

MATEMATICKO-FYZIKÁLNÍ FAKULTA  
PRAHA

**Technická zpráva o vývoji projektu THEaiTRE v roce 2020**  
RUDOLF ROSA

ÚFAL Technical Report  
**TR-2021-65**

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: <http://ufal.mff.cuni.cz/techrep>

# Technická zpráva o vývoji projektu THEaiTRE v roce 2020

Rudolf Rosa

Univerzita Karlova, Matematicko-fyzikální fakulta,  
Ústav formální a aplikované lingvistiky

2. února 2021

## 1 Úvod

V této zprávě shrnuji průběh řešení projektu THEaiTRE za období duben až prosinec 2020, tedy v prvním roce projektu.

Projekt *TL03000348 THEAITRE: Umělá inteligence autorem divadelní hry?* je spolufinancován se státní podporou Technologické agentury ČR v rámci Programu ÉTA 3.<sup>1</sup> Tento text vychází z Odborné zprávy o řešení projektu, vypracované pro TAČR. Zkrácený text původního návrh projektu podaného na TAČR se nachází v příloze A.

## 2 Představení projektu

Tento projekt kombinující počítačovou lingvistiku a divadelnictví má za cíl přímým způsobem reflektovat nástup a zkoumat roli umělé inteligence ve společnosti a v umění a možnost její spolupráce s člověkem. Navrhujeme spojit teatrologický výzkum s komputačně lingvistickým výzkumem, kde aplikace poznatků a výsledků z těchto oborů povede k sestavení a natrénování systému pro automatické generování scénářů divadelních her. Tento systém následně bude naším týmem interaktivně využit pro vytvoření hry, která bude realizována hereckým souborem ke 100. výročí premiéry dramatu R.U.R. Vzniklé dílo a jeho přijetí diváky budeme zpětně zkoumat v rámci obou vědních oborů a získané poznatky využijeme pro další zdokonalení systému, který zpřístupníme na webu, a pro realizaci dalších her.

## 3 Řešitelský tým a jeho role v projektu

Pro každého účastníka uvádím jeho hlavní úlohu či úlohy.

---

<sup>1</sup><http://www.tacr.cz>

### **3.1 Univerzita Karlova, Matematicko-fyzikální fakulta (MFF)**

Hlavní úloha týmu: vývoj aplikace na generování scénářů divadelních her.

- Rudolf Rosa – vedení výzkumné části projektu
- Ondřej Dušek – hierarchické generování, mentoring
- David Mareček – datové zdroje
- Tomáš Musil – generování pomocí GPT-2
- Patricia Schmidtová – hierarchické generování
- Dominik Jurko – práce s postavami
- Jana Hamřlová – administrativa, finance, komunikace
- Jindřich Vodrážka – technické zajištění
- Ondřej Bojar – otec zakladatel, mentoring, PR
- Tom Kocmi (do srpna 2020) – strojový překlad, automatická sumarizace

### **3.2 Švandovo divadlo na Smíchově (ŠD)**

Hlavní úloha týmu: spolupráce na vývoji aplikace formou zpětné vazby, vygenerování scénáře divadelní hry a realizace této hry ve Švandově divadle.

- Daniel Hrbek – vedení divadelní části projektu, realizace divadelní hry, režisér
- David Košťák – generování scénáře, dramaturg
- Martina Kinská – zpětná vazba pro vývoj aplikace, dramaturg
- Marie Nováková – vedení studentů DAMU, anotace dramatických situací

### **3.3 Divadelní fakulta Akademie múzických umění v Praze (DAMU)**

Hlavní úloha týmu: spolupráce na vývoji aplikace formou zpětné vazby, vytváření podkladů pro generování divadelních her.

- Klára Vosecká – zpětná vazba pro vývoj aplikace, anotace dramatických situací
- Josef Doležal – zpětná vazba pro vývoj aplikace, anotace dramatických situací

### **3.4 CEE Hacks**

Hlavní úloha týmu: koordinace celého projektu, marketing a public relations, realizace premiéry, pořádání satelitních akcí.

- Tomáš Studeník – vedení propagačního týmu, duchovní otec projektu
- Petr Žabka – projektový manažer

## 4 Popis řešení projektu

### 4.1 Sdílení poznatků

Práci na projektu jsme zahájili několika setkáními celého týmu, kde jsme se vzájemně vzdělávali – divadelní odborníci vzdělávali tým MFF v oboru divadelních věd a struktury dramatu, a tým MFF vzdělával divadelní odborníky v oboru automatického počítačového zpracování textu. Tím jsme už od začátku vzájemně získali potřebný vhled do dané problematiky.

Díky těmto vzdělávacím setkáním se nám podařilo už od začátku zapojovat do projektu všechny členy konsorcia. Členové MFF započali s vývojem aplikace, přičemž vycházeli z poznatků získaných od divadelních odborníků, zejména ohledně struktury scénáře divadelní hry.

### 4.2 Vývoj aplikace

#### 4.2.1 Řešení pomocí modelu GPT-2

Vývoj aplikace jsme zahájili okamžitě. V první fázi jsme se rozhodli využít pro generování scénářů nejprve nejjednodušší lineární postup (tedy variantu A dle navrhované metodologie, viz A.3.3). Na rozdíl od návrhu projektu, kde jsme navrhovali použít jednodušší RNN modely, jsme se rozhodli využít tehdy relativně nově vydaný sofistikovanější jazykový model GPT-2 (ve variantě XL) od konsorcia OpenAI,<sup>2</sup> založený na pokročilejší architektuře Transformer.<sup>3</sup> Tento model je známý tím, že je možné pomocí něj vygenerovat velmi věrohodné falešné zprávy (*fake news*), které je těžké rozpoznat od pravých. Model se používá tak, že uživatel definuje jako vstup začátek textového dokumentu, a model následně postupně slovo po slově generuje možné pokračování takového dokumentu.

