

MATEMATICKO-FYZIKÁLNÍ FAKULTA
PRAHA

**Srozumitelnost českých právních a administrativních dokumentů
ve výzkumu a praxi**

BARBORA HLADKÁ, SILVIE CINKOVÁ, JAN ČERNÝ, VÍTEK EICHLER, TOMÁŠ KNAP,
IVAN KRAUS, BARBORA KUBÍKOVÁ, IVANA KVAPILÍKOVÁ, JIŘÍ MÍROVSKÝ,
TEREZA NOVOTNÁ, TOMÁŠ POLÁK, ARNOLD STANOVSKÝ, JANA ŠAMÁNKOVÁ,
MICHAL KUK, PŘEMYSL POSPÍŠIL

ÚFAL Technical Report
TR-2025-75

ISSN 1214-5521



UNIVERSITAS CAROLINA PRAGENSIS

Copies of ÚFAL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 25, CZ-11800 Prague 1

Czechia

or can be obtained via the Web: <https://ufal.mff.cuni.cz/techrep>

Srozumitelnost českých právních a administrativních dokumentů ve výzkumu a praxi¹

Barbora Hladká, Silvie Cinková, Jan Černý, Vítek Eichler, Tomáš Knap, Ivan Kraus, Barbora Kubíková, Ivana Kvapilíková, Jiří Mírovský, Tereza Novotná, Tomáš Polák, Arnold Stanovský, Jana Šamánková (Univerzita Karlova)

Michal Kuk, Přemysl Pospíšil (Frank Bold Society)

Kontakt: Barbora Hladká, hladka@ufal.mff.cuni.cz

Představujeme prototyp nástroje PONK pro hodnocení srozumitelnosti českých právních a administrativních textů. Čerpá inspiraci z podobných nástrojů pro jiné jazyky, zejména z polského nástroje Jasnopis. PONK pomáhá právníkům a úředníkům s korekturou jejich textů, a to aktuálně ze čtyř hledisek: klasické metriky srozumitelnosti přizpůsobené češtině, ručně vytvořená lingvistická pravidla pro detekci potenciálně stylisticky problematických pasáží, lexikální předvídatelnost a komunikační záměry.

PONK vznikl v rámci projektu aplikovaného výzkumu *PONK - Asistent přístupné úřední komunikace* (č. TQ01000526) v období 10/2023-12/2025 za finanční podpory Technologické agentury ČR.

Klíčová slova: hodnocení srozumitelnosti, jednoduchý právní jazyk, počítačové zpracování přirozeného jazyka, aplikovaný výzkum

¹ Text je z velké části převzat a upraven z textu Readability of Czech legal and administrative documents in research and practice (Hladká a kol., 2025).

1. Proč srozumitelnost?

- *Právo na srozumitelný jazyk, závazky ČR*
- *Chtěli jsme vytvořit praktický nástroj ke zlepšení srozumitelnosti*
- *Zpráva je pro právníky, úředníky, obecně organizace pracující se srozumitelností*
- *Proč bylo třeba něco nového*

Demokratické společnosti mají odpovědnost zajistit, aby občané rozuměli svým právům i povinnostem a měli k dispozici informace nezbytné pro veřejnou diskusi. Základní právo na přístup k právu a spravedlnosti neznamena pouze formální dostupnost, ale také srozumitelnost právních předpisů. Právní systém, který je jasný a přístupný, posiluje transparentnost výkonu veřejné moci a přispívá k důvěře občanů ve stát. V tomto kontextu představujeme PONK, softwarový nástroj určený k hodnocení srozumitelnosti právních a administrativních textů. Jeho název je zkratkou spojení Psaní Orientované Na Klienta. V češtině označuje slovo *ponk* pracovní stůl; obdobně PONK funguje jako digitální ponk, který analyzuje právní text z různých hledisek srozumitelnosti a upozorňuje na pasáže, jež mohou vyžadovat další pozornost.

Právní pravidla závazná pro adresáty by měla být přiměřeně čitelná. Srozumitelnost (jasnost, přístupnost nebo čitelnost) je textová funkce, která ovlivňuje, jak snadno je text srozumitelný pro zamýšleného čtenáře. Z lingvistického hlediska jde o stylistickou záležitost, zahrnující gramatické a lexikální volby. Pro právníky je stylistika sice také důležitá, ale je pouze jedním aspektem přiměřeného vyjadřování. Ústředními principy přiměřeného právního vyjadřování jsou pojmová srozumitelnost, preciznost, jednoznačnost a logická koherence.

Další důležitou okolností je, že adresátu práva, typicky běžnému občanovi, je právo obvykle komunikováno skrz sdělení úřadů a soudů, tedy prostřednictvím správních či soudních rozhodnutí a podobných dokumentů. V takových případech se obvykle jedná o texty aplikující a interpretující normativní právní pravidla na konkrétní skutkové okolnosti situace, ve které se adresát nachází. Velmi často pak takové dokumenty ze své podstaty bývají adresátovi bez právního vzdělání těžko srozumitelné.

Mnoho zemí, zejména anglicky mluvících a severských, má tradici jednoduchého právního psaní, někdy dokonce zakotveného v zákoně. Česká vláda požaduje jasný jazyk v zákonech (Úřad vlády České republiky, 2025) a zavázala se zavést jasnou komunikaci mezi správou a veřejností do roku 2030 (Mlsna a Sláma, 2019).

Jazykoví odborníci v právní i pedagogické oblasti také aktivně propagují a vyučují jednoduché, klientsky orientované psaní v rámci svých profesních komunit. V českém

kontextu to vedlo k vývoji několika referenčních prací šitých na míru místnímu publiku (Behún a Behún, 2023; Šamánková a Kubíková, 2022; Slejšková, 2013; Šváb, 2021; Vučka, 2019). Zejména Garnerovo *Právní psaní srozumitelným jazykem* (Garner, 2021), klíčový text široce uznávaný napříč právními systémy pro svůj vliv na praktiky srozumitelného jazyka, byl se značnou péčí přizpůsoben českému právnímu a lingvistickému prostředí a rozšířil jeho použitelnost daleko za jeho původní anglo-americký rámec.

