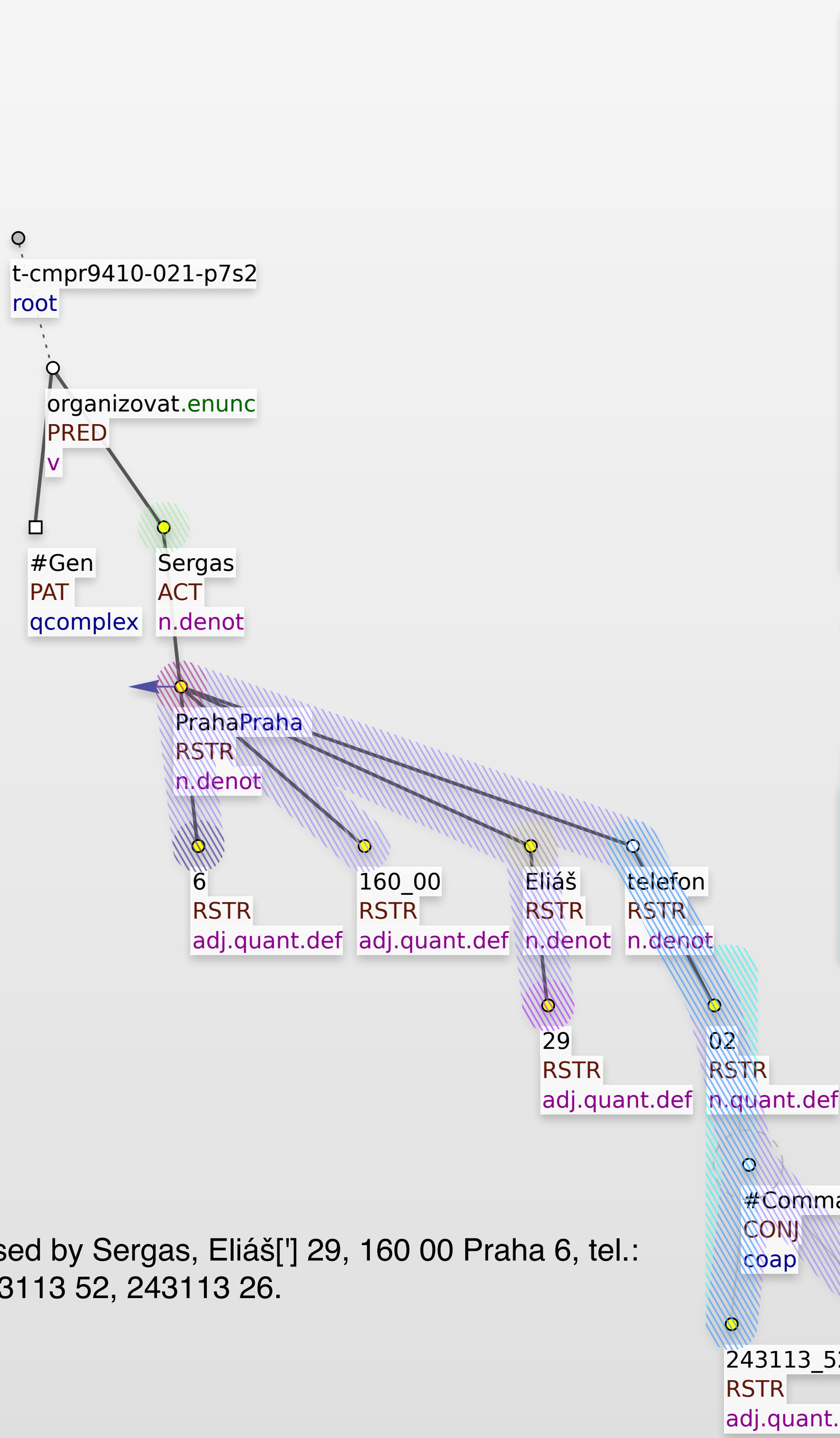


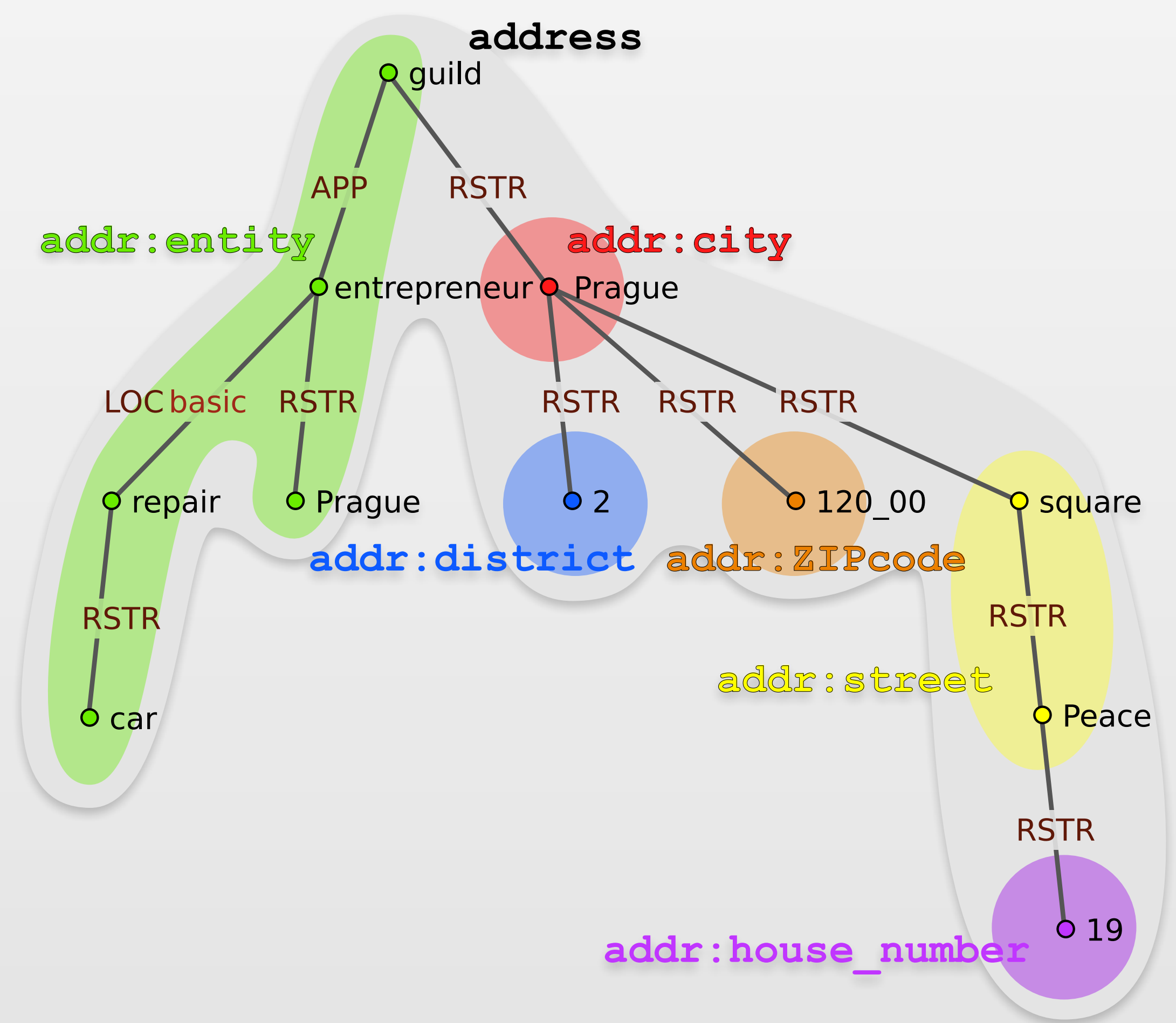


Named Entities within Multiword Addresses



name: The Guild of Prague Entrepreneurs in Car Repair, Peace Square 19, 120 00 Prague 2.
type: address