Model jsme tedy použili tak, že jsme na jeho vstup vložili krátký úsek existujícího scénáře divadelní hry, a nechali jej vygenerovat pokračování. V souladu s návrhem projektu jsme očekávali, že tento způsob generování scénářů bude nedostatečný; model GPT-2 jsme navíc nijak nedotrénovali na scénáře, a očekávali jsme, že v jeho trénovacích datech se scénáře spíše nevyskytovaly, a tedy nebude schopen generovat výstupy, které by vypadaly jako scénáře. Překvapivě se ale ukázalo, že už tento jednoduchý postup vede k výstupům, které obvykle dodržují strukturu i téma scénáře ze vstupu, včetně střídání zadaných postav. Pro první verzi aplikace jsme se proto rozhodli setrvat u tohoto jednoduchého postupu, a pouze ho podle potřeby upravovat.

Protože model GPT-2 je k dispozici pouze v angličtině, generovaný scénář automaticky překládáme do češtiny strojovým překladačem CUBBITT.<sup>4</sup>

#### 4.2.2 Ovládání aplikace

Aplikace pracuje formou interaktivního generování ve spolupráci s člověkem (human-in-the-loop). Lidský operátor definuje vstup aplikace, model následně vygeneruje pokračování čítající 10 následujících replik, a předá řízení zpět operátorovi. Pokud je s výstupem spokojen, nechá vygenerovat dalších 10 replik. Pokud spokojen není, označí libovolnou repliku jako nechtěnou, vygenerovaný text

<sup>2</sup><https://openai.com/blog/better-language-models/>

<sup>3</sup>[https://en.wikipedia.org/wiki/Transformer\\_\(machine\\_learning\\_model\)](https://en.wikipedia.org/wiki/Transformer_(machine_learning_model))

<sup>4</sup><https://ufal.mff.cuni.cz/cubbitt>

počínaje touto replikou se zahodí, a model vygeneruje jinou variantu. Navíc má operátor možnost na zvolené místo ručně vložit vlastní repliku či scénickou poznámku.

Původní verze aplikace pracovala jinak. Model vždy vygeneroval pokračování o délce pouhé jedné repliky, ale rovnou v pěti variantách, a operátor si následně zvolil, které pokračování se má použít (případně všechna pokračování zahodil a nechal si vygenerovat dalších pět pokračování). Na to model opět reagoval vygenerováním dalších pěti možností pokračování po této zvolené replice. Tento způsob generování tedy byl interaktivnější, avšak v praxi se neosvědčil – jednak bylo generování pomocí něj příliš zdlouhavé a časově náročné, a dále nabyli členové týmu ŠD a DAMU dojmu, že mají nad generováním příliš velkou kontrolu, a tedy výsledný scénář je spíše jejich dílem než dílem umělé inteligence.

### 4.2.3 Problémy a jejich řešení

Studenti DAMU pod vedením divadelních odborníků ŠD a s využitím znalostí o způsobu fungování aplikace připravili dramaticky zajímavé vstupy tak, aby byla vysoká pravděpodobnost vygenerování dramaticky zajímavých výstupů, a vygenerované výstupy podrobně analyzovali. Na společných schůzkách jsme pak interpretovali, čím byly pozorované problémy ve výstupech způsobeny, a následně jsme se snažili problémy odstranit jak úpravami aplikace (MFF), tak úpravami podoby a struktury vstupů (DAMU). Úspěšnost provedených úprav jsme opakovaně společně empiricky vyhodnocovali, čímž jsme postupně našli jak ideální podobu vstupu, tak několik úprav aplikace podstatně zlepšujících kvalitu výstupů. Zároveň jsme takto odhalili obtížnější problémy, které se nám v první verzi aplikace nepodařilo uspokojivě vyřešit. Na řešení těchto problémů proto nadále pracujeme.

Uvádím některé problémy, které jsme společně objevili a částečně či plně vyřešili:

**Repertoár postav** Model GPT-2 se při generování vesměs drží postav zadaných na vstupu, ne ale vždy. Pozorovali jsme, že někdy náhodně zavádí nové postavy, někdy zase existující postavy bezdůvodně mizí. Ačkoliv toto může být někdy žádoucí, z analýzy vyplynulo, že k tomu dochází příliš často a beze smyslu. Proto jsme modifikovali predikci následujícího slova, tak aby každý řádek scénáře začínal jménem některé z postav zadaných na vstupu. Také jsme modifikovali pravděpodobnosti vygenerování jednotlivých tokenů, tak aby postava, která se už dlouho nevyskytla, měla větší pravděpodobnost výskytu, a nebyla tedy tak snadno zapomenuta.

**Omezený kontext** Model GPT-2 má omezenou kapacitu na reprezentaci vstupu a výstupu na 1024 tokenů (jako tokeny používá *subwordy*, jde tedy o omezení přibližně na 800 slov). Protože scénář tento limit rychle přesáhne, zapojili jsme do aplikace automatickou sumarizaci (kterou jsme měli v úmyslu v projektu použít, ale jiným způsobem). Pokud délka vstupu pro generování (zadaný vstup + dosud vygenerovaný výstup) přesáhne nastavený limit, tento vstup se nejprve automaticky sumarizuje do nastaveného rozsahu, tak aby se i společně s generovaným výstupem vešel do limitů modelu GPT-2.

**Repetitivnost** V některých případech se stane, že se generování dostane do slepé uličky, a stále se opakuje například jedna dvojice replik. Zavedli jsme proto mechanismus zakázaných řádek, kdy pokud model vygeneruje během 10 replik tutéž repliku více než dvakrát, je tato zahozena, model je přenastaven zpět do stavu před vygenerováním této repliky, a znovu spuštěn. Toto se opakuje tak dlouho, dokud nevygeneruje odlišnou repliku (ale nejvýše desetkrát).

Řešili jsme i další menší problémy, jako jsou chyby v tokenizaci výstupů, chyby v překladu jmen postav a podobně.

Zároveň jsme narazili i na obtížnější problémy, které se nám doposud vyřešit nepodařilo. Ve verzi aplikace použité pro vygenerování scénáře pro první premiéru tedy tyto problémy přetrvávají a bylo obvykle nutné je řešit lidskými zásahy. Pro další verzi aplikace se ale chceme pokusit tyto problémy vyřešit a tak dále minimalizovat množství potřebných lidských zásahů.