Se snahami zdola i shora zlepšit komunikaci mezi veřejnou správou a občany je přirozená poptávka po automatické pomůcce pro psaní, která by dokázala posoudit kvalitu právního/správního dokumentu. Takovou pomůckou je nástroj PONK.

2. Nástroje a data, které můžete použít

V rámci projektu jsme vyvinuli nejen nástroj PONK, ale i další nástroj MasKIT spolu s několika datovými sadami, které mají společný cíl, a sice usnadnit práci s českými právními a administrativními texty.

2.1 PONK – Váš asistent pro kontrolu srozumitelnosti vašich textů

- *Co to je a jak to funguje*
 - *Pravidla*
 - *Umělá inteligence bez velkých jazykových modelů*
 - *Lexikální překvapení s velkými jazykovými modely*
 - *Rétorika s velkými jazykovými modely*
- *Jak používat PONK (webové rozhraní + API)*
- *Bezpečnost vašich dat (žádné ukládání, nezávislost na externích službách)*

PONK je softwarový nástroj, který analyzuje vámi nahraný text z hlediska různých aspektů srozumitelnosti a zvýrazní pasáže, které mohou vyžadovat další pozornost.

PONK přistupuje k analýze textu systematicky v několika krocích. Po nahrání textu nejprve proběhne automatická analýza pomocí nástroje UDPipe,² která rozpozná jednotlivá slova, jejich slovní druhy, pády a syntaktické vztahy mezi nimi. Současně nástroj NameTag identifikuje pojmenované entity, např. jména osob, organizací, zeměpisné názvy a časové údaje.³ Na základě takto obohaceného textu pak PONK provádí čtyři typy hodnocení srozumitelnosti:

² <https://lindat.mff.cuni.cz/services/udpipe/>

³ <https://lindat.mff.cuni.cz/services/nametag/>

1. Klasické metriky srozumitelnosti

PONK počítá pro texty automatizovaný index čitelnosti ARI upravený pro češtinu.⁴ Tento index vychází z průměrné délky slov a vět. Delší slova a delší věty obvykle znamenají méně srozumitelný text. Kromě indexu ARI PONK zobrazuje další charakteristiky: průměrnou vzdálenost mezi slovesy, poměr sloves k podstatným jménům a přídavným jménům či lexikální bohatost textu. Tyto charakteristiky poskytují orientační pohled na celkovou náročnost textu. Samy o sobě však nemohou postihnout specifika právního jazyka, proto slouží především jako doplňkový ukazatel. Jde nicméně o základní a zpravidla snadno odstranitelné překážky čitelnosti.

The screenshot displays the PONK web application interface. At the top, there's a header with the PONK logo, server information (Server: 0.61 20260128 (bez ukládání textů), stav: online), and navigation links (O programu, Spustit, REST API, and a flag icon). Below the header, there's a main content area with a green bar containing 'A. Text', a trash icon, a download icon, and a 'Zpracovat' button. The main text area contains a sample text snippet about a summons. To the right, there's a sidebar with tabs for 'Míry', 'Pravidla', 'Překvapení', and 'Role'. The 'Míry' tab is active, showing 'Míry textu jako celku' with a color-coded background (green for good, orange for average, red for poor). Below this, there's a list of metrics: 'Automatizovaný index čitelnosti: 9.9', 'Vzdálenost sloves: 24', 'Aktivita: 0.43', and 'Slovní bohatství: 0.89'. At the bottom, there's a section for 'PONK 0.61 20260128 (bez ukládání textů)' with processing details like 'počet vět: 3, slov (vč. interp.): 65', 'Doba zpracování: 12.2 s', and a list of sub-tasks and their durations.

2. Lingvistická pravidla

Jádrem lingvistického modulu je sada ručně vytvořených lingvistických pravidel. Tato pravidla byla navržena ve spolupráci s právními a jazykovými experty tak, aby zachytily typické problémy českých úředních textů:

- Syntaktická složitost – pravidla detekují příliš dlouhé věty, velké vzdálenosti mezi podmětem a přísudkem, hluboce vnořené vedlejší věty.
- Gramatické jevy – nadužívání trpného rodu, vícenásobná negace, řetězení předložkových vazeb.
- Lexikální problémy – nadměrné používání abstraktních slov, přehnané hromadění přídavných jmen u jednoho podstatného jména, nadbytečné formální obraty.

⁴ ARI – Automated Readability Index

Každé pravidlo, které text analyzuje, je v rozhraní PONKu zvýrazněno. Uživatel může jednotlivá pravidla zapínat a vypínat podle potřeby. Po najetí myši na zvýrazněnou pasáž se zobrazí název pravidla a vysvětlení, proč byla daná část textu označena.

A Text [Trash] [Download] [Zpracovat]

Jelikož žádost trpěla podstatnými vadami a nebylo možno je odstranit na místě, byl k odstranění vad žádosti úředníkem vyzván žadatel. Žádost v důsledku toho do vypršení zákonné třicetidenní lhůty bohužel vyřízena nebyla.

Dlouhé výrazy: Lépe „proto“
[Fix]

Pravidla | Míry | Překvapení | Role

Gramatická pravidla
Kliknutím můžete pravidlo zapnout či vypnout. Slova zasažená více aktivními pravidly mají šedivé pozadí. (Více informací v dokumentaci.)

- ☒ Dlouhé výrazy
- ☒ Opisné pasívum
- ☒ Vzdálenost mezi přísudkem a podmětem

3. Lexikální překvapení

Další modul měří takzvané lexikální překvapení, tj. jak předvídatelné je pro jazykový model dané slovo vzhledem k předchozímu kontextu. Například ve spojení „S přáním dobrého...“ je slovo „dne“ méně překvapivé než slovo „měsíce“. Čím překvapivější slovo, tím větší kognitivní úsilí musí čtenář vynaložit na jeho zpracování.

