Masarykova univerzita

Filozofická fakulta

Ústav českého jazyka

Bakalářská diplomová práce

2014 Vojtěch Škvařil

Masarykova univerzita

Filozofická fakulta

Ústav českého jazyka

Český jazyk se specializací počítačová lingvistika

Návrh algoritmu pro vyhodnocení bezkontextových pravopisných chyb

Bakalářská diplomová práce

Vojtěch Škvařil

Vedoucí práce: Mgr. Marek Grác, Ph.D.

2014

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s využitím uvedených pramenů a literatury.

.....

Podpis autora práce

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu práce Mgr. Marku Grácovi, Ph.D., a Mgr. Haně Žižkové za cenné rady a komentáře při tvorbě této bakalářské práce a své rodině, přítelkyni a přátelům za podporu.

Anotace

Bakalářské práce se zabývá tvorbou formalizovaných pravidel (ve formě pseudokódu), která automaticky rozliší různé druhy pravopisných chyb v česky psaném diktátu. Principem vyhodnocení je porovnání uživatelova přepisu se vzorovým textem, který byl přiřazen ke zvukové podobě diktátu. Pozice chyby ve slově a větě je poté využita k identifikaci více než 50 typů pravopisných chyb a jejich stručnému vysvětlení.

Klíčová slova

diktáty, český pravopis, pravopisné chyby, oprava jazykových chyb, e-learning

Annotation

The bachelor thesis deals with creating formalized rules (in a form of a pseudocode), which automatically recognizes various types of spelling mistakes in Czech written dictation. Evaluation is based on comparison of a user's transcription with model text which directly corresponds to an audio representation of dictation. Subsequently, the position of the mistake in a word and sentence is used to identify more than 50 types of spelling mistakes and to their brief clarification.

Key words

dictation, misspellings, czech orthography, correction of misspellings, e-learning

Obsah

1. Úvod 8
2. Teorie 9
2.1 Český pravopis 9
2.2 Metody procvičování diktátů v současných aplikacích 10
2.3 Ajka a Majka 12
2.4 Korpus 12
2.4.1 Korpus SYN2010 13
3. Fungování algoritmu 14
3.1 Vstup pro zadavatele diktátu 16
4. Formalizovaná pravidla a jejich vysvětlení 17
4.1 Analýza důležitosti jednotlivých pravopisných jevů 17
4.1.1 Stručné shrnutí problematických aspektů formalizace kontextových chyb 18
4.2 Pravopis – hláska a písmeno 19
4.2.1 Vyjmenovaná slova 19
4.1.1.1 Podobné a sémanticky rozlišitelné podoby vyjmenovaných slov 19
4.1.1.2 Vyjmenovaná slova bez sémantických rozlišení 21
4.1.2 Psaní *i/y* po písmenu *c* 22
4.1.3 Psaní předpon s-, z- 23
4.1.4 Psaní předložek *s, z* (*z postele, s knihou, s sebou*) 24
4.1.5 Pravopis a výslovnost přejatých slov se *s – z* 24
4.1.6 Psaní samohlásek v zakončení přejatých slov 26
4.1.7 Psaní dis-, dys- 27
4.1.8 Psaní slov se zakončením *-manie* 27
4.1.9 Psaní *n – nn* 28
4.1.10 Psaní spřežek a spřahování 28
4.1.11 Složená přídavná jména 31
4.1.12 Velká písmena 32
4.2 Slovotvorba 32
4.2.1 Přídavná jména zakončená na *-icí – -ící* 33
4.2.2 Přídavná jména zakončená na *-ní – -ný* 34
4.2.3 Typ ačkoli – ačkoliv, kdokoli – kdokoliv 35
4.2.4 Vokalizace předložek (*v třetí – ve třetí*) 35
4.2.4.1 Vokalizace předložek *k, s, v, z* před slovem se stejnou souhláskou jako předložka 35
4.2.4.2 Některé případy, ve kterých nedochází k vokalizaci předložek *k, s, v, z* 36
4.2.4.3 Další způsob vokalizace předpony *k* 36
4.2.4.4 Neimplementované pravopisné jevy 37

4.3 Jevy, které nejsou v Akademické příručce českého jazyka přímo uvedeny 37
4.3.1 Zájmena typu vaši/vaší, ji/jí, ni/ní 37
4.3.2 Psaní bě/bje, vě/vje, pě 38
4.3.3 Souhlásky párové 39
4.3.4 Diakritika 40
4.3.5 Psaní y/i po tvrdých, měkkých a některých obojetných souhláskách 41
4.3.6 Zájmena mně x mě a slova obsahující mě a mně 42
5. Vyzkoušení pravidel na autentických diktátech 43
6. Návrhy na funkční a didaktické funkce 46
7. Závěr 47
Seznam použité literatury 49
Seznam obrázků 51
Seznam tabulek 51

1. Úvod

Osvojení pravopisu mateřského jazyka patří ve vyspělé společnosti mezi očekávané dovednosti člověka. Jeho komplexní zvládnutí souvisí i se znalostí přidružených syntaktických, morfologických a dalších systémů, jejichž výsledkem je bezchybný projev, který je čtenářům lépe srozumitelný. Samotné procvičování pravopisu je v České republice nedělitelnou součástí všeobecného vzdělání. Jedním ze způsobů procvičování je ústní diktovaní textu pedagogem, přičemž žáci jej zapisují. Ve 21. století jsou již však vidět nedostatky tohoto způsobu ověřování znalostí: žák se často nedozví, jaký pravopisný jev jeho chyba představuje, a samotné psaní diktátu se děje jen jednou za jistý čas.

Proto byl vyučujícími Filozofické fakulty a Fakulty informatiky Masarykovy univerzity založen projekt na elektronické procvičování pravopisu, který zastřešuje závěrečné práce několika studentů ze zmíněných fakult. Tato práce částečně navazuje na již dokončené bakalářské práce Barbory Trifanové a Jakuba Rumanovského a další práce z ní mohou vycházet. Výsledkem má být webová stránka pracující se statistikou a pravidly pro kategorizaci jednotlivých chyb. V rámci magisterské práce ji tvoří Jakub Rumanovský. Podobné webové stránky již existují, jak je zmíněno v teoretické části, ale nikoli v takto komplexní formě.

Cílem této bakalářské práce je vytvořit formalizovaná pravidla (ve formě pseudokódu[1]), která budou schopna rozlišit povahu nekontextových chyb (jejich rozčleněním se zabývá čtvrtá kapitola) v česky psaném textu. Uživatel poté snadno zjistí, v jakém pravopisném jevu chyboval.

V následující kapitole popisuji teoretický podklad této práce. Ve třetí kapitole vysvětluji systém fungování pravidel. Stěžejní kapitola práce zahrnuje popisy formalizace jednotlivých kapitol dle Internetové jazykové příručky (IJP) a odvozené Akademické příručky českého jazyka (APČJ), ze které vychází Pravidla českého pravopisu (PČP). Zde také zdůvodňuji, proč některé problematiky lze formalizovat snadno a jiné nikoli, popřípadě jaké nástroje by bylo vhodné použít. V páté kapitole pravidla aplikuji na diktát psaný žáky, kteří ukončili první stupeň základní školy. V šesté kapitole navrhuji administrativní a didaktické funkce, které přínos aplikace zvýší. Příloha práce obsahuje samotný pseudokód v přehlednější formě dokumentu a naskenované diktáty žáků použité v páté kapitole.

[1] Pseudokód vychází z konvence zápisu programovacího kódu, avšak nezahrnuje detailní popis některých jeho náležitostí a obsahuje přirozený jazyk.

2. Teorie

2.1 Český pravopis

Znalost pravopisu je ve společnosti brána za znak vzdělanosti a udržování české kulturní tradice. Tato pravidla jsou však nevynutitelná a uživatel jazyka je může v libovolné míře porušovat. V jistých uměleckých disciplínách to může být i žádoucí. Ve většině případů však píšící riskuje, že pro příjemce komunikace bude složitější pochopit jeho sdělení. Proto je ve výuce českého jazyka kladen na zvládnutí pravopisu a gramatiky důraz, a to zejména na základní a v menší míře i na střední škole (Hlavsa, 1999, s. 5). Pravopis je rozdělen do několika pravopisných principů, které jsou důležité i v kontextu této práce.

V lexikálním pravopisu převažuje princip fonologický (ekvivalentně fonetický), tedy že z grafického znázornění slova lze reprodukovat, jak bude slovo znít. Principy historické a etymologické však princip fonologický komplikují a tvoří výjimky z pravidel. Jsou spjaté s historií vývoje jazyka, popř. původními variantami přejatých slov. Např. psaní *i/y* po obojetných souhláskách není z výslovnosti (pouze v některých nářečích) rozpoznatelné (Grepl et al., 2008, s. 61).