**Neschopnost generovat celý scénář** Ukázalo se, že pomocí zvoleného postupu je možné vygenerovat scénáře jednotlivých scén, ale ne už celé divadelní hry. Generátor takto dlouho neudrží kontext, zároveň se po nějaké době už scéna nikam neubírá a je nutné ji ukončit a přejít na jinou scénu. V současné verzi toto řešíme tak, že tým ŠD definuje vstup pro každou scénu zvlášť, a z vygenerovaných scénářů jednotlivých scén pak složí celou hru. Pro druhou verzi aplikace máme v plánu se vrátit k variantě C z návrhu projektu, tedy generovat scénář pomocí hierarchického textového generování.

**Špatná práce s postavami** Model sice ve scénáři vypisuje jména postav, ale postavy se nezdají být dostatečně svébytné. Model generuje text, který je konzistentní s předchozím textem, ale zjevně nepracuje s konceptem toho, že jednotlivé postavy by měly být konzistentní samy se sebou, ale mohou se lišit mezi sebou. Odborníci ŠD toto označují jako jakousi dětskou naivitu modelu, který předpokládá, že vše co je řečeno je pravda, bez schopnosti pracovat s tím, že různé postavy mohou mít různý úhel pohledu, mají v situaci různé postavení, dokonce mohou i lhát a podobně. Zároveň se snadno stane, že jednotlivé postavy splývají či si prohazují pozice (např. dcera se stane matkou a matka dcerou). Pro první verzi se tento problém řešil pomocí zahazování špatných replik operátorem, případně ruční změnou přiřazení postav ve vytvořeném scénáři. Pro novou verzi aplikace chceme zvolit postup, kdy vyladíme jednotlivé varianty modelu pro různé postavy tak, aby repliky každé z postav generoval jinak upravený model.

**Neschopnost myslet v dramatických situacích** Autor divadelní hry typicky neuvažuje v jednotlivých replikách, ale v jednotlivých dramatických situacích, a následně je každá realizována několika replikami. Toto model nedělá, děj se tak postupně přelévá, ale chybí v něm jasné situace a jejich změny. Pro další verzi aplikace proto chceme model obohatit o práci s dramatickými situacemi, jak jsme zvažovali v návrhu projektu v rámci varianty B. Za tímto účelem již ve spolupráci MFF a DAMU probíhá anotace scénářů pomocí dramatických situací.

Z analýz fungování aplikace divadelními odborníky vplývají i další problémy, které jsou ale vážnější a těžko uchopitelné, a tedy v současné době nemáme žádný plán na jejich řešení.

Jde například o fakt, že model pracuje převážně pouze s textovými asociacemi a souvislostmi, chybí mu představa toho, jak se scénář odehrává na pódiu (což je představa, se kterou lidský autor pracuje primárně). Zde by bylo možné uvažovat o nějakých multimodálních a multitaskových postupech a například kromě textu generovat i obrazovou informaci; generování obrazových dat je však značně obtížné, jeho úroveň zatím není příliš vysoká, a neočekáváme proto, že by takovýto (jistě složitý) postup vedl ke kýženému výsledku.

Obecně lze také mluvit o tom, že vygenerovanému scénáři chybí *aboutness* (*o-čem-nost*, tedy zda je hra o něčem, zda má nějaký smysl či nějaké sdělení). Zde poznamenáváme, že umělá inteligence pravděpodobně, na rozdíl od lidského autora, nemá potřebu něco vyjádřit či sdělit, netvoří text s nějakým daným účelem, jde tedy pravděpodobně o principiální problém. Neznamená to ale nutně, že vygenerovaný text nemá žádný smysl. Lidé totiž mají pozoruhodnou schopnost najít smysl i v tom, co žádný apriorní smysl nemá. Zde tedy spoléháme na představivost dramaturga, režiséra, herců a následně diváků, že podobně jako v případě *found art* zvládnou i v třeba nesmyslném textu vlastní smysl najít.

### 4.3 Generování a realizace divadelní hry

Hotovou aplikaci verze 1 převzal do užívání tým ŠD, který na základě svých zkušeností a s využitím znalostí o fungování aplikace definoval celkový rámec divadelní hry (robot v typicky lidských situacích) a připravil sadu asi 12 tématických vstupů (např. na témata *robot a láska*, *robot a práce*, *robot a smrt*, *robot a smysl pro humor*). Pro tyto vstupy pak tým ŠD interaktivně generoval scénáře, z nichž nakonec vybral 8, ze kterých sestavil scénář vlastní hry, čítající 32 stran textu. S textem tým ŠD dále prováděl podobné úpravy, jako by šlo o standardní scénář divadelní hry napsaný lidským autorem. Pro zachování co největší autenticity se ale snažil tyto úpravy minimalizovat. Ve finální verzi scénáře tak tvoří lidské zásahy méně než 10% textu.

Vzniklý scénář nese název *AI: Když robot píše hru* a ŠD pracuje na jeho realizaci. Premiéra je plánovaná na 26. února 2021;<sup>5</sup> z důvodu nemoci režiséra Daniela Hrbka musela být premiéra o měsíc odložena oproti původnímu plánu. Kvůli přetrvávající pandemii Covid-19 a uzavření divadel bude premiéra streamovaná online.

Při práci s prvotními verzemi aplikace byl tým ŠD spíše zklamán nízkou kvalitou výstupů a hru plánoval realizovat méně nákladnou formou scénického čtení, bez dekorací, kostýmů a bez repríz. Postupným vývojem aplikace a také postupným bližším seznamováním týmu ŠD s aplikací a lepším pochopením jejího fungování a jejích silných a slabých stránek se ale podařilo dosáhnout stavu, kdy tým ŠD pomocí aplikace vygeneroval, dle vlastního názoru, dostatečně kvalitní scénář. Hra proto bude realizována v plné podobě a bude opakovaně reprízována, jakmile to epidemiologická situace umožní.

<sup>5</sup><https://www.svandovodivadlo.cz/inscenace/673/ai-kdyz-robot-pise-hru>



## 4.4 Anotace dramatických situací

Z analýz nedostatků vygenerovaných scénářů vyvstal návrh týmu ŠD, aby aplikace pracovala na úrovni dramatických situací, což jsme v návrhu projektu plánovali, ale dosud nerealizovali. Tým MFF proto prozkoumal existující datové zdroje a konstatoval jejich nedostatečnost. Tým MFF tedy nyní spolupracuje s týmem DAMU na anotaci dramatických situací ve scénářích existujících divadelních her – tým MFF připravuje data a vyhodnocuje anotace, tým DAMU anotuje scénáře, a na společných schůzkách diskutujeme o anotacích a jejich problémech a následně upravujeme anotační schéma a postupy.

