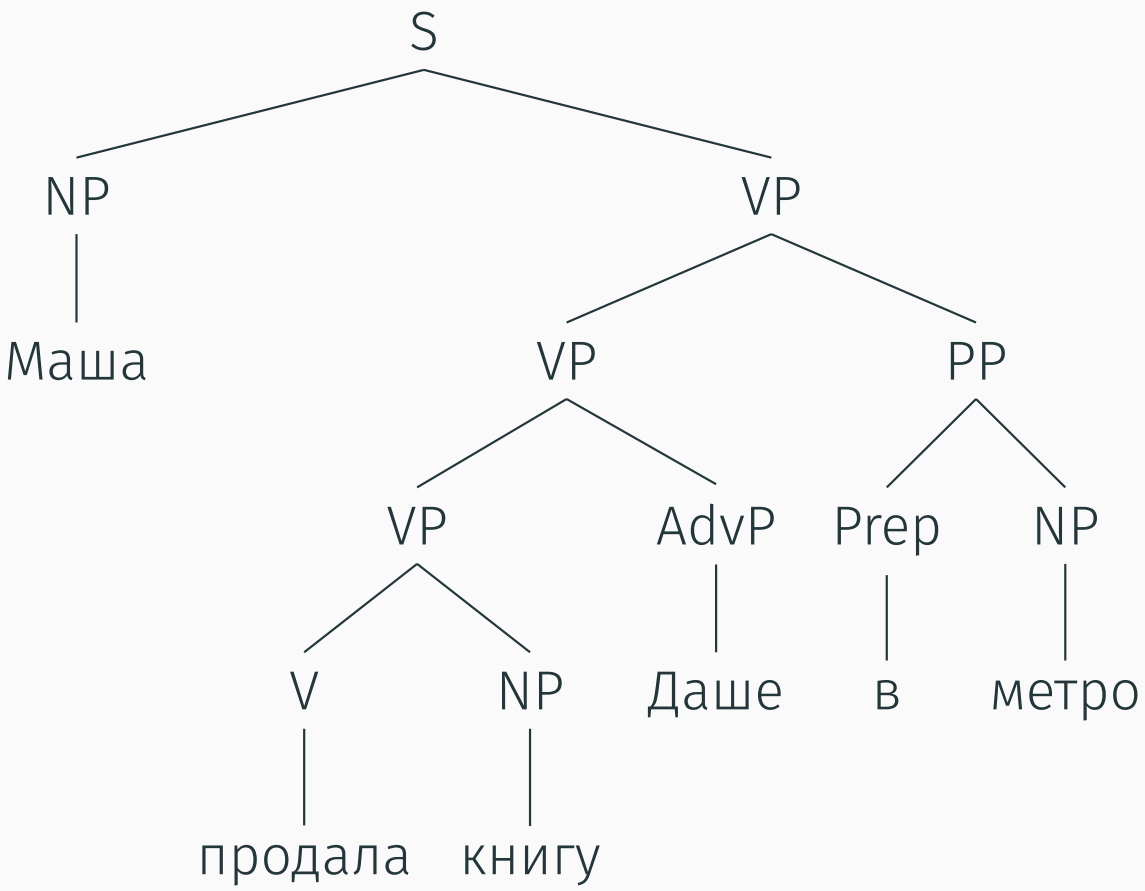
Introduction

Кто ?	сделал что ?	кому ?	где ?
Маша	продала книгу	Даше	в метро
Maša	*sold the book*	*to Daša*	*on the metro*