Pravopis morfologický se zabývá koncovkami u ohebných slovních druhů. Zde je pro uživatele jazyka klíčové určit správný skloňovací vzor, který se k danému slovu váže. Pravopis syntaktický je zaměřen na vztahy mezi slovy v rámci věty, přičemž tyto vztahy větu často přesahují (Grepl et al., 2008, s. 62).

S užíváním jazyka se pojí pojmy: úzus, norma a kodifikace. Úzus je „soubor jazykových prostředků, které jsou ve vžité podobě užívány jazykovým společenstvím," a to bez ohledu na správnost (Karlík, Nekula a Pleskalová, 2002, s. 516).

V rámci normy vydělujeme termíny jazyková norma a spisovná norma (event. norma spisovného jazyka). Jazyková norma představuje soubor jazykových prostředků, které uznává a užívá dané jazykové společenství. Patří sem však i vazby zejména předložek, sloves, zájmen na určité tvary, pravidla používání slov atp. Jazyková norma se opírá o úzus, má však vyšší závaznost pro jeho uživatele. Norma spisovného jazyka je vázána na jazykovou kulturu (kultivovanost jazykového vyjadřování). Skládá se z prvků spisovného jazyka, které jsou považovány za závazné. (Karlík, Nekula a Pleskalová, 2002, s. 288–289).

Kodifikace jazykové normy „je vědecké poznání a utvrzení objektivně existující normy spisovného jazyka i zachycení výsledků této poznávací činnosti v slovnících, mluvnicích, pravopisných a výslovnostních pravidlech n. jiných příručkách." (Karlík, Nekula a Pleskalová, 2002, s. 216) Kodifikace je zaměřena především na spisovný jazyk a sjednocení jeho pravidel, snaží se o statické popsání normy. Norma se však vyvíjí, což je řešeno změnou kodifikace nebo různými formami dodatků. Zpřesnění a rozlišení termínů norma a kodifikace vypracovali členové tzv. Pražské školy ve 30. letech. Jako kodifikační příručky jsou brány Pravidla českého pravopisu, Příruční

mluvnice češtiny a Slovník spisovné češtiny, které vycházejí z rozsáhlejších publikací jako Mluvnice češtiny, Akademická příručka českého jazyka atp. (Cvrček, 2006, s. 19).

Současná Pravidla českého pravopisu se řídí podobou z roku 1993 a jeho dodatku (tzv. Pilipův dodatek) ze stejného roku.

2.2 Metody procvičování diktátů v současných aplikacích

V následující podkapitole uvedu webové stránky, které se objevily mezi prvními odkazy při zadání hesel „diktát“, event. „diktáty online“ do internetových vyhledávačů seznam.cz a google.cz.

Obrázek 1 pravopisne.cz



Web pravopisne.cz je jako jediný z uvedených odkazů garantován metodickým portálem RVP.cz (Tuček a Martinek, 2011). Byl založen dvěma studenty Pedagogické fakulty Univerzity Hradec Králové v roce 2011. Stránka obsahuje stovky diktátů rozdělených podle jednotlivých pravopisných jevů, jako psaní vyjmenovaných slov nebo koncovek u podstatných a přídavných jmen. Na začátku těchto kapitol je ve většině případů odkaz s vysvětlením jevu a následně cvičení. Těchto vysvětlení je na webu přes 200, od komplexních, zabývajících se shodou podmětu a přísudku, až po stručná vysvětlení původu některých slov (např. *garnýž*, rozdíl mezi *sezení* a *sedění* atp.). Součástí webu jsou i testy na větné rozbory a českou literaturu. Diktáty vytvářejí v některých případech tvůrci webové stránky, jiné posílají sami uživatelé, přičemž jsou následně zkontrolovány a upraveny pro potřeby webu (Černá, 2012). Ve zvláštní kategorii se vyskytuje i 21 zvukových diktátů, které jsou dostupné ve dvou až třech rychlostech diktování (např. pomalý přednes textu trvá 6 minut a 54 sekund, zatímco rychlý kontrolní pouze 43 sekund). Tyto diktáty nejsou zaměřeny na určitý pravopisný jev, ale na jednotlivé ročníky žáků základních škol dle metodického portálu RVP.cz. Následnou kontrolu správnosti a vyhodnocení diktátu si uživatel dělá samostatně, což je největší rozdíl oproti řešení, kterému se věnuje tato bakalářská práce. Pravopisne.cz také roku 2011 získalo

Cenu rektora Univerzity Hradec Králové 2011. Přes 2700 uživatelů sociální sítě Facebook stránku pravopisně.cz zařadilo do kategorie To se mi líbí, což svědčí o popularitě a nemalé komunitní základně spjaté s tímto webem.

Webovou stránku mojecestina.cz spravuje Mgr. Jana Skřivánková od roku 2008. Obsahuje podobný počet cvičení i vysvětlení pravopisných jevů jako pravopisne.cz. Liší se v možnosti založit uživatelský profil, což na pravopisne.cz není možné. Profil ukládá celkovou úspěšnost z daného diktátu, což pomáhá uživateli mít přehled o zvládnutí daných pravopisných jevů.

Pravděpodobně nejnavštěvovanější webovou stránkou zabývající se diktáty jsou diktaty.cz[2], na které bylo dle uvedeného počítadla spuštěno téměř 19 miliónů diktátů. Jako autor je uveden Jiří Husa. Počet pravopisných a gramatických cvičení je 251. Web neobsahuje žádná vysvětlení pravopisných jevů a není zde garance správnosti předkládaných diktátů.

Rumanovský (2013, s. 6) ve své práci zmiňuje počítačové programy PC Diktáty Holubec a Diktátor, které však mají podobný rozsah a funkce jako uvedené bezplatné webové stránky.

Pro zajímavost uvádím i televizní pořad Diktát, ve kterém Zdeněk Svěrák každoročně předčítá text s obsahem mnoha pravopisných jevů. Pořad není zaměřen na určitou cílovou skupinu. Text je kultivovaným způsobem diktován ve schématu: přečtení celého textu, poté jednotlivé věty, které jsou opakovány, a na závěr opět celý text. Diktující místy upozorňuje na vedlejší okolnosti, které mohou píšícímu s jistou znalostí jevu pomoci. Tento způsob považuji za vzorový, jelikož uživateli poskytuje maximální komfort pro psaní náročného textu. Na obrazovce je vždy vidět aktuální věta bez některých písmen, což redukuje počet pravopisných jevů a zaměřuje se na obtížnější jevy. Od roku 2012 další pokračování pořadu nevzniklo.

Obrázek 2 Televizní pořad Diktát



Jednotlivé diktáty na uvedených webových stránkách jsou ve většině případů zaměřeny jen na jeden pravopisný jev. Vysvětlení daného jevu je buď v jiné části webové stránky, nebo není přítomno nijak. Diktáty s obsahem více typů pravopisných jevů jsou uváděny jako kontrolní nebo souhrnné a jsou rozlišeny podle ročníků základních a středních škol. V případě chyby však musí sám uživatel

[2] Weby www.ewa.cz a diktaty.ewa.cz mají totožný obsah stránek jako diktaty.cz.

vyhodnotit, o jaký pravopisný jev jde, přičemž cílem této práce je vytvořit formalizovaná pravidla, která rozpoznají chybu automaticky.

2.3 Ajka a Majka

„Majka je rychlý morfologický analyzátor ve formě knihovny a řádkového programu. Systém Majka navazuje na předchozí systém pro morfologickou analýzu Ajka, se kterým dává zhruba stejné výsledky. Majka je ale zcela nová implementace kompletně založená na konečných automatech, proto je také mnohonásobně rychlejší a flexibilnější. V základní podobě systém k zadanému slovnímu tvaru přiřadí: základní tvar a gramatickou značku, všechna slova patřící ke stejnému lemmatu, všechna možná slova s diakritikou.“ (Šmerk a Rychlý, 2009) Ajka je používána např. v Informačním systému Masarykovy univerzity nebo internetovém vyhledávači Seznam.cz (Šmerk, 2007). Pokud se v této práci zmiňuji o možnostech Ajky, vycházím z užívání její webové reprezentace[3].

2.4 Korpus

Korpus je vybraný vzorek užívání jazyka, který se pokouší o co nejlepší reprezentaci používání jazyka (Kocek, Kopřivová a Kučera, 2000, s. 5). Jde o hlavní nástroj korpusové lingvistiky a používá se i v dalších oblastech počítačové lingvistiky. Korpus má formu souboru textových dat, která jsou zpracovávána elektronickou cestou. Používá se např. k testování jazykovědných hypotéz, vyhodnocení četnosti pravopisných, tvarových nebo slovotvorných variant. Jeho základní funkcí je vyhledávání četnosti libovolných slov a slovních tvarů, proto slouží i při tvorbě slovníků a korektorů. První vytvořený korpus je britský Brown Corpus z roku 1964 s jedním milionem slov.