Anotace vychází z Poltiho 36 dramatických situací, které jsou v teatrologii velmi dobře známé a zavedené. V průběhu anotací se nicméně empiricky potvrdil předpoklad, že schéma těchto 36 situací, vycházející z antického dramatu, není pro současná dramata příliš vhodné, neboť zahrnují situace, které nejsou v současném dramatu příliš běžné (např. *Nevědomky učiněné hříchy v lásce*) a naopak neobsahují některé situace, se kterými dnešní autoři běžně pracují (např. *Výslech*). Tým DAMU proto průběžně schéma situací upravuje. Aktuální pracovní verze obsahuje 57 situací, z nichž ale některé ještě máme v úmyslu odstranit či spíše sloučit tak, abychom pro každou situaci ve finálním schématu měli dostatečnou oporu v datech; námi vytvářené schéma dramatických situací tak vychází z kombinace teatrologických poznatků a statistického zpracování dat. V současné době máme vytvořené anotace pro scénáře 16 divadelních her.

## 5 Publicita

Jedním z našich cílů v projektu THEaiTRE je přiblížit umělou inteligenci širší veřejnosti a podpořit debatu o její roli ve společnosti. Proto o projektu různými kanály průběžně informujeme odbornou i laickou veřejnost, přijímáme pozvánky na rozhovory a diskuze, komunikujeme s médii, a podobně.

### 5.1 Vlastní kanály

Mezi hlavní kanály informování o projektu patří níže uvedené:

- Webová stránka projektu: <https://theaitre.com/>
- Twitter: <https://twitter.com/theaitre>
- Facebook: <https://www.facebook.com/Theaitre/>
- Instagram: <https://www.instagram.com/the.ai.tre/>

Dle potřeby a možností příležitostně využíváme i další kanály, jako například webové stránky a sociální síť MFF a ŠD.

### 5.2 Články v médiích

Projektu se podařilo vzbudit jistou mediální pozornost, zejména po zveřejnění pre-printu článku (*THEaiTRE: Artificial Intelligence to Write a Theatre Play*) na serveru *arXiv.org* (<https://arxiv.org/abs/2006.14668>), odkud informace o projektu převzal server *TechXplore* a následně řada zahraničních médií.

O projektu tak psala média americká, britská, italská, španělská, polská, ruská, japonská, čínská, ruská, apod.

Radě médií jsme poskytli rozhovory, na jejichž základě byly publikovány například tyto články:

- TechXplore: *THEaiTRE: A theatre play written entirely by machines.*<sup>6</sup>
- Respekt: *Ze života robotů.*<sup>7</sup>
- Hospodářské noviny: *Scénář: Robot. Ve Švandově divadle píše hru k výročí R.U.R. umělá inteligence.*<sup>8</sup>
- Kultura 21: *AI: Když robot píše hru. Švandovo divadlo uvede hru napsanou umělou inteligencí.*<sup>9</sup>
- Goethe Institut: *Umělá, ale přesto kreativní.*<sup>10</sup>
- AbbaNews: *Io robot. La linguistica computazionale arriva a teatro.*<sup>11</sup>
- Idnes: *Novou hru Švandova divadla píše počítač. Letos bude první verze, říká jeho šéf.*<sup>12</sup>
- Startupjobs: *RUR v dobách umělé inteligence: Divadelní hru k 100 letům Čapkova díla píše robot z Matfyzu.*<sup>13</sup>
- Matfyz.cz: *Umělá inteligence píše divadelní hru.*<sup>14</sup>
- TA.DI: *Dočkáme se digitálního Shakespeara? AI jako autor divadelní hry.*<sup>15</sup>

### 5.3 Zvané přednášky a rozhovory

Byli jsme pozváni k představení projektu při různých příležitostech, v roce 2020 šlo zejména o následující:

- Innocamp: *Umělá inteligence a kreativita.*<sup>16</sup>
- Noc vědců: *THEaiTRE: Umělá inteligence píše divadelní hru.*<sup>17</sup>
- Science Café: *S Rudolfem Rosou o umělé inteligenci a projektu THEaiTRE.*<sup>18</sup>

<sup>6</sup><https://techxplore.com/news/2020-08-theatre-theatre-written-machines.html>

<sup>7</sup><https://www.respekt.cz/tydenik/2020/46/ze-zivota-robotu>

<sup>8</sup><https://procne.ihned.cz/c1-66819080-scenar-robot-ve-svandove-divadle-pise-hru-k-vyroci-r-u-r-umela-inteligence>

<sup>9</sup><https://kultura21.cz/divadlo/21255-ai-kdyz-robot-pise-hru-svandovo-divadlo>

<sup>10</sup><https://www.goethe.de/ins/cz/cs/kul/mag/21955498.html>

<sup>11</sup><https://www.abbanews.eu/educazione-lavori-e-ricerca/io-robot-la-linguistica-computazionale-arriva-a-teatro/>

<sup>12</sup>[https://www.idnes.cz/kultura/divadlo/daniel-hrbek-svandovo-divadlo.A200306\\_155904\\_divadlo\\_rik](https://www.idnes.cz/kultura/divadlo/daniel-hrbek-svandovo-divadlo.A200306_155904_divadlo_rik)

<sup>13</sup><https://www.startupjobs.cz/newsroom/rur-v-case-umele-inteligence-aneb-kdyz-divadelni-hru-pise-robot-z-matfyzu>

<sup>14</sup><https://www.matfyz.cz/clanky/1505-umela-inteligence-pise-divadelni-hru>

<sup>15</sup><https://www.tacr.cz/propagace/>

<sup>16</sup><https://www.innocamp.cz/>

<sup>17</sup><https://www.nocvedcu.cz/udalost/1787-theatre-umela-inteligence-pise-divadelni-hru>

<sup>18</sup><https://www.sciencecafe.cz/sc-rozhovor/science-cafe-rozhovor-2/>

- Český rozhlas: *Robot píše divadelní hru*.<sup>19</sup>

## 6 Závěr

V roce 2020 jsme na projektu THEaiTRE splnili všechny vytyčené cíle a úspěšně se blížíme ke splnění prvního výsledku, premiéře první automaticky vygenerované divadelní hry. Zároveň už pracujeme na nové verzi aplikace, která bude řešit některé nedostatky předchozí verze a měla by tak vést k ještě lepšímu scénáři druhé divadelní hry.