PONK zobrazuje lexikální překvapení jako teplotní mapu: teplejší barvy značí vyšší překvapení, tedy potenciálně problematická místa. Tento modul využívá jazykový model GPT-2, který byl natrénován na velkém množství obecných textů. Právě proto může označit jako „překvapivá“ i slova, která jsou v právním kontextu běžná, což je informace sama o sobě cenná, protože ukazuje, kde se právní jazyk vzdaluje běžné češtině.

PONK Server: 0.61 20260128 (bez ukládání textů), stav: online [O programu] [Spustit] [REST API] [UK]

A Text [Trash] [Download] [Zpracovat]

Jelikož žádost trpěla podstatnými vadami a nebylo možno je odstranit na místě, byl k odstranění vad žádosti úředníkem vyzván žadatel. Žádost v důsledku toho do vypršení zákonné třicetidenní lhůty bohužel vyřízena nebyla.

Překvapení | Míry | Pravidla | Role

Lexikální překvapení
Distribuce a barevný kód různých úrovní lexikálního překvapení. Kliknutím v tabulce dole vyberete minimální úroveň pro zobrazení v textu. (Více informací v dokumentaci.)

Bar chart showing surprise levels across 16 segments. The heatmap below shows the distribution of surprise levels across 16 segments, with colors ranging from blue (low) to red (high).

4. Rétorika

Rétorický modul rozpoznává základní typy komunikačních záměrů (rolí), například radu, výzvu, nebo argumenty a jejich podklady v zákonných normách a ve skutkové podstatě. Tato úloha zahrnuje jednak určení správného rozsahu textové pasáže, jednak její správnou klasifikaci. Je příliš komplexní na to, aby se dala zachytit pravidlově. Je to jediný modul PONKu, do kterého je experimentálně zapojen generativní jazykový model.

Server: 0.61 20260128 (bez ukládání textů), stav: online

O programu Spustit REST API

A Text

Obdrželi jste výzvu k úhradě sankce za porušení veřejného pořádku podle zákona č. 251/2016 Sb., o některých přestupcích. Výzva se týká údajného přestupku spáchaného dne 15. listopadu 2025 v Praze 3, konkrétně na Jiřího z Poděbrad. Správní orgán vás vyzývá k zaplacení pokuty ve výši 3 000 Kč do 30 dnů od doručení tohoto rozhodnutí.

Pokud s pokutou nesouhlasíte, máte právo podat odpor proti příkazu. Odpor musíte podat písemně do 15 dnů od doručení příkazu k Magistrátu hlavního města Prahy, odboru přestupků. V odporu uveďte identifikační údaje věci, své osobní údaje a stručné zdůvodnění, proč s pokutou nesouhlasíte.

Po podání odporu správní orgán zruší příkaz a zahájí řádné správní řízení. V rámci tohoto řízení budete mít možnost vyjádřit se k věci, navrhnout důkazy a seznámit se s celým spisem. Řízení musí být ukončeno do 30 dnů, ve složitějších případech do 60 dnů od jeho zahájení.

Doporučujeme vám shromáždit veškeré dostupné důkazy, které mohou podpořit vaši obhajobu. To může zahrnovat fotografie, videa, svědecké výpovědi nebo jiné dokumenty prokazující vaši nevinu či polehčující okolnosti. Je také vhodné konzultovat váš případ s advokátem specializujícím se na správní právo.

Podle judikatury Nejvyššího správního soudu platí, že pokuta musí být přiměřená závažnosti spáchaného přestupku. Soudy v minulosti několikrát zrušily rozhodnutí správních orgánů, kde pokuta byla nepřiměřeně vysoká nebo kde nebyly dostatečně prokázány všechny skutečnosti.

Řečové akty (role)

Barevný kód pro různé řečové akty. Kliknutím na kategorii v tabulce ji z(ne)aktivníte. (Více informací v dokumentaci.)

Situace
Kontext
Postup
Proces
Podmínky
Doporučení
Odkazy
Prameny
Nezařaditelné

Tato technologie má omezení, která je třeba mít na paměti při posuzování výstupů modulu:

- Bezpečnost dat – texty odesíláme do modelů, které jsou lokálně nainstalovány na serverech pracoviště ÚFAL MFF UK, během zpracování neopustí lokální servery a po zpracování se neukládají.
- Reprodukovatelnost – stejný text může při různých pokusech být obarven různě.
- Transparentnost – u generativních modelů není zpětně dohledatelné, proč systém navrhl konkrétní změnu.

Navzdory uvedeným omezením je provádění takových experimentů nezbytné. Generativní modely jsou již dnes reálně využívány v nástrojích, které ovlivňují práci uživatelů, a není možné je hodnotit pouze na základě teoretických předpokladů. Systematické experimentování je nutné k tomu, abychom porozuměli jejich chování, variabilitě výstupů a limitům interpretovatelnosti v praxi. Pouze prostřednictvím opakovaných experimentů s různými nastaveními lze odhalit nestabilitu, zkreslení či nečekané efekty modelů. Neprovádět takové experimenty by znamenalo přijímat výstupy těchto systémů bez dostatečného porozumění jejich povaze a rizikům.

Kde je PONK dostupný

Webové rozhraní na adrese <https://quest.ms.mff.cuni.cz/ponk> umožňuje jednoduše vložit text nebo nahrát soubor a prohlédnout si výsledky analýzy PONKu v interaktivním prostředí.

REST API umožňuje volání PONKu z vlastních aplikací. To je užitečné pro organizace, které chtějí integrovat kontrolu srozumitelnosti do svých interních systémů nebo zpracovávat větší množství dokumentů automatizovaně.

2.2 MaskIT

- *Proč jsme to taky dělali*
- *Jak funguje (vyhláška)*

Při práci s právními dokumenty často vyvstává potřeba zveřejnit nebo sdílet texty, které obsahují osobní údaje. Soudy jsou povinny publikovat svá rozhodnutí online, ale současně musí chránit soukromí účastníků řízení. Podobné dilema řeší i další instituce veřejné správy a řešili jsme jej i my při přípravě textů pro testování nástroje PONK.