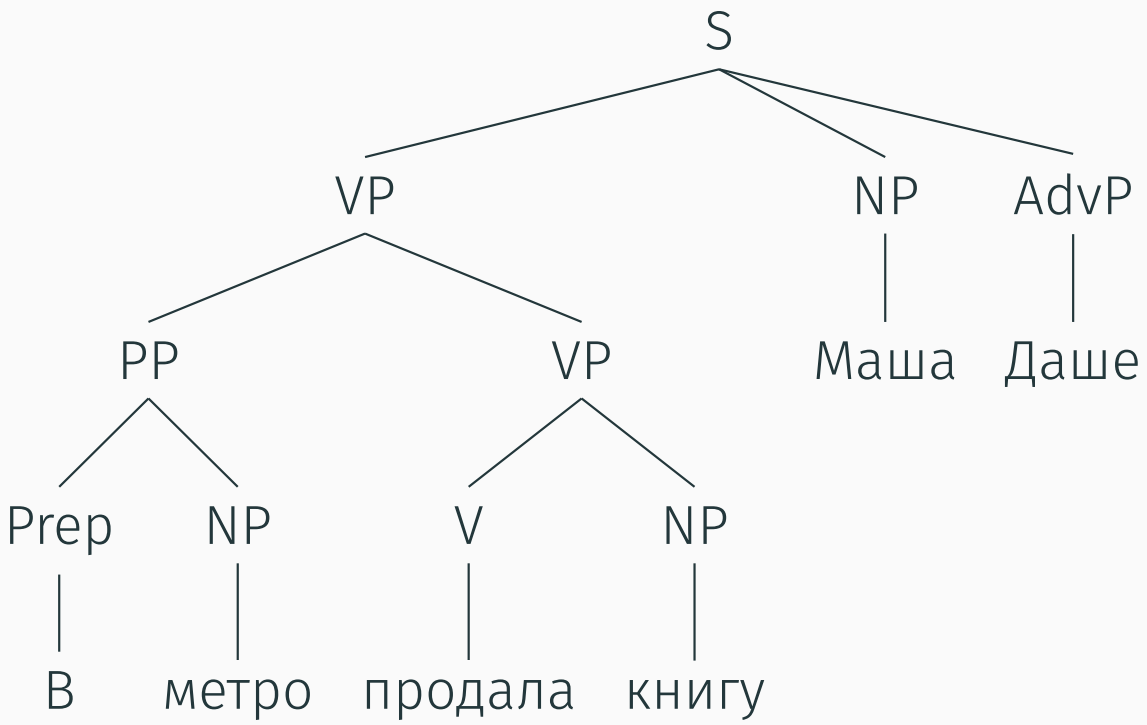
- Что кто-то сделал ?
- Кто продал книгу?
- Кому продала Маша книгу ?
- Где Маша продала книгу ?

