

What is Inside Llewyn Davis? Inner Structure of MWEs



Eduard Bejček, Pavel Straňák
Charles University in Prague, MFF,
Institute of Formal and Applied Linguistics
{bejcek, stranak}@ufal.mff.cuni.cz

Deep syntactic trees...

Are all the relations of the same kind?

such a lovely flower
great experience
at least once
salt and pepper

South Carolina

ad hoc

high school

George Washington

Metro-Goldwyn-Mayer from 5 to 8%

criss cross

dead tired

New York

housing project

Goethe Universität
Grüneburgplatz 1
Frankfurt am Main,
Germany

NASA

Kim Ir Sen

Venice Film Festival

September 2014

Dependency relations inside a collocation (MWE)

Different or no relations inside MWE

Collocation

MWE

Location
Object
Institution

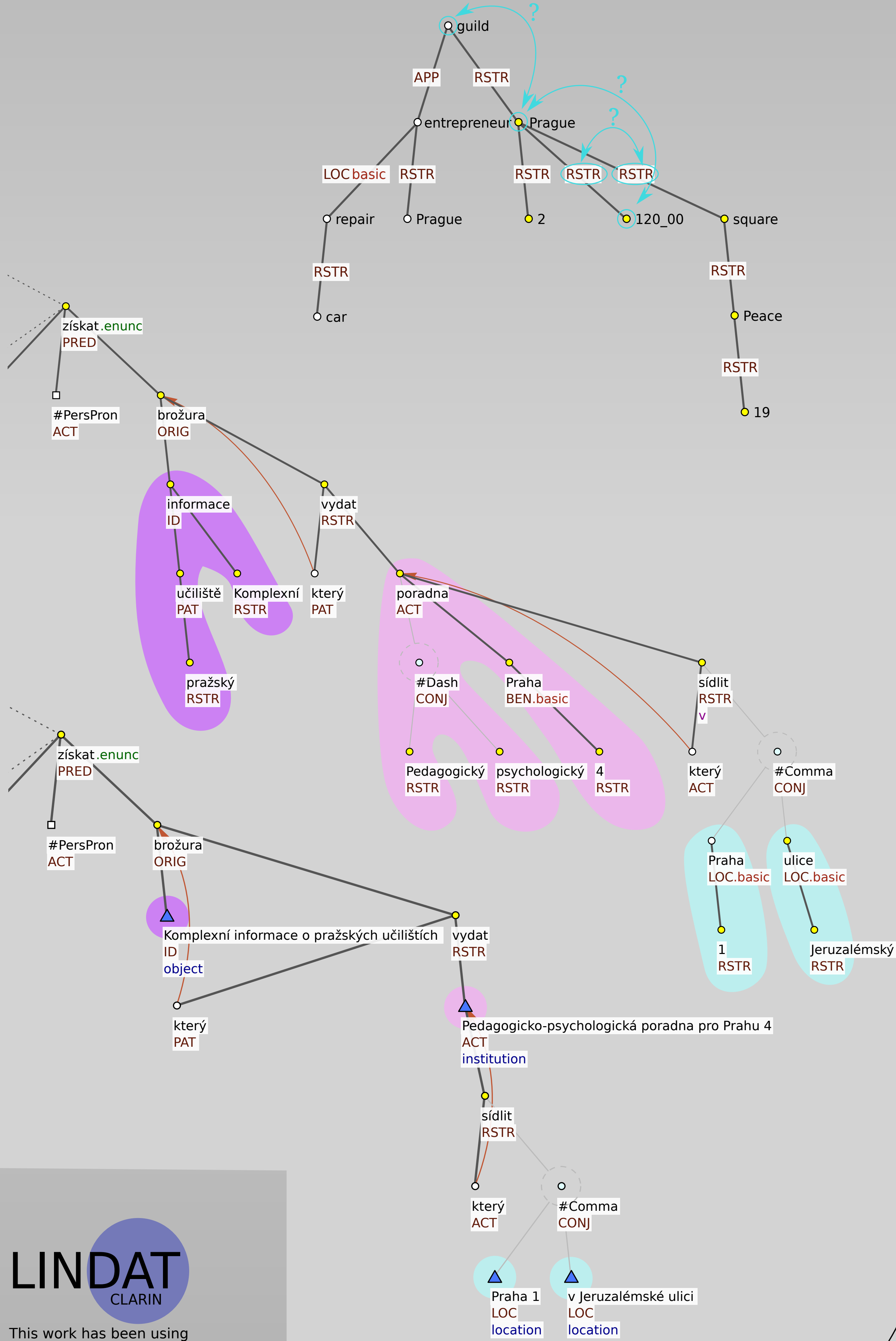
Lexeme

Person
Foreign

Numbers
Time

Address
Biblio

Prague Dependency Treebank (past, present, future)



part of the lexicon
no standard components
represented as: an entry
- atomic
- or with nested NE -- links
- pragmatic attributes:
- type (river, country, street, ..., or book, law, chemical, unit, ...)
- location (GPS coord.)
- link to an entity handle (e.g. Wikipedia)

part of the lexicon
some standard components
represented as: a (table) entry
- slots (family name, first name, title, ...)
- pragmatic attributes (e.g. link to Wikipedia)

part of the grammar
regular structure
represented as: a table
- constitutes the node in the tree
- many (empty?) slots
- embeds other NE: link to the lexicon
no "lexicon of addresses"