Mezi ceněné vlastnosti většiny korpusů patří reprezentativnost. Ta „zdůrazňuje odpovídající zastoupení jednotlivých typů jazyka z hlediska recepce.“ (Kocek, Kopřivová a Kučera, 2000, s. 6) U vyvážených korpusů se tedy bere ohled spíše na pozornost (resp. čtenost), která se textům dostává, než na absolutní množství textů určitých typů. Důležitá je také pestrost zdrojů. Užitečnost korpusu zvyšuje přítomnost tzv. vnější anotace, jež sdružuje informace o roku vydání, jménu autora atp.

Vnitřní anotací (resp. tzv. tagování) se většinou rozumí přiřazení morfologických značek. Tedy určení slovnědruhových charakteristik a lemmatizace[4]. Může probíhat automaticky, manuálně, event. poloautomaticky.

V České republice se korpusům věnuje projekt Český národní korpus (ČNK), v rámci kterého jsou budovány konkrétní korpusy. ČNK je spravován Ústavem Českého národního korpusu (ÚČNK) při Filozofické fakultě Univerzity Karlovy (Kocek, Kopřivová a Kučera, 2000, s. 10). V současné době ÚČNK vede doc. Václav Cvrček. Největší složkou ÚČNK jsou korpusy řady SYN, které mají

[3] Dostupné z nlp.fi.muni.cz/projekty/wwwajka

[4] Základní podoba slova, shodná s podobou ve slovnících, která zahrnuje všechny tvary tohoto slova.

dohromady přes dvě miliardy textových slov (zatím posledním nově vytvořeným korpusem je publicistický SYN2013PUB).

2.4.1 Korpus SYN2010

Korpus SYN2010 je reprezentativní synchronní korpus současné psané češtiny. Obsahuje přibližně 100 milionů textových slov. Navazuje na korpusy SYN2000 a SYN2005, oproti kterým byla „výrazně vylepšena lemmatizace a slovnědruhové značkování.“ (Ústav Českého národního korpusu FF UK, 2010b) Původ textů je ze 40 % beletrie, 27 % odborná literatura a 33 % publicistické texty.

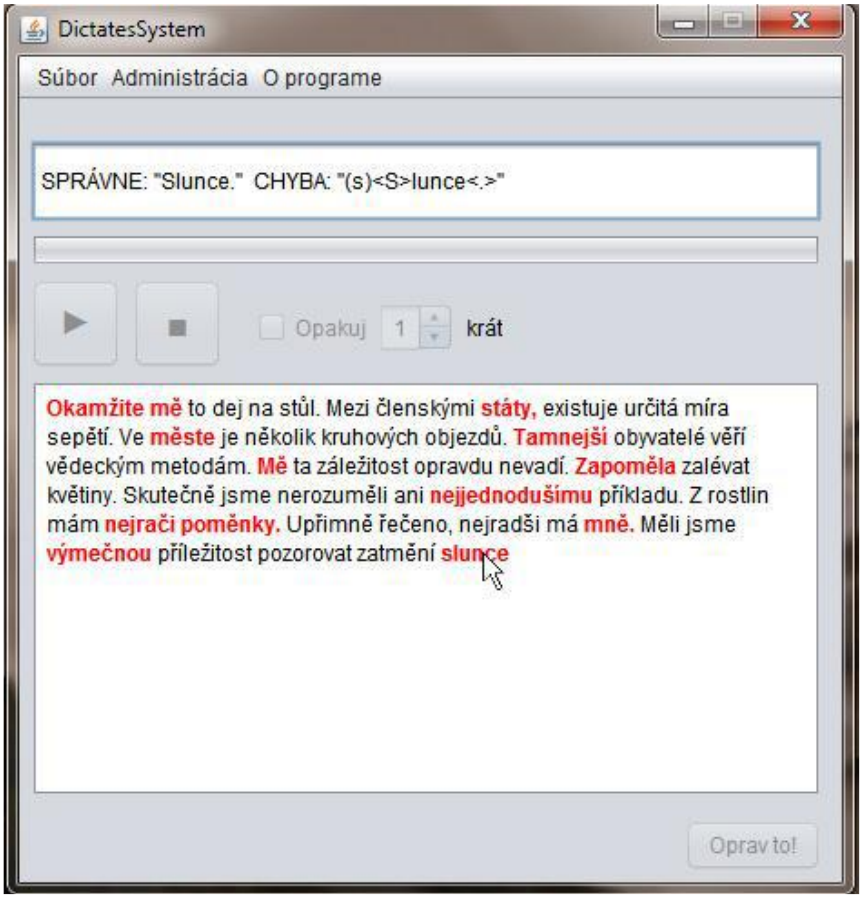
Jelikož jsou pravidla psána v pseudokódu, tak testuji některé jejich části na vyváženém korpusu SYN2010. Tento korpus jsem si vybral pro obsah složený pouze z beletrie, publicistických a odborných textů, u kterých je větší pravděpodobnost, že prošly jazykovou korekturou, na rozdíl od např. webových korpusů czTenTen.

3. Fungování algoritmu

Výchozí aplikací pro implementaci pravidel popsaných v pseudokódu je DictatesSystem, který je popsaný v práci Jakuba Rumanovského (2013, s. 18). V dalším textu se tedy nezabývám značkováním chybného textu (tedy uživatelova vstupu), jelikož považuji tento problém za vyřešený.

Původní systém záznamu chyby: „Značky sú pridelené nasledujúcim spôsobom: ak je slovo, znak alebo skupina znakov navyše alebo chybných, označí sa token alebo skupina tokenov párovými zátvorkami „(“ a „)“. Ak chýba, označkuje sa chýbajúci resp. správny znak/token/tokeny párovými značkami „<“ a „>“. Práca metódy je demonštrovaná príkladom nižšie.“ (Rumanovský, 2013, s. 23) V následujícím kódu jsem provedl úpravu ve značkování chyb. Odlišení chyb a správných znaků je podobné. Pokud se však vedle sebe vyskytne více chybných znaků, tak je každý z nich brán samostatně. Vyhodnocení, zda jde o jednu chybu, je poté rozhodnuto v rámci příslušného pravidla.

Obrázek 3 Uživatelské rozhraní aplikace DictatesSystem (Rumanovský, 2013, s. 30)



Jako způsob formalizovaného zápisu chyby jsem zvolil: *[v,w], x, y, z*. Kde *v* je chybný znak, *w* je správný znak, *x* je číslo pozice neshody znaků (číslované od čísla 1), *y* je chybné slovo a *z* samotná věta, ve které k chybě došlo. Těchto pět proměnných považuji za minimum pro možné vyhodnocení

dotazu. V dále popisovaném pseudokódu operuji přímo s prvními dvěma proměnnými, vyskytují se pro přehlednost ve dvou tvarech:

> **[v,w]** – označují pevnou pozici znaků, jsou nezaměnitelné [špatně, správně]
> **(v,w)** – pozice jsou zaměnitelné, např. v určitém kontextu má slovo obsahovat *n* (a uživatel by napsal *nn*, zápis *[n,0]*[5]), a v jiném *nn* (uživatel by napsal jen *n*, zápis *[0,n]*), pravidlo nadepsané *(n,0)* takto zahrnuje oba typy chyb. V následném vysvětlení se zaměřuji na popis rozdílu daného pravopisného jevu, který je platný pro obě varianty. Pravidla jsou poté vyhodnocena vzhledem k vzorovému textu, z hlediska chybného výrazu je důležitý pouze chybný znak a jeho pozice ve slově.

Proměnné *w* a *x* mohou mít hodnotu *0*, tedy prázdný znak, např. když dojde k vepsání znaku, který je navíc (ve slově *vran(n)í*). Pokud vznikne chyba typu *[e,0]*, tak na místě chybného znaku přijme dočasně vzorový výraz prázdný znak *0*. Např.: VZOR: *k* (*autu*), PSÁNO: *ke* (*autu*), NOVÝ VZOR: *k0* (*autu*); s tímto vzorem je kód následně vyhodnocován. Je bráno v potaz, zda jde o velké, či malé písmeno, avšak je toho využito jen u pravidla, které se týká velkých písmen.

Ohraničení seznamu je znázorněno – *{}*, seznam se skládá z jednotlivých položek, většinou lemmat, a pokud není psáno jinak, seznam vyhledává i v závorkách. Vyhledávání lemmat v analyzátoru Ajka však nezajistí, že dojde k nalezení i odvozených tvarů slova. A např. problematika deverbativ je příliš komplexní na to, aby se dala obsáhnout v některém z pravidel. Proto jsem se zaměřil spíše na základní tvary slov bez odvozených variant, jejichž ověřování není zcela spolehlivé.

Vysvětlení je v kódu značeno *>>* a u každého pravidla obsahuje přesný text, ze kterého by měl uživatel pravopisný jev (nebo alespoň část jevu) pochopit. Některá vysvětlení jsou pro větší přehlednost rozdělena do odstavců. Zdrojem je Internetová jazyková příručka (IJP), pokud není uvedeno jinak. Texty jsem však místy upravil, aby měly spíše podobu pouček. „Poučka vystihuje podstatné vlastnosti pojmů, avšak na rozdíl od definice se u ní připouští nižší míra úplnosti a přesnosti (…).“ (Čechová a Styblík, 1998, s. 47) Společně se zmíněným vysvětlovacím textem dojde vždy k výpisu aktuální hodnoty proměnné *p* (např. *p = „Psaní předpon s-, z-“*). Tato proměnná má za cíl uživateli dát přehled, v jaké problematice se nachází a vysvětlení jevu tak mohl hledat i v případných jiných zdrojích. Zároveň může sloužit jako štítek pro kategorizaci chyb pro zpracovatele dat.