- Question answering
- Machine translation

Syntax/1

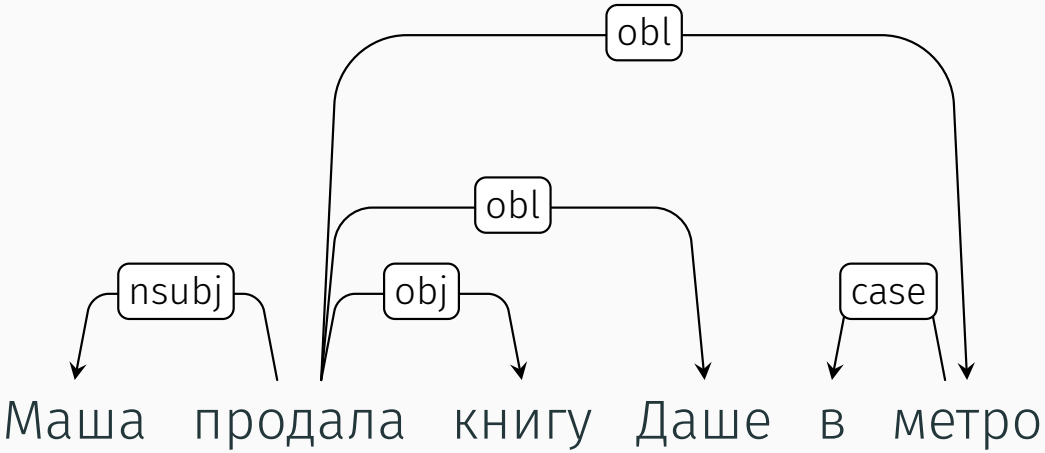

S
NP
VP
Маша
VP
PP
VP
AdvP
Prep
NP
V
NP
Даше
в
метро
продала
книгу

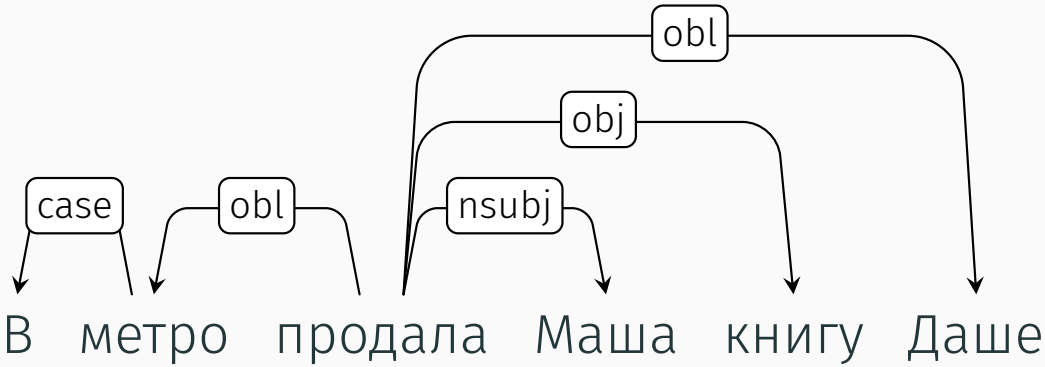
Syntax/1

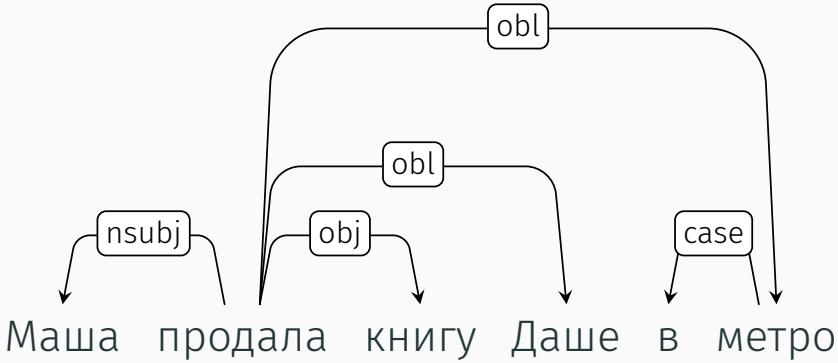


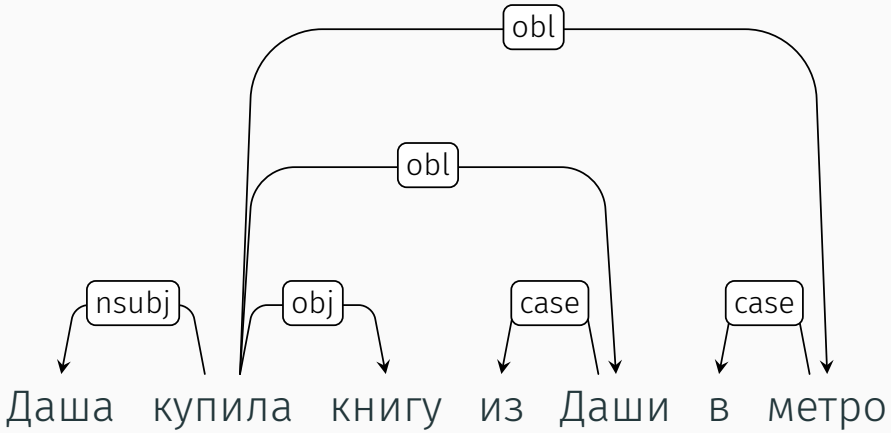
Syntax/2

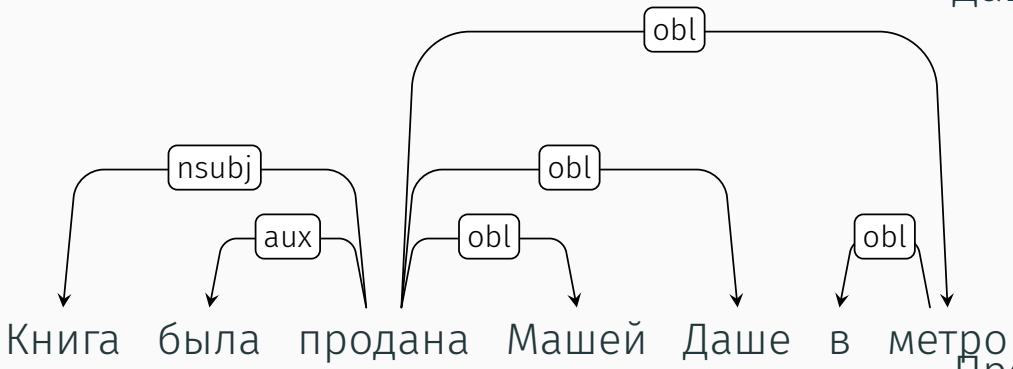
Doesn’t dependency parsing solve this ?