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

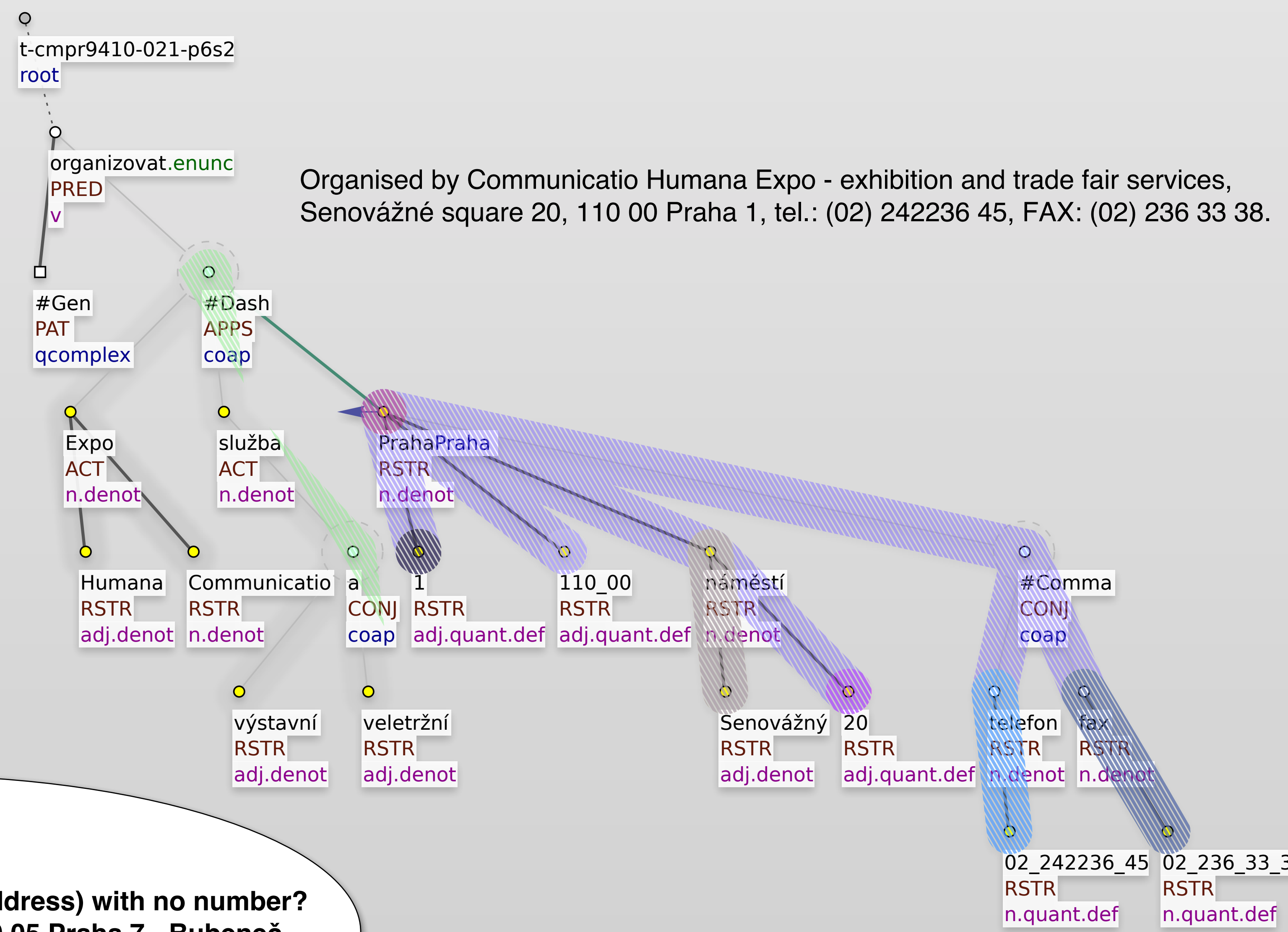


name: Peace Square
type: location
info: square, atomic

name: Prague 2
type: location
info: city district, atomic, part of 'Prague'