Lexicon (future version of SemLex)

name: George of Cambridge statue
type: object
components: George of Cambridge statue

name: George of Cambridge
type: person
components: George of Cambridge

name: Cambridge
type: location
info: city, atomic

name: George
type: person
info: first name, atomic

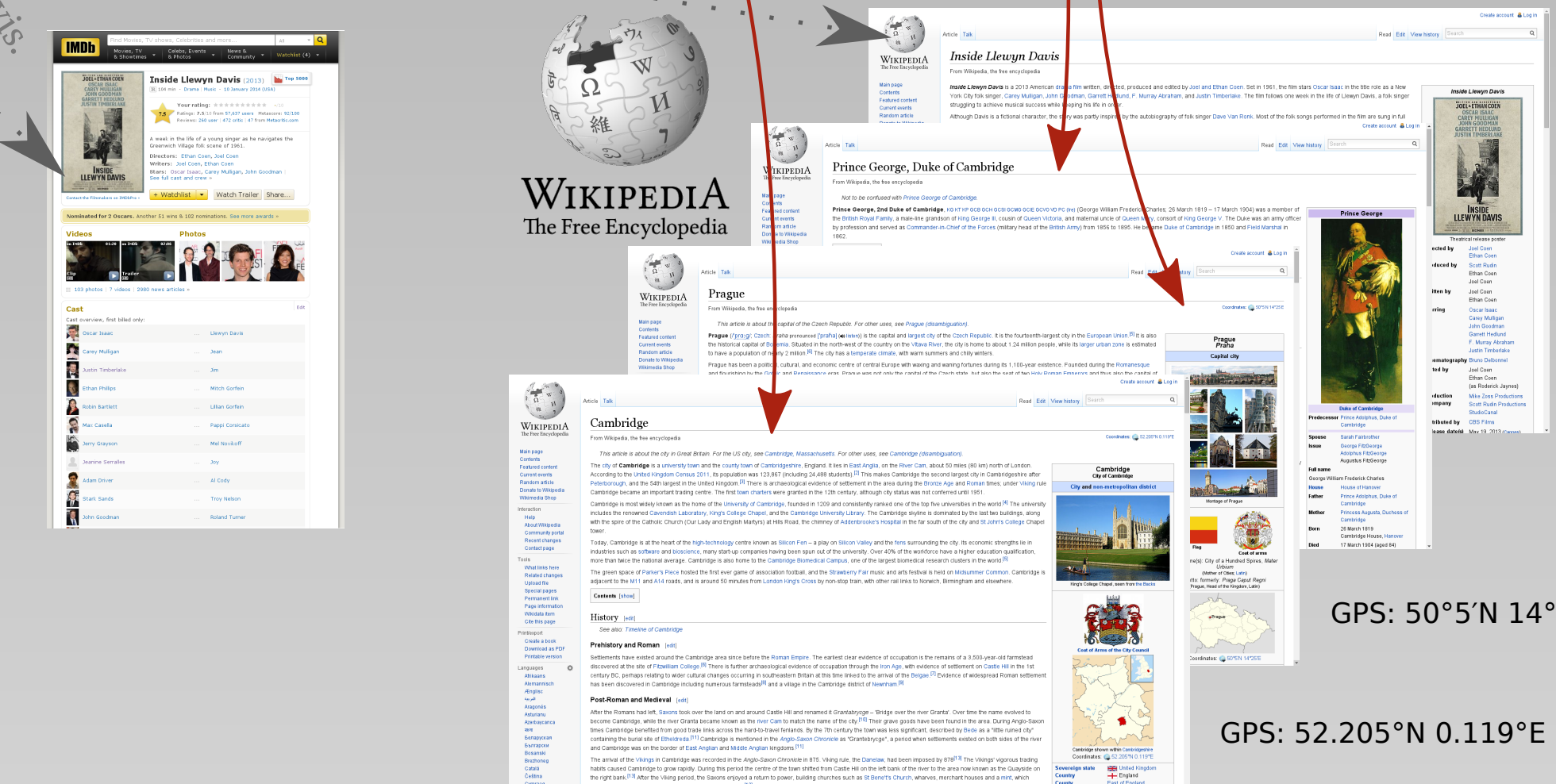
name: The Guild of Prague Entrepreneurs in Car Repair, Peace Square 19, 120 00 Prague 2.	
type:	address
components:	entity: The Guild of Prague Entrepreneurs in Car Repair
salutation:	---
street:	Peace Square
house number:	19
ZIP code:	120 00
city:	---
district:	Prague 2
state:	---
country:	---

name: Peace Square
type: location
info: square, atomic

name: The Guild of Prague Entrepreneurs in Car Repair
type: institution
components: guild, Prague entrepreneur car repair

name: Prague
type: location
info: city, GPS, atomic

External sources
(Wikipedia, IMDb, calendars, ...)



This work has been using language resources developed and stored and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).

This research has been supported by grant COST CZ LD14117 of the Ministry of Education, Youth and Sports of the Czech Republic.

The Guild of Prague Entrepreneurs in Car Repair, Peace Square 19, 120 00 Prague 2

GPS: 50°5'N 14°25'E

GPS: 52.205°N 0.119°E

WG4