## A Příloha: návrh projektu

Zkrácený text původního návrh projektu podaného na TAČR. Návrh byl agenturou přijat a vyhodnocen jako nejlepší z podaných návrhů ve výzvě ÉTA 3 se ziskem plného počtu bodů.

### A.1 Souhrn informací o projektu

**Název projektu** THEAITRE: Umělá inteligence autorem divadelní hry?

**Začátek projektu** 04/2020

**Konec projektu** 04/2022

**Hlavní řešitel** Mgr. Rudolf Rosa, Ph.D.

### A.2 Cíle návrhu projektu a relevance k programu

#### A.2.1 Cíle návrhu projektu česky

Tento projekt kombinující počítačovou lingvistiku a divadelnictví má za cíl přímým způsobem reflektovat nástup a zkoumat roli umělé inteligence ve společnosti a v umění a možnost její spolupráce s člověkem. Navrhujeme spojit teatrologický výzkum s počítačově lingvistickým výzkumem, kde aplikace poznatků a výsledků z těchto oborů povede k sestavení a natrénování systému pro automatické generování scénářů divadelních her. Tento systém následně bude naším týmem interaktivně využit pro vytvoření hry, která bude realizována hereckým souborem ke 100. výročí premiéry dramatu R.U.R. Vzniklé dílo a jeho přijetí diváky budeme zpětně zkoumat v rámci obou vědních oborů a získané poznatky využijeme pro další zdokonalení systému, který zpřístupníme na webu, a pro realizaci dalších her.

#### A.2.2 Cíle návrhu projektu anglicky

This project, combining computational linguistics and theatre, aims to directly reflect the onset and explore the role of artificial intelligence in society and art and the possibility of its cooperation with humans. We propose to combine theatrical research with computational linguistic research: the application of results from these fields will lead to the creation and training of a system for

<sup>19</sup><https://program.rozhlas.cz/zaznamy#/radiozurnal/23/2020-10-10>

automatic generation of theatre play scripts. This system will be interactively used by our team to create a play that will be put on by a theatre on the occasion of the 100th anniversary of the R.U.R drama premiere. The resulting work of art and its acceptance by viewers will be examined, and the knowledge gained will be used to improve the system and to put on another play.

### A.2.3 Naplnění cílů programu

V současné době se rozvoj umělé inteligence (AI) začíná promítat do kulturního života člověka. Konkrétně v oblasti automatického generování textu vědci již dokáží generovat například básně či fake news. V projektu se pokusíme vytvořit komplexnější dílo: scénář divadelní hry. Ve spolupráci odborníků z DAMU, MFF UK a Švandova divadla prozkoumáme možnosti tvorby scénáře, do jaké míry můžeme zapojit umělou inteligenci a kde bude třeba spolupráce lidského faktoru. Vyjdeme z metod, poznatků a výzkumu v oboru divadelních věd i počítačové lingvistiky a využijeme je pro vývoj a učení AI systému pro interaktivní automatické generování scénářů divadelních her. Přivedeme tak AI do oboru divadelních věd a zároveň využijeme teatrologické poznatky pro vývoj AI. Aplikace těchto dosud počítačově nevyužitých poznatků umožní počítačovým lingvistům pracovat se složitějšími literárními díly a napomůže divadelním vědcům částečně novým způsobem nahlížet na strukturu divadelních her. Vyvinutý systém ale nemusí být zajímavý jen pro vědce a umělce, vidíme také potenciál jeho využití pro zvláštní účely, například pro generování scénářů přizpůsobených divákům se speciálními potřebami.

Na základě vygenerovaného scénáře bude uvedena divadelní hra a v následném dalším vývoji automatického generování scénáře budeme reflektovat zpětnou vazbu jak od divadelníků, tak od samotného diváka. Divadlo jako aplikační garant zajistí projektu publicitu a propagaci v rámci široké veřejnosti, což povede k celospolečenské diskusi. Předpokládáme, že takto vzniklá divadelní hra a následná reflexe obohatí kulturní prostor v oblasti divadelního umění a zároveň otevře diskusi o potenciálu, úloze a úskalích umělé inteligence v této oblasti. Nejširší veřejnosti pak chceme tímto způsobem přiblížit AI, ukázat její možnosti a nastínit její možnou roli ve společnosti jako asistenta člověka.

### A.2.4 Aspekty inovačního ekosystému SHUV

**Přínosy multidisciplinárních přístupů.** V rámci divadelních věd existuje rozsáhlý výzkum struktury a interpretace divadelních her, jehož výstupy ale nejsou k dispozici v podobě a formátu vhodných k automatickému počítačovému zpracování pro účely strojového učení. Metody a nástroje počítačové lingvistiky zvládají dobře zpracovávat jednoduše strukturované faktické texty (typicky novinové články), ne ale texty s komplexní strukturou či s významem vyjádřeným spíše nepřímou, jako jsou divadelní hry. V rámci projektu se spojí odborníci z obou oborů a vzájemně se obohatí o své poznatky. Společně pak navrhnu a vyvinou nástroje pro práci se scénáři divadelních her a pro tyto nástroje vhodný způsob zachycení rozborů scénářů. Tento výzkum zároveň umožní zrevidovat dosavadní teorie popisující divadelní hry a otevře cestu k podrobnému kvantitativnímu zkoumání scénářů vycházejícímu z metod korpusové lingvistiky.

**Propojení výzkumu technického a netechnického charakteru.** Budeme se přímo zabývat otázkou role AI (umělé inteligence) ve společnosti a v umění, přičemž metodu generování divadelní hry postavíme na symbiotické spolupráci člověka (dramaturg) a AI (hluboké učení, umělé neuronové sítě). AI tak není nepřítelem člověka, který ho má nahradit, ale jeho společníkem, který mu má ulehčit práci a jež má člověk pod kontrolou.