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

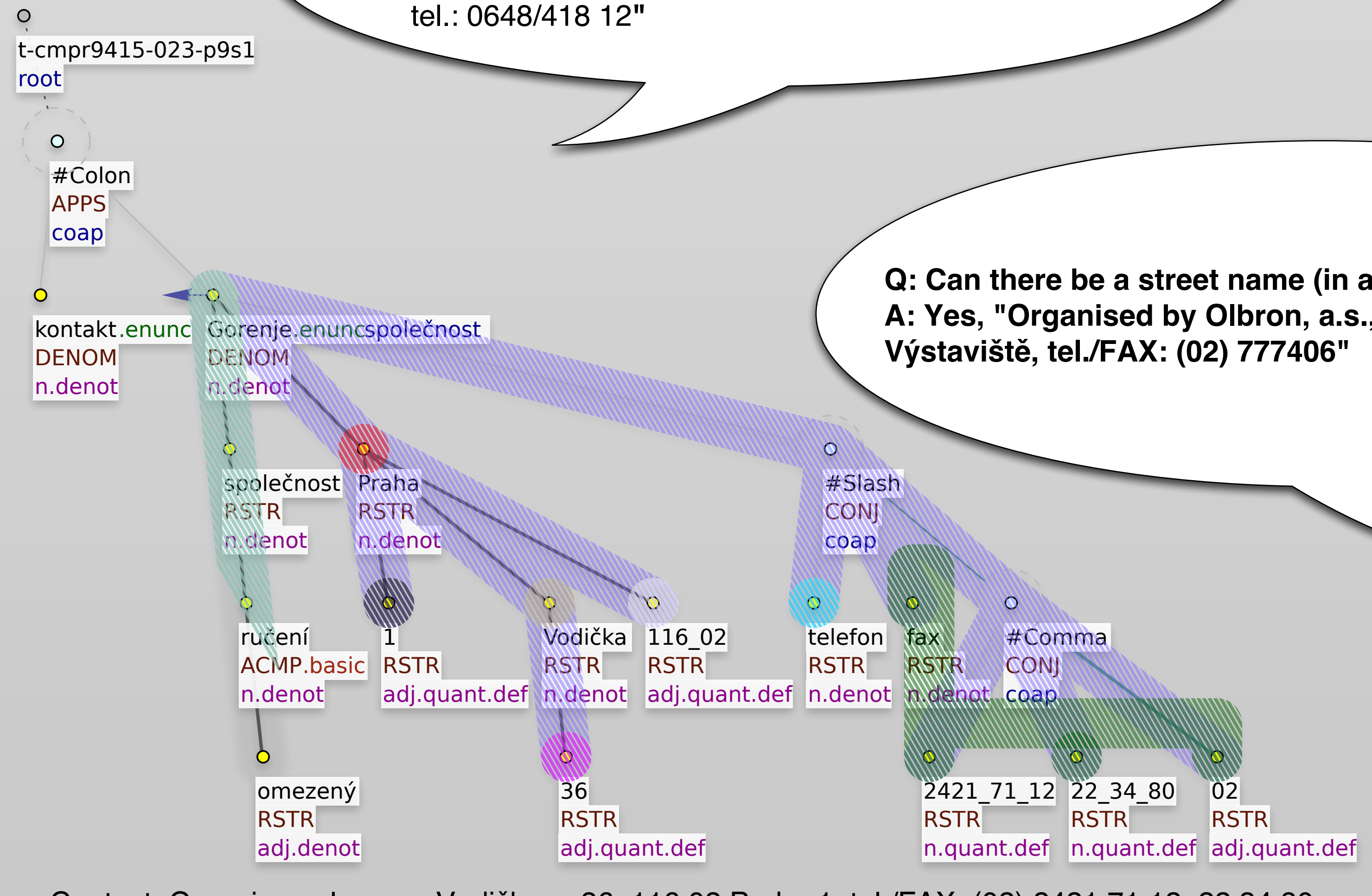
name: The Guild of Prague Entrepreneurs in Car Repair
type: institution
components: guild entrepreneur car repair



Q: What components can be repeated in 1 address?
A: Practically everything: company, name, street number, phone ...

Q: Is there something an address MUST include?
A: Not really, e.g. here firm + town + telephone:
"Produced and marketed by ZD Úsovsko established in Klopino tel.: 0648/418 12"

Q: Can there be a street name (in an address) with no number?
A: Yes, "Organised by Olbron, a.s., 170 05 Praha 7 - Bubeneč, Výstaviště, tel./FAX: (02) 777406"



Contact: Canstar Sport Inc., 5705, rue Ferrier, Suite 200, Ville Mont-Royal, Québec, Canada H4P 1N3, tel.:(001 514) 738 51 78.

124 addresses analysed (full PDT 3.0 – incl. dev and eval)

Address	Notes / experience
123	telephone ellipses, coordination problems, tel+fax one number
112	house_number not <i>street</i> number; ranges; 6/7 (what?); <i>house_name</i>
112	street
111	city
109	name-inst not atomic, very hierarchical, often unclear borders
93	zipcode missing "P.O. BOX";
86	fax telex, email, web, Skype, ???; also see telephone
75	district very hierarchical (Praha 6 (distr.) – Břevnov (sub.))
4	state country, province, county, etc.?
0	name-person
34	other apartment_number; office
859	average 7 components per address



This work has been supported by grant COST CZ LD14117 of the Ministry of Education, Youth and Sports of the Czech Republic and it has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarín project of the Ministry of Education of the Czech Republic (project LM2010013).