Pokud je vysvětlení doplněno prefixem *NECHYBA*, tak dané odchylky nejsou brány jako běžné chyby, uživateli se však dostane vysvětlení o rozsahu platnosti dublet.

Používané výrazy znaková neshoda a chyba mají podobný význam a jsou zaměnitelné. Pravidlo označuje celý úsek popisovaného kódu (začíná vždy *(v,w)* nebo *[v,w]*). Důležité je i odsazení jednotlivých částí pseudokódu, které rozhoduje o pořadí vyhodnocování.

[5] Princip prázdného znaku „0“ je vysvětlen v následujícím odstavci.

Pokud se v pseudokódu zmiňuji o lemmatu, jde o automatizovaný dotaz v podobě původní (vzorové) varianty slova, ve kterém byla objevena neshoda znaků. Výrazy jsou v uvozovkách uvedeny pro přehlednost. Číslice jsou ve většině případů psány číslicemi, nikoli slovy. V pseudokódu lze vyhledávat jen za pomocí znalosti hlavičky pravidla (např. *[y,i], [b,p]* apod.)

Pokud k úspěšnému vyhodnocení chyby nedojde, chyba je pouze vyznačena (pravděpodobně červenou barvou), bez vysvětlení. Pro další rozvoj pravidel je však užitečné tento výpis uchovat. Ze statistik těchto výsledků je možné dělat závěry o chybovosti jistých jevů. Využití statistik nastiňuji v kapitole 6 – Návrhy na funkční a didaktické funkce.

Zkratky, značky a vysvětlení dalších zákonitostí užitého pseudokódu:

!= – nerovná se (platí „a“ != „b“)

== – zkouška, jestli se dvě hodnoty shodují (platí „a“ == „a“)

AND, **OR** – logické operátory, jsou nadřazené ostatním příkazům, krom obyčejných závorek ()

ADJ, **SUBS**, … – zkrácené označení slovních druhů v latině

Zdroje pravidel a seznamů:

PČP – zdrojem jsou Pravidla českého pravopisu od nakladatelství Academia (Hlavsa, 1999)

IJP – zdrojem je Internetová jazyková příručka

VOJ – autorem je autor textu

Pravidla a seznamy z jiných textů jsou buď nezměněné, nebo lehce upravené. Jsou uvedena vždy za určitým seznamem nebo pravidlem v závorce (např.: „(VOJ)“, „(IJP)“).

3.1 Vstup pro zadavatele diktátu

Vstupem je soubor mp3 s uloženým textem. „[Diktát] sa následne vyhodnotí a opraví podľa prepisu diktátu uloženého priamo v ID3 tagu audio súboru.“ (Rumanovský, 2013, s. 8) Níže popisovaný pseudokód není vázán na určitou formu diktovaných položek a je možné použít diktát jednotlivých slov, slovních spojení i vět. Zvukový soubor je rozdělený do jednotlivých diktovaných vět.

Vkladatel diktátu má možnost vložit dva typy textů. Buď texty přizpůsobené znalostem a schopnostem algoritmu, u nichž může očekávat, že problémové jevy uživateli vysvětlí. U nepřizpůsobených textů je výhoda v získání odlišných dat o chybovosti mimo omezený okruh algoritmizovaných pravidel.

4. Formalizovaná pravidla a jejich vysvětlení

V této kapitole popisuji pravidla, která vyhodnotí, jaké chyby se uživatel dopustil. U každého z pravidel je také uveden popis běhu pravidla a jeho zákonitosti, pokud to není zcela zřejmé. Každé z pravidel má za úkol rozpoznat vhodné výrazy, kterých se daná chyba týká.

Pokud se v pravidlu vyskytuje seznam, tak tu pro přehlednost uvádím pouze jeho část. Tato neúplnost je značena třemi tečkami (...). Celé seznamy jsou k dispozici v samostatném pseudokódu. Uvedené seznamy jsou ve většině případů ze stránek IJP, pokud není uvedeno jinak.

Většina z tzv. vysvětlení (tedy >>) je dále rozvedena v jedné z očíslovaných položek výčtu. Názvy a pořadí kapitol jsou ve většině případů shodné s těmi v APČJ pro snazší orientaci.

Aktuální pravidla v pseudokódu neberou v potaz zvláštní potřeby dysgrafiků, dyslektiků a dalších problémových skupin, kterým pravopis činí větší potíže (Čechová a Styblík, 1998, s. 172). Při získání reprezentativního množství vyhodnocených diktátů od těchto skupin by bylo možné vyhodnotit, které typy chyb je potřeba uživateli vysvětlit.

4.1 Analýza důležitosti jednotlivých pravopisných jevů

Jedním z cílů v práci Barbory Trifanové, na kterou navazuji, bylo zjistit, jaké typy chyb a v jaké míře dělají žáci základních škol. Diktát, který žákům předložila, obsahuje 15 vět o 132 slovech a byl sestaven vyučující na základní škole.

Tabulka 1 Statistika chyb dle pravopisných jevů (Trifanová, 2013, s. 42)

Pravopisný jev	**Počet chyb**	**Četnost [%]**
Interpunkce	69	4,70
Velká písmena	51	3,47
Diakritika	113	7,69
Psaní i/y po měkkých a tvrdých souhláskách	23	1,57
Psaní i/y po obojetných souhláskách	44	3,00
Správné užívání párových souhlásek	22	1,50
Psaní předložek a předpon, rozlišení předložky a předpony	117	7,96
Psaní skupiny bě – bje, pě – pje, vě – vje, mě – mně	5	0,34
Koncovky při skloňování substantiv	248	16,88
Skloňování adjektiv	227	15,45
Shoda podmětu s přísudkem	330	22,46
Chyby z nepozornosti	220	14,98

Diktát by měl obsahovat všechny pravopisné jevy, počet chyb v tabulce však mohl být ovlivněn výběrem výrazů, které nemusely dostatečně prověřit zvládnutí pravopisných jevů, proto beru tyto výsledky spíše orientačně. Celkový počet chyb je 1469. Diktát byl psán žáky, kteří ukončili první stupeň základní školy.

Tyto pravopisné jevy rozděluji na kontextové (v tabulce označené šedou barvou), jejichž správné vyhodnocení je závislé na dodatečné interpretaci okolního textu (v rámci věty, resp. klauze), a nekontextové, jejichž podoba je méně závislá na okolním textu. Kontextovými jevy se zabývám v další podkapitole a nekontextovými už podrobněji v následující kapitole (názvosloví kapitol jsem přebral z IJP, proto se liší).

4.1.1 Stručné shrnutí problematických aspektů formalizace kontextových chyb

Kontextové chyby nejsou hlavním předmětem této práce a v této podkapitole pouze nastíním některá specifika jejich implementace pro správné vyhodnocení chyby. V této práci řešená podoba pseudokódu počítá s dodatečným doplněním těchto pravidel, protože představují významnou část pravopisných chyb.

Koncovky skloňování substantiv a adjektiv patří mezi jevy s největší mírou chybovosti. Jejich správné vyhodnocení je závislé na morfologické a syntaktické analýze. Např. k určení pádu je často nutné provést i syntaktickou analýzu (např. neživotná substantiva se neliší v prvním a čtvrtém pádě). Analyzátor Ajka navíc slova rozlišuje do mnoha vzorů (skloňovacích paradigmat), které se neshodují s vyučovanými vzory, jichž je řádově menší počet. Vzory v analyzátoru Ajka jsou tedy spíše „podvzory“ vyučovaných vzorů. Pro správné vysvětlení problematiky je důležité vzor a další slovní kategorie určit a uživateli vypsat. Skloňování substantiv a adjektiv se navíc pojí se spoustou výjimek.

Shoda podmětu s přísudkem je dalším četným problémovým jevem. Pro správnou shodu je nutné brát v potaz několik faktorů (např. mluvnické kategorie sloves a substantiv, valence, případné několikanásobné podněty, …).

Interpunkce je reprezentována především čárkami a uvozovkami. Čárka ve větě má často sémantický význam, což ztěžuje její formalizaci. V tomto případě je nutné zaznamenat vztahy mezi čárkou a jednotlivými spojkami, případně syntaktickými vazbami, a tak případné vysvětlení zpřesnit, aby šlo daný jev rychleji pochopit. Uvozovky jsem do návrhu pseudokódu nezařadil. Při diktování jsou užívány totiž různé způsoby, od explicitních typu „pozor žáci, přímá řeč“, kdy si žák musí uvědomit, kde v textu přímá řeč začíná a končí a jak se značí, až po sémantické, když přímá řeč přímo vyplývá z textu (např.: *Zeptal se ho: „Byls už v Brně?“*).