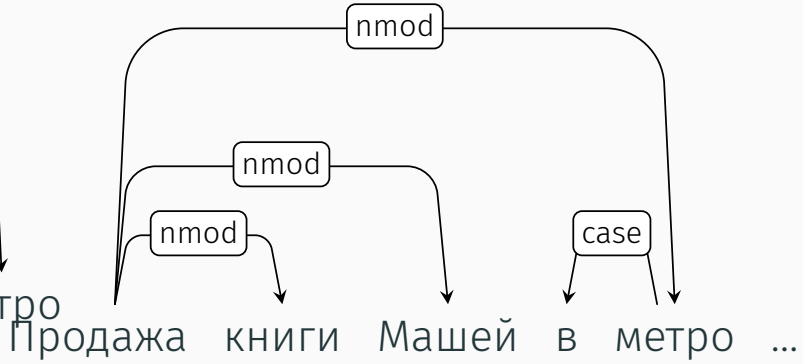





obl
obl
nsubj
obj
case
Маша продала книгу Даше в метро


obl
obl
nsubj
obj
case
case
Даша купила книгу из Даши в метро


obl
nsubj
obl
aux
obl
obl
Книга была продана Машей Даше в метро


nmod
nmod
nmod
case
Продажа книги Машей в метро ...

Semantic roles

Shallow representation

Predicates (like продать, купить) represent an **event**.

Semantic roles (like Agent, Theme) express the abstract role of the arguments of the predicate.

Buyer	Agent	Proto-Agent
$\leftarrow$		$\rightarrow$
More specific		**More general**

Deep roles

Specific for a predicate,

- Maša broke the window
- Saša opened the door

Subjects of *break* and *open*: **Breaker** and **Opener**

The objects are: **BrokenThing** and **OpenedThing**

Hard to reason with for applications

Thematic roles/1

But both **Breaker** and **Opener** have something in common:

- Volitional actors
- Often animate
- Direct causal responsibility for their events

Thematic roles capture this similarity,

- **Breaker** and **Opener** are both AGENTS
  - Volitional actors with causal responsibility for an event
- **BrokenThing** and **OpenedThing** are both THEMES
  - Inanimate objects affected in some way by an action

Thematic roles/2

One of the first linguistic models:

- Introduced by the grammarian Pāṇini between the 7th and 4th centuries BCE
- Called *kāraka* in Sanskrit/Indo-Aryan linguistics

Modern formulation by Fillmore (1966):

- Influenced by Tesnière (1959)'s dependency syntax
- Called first *actants* (following Tesnière) and then later *case*

The terminology is confusing.

Thematic roles/3

Role	Definition
AGENT	The volitional causer of an event
	Маша разбила окно
EXPERIENCER	The experiencer of an event
	Саше болеет голова
FORCE	Non-volitional causer of an event
	Ветер сдувал снег
THEME	Participant most directly affected by an event
	Маша продала **книгу**
INSTRUMENT	An instrument used in an event
	Она написала письмо **ручкой**
BENEFICIARY	The beneficiary of an event
	Я купил **тебе** кофе
SOURCE	Origin of a transfer event
	Ты не приехала **из Кызыла?**
GOAL	The destination of a transfer event
	Я хочу **в Якутск**

Thematic «grid»

разбить:

- AGENT
- THEME
- INSTRUMENT

Realisations:

- AGENT/Subject THEME/Object
- AGENT/Subject THEME/Object INSTRUMENT/NP_{ins}
- THEME/Subject

Маша	*разбила*	*окно*	
AGENT		THEME	
Маша	*разбила*	*окно*	*молотком*
AGENT		THEME	INSTRUMENT
? *Молоток*	*разбил*	*окно*	
INSTRUMENT		THEME	
Окно	*разбилось*		
THEME			
Окно	*было*	*разбито*	*Машей*
THEME			AGENT
Окно	*было*	*разбито*	*молотком*
THEME			INSTRUMENT

Problems

Very hard to create a standard set of roles or formally define them.