MaskIT je nástroj pro automatickou anonymizaci a pseudonymizaci českých právních textů, který jsme vyvinuli právě jako doplněk k PONKu.

Ruční anonymizace je časově náročná a náchylná k chybám. Automatizovaný nástroj může významně ušetřit čas a zajistit konzistentní úroveň ochrany. MaskIT byl navržen tak, aby odpovídal požadavkům české legislativy – zejména vyhlášky č. 403/2022 Sb., o zveřejňování soudních rozhodnutí. Zároveň pseudonymizace textu vede k výstupům lépe použitelným pro trénování nástrojů umělé inteligence. MaskIT aktuálně pokrývá přibližně 20 z 35 kategorií definovaných vyhláškou č. 403/2022 Sb. Vývoj pokračuje a další kategorie budou doplňovány.

Jak MaskIT funguje

MaskIT přijímá prostý text a zpracovává ho v několika krocích:

1. Morfosyntaktická analýza – pomocí nástroje UDPipe se text rozdělí na věty a slova, každému slovu se přiřadí slovní druh a syntaktická role.
2. Rozpoznání pojmenovaných entit – nástroj NameTag identifikuje jména osob, organizací, adresy, data a další entity.
3. Sjednocení entit – protože automatické rozpoznávání entit není stoprocentní, MaskIT provádí dodatečnou kontrolu a sjednocuje klasifikaci stejných entit v rámci dokumentu.

4. Pravidlová analýza – pomocí ručně kódovaných pravidel se identifikují citlivé výrazy, které mají být anonymizovány.
5. Generování výstupu – citlivé údaje jsou nahrazeny anonymizovanými nebo pseudonymizovanými verzemi.

Vstup: *Paní Marie Nováková z Myslíkovy ulice č. 25 dostala dopis od firmy Škoda Auto, a.s.*

Příklad anonymizace

Anonymizovaný výstup: *Paní M-ŽENA-JMÉNO-1 M-ŽENA-PŘÍJMENÍ-1 z M-ULICE-1 ulice č. M-ČÍSLO-ULICE-1 dostala dopis od firmy M-FIRMA-1.*

Příklad pseudonymizace

Pseudonymizovaný výstup: *Paní Alena Pospíšilová z Květinové ulice č. 43 dostala dopis od firmy UniTechna.*

Při pseudonymizaci jsou citlivé údaje nahrazeny náhodnými, ale realisticky vypadajícími texty. Příjmení mužů a žen jsou provázána – pokud se v textu objeví pan Novák i paní Nováková, budou nahrazeni panem Pospíšilem a paní Pospíšilovou. Při navazující práci s textem je tak výhodou, že se z textu nevytratí jeho původní význam..

MaskIT byl evaluován na 53 právních dokumentech (soudní rozhodnutí, správní rozhodnutí, právní rady, žaloby, zprávy ombudsmana) ručně anotovaných studenty Právnické fakulty Masarykovy univerzity. Nástroj dosáhl přesnosti 0,64 a úplnosti 0,80 při rozpoznávání výrazů k anonymizaci. Přesnost klasifikace u správně rozpoznaných výrazů činila 0,91.

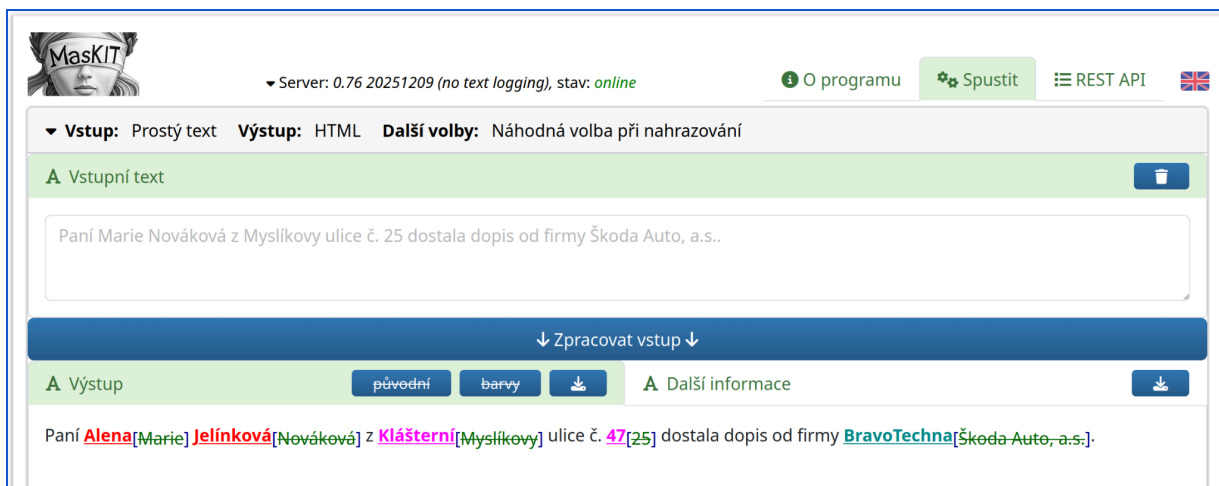
MaskIT je dostupný třemi způsoby:

Webové rozhraní na adrese <https://quest.ms.mff.cuni.cz/maskit> pro běžné uživatele a testování.

REST API pro programové volání, ideální pro hromadnější využití ze strany organizací.

Nástroj příkazové řádky pro lokální instalaci, pro využití v rámci vlastní infrastruktury.

Stejně jako PONK, ani MaskIT neukládá zpracovávané texty a garantuje jejich anonymitu.



2.3 Korpus KUK a KUKY

Pro vývoj a vyhodnocování nástrojů jako PONK a MaskIT byla nezbytná kvalitní vstupní data. V rámci projektu jsme proto vytvořili dva specializované korpusy českých právních textů.