Zmíněné pravopisné jevy jsem po domluvě s vedoucím práce z práce vyjmul, zejména z důvodu potřeby jejich komplexního řešení, které by v součtu s níže řešenými jevy bylo pro potřeby bakalářské práce příliš rozsáhlé. Znalost vzorového textu, však bude jistě zlehčující okolností pro formalizaci i v tomto případě.

4.2 Pravopis – hláska a písmeno

4.2.1 Vyjmenovaná slova

Vyjmenovaná slova patří mezi pravopisné jevy s průměrnou mírou četnosti chyb, alespoň mezi nekontextovými chybami (Tabulka 1). Z tohoto důvodu jsou pravidla rozdělena do několika úrovní. Tyto úrovně se liší komplexností poskytnutého vysvětlení a důležitá je zde i posloupnost vyhodnocení uvedená v pseudokódu a v textu níže. Při jeho nedodržení by chyba nemusela být vyhodnocena s nejpřesnějším možným vysvětlením.

První část kódu se zabývá vyjmenovanými slovy, u nichž *i/y* mění význam slova, v určitém kontextu tedy může být správně i varianta s *i/í*. Seznam u písmene *b* je převzatý z kapitoly Předponová slovesa odvozená od *být – bít* na IJP. V druhé části se zabývám pouze variantami s *y/ý* po obojetné souhlásce (*b, f, m, l, p, s, v, z*). Pro snazší vyhodnocení jsou do problematiky zahrnuta i slova přejatá, která po obojetné souhlásce mají *y/ý*, ačkoli jsou vyjmenovaná slova formálně vázaná na slova domácí. Pravopisný princip ovšem zde považuji za podobný, a tak to uvádím i ve vysvětlení. V Pravdová a Svobodová (2014, s. 42) je navíc uvedeno, že se slova přejatá uvádějí v některých učebnicích např. u vyjmenovaných slov po f.

4.1.1.1 Podobné a sémanticky rozlišitelné podoby vyjmenovaných slov

```
p = "Vyjmenovaná slova"
(í,ý),(i,y)
      ((lemma != ADJ AND patří do seznamu, mimo závorky) OR (lemma == ADJ
      AND začátek slova se shoduje s některým ze slov v seznamu, mimo
      závorky, v němž je ignorována kvantita u "í" a "ý" AND chyba nesmí
      být na posledním znaku)) AND lemma tohoto slova má tato 2 písmena na
      stejném místě jako chybný tvar, počítáno od začátku slova:
            předchozí písmeno je "b":
                  {být (ve škole) x bít (nepřítele), …} (IJP):
                        >> „Ke slovesům být (= existovat) i bít (= tlouct)
                        tvoříme řadu předponových sloves (popř. přídavných
                        jmen). Rozlišujeme: “ (IJP) + řádek s daným
                        rozlišením
            předchozí písmeno je "l":
                  {
                  vlis (k vlisovat) x vlys (ozdobný pruh na stěně)
                  …
                  } (PČP):
            předchozí písmeno je "m":
            …
                        >> vypíše řádek s daným rozlišením
```

Pravidlo je uvozeno *(í,ý), (i,y)* z důvodu, abych do něj mohl zařadit i výrazy typu *zbít/zbýt*, u kterých je v určitém kontextu správné *i*.

Další část pravidla se týká užívání adjektiv, která tvoří velkou část vyjmenovaných slov. Z 6 675 318 všech výskytů *y/ý* po obojetných souhláskách je 1 163 929 adverbií (SYN2010). Pravidla však nemají schopnost rozpoznat koncovku, která by mohla ve slovech typu *bosý* evokovat, že jde o vyjmenované slovo. Proto je zde pravidlo, že adjektiva nesmí mít chybu na posledním znaku slova. Jak pravidlo funguje, je popsáno níže v modelovém běhu. U slov typu *aby*, patřící mezi spojky, naopak být může. Rozlišení slovních druhů je u Ajky dostatečně spolehlivé.

Jsou zahrnuty i tvary adjektivních deverbativ, u nichž je první část slova shodná s lemmatem slovesa (např. *zbít – zbit-ý*). Mění se pouze kvantita, která přechází na koncovku adjektiva. Proto je implementováno pravidlo, že v případě, že jde o adjektivum, nedochází ke kontrole shody kvantity u samohlásek *i* a *y*. V takovém případě *zbitý* a *zbít* mají stejný počátek slova a dojde ke správnému vyhodnocení.

Některé výrazy však pravidlo nedokáže ideálně zařadit. Např. u slova s předponami jako *umlít*, které analyzátor Ajka rozpozná, není schopný rozpoznat prefix *u*. Bylo by možné přijít s řešením, v němž je výraz ze skupiny prefixů oddělen od slova, které začíná na posloupnost písmen (nebo i jeden znak jako *u, z, s*…) shodnou s některým z prefixů. Toto naivní řešení by však přineslo více problémů než řešení, např. kvůli kořenům, které se bez prefixů nevyskytují, u některých homonymních výrazů nebo u slov typu *mlít* a *mletý*, kde dochází ke hláskovým změnám. Složitost takového řešení je nad rámec zadání této práce.

Předchozím znakem je myšleno písmeno, které je před chybou. Např. v chybně psaném slově *visoký* je výpis *[i, y], 2, vysoký* a předcházející znak má pozici 1, tedy *v*.

Část pravidla, která kontroluje, zda se pozice znaků shodují i v lemmatu, předchází chybným vyhodnocením typu *[i,y], 4, byl**y***. Zde dochází ke špatné shodě podnětu s přísudkem, jelikož je *y* za *l*, algoritmus by vyhodnotil, že jde o problematiku *y/ý* u obojetných souhlásek. Tímto pravidlem Ajka vygeneruje lemma (*být*) a porovná, zda se znaky *b* a jeden z *(í, ý)* shodují v pozici, která je společná pro lemma ***by**l* a daný tvar ***by**ly*, tedy první dva znaky.

Modelový běh pravidla:

Věta: *Dob**y**tý chlapec nestihl autobus.*

1) Výpis: [y,i], 4, dobitý
2) Ajka: dotaz = dobitý, lemma = dobitý, slovní druh = k2eAgMnSc1d1 = ADJ
3) Předchozí znak je *b*, v seznamu byla nalezena odpovídající položka (při ignorování kvantity u *í* a *ý*), začátky slov u *dobít* a *dobit-ý* se shodují: `dob**ý**t (hrad, úspěch) x dob**í**t (baterii, telefon, kredit)`
4) Výpis odpovídající položky uživateli.

4.1.1.2 Vyjmenovaná slova bez sémantických rozlišení

```
[í,ý],[i,y]
      předchozí písmeno je z {b, f, m, l, p, s, v, z} AND lemma tohoto
      slova má tato 2 písmena na stejném místě jako chybný tvar, počítáno
      od začátku slova AND pokud se jedná o ADJ, chyba nesmí být na
      posledním znaku):
            pokud patří lemma do seznamu, včetně závorek:
                  předchozí písmeno je "b":
                        {
                        obyčej (obyčejný), bystrý (bystře, bystrost,
                        bystřina, Bystřice), …
                        } (IJP):
                  předchozí písmeno je "l":
                        …
                              >> vypiš dané lemma včetně závorky + vypiš
                              obecné vysvětlení níže

                  první 4 znaky jsou „fyto“:
                        >> „Slova s předponou fyto- patří mezi vyjmenovaná
                        slova po f.“ + vypiš obecné vysvětlení níže

                  první znak je "v" AND 2. znak {y,ý}:
                        >> „Slova s předponou vy-/vý- patří mezi
                        vyjmenovaná slova po v.“ + vypiš obecné vysvětlení
                        níže
            jinak:
                  >> obecné vysvětlení: "Jde o vyjmenované nebo z cizího
                  jazyka přejaté slovo. Tato slova mají společné psaní i/í
                  po souhláskách (b, f, m, l, p, s, v, z), které jsou tzv.
                  obojetné, a ke každému ze zmíněných písmen existuje
                  seznam výjimek, u nichž se na místo i/í píše y/ý jako
                  v tomto případě. Důvod je historický a u každého slova
                  jiný, je nutné si jej pamatovat." (VOJ)
```

Průchod algoritmem je podobný jako v předchozím případě. Liší se především ve vysvětlení, které je vždy stejné („obecné vysvětlení“), a případné shody s položkami v seznamech vyjmenovaných slov vysvětlení pouze doplňují. To by mělo v určitých případech pomoci si uvědomit původ slova, což je pro pochopení daného pravopisného jevu klíčové.

Slabinou je absence morfologické segmentace, takže pravidlo nalezne např. tvar *lyžař*, avšak *lyžařka* nikoli. Nejde o zásadní problém, chyba bude vyhodnocena jako vyjmenované slovo i v tomto případě, vysvětlení bude pouze méně obsáhlé.