For example for INSTRUMENT,

- **intermediary instruments** can appear as subjects:
  - The cook opened the jar with the new gadget
  - The new gadget opened the jar
- **enabling instruments** cannot:
  - They ate rice with chopsticks
  - *The chopsticks ate rice

Alternatives

Buyer Agent Proto-Agent

$\leftarrow$ $\rightarrow$

FrameNet **More roles** **Fewer roles** PropBank

PropBank:

- Generalised roles defined as prototypes

FrameNet:

- Define roles specific to a group of predicates

Pause for thought:

- If we want to use this in a practical NLP system, does the label matter or does the distribution matter?
- If we can generalise over different things that look different but refer to the same event (buy, sell; kick, is kicked) does the precise formalism matter?

PropBank and FrameNet

PropBank/1

A **PropBank**[1] is a corpus annotated with predicates and arguments

The English PropBank:

- Annotated on top of the Penn Treebank
- Not freely available

Uses numbered arguments:

- Arg0: PROTO-AGENT
- Arg1: PROTO-PATIENT
- Arg2: BENEFACTIVE, INSTRUMENT, ATTRIBUTE END STATE
- ...

PropBanks exist for: English*, Chinese*, Arabic*, Finnish, Russian?

[1] Martha Palmer, Daniel Gildea and Paul Kingsbury (2005) "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics* 31(1):71–106

PropBank/2

Proto-Agent:[2]

- Volitional involvement in event or state
- Sentience (and/or perception)
- Causes an event or change of state in another participant
- Movement (relative to position of another participant)

Proto-Patient:

- Undergoes change of state
- Causally affected by another participant
- Stationary relative to movement of another participant

[2] David Dowty (1991) "Thematic Proto-Roles and Argument Selection". *Language*, 67(3) pp. 547–619.

PropBank/3

PropBank/4

PropBank comes with **frame files** which contain predicates and their argument structure.

ostaa.1

Buy, purchase

tags:model:buy.01

Ostaa

A0	Entity, who buys	Ostaja
A1	Thing bought	Ostettu esine, asia tai palvelu
A2	Entity, who sells	Myyjä
A3	Price paid	Ostohinta
A4	Beneficiary	Hyötyjä

- Finnish PropBank is freely available
- `https://github.com/TurkuNLP/Finnish_PropBank` (data branch)

PropBank/4

PropBank comes with **frame files** which contain predicates and their argument structure.

tykätä.1

To like someone or something

(tags: model:like.01, seed:tykätä, colloquial)

Pitää jostakin

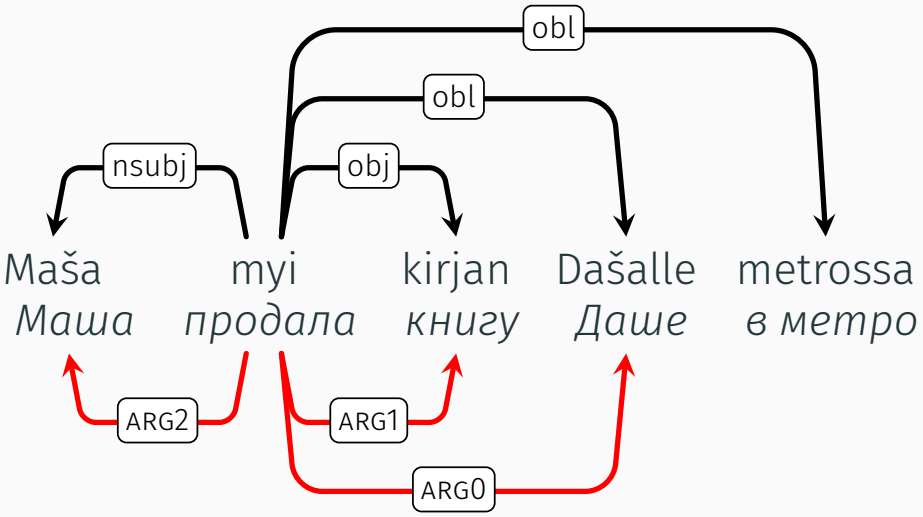
(tags: model:like.01, seed:tykätä, colloquial)