KUK 1.0 – Korpus českých právních textů

KUK 1.0 (Korpus úřední komunikace) obsahuje 6 275 právních a administrativních dokumentů o celkovém rozsahu více než 19 milionů slov. Všechny texty mají automatickou morfosyntaktickou anotaci a anotaci pojmenovaných entit.

Texty pocházejí z několika zdrojů:

- Veřejné materiály právní poradny Frank Bold Society
- Vyjádření veřejného ochránce práv (databáze ESO)
- Informační letáky veřejného ochránce práv
- Další texty z různých zdrojů

Každý dokument je doplněn o rozsáhlá metadata navržená právními experty:

- Žánrová klasifikace – rozlišení normativních textů (zákony) od argumentačních (soudní rozhodnutí)
- Závaznost – zda je text právně závazný (soudní rozhodnutí, zákon) nebo nezávazný (informační leták, právní rada)
- Charakteristika autorů a příjemců – zda jsou právníky, zda znají kontext případu
- Snaha o srozumitelnost – u institucí se známou politikou řízení srozumitelnosti

KUKY 1.0 – Korpus s ručními anotacemi

KUKY 1.0 obsahuje téměř 250 právních textů s ručními anotacemi:

- Rétorické role – označení funkce jednotlivých částí textu (skutkový stav, právní argumentace, výrok)
- Hodnocení čitelnosti – experti posoudili srozumitelnost textů

Tento korpus slouží k trénování a vyhodnocování statistických modelů pro automatické hodnocení srozumitelnosti.

Dostupnost pro další výzkum

Oba korpusy jsou volně dostupné v repozitáři LINDAT/CLARIAH-CZ:

- KUK 1.0: <http://hdl.handle.net/11234/1-5821>
- KUKY 1.0: <http://hdl.handle.net/11234/1-5812>

Korpusy mohou sloužit pro další výzkum, ale také jako doklad toho, z jakých textů jsme vycházeli při přípravě nástrojů hodnocení srozumitelnosti.

3. Jak vám to může pomoci

3.1 Co nástroje zvládají

- *Metriky a pravidla – funkční, otestované*
- *Konkrétní příklady, které nástroj odhalí*
- *Jak začít*

Pro konkrétní představu představujeme pár příkladů problému, které dokáže PONK v textu odhalit.

Příliš dlouhé věty

"Žalovaný ve své reakci na žalobu uvedl, že s žalobcem uzavřel dne 15. 3. 2023 smlouvu o dílo, na základě které se zavázal provést práce spočívající v rekonstrukci koupelny v rodinném domě žalobce nacházejícím se v katastrálním území obce Brno, přičemž tyto práce měly být dokončeny nejpozději do 30. 6. 2023..."

Věta obsahuje přes 50 slov a obsahuje čtyři různé informace. Takové věty jsou extrémně náročné na to, aby čtenář na jeden „dech“ udržel všechny relevantní informace v hlavě.

Nadužívání trpného rodu

"Smlouva byla uzavřena dne 1. 1. 2023. Následně byla předložena k registraci."

PONK zvýrazní pasivní konstrukce, které lze nahradit aktivním tvarem: *"Strany uzavřely smlouvu... Žalobce ji předložil k registraci."*

Vzdálenost přísudku a podnětu

„Žádost v důsledku toho do vypršení zákonné třicetidenní lhůty bohužel vyřízena nebyla.“

Mezi hlavní myšlenku věty (*žádost nebyla*) je vloženo tolik dalšího textu, že je šance, že ke konci už čtenář úvod myšlenky zapomněl.

Problematická lexikální překvapení

Modul upozorní na slova neobvyklá v obecném kontextu. Může upozornit na nečekaně odborný pojem v místě, kde existuje jednodušší ekvivalent.

Konkrétní scénáře použití

Revize rozhodnutí před odesláním občanovi

Úředník vloží rozhodnutí o stavebním povolení do PONKu. Nástroj odhalí průměrnou délku věty 42 slova, 5 vět delších než 50 slov a 8 případů trpného rodu. Úředník text upraví: rozdělí nejdelší věty, některé pasáže převede do činného rodu.

Anonymizace rozhodnutí před zveřejněním

Soud potřebuje zveřejnit rozhodnutí. MaskIT automaticky najde a anonymizuje jména účastníků, adresy, data narození a názvy firem. Úředník zkontroluje výsledek a ručně doplní případné detaily, které nástroj přehlédl.

Příprava informačního letáku

Instituce připravuje leták o právech spotřebitelů. PONK pomůže eliminovat zbytečné formalismy a najít místa s vysokým lexikálním překvapením, kde lze odborné termíny nahradit běžnějšími výrazy.

Limity nástrojů

PONK ani MaskIT nejsou dokonalé a pracují podle principu "human-in-the-loop", tedy člověk má vždy poslední slovo.

PONK nemůže rozhodnout, zda složitá právní formulace je skutečně nutná, ani zda lze odborný termín nahradit bez ztráty přesnosti. **MaskIT** neodhalí 100 % citlivých údajů

(úplnost 80 %, přesnost 64 %) a nepokrývá zatím všechny kategorie z vyhlášky č. 403/2022 Sb. Proto platí: **nástroje navrhuji, člověk rozhoduje**. PONK zvýrazní potenciální problém, ale vy rozhodnete, zda je třeba něco měnit. MaskIT označí údaje k anonymizaci, ale vy provedete finální kontrolu. Jedná se tedy o nástroje pomoci při práci, nikoli automatizace procesu.

Jak začít

PONK (<https://quest.ms.mff.cuni.cz/ponk>)

1. Zkopírujte text nebo nahrajte soubor.
2. Klikněte na „Zpracovat“.
3. Prohlédněte si čtyři záložky: Míry (celkový přehled), Pravidla (konkrétní problémy), Překvapení (teplotní mapa), Akty (řečové akty).

MaskIT (<https://quest.ms.mff.cuni.cz/maskit>)

1. Vložte text k anonymizaci.
2. V nabídce voleb vyberte, zda chcete ponechat pseudoanonymizaci nebo nahradit údaje obecnými třídami.
3. Zkontrolujte výsledek.