Při použití Ajky je nutné počítat s rozporem, že v některých případech dojde k vyhodnocení více slovních druhů u jednoho slova z důvodu absence desambiguace[6], např. slovo *mladí* může být jak substantivem, tak adjektivem. V takových případech je lepší slovo brát jako adjektivum a nepočítat tak s posledním znakem.

4.1.2 Psaní *i/y* po písmenu *c*

```
p = "Psaní i/y po písmenu c"
(í,ý),(i,y)
      předchozí písmeno je "c" AND lemma tohoto slova má tato 2 písmena na
      stejném místě jako chybný tvar, počítáno od začátku slova AND pokud
      se jedná o ADJ, chyba nesmí být na posledním znaku:
            >> „Po c se v domácích slovech píše měkké i (např. cit, církev
            atd.). U cizích slov pravidlo o tom, že se po c píše měkké i,
            neplatí. Na jedné straně máme slova jako citron, nacista atp.,
            na straně druhé máme slova, v nichž se po c píše tvrdé y –
            např. cyklon, cynik apod. Pro psaní i a y v těchto slovech
            neexistuje žádné pravidlo, musíme si pamatovat jednotlivé
            příklady.“ (IJP)
```

Průběh vyhodnocení je velice podobný jako u vyjmenovaných slov.

[6] Přiřazení značky (tagu) danému výrazu podle významu a kontextu.

4.1.3 Psaní předpon s-, z-

```
(z,s)
      jde o 1. znak slova:
            lemma je přítomno v seznamu:
                  pokud v {zcestovat, scestovat, zkrápět, …}:
                        >> NECHYBA „U skupiny sloves mezi podobami
                        s předponou s- a z- zásadní významový rozdíl není,
                        a můžete je psát oběma způsoby: zcestovat i
                        scestovat, zkrápět i skrápět...“ (IJP)
                  pokud v {
                  sběh (lidí) x (válečný) zběh,
                  sjednat (přednášku) x zjednat (nápravu),
                  …
                  } (IJP):
                        >> „Některá slova lze psát jak s předponou s-, tak
                        s předponou z-. U většiny z nich je mezi oběma
                        podobami zřetelný významový rozdíl.“ (IJP) + tisk
                        řádku s chybou
            chyba je [s,z] AND lemma daného slova končí na řetězec "ovat":
                  >> „U sloves zakončených na -ovat (aktualizovat,
                  kopírovat) je podoba s předponou z- vždy správná. Jen u
                  některých z nich lze zvolit i předponu s- (zestylizovat
                  i sestylizovat, zkontaktovat i skontaktovat). Výjimkou
                  jsou pouze ta slovesa, která mají s- už v původním jazyku:
                  skandalizovat (neexistuje samotné kandalizovat),
                  skandovat.“ (IJP)
            jinak:
                  >> „Předpona s(e)- naznačuje a) směřování dohromady:
                  scelit, shromáždit se, spojit; b) směřování shora dolů,
                  po povrchu pryč: sklonit, sklopit, shýbnout se; c)
                  zmenšení objemu, až zánik (scvrknout se, shořet). U
                  mnohých slov je třeba si způsob psaní pamatovat, neboť
                  původní význam si dnes již neuvědomujeme: skončit, slevit
                  (sleva).
                  Předponou z(e)- se tvoří: a) od nedokonavých sloves
                  dokonavá, která nemají žádný z významů (moknout ›
                  zmoknout); b) od podstatných nebo přídavných jmen slovesa
                  mající význam ‚učinit nebo stát se tím, co znamená slovo
                  základové‘: ocel › zocelit. U mnohých slov si je třeba
                  způsob psaní pamatovat: zkoumat, zkoušet.“ (IJP)
```

Pravidlo obsahuje poučky pro psaní předpon s(e)-, z(e)- ve slovech domácích a přejatých. U tohoto pravidla je opět důležitá posloupnost vyhodnocení.

1. Jde o dubletní slova, která nesou žádný nebo jen minimální stylový příznak při psaní jedné z variant.

2. Selektivní výběr slov, která nemají jen jednu správnou formu s počátečním *s* nebo *z*. Ze vzorového textu můžeme rozpoznat, která varianta je správná, avšak slova typu *sběh/zběh, sjednat/zjednat* se liší ve významu. Vyjádření obecnějšího pravidla, které by vysvětlilo více než jednu zmíněnou neshodu, je nemožné. Výše je uveden kontext, ve kterém jsou daná slova užita správným způsobem.

3. U odvozených slov (*aktualizovaný*, které nekončí na *-ovat*) je nutné počítat s jistou mírou tzv. false negatives.[7]

4. Poslední vyhodnocení pokryje případy, které „zbyly" a poskytuje obecné vysvětlení. Týká se i slov, u nichž je pravopis už neprůhledný.

4.1.4 Psaní předložek *s, z* (*z postele, s knihou, s sebou*)

```
p = "Psaní předložek s, z (z postele, s knihou, s sebou)"
(z,s)
      slovo má buď 1 znak OR (2 znaky AND 2. znak je "e"):
            >> „S druhým pádem se pojí předložka z (ze). Např. vstal
            z postele; vyšel z banky. Se sedmým pádem se pojí jedině
            předložka s (se). Např. trávil čas s knihou; s pilinami si lze
            hrát." (IJP)
```

Zmíněné pravidlo se týká jen čtyř slov (*z, ze, s, se*), která však mají obrovskou četnost (celková četnost 4 114 426 v SYN2010), což jen málo sráží fakt, že je předložka *se* homonymní se zvratným slovesem *se*. Vysvětlení je možné podat i bez znalosti pádu následujícího výrazu.

4.1.5 Pravopis a výslovnost přejatých slov se *s – z*

Při komponování diktátu není nutné do textu zahrnovat pouze doporučené a základní varianty, aby byla chyba vyhodnocena. V libovolné kombinaci např. slov *charisma*/*charizma* a změn s/z půjde o

[7] False negative, jinak řečeno chyba typu II, je situace, ve které je vyhodnocení chybně negativní.

stejný jev s totožným vysvětlením. Vyhodnotit, ve kterých znacích se mohou výrazy lišit, pomáhá lemma chybného (!) výrazu. Používám je pouze v tomto pravidlu z důvodu shody stylových příznaků. „*Dodatek k Pravidlům uvádí v poznámce doporučení, že tam, kde je v jednom slově spojeno více dubletních možností, je třeba dbát na to, aby se jejich slohový příznak shodoval, tj. aby se nemísily prvky vlastní stylově neutrální podobě s prvky stylově zabarvenými. Píše se tedy báze nebo base, ale ne báse.*“ (Pravdová a Svobodová, 2014, s. 61) V uvedených tvarech báze a base dochází k záměně více než jednoho znaku a tyto změny musí být ve shodě.

```
p = "Pravopis a výslovnost přejatých slov se s – z"
(z,s)
      nejde o 1. znak slova:
            lemma chybného slova je přítomno v seznamu
                  {kurz, kurs, diskurz, diskurs, …} (IJP):
                        >> NECHYBA "Ve slovech zakončených v 1. pádě na
                        skupinu vyslovovanou [-ns, -rs, -ls, např. kurs],
                        v ostatních pádech a ve slovech odvozených
                        vyslovovanou [-nz-, -rz-, -lz-, např. bez kurzu],
                        jsou obě možnosti psaní rovnocenné, a to bez
                        stylového rozlišení." (IJP) AND další znakovou
                        neshodu v chybném lemmatu, která se týká kvantity
                        vokálů, ignorovat

            lemma chybného slova je přítomno v seznamu
                  {analýza – analysa, azyl – asyl, báze – base, …} (IJP):
                        >> „Ve slovech zdomácnělých, kde se původní s vždy
                        v češtině vyslovuje jako [z], se podoby se z
                        považují za základní, tedy stylově neutrální.
                        Podoby se s jsou stylově příznakové, proto se
                        užívají ve specifických (např. úzce odborných)
                        textech. Např.: analýza – analysa, azyl – asyl“
                        (IJP) AND další znakovou neshodu v chybném lemmatu,
                        která se týká kvantity vokálů, ignorovat

            lemma chybného slova je přítomno v seznamu
                  {neologismus, neologizmus, …} (IJP):
                        >> "Ve slovech s příponou vyslovovanou jako
                        [-izmus]/[-zmus] a [-zmus]/[-zma] můžeme psát s i
                        z, přičemž podoba se s se považuje za základní."
                        (IJP) AND další znakovou neshodu v chybném lemmatu,
                        která se týká kvantity vokálů, ignorovat
```

1. Uvedené varianty jsou stylově rovnocenné. (Pravdová a Svobodová, 2014, s. 59) Proto by neměla být případná záměna *s*/*z* vyhodnocena jako chyba.