**Využití potenciálu základního výzkumu k aplikacím.** Budeme vycházet z výzkumu struktury divadelních her, jako např. z Poltiho teorie 36 dramatických situací, a z výzkumu automatického generování příběhů. Rozvinutím a úpravou stávajících přístupů vyvineme systém použitelný pro automatické interaktivní generování scénářů divadelních her. Ten bude následně využit pro vygenerování scénářů divadelních her, které budou skutečně uvedeny v divadle. Projdeme tedy celou cestu od teorií až k aplikaci.

### A.3 Věcná náplň návrhu projektu

#### A.3.1 Současný stav poznání nebo existující způsoby řešení podstaty návrhu projektu

Generování literárních děl často využívá nějakou abstrakci nad dějovými situacemi a postavami. Obvykle se vychází ze 36 dramatických situací G. Poltiho (1916), či 31 funkcí a 7 typů postav V. Proppa (1968). Pro konkrétní doménu, danou zadáním projektu a typem zpracovávaných literárních děl, je pak na základě těchto klasických klasifikací sestavena ontologie situací a charakteristik postav, která se použije pro anotaci podkladových dat (literárních textů).

Klasický postup automatického generování příběhu je rozdělený do dvou fází: vymyšlení příběhu, kdy je pomocí technik rozvrhování a plánování sestaven popis postav a příběhu na abstraktní úrovni (například v podobě grafu postav a situací) a vyprávění příběhu, kdy je z abstraktní reprezentace vygenerován text pomocí technik generování jazyka.

Nově se objevují přístupy využívající hlubokých neuronových sítí. Některé využívají klasické abstrakce nad postavami a příběhem, ale techniky rozvrhování a plánování jsou nahrazeny hlubokým strojovým učením. Další experimentální variantou je hierarchické generování, kdy abstrakce nad příběhem je nahrazena jeho stručným shrnutím v přirozeném jazyce. Jiné přístupy generují přímo text literárního díla bez dalších mezikroků a explicitních abstrakcí.

Na základě extenzivních rešerší jsme přesvědčeni, že neexistuje výzkum zaměřený přímo na generování scénářů divadelních her. Mezi nejbližší existující díla vytvořená ve spolupráci člověka a umělé inteligence patří:

- sbírka básní Poezie umělého světa (básně vygenerované AI, člověkem provedený výběr)
- sedmiminutový film Sunspring (AI autorem scénáře, ten byl dále upraven člověkem)
- muzikál Beyond The Fence (AI autorem asi 25% scénáře, dotvořeno člověkem)
- hra Lifestyle of the Richard and Family (návrhy predikce textu SwiftKey Note užity k dokončení některých jednotlivých vět)

Všechna tato díla jsou značně experimentální a mají mnohé problémy.

### A.3.2 Popis novosti navrhovaného výzkumného řešení

Generování literárních děl je značně experimentální oblastí výzkumu a existující postupy mají dosud mnohá omezení. Zejména u generování delších textů je obvyklá nekonzistence či vnitřní spornost textu, odchylování od tématu, ale i nízká zajímavost a různorodost textu. Při řešení budeme vycházet z předchozích poznatků, avšak výše uvedené problémy dosud nebyly uspokojivě vyřešeny, a tedy se je pokusíme řešit novým způsobem, například hierarchickým generováním.

Konkrétní zaměření na divadelní hry pak představuje nový, dosud neřešený problém. Původní Poltího situace sice popisují dramata, ale ontologie sestavené na jejich základě byly dosud užívány zejména pro generování příběhů, nikoli divadelních her.

Zásadní výzvou oproti generování příběhů je absence vypravěče. Zatímco při generování příběhu stačí jednotlivé situace explicitně popsat, u divadelních her vyplývá situace implicitně z replik postav. Bude nutné proto zcela nově prozkoumat, jak se v textu scénáře realizují jednotlivé dějové situace a do jaké míry jsou vůbec v textu rozpoznatelné; to samozřejmě bude mít zpětně vliv na ontologii těchto situací.

Dostupnost vhodných dat pro učení systému je slabá. Data buď neobsahují potřebná metadata (např. shrnutí děje), anebo se jedná o data z příbuzných ale odlišných žánrů (např. filmy a seriály). Pokusíme se data automaticky sesbírat a připravit, nevyhnutelné ale bude i dotváření potřebných metadat, jak v textové podobě, tak s využitím navržených ontologií.

Pro samotné generování scénáře navrhujeme hierarchické postupy, které jsou myšlenkově velmi nové a v praxi dosud málo vyzkoušené. Konkrétní realizaci této již existující myšlenky tedy budeme muset v projektu rozpracovat sami.

Explicitní zapojení člověka do procesu generování textu (přístup “human in the loop”) je myšlenka částečně rozpracovaná teoreticky, ale málo užívaná v praxi. V projektu se pokusíme navrhnout a implementovat efektivní způsob spolupráce AI a člověka pro účely generování scénáře.

### A.3.3 Metodologie řešení projektu

Existující výzkum ukazuje různé možné cesty, ale neříká, který postup je nejvhodnější pro generování scénářů divadelních her. Navrhujeme proto tři různé postupy využívající hluboké neuronové sítě, které budeme v rámci projektu zkoumat a ověřovat a z nichž následně sestavíme výsledný systém.

**Varianta A.** Naivní lineární postup. Jde o základní postup, vhodný pro generování velmi krátkých textů. Je založený na použití jazykových modelů (RNN), které generují text slovo po slově a větu po větě, tak aby nově vygenerované slovo či věta navazovaly na předcházející. Takový postup nevyžaduje hlubší porozumění textu, systém je natrénován přímo na scénářích jednotlivých her bez využití dalších metadat; je ale sám o sobě pro generování celé divadelní hry zcela nevhodný, neboť nezvládá u delšího textu udržet myšlenku a konzistenci. Navrhujeme jej jednak jako základní kontrastivní postup, oproti kterému budeme porovnávat pokročilejší varianty, a dále jako možnou komponentu pro lokální generování kratších úseků scénáře.