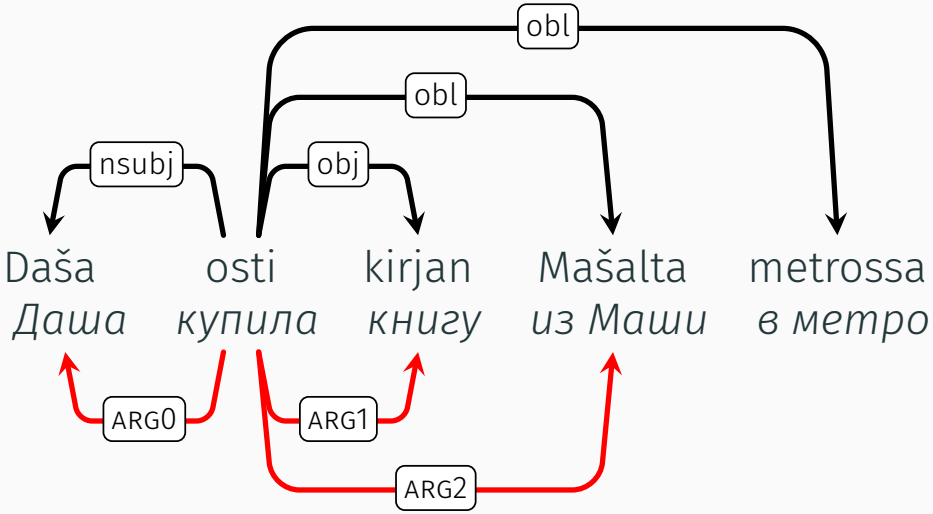
A0	Creature feeling affection	Olento, joka tuntee kiintymystä
A1	Object of affection	Kiintymyksen kohde

- Finnish PropBank is freely available
- `https://github.com/TurkuNLP/Finnish_PropBank` (data branch)

PropBank/5

PropBank-style annotation allows us to see commonalities:





PropBank/6

Summary:

- A propbank is a corpus annotated with predicate–argument structure
- Predicate–argument structure generalises over syntax
- There is a free PropBank for Finnish

But how about Russian?

- There is a semantically-annotated corpus based on FrameNet
- It could be converted into a PropBank
- For more info ask Olya Lyashevskaya

FrameNet/1

FrameNet is very popular:

- Semantically-annotated database/electronic resource

It contains (for English):

- 1,200 frames
- 13,000 lexical units (word–meaning correspondence)
- 202,000 example sentences

FrameNet/2

Frames:

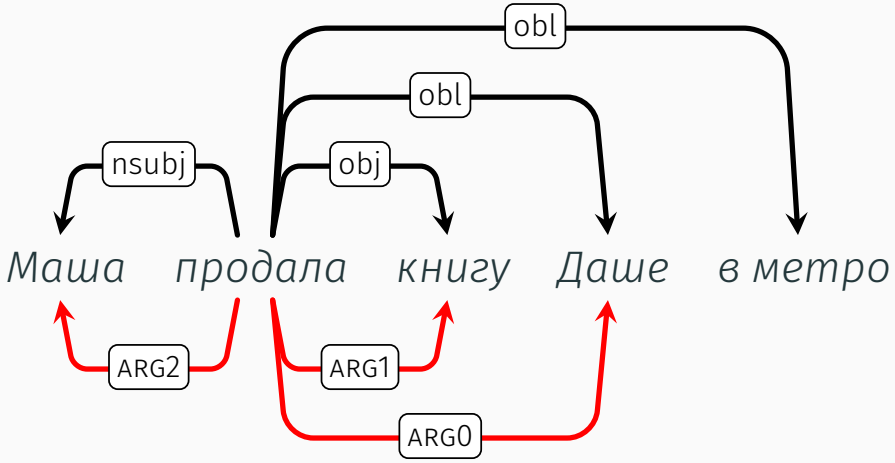
- Conceptual structure involving participants, events and background knowledge
- Extremely specific, e.g.
  - `Commerce_goods-transfer`
  - `Being_born`
  - `Criminal_process`