2. V tomto případě už stylová hodnota není rovnocenná a za základní podobu se považuje varianta se *z* (konzervatoř, 787 výskytů; konservatoř, 8 výskytů v SYN2010). „*Podoba se s je také správná, považuje se však za knižní*“ (Pravdová a Svobodová, 2014, s. 58), proto je možné toto pravidlo vyhodnotit jako tzv. nechybu.

3. U slov zakončených na *[-izmus]/[-zmus]* a *[-zmus]/[-zma]* je za základní považována varianta se *s* (*charisma*). Lze však nalézt výrazy, zejména končící na *[-zma]*, u nichž je v reálném použití preferována varianta se *z* jako například *plazma* (88 výskytů v SYN2010) a *plasma* (6 výskytů v SYN2010). Zmínka je o nich v PČP od Lingey, avšak Pravdová a Svobodová (2014, s. 60) takové případy neuvádějí a v rámci této práce vycházím především z této publikace.

Modelový běh pravidla:

Věta: *Proslavil se psaním pr**os**y.*

1) Výpis první odchylky: *[o,ó], 3, próza*
2) Dojde k vyhodnocení chyby
3) Výpis druhé odchylky: *[s,z], 4, próza*
4) Lemma chybného slova: *prosa*
5) Hledání chybného lemmatu v jednom ze seznamů
6) Výpis odpovídajícího vysvětlení uživateli.
7) Ignorování případné neshody v kvantitě. (v tomto případě *[o,ó], 3, próza*)

4.1.6 Psaní samohlásek v zakončení přejatých slov

Tento pravopisný jev jsem se rozhodl neimplementovat. Koncovky *-en, -in, -iv, -iva, -ivum, -ivní, -emie, -erie, -on, -onek, -ona, -onka, -ped* jsou považovány u slov přejatých z jiných jazyků za dubletní, avšak existuje ustálená množina slov s těmito koncovkami, která dubletní není, např.: *terén, suterén, fén.*

4.1.7 Psaní dis-, dys-

```
p = "Psaní dis-, dys-"
(i,y)
      chyba je přítomna i v lemmatickém tvaru AND lemma chybného slova je
      přítomno v seznamu
            {
            disharmonie    (nesoulad),    diskontinuita    (nesouvislost,
            rozpojenost), …
            dyslexie (čtení), dysortografie (pravopis), …
            } (IJP):
                  >> "Významově poměrně blízké předpony dis- a dys- mají
                  rozdílný původ. O jejich užití obvykle rozhoduje tradice.
                  Latinské dis- odpovídá české předponě roz-, popř. ne-.
                  Řecké dys- znamená ‚zeslabený, vadný, porušený', předpona
                  vyjadřuje něco negativního, špatného." (IJP) + vypiš
                  nalezenou položku
```

Pravopis slov s předponou *dys-* a *dis-* je nutné si pamatovat, jelikož se odvíjí od původu slova. Jelikož není uvedeno jinak, jsou vyhledávány pouze výrazy mimo závorky.

4.1.8 Psaní slov se zakončením *-manie*

```
p = "Psaní slov se zakončením -manie"
(a,á)
      lemma končí na ("manie" OR "ménie") AND má více než 5 znaků:
            >> NECHYBA „Ve složeninách obsahujících -manie je možné psát -
            ménie i -manie.“ (VOJ)
      lemma slova je "ménie":
            >> „Přípustná je pouze varianta ménie. Pouze pokud jde o slovo
            složené (např. cestoménie, pokemanie) je možné psát krátce i
            dlouze.“ (VOJ)
```

Slovo ménie mění podle své pozice ve slově kvantitu samohlásky *a*. Pokud je samostatné, je přípustná pouze varianta s *á*, tedy *ménie*. Pokud je však pouze součástí slova, tak jsou přípustné varianty *-ménie* i *-manie*, odchylka tedy není považována za chybu.

4.1.9 Psaní *n – nn*

```
p = "Psaní n – nn"
(n,0)
      chyba není na 1. znaku slova AND vedle chyby je "n":
            >> „Dvě n se píšou u přídavných jmen odvozených příponou -ní, -ný od podstatných jmen, jejichž kořen končí na -n, -ň (den – denní, holeň – holenní). Dvě n se píšou i u přídavných jmen (a u výrazů od nich odvozených), u nichž výchozí podstatné jméno neexistuje nebo se neužívá, ale kořen je zakončen na -n (bezcenný (bezcennost)). Píše se také v druhých a třetích stupních přídavných jmen a slovech od nich odvozených.
            Jedno n se píše v případech:
            a) u přídavných jmen odvozených příponou -í od jmen pojmenovávajících zvířata, jejichž kořen končí na -n: (havran – havraní), a od příčestí trpných: dán – daný;
            b) slovo raný ve významu ‚brzký, časný' (raná gotika) a od r. 1957 též dceřiný;
            c) u přídavných jmen utvořených od podstatných jmen příponou -ěný: vlna – vlněný (stejně jako sláma – slaměný);
            d) podstatná jména odvozená od přídavných jmen na -ní, -ný příponami -ík, -ice, -ina: deník, okenice, cenina.“ (IJP)
```

Pravděpodobnost, že dojde k absenci na začátku slova je sice mizivá, ovšem při vyhodnocení by šlo o nepřesnost. Chybou by byl překlep, a nikoli problematika psaní *n – nn*. Pravidlo dále kontroluje, zda se v chybném slově vyskytuje ještě další *n*, aby pravidlo úspěšně vyhodnotilo slova stylu *vraní* (a jeho chybnou variantu *vranní*), avšak nikoli chyby, resp. překlepy, u slov typu *vojenský* při variantě *vojeský*.

4.1.10 Psaní spřežek a spřahování

Psaní spřežek je obtížně formalizovatelný pravopisný jev. Vyskytují se zde totiž tvary typu *zmrtvýchvstání*, u kterých je těžké rozhodnout, zda je jejich rozdělená podoba, jako z *mrtvých vstání*, přípustná. Některé výrazy se mohou vyskytovat rozděleně i dohromady beze ztráty významu (*například* i *na příklad*), jiné jsou sémanticky odlišeny (*na to* a *nato*) a u dalších je možná pouze jedna varianta (*na shledanou*). Proto pravidla redukuji na několik nejčastějších výrazů.

Podobný problém je u číslovek, které mohou být psány způsoby *pětadvacet, pět a dvacet, dvacet pět*, (nikoli však *dvacetpět*) a *25*. Množství možných variant a tvarů se také mění u řadových číslovek. Jako alternativu navrhuji, aby zadavatel diktátu u některých slov uvedl (při vkládání textu) více možných způsobů psaní, dělo by se tak např. formou zvláštní závorky.

A.

```
p = "Psaní spřežek a spřahování"
(0,“ “)
      správně je „zato“ OR „za to“:
            >> „Příslovce zato užíváme ve významu ‚náhradou za něco, místo
            něčeho‘ (Vrátil se s nepořízenou, zato rozčílený.). Zato také
            funguje ve významu spojky odporovací, synonymickým výrazem je
            však: Učit se nechtěl, zato se uměl prát. V ostatních případech
            píšeme výraz za to zvlášť: Mám za to, že je schůze už zbytečně
            dlouhá.“ (IJP)
      správně je „nato“ OR „na to“:
            >> „Příslovce nato znamená ‚potom, poté‘: Vypil láhev vitriolu,
            brzy nato zemřel. Hned nato začala předstírat, že se nic
            neděje. Zvlášť píšeme spojení předložky a zájmena: Jdeme na
            to.“ (IJP)
      správně je jeden z {„přitom“, „při tom“, "potom", "po tom"}:
            >> "Zde není vymezení tak jednoznačné. Oslabuje se totiž
            chápání tohoto výrazu jako spojení předložky a ukazovacího
            zájmena a ve většině textů lze volbu odůvodnit jak významem
            ‚zároveň‘ (přitom), tak vztáhnutím na děj předchozí věty, ke
            kterému zájmeno odkazuje (při tom).
            Podobný případ je předložka + zájmeno po tom a příslovce potom.
            Hrál a potom zpíval (‚následovně‘). Hrál a po tom zpíval. (=
            Po hraní na hudební nástroj začal zpívat.)" (IJP)
      správně je „poté“ OR „po té“:
            >>    "Pokud ukazovací zájmeno odkazuje k entitě ženského rodu,
            jsou ve většině případů možné obě varianty: Jako předkrm si
            dáme šunku a po té (‚po šunce‘) hlavní chod. Jako předkrm si
            dáme šunku a poté (‚následovně‘) hlavní chod. Kde však tato
            situace nenastává, tj. pokud odkazujeme k jinému rodu, je možná
            pouze jednoslovná podoba – Dáme si čaj a poté (v časovém významu
            – ‚až dopijeme čaj‘) půjdeme spát." (IJP)
```

Příslovce, která vznikla spřažením předložky a zájmena jako *nato*, *zato*, *přitom*, *potom*, *poté* a jejich původní varianty, jsou poměrně četné (u obou variant: nejméně *poté* 16 696 výskytů, nejvíce *nato* 49 195 výskytů v SYN2010). Zahrnuji tedy všechny varianty uvedené v Akademické příručce českého jazyka. U slov *přitom*, *potom* a jejich původních variant je možné zvážit, zda nepatří mezi tzv. nechyby. Různé varianty spojení *kdovíco* nezahrnuji pro jejich nízkou četnost (103 výskytů u *kdovíco*, 34 výskytů u *kdoví co*).