3.2 S čím jsme se potýkali a co jsme se naučili

- *Stručně hlavní výzvy (ne vyčerpávající katalog):*
 - *Přesnost automatické analýzy není 100%*
 - *Lexikální překvapení – zajímavý směr, ale ještě si hledá cestu*
 - *Proč nepoužíváme ChatGPT (halucinace, bezpečnost, reprodukovatelnost)*
- *Co nástroj nenahradí (lidskou korekturu, právní znalost)*

Srozumitelnost jakéhokoliv textu je vysoce subjektivní. Srozumitelnost právního textu je navíc svázána řadou pravidel, která je nezbytné respektovat, ale pro čtenáře ne-právnicka mohou být nesrozumitelná. Jedná se například o přesnou terminologii, ustálené výrazy a ustálené použití konkrétních i neprávních termínů, komplexnost argumentace apod. Uvědomujeme si, že toto bude vždy limitem jakékoli automatické či semiautomatické analýzy srozumitelnosti právního textu. I z tohoto důvodu jsou běžně používané automatické metriky srozumitelnosti textu v rámci PONKu (kapitola 2.1) pouze doprovodným ukazatelem a důraz je kladen hlavně na metody založené na detailním rozboru právního textu, ať už pomocí specificky navržených lingvistických pravidel, nebo promptů.

Lingvistická pravidla jsou založena na syntaktické analýze textu, která však nemusí vždy plně reflektovat specifika právního textu. Například některá pravidla se týkají délky termínu (počtu přídavných jmen nebo příslovcí spojených s podstatným jménem). Avšak některé právní termíny jsou definičně postaveny na takovém řazení několika slov za sebou. Limity těchto pravidel jsou v PONKu zachyceny tak, že příslušné jevy jsou v textu pouze zvýrazněny. Rozhodnutí o případné editaci je vždy na autorovi textu se znalostí popisovaného právního problému.

Lexikální překvapení je metoda postavená na jazykových modelech natrénovaných na obecném textu. Při použití na právních textech může tedy vykazovat vysoké hodnoty překvapení (a tedy potenciální problém se srozumitelností) u slov, která jsou právnímu textu vlastní a přirozená, ale v textu obecném se často nevyskytují. I toto je na rozhodnutí autora textu, zda lze takové slovo nahradit srozumitelnějším (běžnějším) synonymem, či zda je nutné trvat na konkrétním právním termínu.

Přirozenou námitkou může být, zda by nebylo jednodušší použít přímo systém generativní umělé inteligence, např. ChatGPT. Aktuálně se domníváme, že rizika spojená s takovým využitím převyšují možné výhody využití takového systému, což jsou uživatelská přívětivost a jednoduchost. Prvním obecným rizikem je bezpečnost – každý text, který je do ChatGPT nahrán, může sloužit k následnému trénování modelů, tzn. takové texty by určitě neměly obsahovat osobní údaje ani jiné neveřejné informace. PONK žádná data nikam neukládá ani dále nepoužívá, text, který je nahrán, se pouze propíše zpracovaný do výstupního okna. Dalším rizikem jsou halucinace, které jsou s používáním ChatGPT neoddelitelně spjaty. I v případě, že bychom instruovali ChatGPT ke konkrétním úpravám nahraného textu za účelem srozumitelnosti, každý výstup je nutné pečlivě zkontrolovat s ohledem na konzistenci s původním zněním.

Nejdůležitějším principem, na kterém je PONK postaven, je však tzv. *human-in-the-loop*. Znamená to, že žádné rozhodnutí v jeho používání se neděje automaticky nebo bez zásahu člověka. Účelem PONKu je asistovat v právním psaní odborníkovi, který vždy rozhoduje o aplikaci navržených úprav. Tento princip plně respektuje právní znalost a doménovou odbornost uživatelů a nechává konečná rozhodnutí o textu na jejich uvážení.

4. Závěr

4.1 Výzkum pokračuje

Na projekt PONK navazuje od 1.1.2026 projekt **SPRINT**: *Srozumitelné právo s umělou inteligencí – Trénink pro úředníky a právníky* (č. TQ11000058).⁵ Cílem projektu je navrhnout a ověřit metodiku SPRINT, která slouží ke kontrole srozumitelnosti právních textů. Projekt si klade za cíl prokázat, že umělá inteligence může fungovat jako účinný vzdělávací nástroj pro právníky a státní úředníky a současně jako praktický asistent při tvorbě jasných a srozumitelných právních dokumentů. Součástí projektu jsou školení, v jejichž rámci budou účastníci seznámeni s postupy, jak ručně navržená pravidla předkládat jazykovým modelům tak, aby si je dokázaly osvojit, a zároveň jak hodnotit úspěšnost velkých jazykových modelů po natrénování těchto pravidel. Prezentace bude demonstrovat proces učení také na pravidlech vytvořených v rámci nástroje PONK.

4.2 PONK pro jiné domény

PONK je relativně snadno přenositelný i do jiných domén než je právo a veřejná správa. Základní architektura – kombinace kvantitativních metrik, pravidlových lingvistických analýz a lexikálních či pragmatických ukazatelů – zůstává zachována, zatímco se mění doménově specifické zdroje a parametry. V praxi by to znamenalo zejména úpravu pravidel a modelů tak, aby reflektovaly terminologii, styl a komunikační cíle dané oblasti (např. zdravotnictví, technickou dokumentaci nebo vzdělávací texty). Díky své modularitě může PONK sloužit jako obecná platforma pro podporu srozumitelnosti textů napříč doménami.

4.3 Kde najdete víc

Webová stránka projektu <https://ufal.mff.cuni.cz/grants/ponk> slouží jako rozcestník k nástrojům a datům popsaným v této zprávě. Zveme vás k její návštěvě a k vyzkoušení nástrojů. Budeme rádi za jakoukoli zpětnou vazbu.

Literatura

Behún, D., & Behún, P. (2023). *Pište čtivě a srozumitelně*. Zoner Press.