**Varianta B.** Hierarchický ontologický postup, inspirovaný klasickými přístupy ke generování příběhu. Nejprve prozkoumáme zavedené ontologie abstrakcí nad dějem a postavami, určené obvykle pro analýzu a generování příběhů, kontrastujeme je s divadelními hrami a následně navrhujeme novou ontologii odpovídající divadelním hrám, s důrazem na jevy podstatné pro generování nových her. Poté rozebereme a označujeme jednotlivé divadelní hry pomocí vytvořené ontologie, a to jak na úrovni jednotlivých replik, tak na úrovni scén, dějství a celé hry. Systém pak naučíme hierarchicky vygenerovat nejprve abstraktní shrnutí celé hry, z něž vygeneruje shrnutí jednotlivých dějství, pak scén a až v posledním kroku realizuje abstraktní situace pomocí konkrétních posloupností textových replik. Takový postup může vhodně abstrahovat nad textem hry, což mu umožní generovat smysluplné příběhy jako posloupnosti jednotlivých na sebe navazujících situací. Zároveň je ale tento postup velmi náročný na vytvoření podkladových dat s bohatou anotací, která v požadované podobě neexistují. Pro plné nasazení tohoto postupu je proto nutné buď hledat metody částečně automatického vytváření potřebných dat, anebo úlohu zjednodušit natolik, aby si vystačila s velmi malým množstvím dat.

**Varianta C.** Hierarchický textový postup (text expansion). Tento postup vychází z varianty B, ale abstraktní shrnutí s využitím ontologií nahrazuje shrnutími textovými. Vytvoření příslušné ontologie stále může být užitečným mezikrokem při zkoumání toho, které části scénáře jsou podstatné pro posun děje a které situace je třeba popsat ve shrnutí děje na různých úrovních detailu. Samotné hry pak ale stačí opatřit textovými shrnutími, což je přívětivější pro výzkumníky věnující se tomuto úkolu a zároveň umožňuje využití již existujících shrnutí tohoto typu z Wikipedie či odborné divadelní literatury. Textová shrnutí ale neposkytují tak vysokou úroveň abstrakce a naučení automatického systému proto bude obtížnější a bude vyžadovat větší množství dat. Zároveň ale může být v tomto případě snazší zapojení dat z jiné než divadelní domény.

Nezávisle na zvoleném postupu plánujeme uplatnit myšlenku přímého interaktivního zapojení člověka v roli dramaturga do procesu generování, tzv. “human in the loop”. Jde o koncept spolupráce člověka a AI, kdy AI vykonává úkony obtížné pro člověka a člověk vykonává úkony obtížné pro AI. Práci řídí AI, která v určitých chvílích požaduje od člověka nějaký vstup; např. AI vygeneruje několik možností, jak hru rozvinout, a člověk z nich vybere tu, kterou považuje za nejlepší.

Výstupy systému budeme průběžně odborně analyzovat, zkoumat v čem spočívají jejich nedostatky a následně navrhovat příslušná vylepšení systému. Základní význam pak bude mít skutečná realizace vygenerované hry divadelním souborem a její sehrání před publikem, což projektu poslouží jako extenzivní otestování a poskytne rozsáhlou zpětnou vazbu pro další vylepšování systému.

#### **A.3.4 Harmonogram řešení projektu**

**duben - prosinec 2020** sběr a příprava dat, vývoj první verze systému (všichni účastníci: výzkumníci DAMU, dramaturgové Švandova divadla, výzkumníci MFF UK)

**září - listopad 2020** průběžná evaluace výsledků a poskytování zpětné vazby

pro úpravu systému (všichni)

**prosinec 2020 - leden 2021** dramaturgizace a nazkoušení hry (Švandovo divadlo)

**konec ledna 2021** premiéra pilotní hry (Švandovo divadlo) – z důvodu onemocnění režiséra odsunuta na únor 2021

**leden - březen 2021** sbírání zpětné vazby na realizaci a uvedení hry a její interpretace (všichni)

**únor - prosinec 2021** vývoj druhé verze systému, vylepšování systému na základě zpětné vazby, zvyšování autonomie systému, získávání a tvorba dalších potřebných dat (všichni)

**září 2021 - březen 2022** generování testovacích her a jejich minimalistická realizace v rámci kabaretu Švandova divadla jako forma průběžného testování a evaluace systému, další zlepšování systému na základě zpětné vazby (všichni)

**prosinec 2021 - leden 2022** dramaturgizace a nazkoušení druhé hry (Švandovo divadlo)

**konec ledna 2022** premiéra druhé hry (Švandovo divadlo)

**leden - duben 2022** sbírání zpětné vazby, analýza a vyhodnocení druhé hry a reakce na ni, vyhodnocení celého projektu (všichni)

**průběžně** publikace a prezentace průběhu a výsledků výzkumu na konferencích a v časopisech (všichni)

### **A.3.5 Výzkumná nejistota**

Postupy a technologie, které hodláme v projektu využít, jsou z části dobře zavedené, z části poměrně nové, ale v každém případě dosud nebyly nasazeny tímto způsobem. Existuje proto riziko, že se během práce na projektu ukáže, že použití předpokládaného postupu není pro tuto úlohu dobře možné, či by vyžadovalo delší období výzkumu, než kolik je v harmonogramu projektu vyhrazeno. V první řadě projekt chápeme jako ověřovací. Samotným cílem projektu tedy je ověřit, zda a do jaké míry je možné vygenerovat automaticky divadelní scénář. Zároveň ovšem navrhujeme několik různých metod (Varianty A-C v sekci Metodologie řešení projektu), kdy varianta A je zcela jistě proveditelná, ovšem za cenu pravděpodobně nízké kvality výstupů. V tomto směru spoléháme na zkušený tým Švandova divadla ve schopnosti realizovat hru i podle špatného scénáře; cíl projektu ve smyslu uvedení automaticky vygenerované hry tedy bude v každém případě splněn, ale je možné, že projekt ukáže, že současný stav poznání ještě neumožňuje efektivní využití AI v dramatické tvorbě.

Zásadním problémem pro počítačně lingvistické systémy může být nedostatek vhodných dat pro naučení systému. Z toho důvodu je klíčové jak zapojení odborníků na sběr a úpravu dat (MFF UK), tak divadelních odborníků schopných připravit data (například strukturované popisy či srhnutí konkrétních divadelních her) dle potřeb projektu vytvořit (DAMU). Nedostatek existujících dat



proto může vést ke zpožděním projektu oproti plánu, ale neměl by vést k celkovému selhání projektu, neboť potřebná data budou řešitelé schopni sesbírat či vytvořit.