Frame elements:

- **Core:** essential to the meaning of the Frame
  - `Seller`, `Buyer`, `Goods`
- **Non-core:** descriptive, e.g. time, place, manner
  - `Place`, `Purpose`

vs. PropBank

PropBank:



FrameNet:

Маша	*продала*	*книгу*	*Даше*	*в метро*
`Seller`		`Goods`	`Buyer`	
	продать.1			
	`Commerce_goods-transfer`			

Semantic role labelling

Semantic role labelling

A generic algorithm:

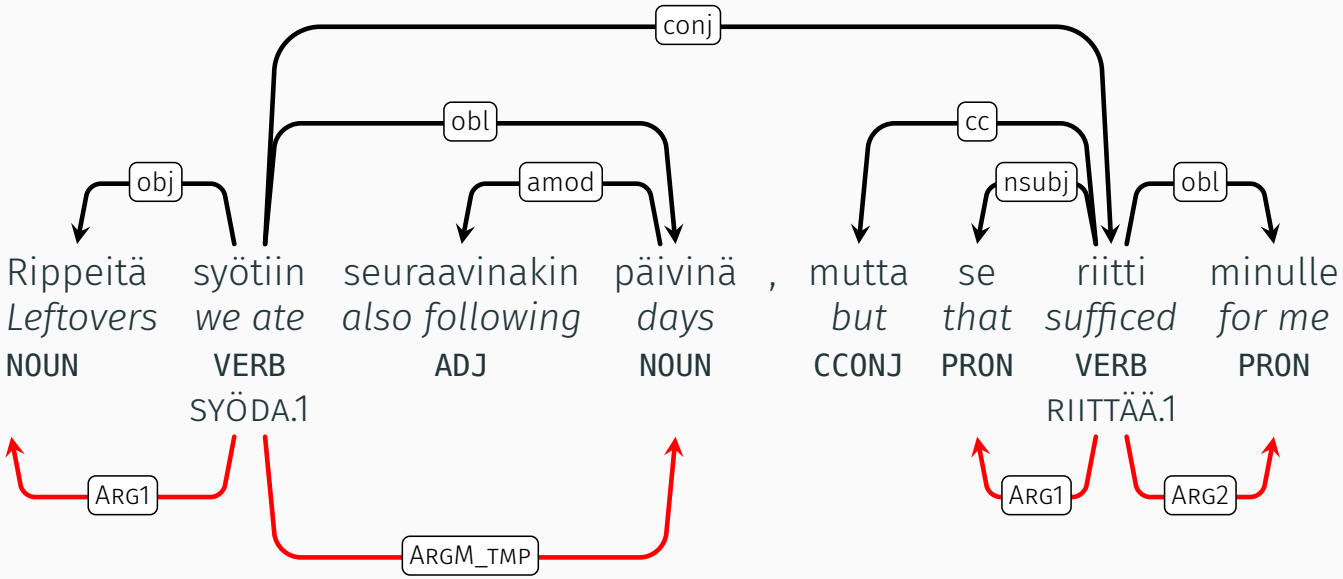
```
function SEMANTICROLELABEL(words) returns labeled tree

    parse ← PARSE(words)
    for each predicate in parse do
        for each node in parse do
            featurevector ← EXTRACTFEATURES(node, predicate, parse)
            CLASSIFYNODE(node, featurevector, parse)
```

How do we decide what is a predicate ?

- **PropBank**: Use the verbs
- **FrameNet**: Use what was labelled as such in the training data

Features



Headword of constituent	Rippeitä
Headword POS	NOUN
Headword Morph. features	Case=Par
Voice of clause	Active
Linear position (wrt. predicate)	before

Practical

- Download Finnish PropBank
  - `https://github.com/TurkuNLP/Finnish_PropBank` (data branch)
- Write a semantic role labeller
- Train on `train`, find good feature combination on `dev` and test on `test`.