Bendová, K., & Cinková, S. (2021). Adaptation of Classic Readability Metrics to Czech. *24th International Conference on Text, Speech and Dialogue*, 159–171.

⁵ <https://ufal.mff.cuni.cz/grants/sprint>

Dębowski, Ł., Broda, B., Nitoń, B., & Charzyńska, E. (2015). Jasnopis – A Program to Compute Readability of Texts in Polish Based on Psycholinguistic Research. *Natural Language Processing and Cognitive Science*, 51–61.

DuBay, W. (2007). *Smart Language. Readers, Readability, and the Grading of Text*. Impact Information.

Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology*, 32(3), 221–233.

Flesch, R. (1974). *The art of readable writing* (2nd ed.). Harper.

Garner, B. A. (2021). *Právní psaní srozumitelným jazykem* (J. A. Gealfow, A. Valentová, T. Königová, P. Hubková, Š. Janků, & Bejčková, Michaela, Trans.). Nugis Finem Publishing.

Gunning, R. (1952). *The Technique of Clear Writing*. McGraw-Hill.

Hladká, B., Cinková, S., Černý, J., Eichler, V., Knap, T., Kraus, I., Kubíková, B., Kuk, M., Kvapilíková, I., Mírovský, J., Novotná, T., Polák, T., Pospíšil, P., Stanovský, A., Šamánková, J. (2025): Readability of Czech legal and administrative documents in research and practice. In: *Nová čeština doma a ve světě*, ISSN 1805-367X, vol. 2025, no. 2, pp. 47-57

Lov om språk (språkløva), Pub. L. No. LOV-2021-05-21-42, § 9. Klart språk (2021).
<https://lovdata.no/dokument/NL/lov/2021-05-21-42?q=Spr%C3%A5kløv>

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Zhiqiang, C. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.

Mlsna, P., & Sláma, D. (2019). *Klientsky orientovaná veřejná správa 2030. Koncepce—Verze pro vnitřní i vnější připomínkové řízení*. Ministerstvo vnitra ČR.

Plain Writing Act of 2010, Pub. L. No. 111–274, Public Law 2861 (2010).
<https://www.govinfo.gov/content/pkg/PLAW-111publ274/pdf/PLAW-111publ274.pdf>

Šamánková, J., & Kubíková, B. (Eds.). (2022). *Jak psát srozumitelné úřední texty. Příručka srozumitelného psaní pro úředníky* (1st ed.). Kancelář veřejného ochránce práv.
https://www.ochrance.cz/uploads-import/ESO/p%C5%99%C3%ADru%C4%8Dka/Prirucka_srozumitelneho_psani_tisk.pdf

Slejšková, L. (2013). *Čeština za pracovním stolem. Praktické lekce z úřední korespondence*. Grada.

Smejkalová, T. (2013). Srozumitelnost práva. *Právník*, 2013(5), 447–472.

Språklag (2009:600), Pub. L. No. 2009:600, SFS nr: 2009:600 Svensk författningssamling (2009). <https://rkrattsbaser.gov.se/sfst?bet=2009:600>

Šváb, J. (2021). *Jak psát, aby se to dalo číst: Příručka přístupného psaní* (Vydání první). Leges.
<https://kramerius.lib.cas.cz/uuid/uuid:6dac8db8-6dfa-4d97-93e8-d31350679147>

Vučka, J. (2019). *Právní psaní*. C.H. Beck.

THE ÚFAL TECHNICAL REPORT SERIES

ÚFAL (Ústav formální a aplikované lingvistiky; <https://ufal.mff.cuni.cz>) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the “sister” Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

The ÚFAL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. Since November 1996, the following reports have been published.

- ÚFAL TR-1996-01 Eva Hajičová, *The Past and Present of Computational Linguistics at Charles University*
Jan Hajič and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*
- ÚFAL TR-1997-02 Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*
- ÚFAL TR-1997-03 Alla Bémová at al., *Anotace na analytické rovině, Návod pro anotátory (in Czech)*
- ÚFAL TR-1997-04 Jan Hajič and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*
- ÚFAL TR-1998-05 Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*
- ÚFAL TR-1999-06 Vladislav Kuboň, *A Robust Parser for Czech*
- ÚFAL TR-1999-07 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*
- ÚFAL TR-2000-08 Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*
- ÚFAL/CKL TR-2000-09 Eva Hajičová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-10 Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-11 Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*
- ÚFAL/CKL TR-2001-12 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*
- ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*
- ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*
- ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová, *Tektogramaticky anotovaný valenční slovník českých sloves*
- ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*
- ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*
- ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*
- ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolářová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*
- ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*
- ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*
- ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Uřešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Uřešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*

- ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*
- ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*
- ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*
- ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*
- ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*
- ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otrádočková, Petr Pajas, Petr Podveský, Zdeňka Urešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*
- ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*
- ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*
- ÚFAL/CKL TR-2007-36 Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*
- ÚFAL/CKL TR-2008-37 Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*
- ÚFAL/CKL TR-2008-38 Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*
- ÚFAL/CKL TR-2008-39 Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*
- ÚFAL/CKL TR-2008-40 Lucie Mladová, *Diskurzní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*
- ÚFAL/CKL TR-2009-41 Marie Mikulová, *Pokyny k překladu určené překladatelům, revizorům a korektorům textů z Wall Street Journal pro projekt PCEDT*
- ÚFAL/CKL TR-2011-42 Loganathan Ramasamy, Zdeněk Žabokrtský, *Tamil Dependency Treebank (TamilTB) – 0.1 Annotation Manual*
- ÚFAL/CKL TR-2011-43 Nguy Giang Linh, Michal Novák, Anna Nedoluzhko, *Coreference Resolution in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2011-44 Anna Nedoluzhko, Jiří Mirovský, *Annotating Extended Textual Coreference and Bridging Relations in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2011-45 David Mareček, Zdeněk Žabokrtský, *Unsupervised Dependency Parsing*
- ÚFAL/CKL TR-2011-46 Martin Majliš, Zdeněk Žabokrtský, *W2C – Large Multilingual Corpus*
- ÚFAL TR-2012-47 Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mirovský, Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová, *Manual for annotation of discourse relations in the Prague Dependency Treebank*
- ÚFAL TR-2012-48 Nathan Green, Zdeněk Žabokrtský, *Ensemble Parsing and its Effect on Machine Translation*
- ÚFAL TR-2013-49 David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zeman, Zdeněk Žabokrtský, Jan Hajič *Cross-language Study on Influence of Coordination Style on Dependency Parsing Performance*
- ÚFAL TR-2013-50 Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović, Daniel Zeman, *Tools for Machine Translation Quality Inspection*
- ÚFAL TR-2013-51 Marie Mikulová, *Anotace na tektogramatické rovině. Dodatky k anotátorské příručce (s ohledem na anotování PDTSC a PCEDT)*