Může se také ukázat, že některé aspekty, jevy či situace je těžké postihnout a jednoznačně identifikovat ve scénáři; v takovém případě jsme připraveni omezit repertoár těchto jevů, a tedy generovat jen určitou podkategorii divadelních her.

## **A.4 Plánované výsledky projektu**

### **A.4.1 Premiéra pilotní divadelní hry**

V rámci projektu bude ve spolupráci umělé inteligence a divadelních odborníků vygenerován scénář divadelní hry. Dramaturgové Švandova divadla posoudí, zda je scénář vhodný k realizaci formou klasické divadelní hry či zda bude vhodnější využít formu scénického čtení, případně navrhnou jinou formu. Následně opatří scénář scénickými a režijními poznámkami a budou spolupracovat na nazkoušení a uvedení hry v divadle. Následně budou sbírat reakce odborné i laické veřejnosti na uvedení hry.

### **A.4.2 Premiéra druhé divadelní hry**

Po dalším roce výzkumu a vývoje připravíme druhou hru, která bude opět vygenerována za pomoci vytvořeného systému ve spolupráci s dramaturgy Švandova divadla, realizována souborem Švandova divadla, a vyhodnocena projektovým týmem. V tuto chvíli projekt již bude v závěrečné fázi, díky realizaci druhé hry a následné analýze proto vyhodnotíme konečný stav vyvinutého systému a potažmo celý projekt. Uvedením druhé hry se také ukáže potenciál praktického využití automatického generování scénářů.

### **A.4.3 Aplikace na generování divadelních her**

Systém na generování divadelních her vyvinutý v rámci projektu zveřejníme v podobě webové aplikace, pomocí které si libovolný zájemce bude moci vygenerovat scénář divadelní hry dle zadaných parametrů. To umožní zejména dalším dramatikům nepřímou pokračovat v projektu realizací dalších divadelních her stvořených za pomoci AI a tím dále rozvíjet otázku zapojení AI do umělecké tvorby. Aplikace zároveň umožní široké veřejnosti se detailně seznámit s výsledky projektu, a potažmo s AI jako takovou.

### **A.4.4 Odborná kniha shrnující výzkum v projektu a jeho výsledky**

V knize popíšeme od začátku až do konce celý výzkumný proces spojený s vytvořením aplikace na automatické generování divadelního scénáře. Představíme použité výzkumné metody, osvětlíme důvody, proč jsme se rozhodli právě pro tyto metody a navrhujeme možnosti, kterými by se mohl ubírat navazující výzkum a vývoj. Nedílnou součástí knihy bude i představení všech výsledků projektu a jejich zhodnocení a interpretace. Knihu plánujeme vydat v elektronické podobě.

#### **A.4.5 Uplatnění a očekávané pozitivní společenské přínosy a dopady výsledků a výstupů projektu**

Náš projekt se zaměřuje na nový způsob dramatické tvorby, je tedy určen v první řadě dramatikům a divadelním společnostem. V současné době divadlo čím dál hůř konkuruje atraktivnějším formám zábavy, které s sebou nese moderní doba a moderní technologie; je přitom na místě se zamyslet nad společenským dopadem takového vývoje. Náš projekt vytváří pro divadelní tvorbu novou experimentální platformu stavějící na moderním tématu umělé inteligence (AI), která může poskytnout divadlu určité oživení, aktuálnost a přitažlivost pro obecnost se zaujetím pro moderní technologie. Vyvinutou aplikaci dáme volně k dispozici, což umožní její užití pro generování divadelních her za různými účely, zejména pro realizaci dalšími divadelními soubory a pro zkoumání ze strany odborníků z oboru divadelních věd i umělé inteligence. Volby nastavení aplikace mohou také umožnit např. generování her s omezenou slovní zásobou a jednoduchou větovou stavbou, určených pro znevýhodněné diváky se sníženými kognitivními a komunikačními schopnostmi. Další možnosti využití aplikace mohou být pro výuku, dramaterapii či artefietiku.

Hlubším cílem našeho projektu je ale pracovat s tématem role AI ve společnosti a vztahu člověka a AI. V poslední době ve společnosti narůstají obavy ze stále sílícího postavení umělé inteligence, přičemž mnoho lidí, zejména s nižším vzděláním, má jen velice mlhavou představu o tom, co to AI je, jak funguje a čeho je schopná. V projektu budeme pracovat s výsledky nejnovějšího výzkumu v oblasti AI a budeme posouvat současné hranice poznání v oboru generování literárních děl. Na vytvořených divadelních hrách, které budou hlavními výstupy projektu, proto názorně ukážeme, co již dnes umělá inteligence dokáže a co je naopak dosud mimo její schopnosti. Formou her samotných ve spojení s prezentací způsobu vzniku těchto her chceme zpřístupnit a přiblížit AI široké veřejnosti a zároveň nabídnout vizi role AI jako spolupracovníka člověka, nikoli jako konkurenta či nepřítele.

## ÚFAL

ÚFAL (Ústav formální a aplikované lingvistiky; <http://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.

## CKL

As of 1 June 2000 the Center for Computational Linguistics (Centrum počítačové lingvistiky; <http://ckl.mff.cuni.cz>) was established as one of the centers of excellence within the governmental program for support of research in the Czech Republic. The center is attached to the Faculty of Mathematics and Physics of Charles University in Prague.

## TECHNICAL REPORTS

The ÚFAL/CKL 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. The technical reports published in this Series are results of the research carried out in the research projects supported by the Grant Agency of the Czech Republic, GAČR 405/96/K214 (“Komplexní program”), GAČR 405/96/0198 (Treebank project), grant of the Ministry of Education of the Czech Republic VS 96151, and project of the Ministry of Education of the Czech Republic LN00A063 (Center for Computational Linguistics). 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 tectogramatické 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 Urešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Urešová, 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 Otradovcová, 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 Ngųy 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 Urešová, 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 Urešová, Eva Fučíková, Eva Hajičová, *CzEngClass: Contextually-based Synonymy and Valency of Verbs in a Bilingual Setting (CzEngClass: Kontextová synonymie a valence sloves v bilingvím 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*