- ÚFAL TR-2013-52** Marie Mikulová, *Annotation on the tectogrammatical level. Additions to annotation manual (with respect to PDTSC and PCEDT)*
- ÚFAL TR-2013-53** Marie Mikulová, Eduard Bejček, Jiří Mirovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *Úpravy a doplňky Pražského závislostního korpusu (Od PDT 2.0 k PDT 3.0)*
- ÚFAL TR-2013-54** Marie Mikulová, Eduard Bejček, Jiří Mirovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *From PDT 2.0 to PDT 3.0 (Modifications and Complements)*
- ÚFAL TR-2014-55** Rudolf Rosa, *Depfix Manual*
- ÚFAL TR-2014-56** Veronika Kolářová, *Valence vybraných typů deverbativních substantiv ve valenčním slovníku PDT-Vallex*
- ÚFAL TR-2014-57** Anna Nedoluzhko, Eva Fučíková, Jiří Mirovský, Jiří Pergler, Lenka Šíková, *Annotation of coreference in Prague Czech-English Dependency Treebank*
- ÚFAL TR-2015-58** Zdeňka Uřešová, Eva Fučíková, Jana Šindlerová, *CzEngVallex: Mapping Valency between Languages*
- ÚFAL TR-2015-59** Kateřina Rysová, Magdaléna Rysová, Eva Hajičová, *Topic-Focus Articulation in English Texts on the Basis of Functional Generative Description*
- ÚFAL TR-2016-60** Kira Droganova, Daniel Zeman, *Conversion of SynTagRus (the Russian dependency treebank) to Universal Dependencies*
- ÚFAL TR-2018-61** Lukáš Kyjánek, *Morphological Resources of Derivational Word-Formation Relations*
- ÚFAL TR-2019-62** Zdeňka Uřešová, Eva Fučíková, Eva Hajičová, *CzEngClass: Contextually-based Synonymy and Valency of Verbs in a Bilingual Setting (CzEngClass: Kontextová synonymie a valence sloves v bilingvním prostředí)*
- ÚFAL TR-2019-63** Ján Faryad, *Identifikace derivačních vztahů ve španělštině*
- ÚFAL TR-2020-64** Marie Mikulová, Jan Hajič, Jiří Hana, Hana Hanová, Jaroslava Hlaváčová, Emil Jeřábek, Barbora Štěpánková, Barbora Vidová Hladká, Daniel Zeman, *Manual for Morphological Annotation. Revision for Prague Dependency Treebank – Consolidated 2020 release*
- ÚFAL TR-2021-65** Rudolf Rosa, *Technická zpráva o vývoji projektu THEaiTRE v roce 2020*
- ÚFAL TR-2021-66** Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Daniel Zeman, *Coreference meets Universal Dependencies – a pilot experiment on harmonizing coreference datasets for 11 languages*
- ÚFAL TR-2021-67** THEaiTRobot 1.0, David Košťák, Daniel Hrbek, Rudolf Rosa, Ondřej Dušek, *AI: When a Robot Writes a Play*
- ÚFAL TR-2021-68** Markéta Lopatková, Václava Kettnerová, Anna Vernerová, Eduard Bejček, Zdeněk Žabokrtský, *Valenční slovník českých sloves VALLEX*
- ÚFAL TR-2021-69** Niyati Bafna, Jan Bodnár, Lukáš Kyjánek, Emil Svoboda, Madga Ševčíková, Jonáš Vidra, Zdeněk Žabokrtský, *Towards Universal Segmentations: Survey of Existing Morphosegmentation Resources*
- ÚFAL TR-2021-70** Zdeňka Uřešová, Eva Fučíková, Jan Hajič, Karolina Zaczynska, *Annotation guidelines for German verbal synonyms included in SynSemClass Lexicon*
- ÚFAL TR-2022-71** THEaiTRobot 2.0, Josef Doležal, Klára Vosecká, Tomáš Musil, David Mareček, Rudolf Rosa, *Permeation / Prostoupení*
- ÚFAL TR-2022-72** Cristina Fernández-Alcaina, Eva Fučíková, Zdeňka Uřešová, *Annotation guidelines for Spanish verbal synonyms in the SynSemClass Lexicon*
- ÚFAL TR-2023-73** Zdeňka Uřešová, Cristina Fernández-Alcaina, Eva Fučíková, Jan Hajič, *SynSemClass Czech and English Annotation Guidelines*
- ÚFAL TR-2025-74** Markéta Lopatková, Eva Fučíková, Federica Gamba, Jan Hajič, Hana Hledíková, Marie Mikulová, Michal Novák, Jan Štěpánek, Daniel Zeman, Šárka Zikánová, *UMR 2.0 - Czech: Release Notes*
- ÚFAL TR-2025-75** Barbora Hladká, Silvie Cinková, Jan Černý, Vítek Eichler, Tomáš Knap, Ivan Kraus, Barbora Kubíková, Ivana Kvapilíková, Jiří Mirovský, Tereza Novotná, Tomáš Polák, Arnold Stanovský, Jana Šamánková, Michal Kuk, Přemysl Pospíšil, *Srozumitelnost českých právních a administrativních dokumentů ve výzkumu a praxi*