B.

```
[0,“ ”]
      patří do seznamu
            {
            na shledanou, na viděnou…
            }:
                  >> „Zvlášť píšeme zpodstatnělá přídavná jména, která jsou
                  užívaná pouze či převážně s předložkou na a pojí se
                  s 4. pádem. Jsou to spojení jako: na shledanou, na viděnou
                  apod.“ (IJP)
```

Výrazy, které se mohou plést se spřežkami, v tomto seznamu je možné psát pouze jedním způsobem. Přídavná jména vzniklá spřahováním (typu *níže uvedený – nížeuvedený*) do vyhodnocení nezahrnuji, protože jejich správnost nelze automatizovaným způsobem posoudit. Morfologický analyzátor Ajka výraz *nížeuvedený* nedokáže identifikovat, v korpusu SYN2010 se varianta *níže uvedený* vyskytuje v 10 případech, *nížeuvedený* v žádném z případů.

Mezery u příslovečných spřežek nejsou zahrnuty také. Např. při automatickém oddělení předložek *do, za, na* od začátku slov a zkoušení na přítomnost ve slovníku typu Slovník spisovné češtiny by bylo těžké oddělit slova jako *zařízení*, která by se i po oddělení skládala ze slov, která jsou gramatická, avšak v kontextu věty by nedávala smysl. To by platilo i při případné kontrole slovních druhů (*za* – předložka, *řízení* – substantivum).

4.1.11 Složená přídavná jména

Tento pravopisný jev jsem do pseudokódu zahrnul v omezené míře kvůli těžší rozpoznatelnosti povahy chyby. Zda část slova před spojovníkem končí na *-sko, -cko, -ně, -ově* lze snadno ověřit, avšak další případ už je obtížnější. Např. u slova *zahraničněpolitický* považuji za pravděpodobné varianty *zahraničně politický* (tedy chyba *[„ “,0]*), která není správná v žádném kontextu, a *zahraničně-politický* (chyba *[-,0]*), která může být správná v kontextu *zahraniční a politický*. První složka podobných složenin však může mít různé koncovky. Při případném dalším zahrnutí výrazů obsahujících spojovník je taky nutné počítat s názvy (Brno-střed) a dobře rozpoznat části přídavných jmen.

```
p = "První část přídavného jména je zakončena na -sko, -cko, -ně nebo -
ově"
[0,-]
      chyba není na začátku nebo konci slova AND část před „-“ je z {sko,
      cko, ně, ově}:
            >> „V těchto případech používáme pro oddělení obou složek
            spojovník (ten zde v podstatě plní funkci spojky a). Např.:
            zemědělsko-potravinářský (týkající se zemědělství a
            potravinářství). Někdy je užíván k rozlišení: politicko-
            ekonomický (souřadicí; týkající se politiky a ekonomie) a
            politickoekonomický (podřadicí; týkající se politické
            ekonomie).“ (IJP)
```

Slovní tvar .**(sko|cko|ně|ově)-.** je v korpusu přítomen v 11 621 případech. V každém z nich jde o složené přídavné jméno.

4.1.12 Velká písmena

```
p = "Velká písmena"
(a-z,A-Z)
      obě písmena jsou stejná:
            pokud jde o 1. znak věty:
                  >> „Věta vždy začíná velkým písmenem.“
            pokud jde o 1. znak slova:
                  >> „Názvy jednoslovné (např. jméno Miloš, město Kladno)
                  začínají velkým písmenem. U víceslovných názvů (např.
                  Mrtvé moře) se píše jako velké pouze první písmeno, pokud
                  už neobsahují další název (např. Ústí nad Labem, tedy
                  řeka Labe). K těmto základním pravidlům však existuje
                  množství dodatků (k nalezení např. v Pravidlech českého
                  pravopisu), např. o psaní názvů ulic, živých bytostí,
                  institucí atd.“ (VOJ)
```

Problematice velkých písmen je v APČJ věnováno 42 stran. Pro správné zařazení do jedné z pravopisných zásad je nutné znát povahu pojmenované entity, což se neobejde bez inteligentního vyhledávání v některé z databází zeměpisných, osobních apod. jmen. Tvorba adekvátního vyhledávače a propojení pravidel by však pravděpodobně vydalo na celou další práci, a tak vysvětlení zužuji pouze na dva případy.

V prvním lze snadno ověřit, zda je chybný znak prvním znakem věty, což se váže na nutnou přítomnost velkého písmene na začátku věty. V druhém vysvětlení jsou všechny ostatní případy, tedy ty týkající se pojmenování. Zde krom základních pravidel odkazuji na Pravidla českého pravopisu (event. odkaz na IJP), kde si uživatel bude moci potřebnou kapitolu nalézt sám, jelikož ví, o jaký typ chyby jde.

4.2 Slovotvorba

Do této kapitoly jsem zahrnul jen jevy, které měly potenciál dělat problém kvůli podobné výslovnosti. Např. v konkurenci koncovek typu *-ční – -ivní* či přídavných jmen *typu puškinský – puškinovský* je záměna znaků při ústním diktování málo pravděpodobná.

4.2.1 Přídavná jména zakončená na *-icí – -ící*

```
p = "Přídavná jména zakončená na -icí – -ící"
(i,í)
      pokud lemma končí na ("icí" OR "ící") AND chyba je na 3. pozici od
      konce lemmatu:
            >> „Podoby typu měřicí, řídicí jsou přídavná jména účelová.
            Vyjadřují, k čemu věc slouží: měřicí zařízení je zařízení
            určené k měření, řídicí (ovládací) panel je pojmenován podle
            toho, že slouží k řízení (ovládání).
            Přídavná jména typu měřící, řídící se nazývají dějová. Jimi se
            vyjadřuje děj (činnost) právě probíhající: např. řídící žena
            je žena, která právě řídí. U osob se podoby na -ící používají
            výhradně: řídící pracovník, místodržící. Pomůcka: Do daného
            spojení dosadit jeden člen z dvojic typu: vzdělávací –
            vzdělávající. Je-li vzdělávací středisko, mělo by být i školicí
            středisko, je-li útvar provádějící činnost, pak je i útvar
            řídící činnost.“ (IJP)
```

Přídavná jména zakončená na *-icí* nebo *–ící* v každém skloňovaném tvaru obsahují nijak nepozměněné lemma. Např. slovo *hasicí* v základním tvaru a instrumentál *hasicími*. Takto lze bez komplikací zkontrolovat, zda lemma obsahuje patřičnou koncovku. Zda je výraz adjektivem není nutné ověřovat. V korpusu SYN2010 se lemma s dotazem *.*(i|í)cí* vyskytuje z 97,6 % právě jako adjektivum.

4.2.2 Přídavná jména zakončená na *-ní – -ný*

```
p = "Přídavná jména zakončená na -ní – -ný"
(í,ý)
      chyba je na posledním znaku AND předposlední znak == „n“:
            lemma patří do seznamu
                  {
                  Adresní (např. štítek určený pro napsání adresy) × adresný (‚zcela konkrétní‘, např. výtka konkrétní osobě), Čelní (od podstatného jména čelo, např. kost) × čelný (‚stojící v čele‘, např. představitel státu),
                  …
                  } (IJP):
                        >> „Nevelká skupina jmen, u nichž si -ní a -ný buď konkurují, nebo slouží k rozlišení různých významů. Zakončení na -ný mají přídavná jména s významem hodnotícím, proto např. spojením zábavný pořad vyjadřujeme svůj názor na charakter pořadu, hodnotíme jej, kdežto jako zábavní pořad může být tvůrci označen i pořad, který diváci po zhlédnutí možná ani za zábavný považovat nebudou.“ (IJP) + vytisknout řádek, kde se chybný výraz vyskytuje
            lemma patří do seznamu
                  {
                  Skalní / řid. skalný (ve význ. ‚geologický útvar‘),
                  Niterný / řid. niterní,
                  …
                  } (IJP):
                        >> NECHYBA „Významově totožné varianty.“ + vytisknout řádek, kde se chybný výraz vyskytuje
            lemma patří do seznamu:
                  {Nosní (od podstatného jména nos) × nosný (ke slovesu nosit)}:
                        >> „Tato jména jsou tvořena od různých základů slov.“ + vytisknout řádek, kde se chybný výraz vyskytuje
```

Všechna vysvětlení v tomto slovotvorném jevu jsou odvozena od přítomnosti v některém ze seznamů. Některá adjektiva jsou koncovkami *-ní/-ný* významové rozlišena, avšak jiná nikoli. Měřítko pro rozdělení do jedné ze skupin poskytující odlišné vysvětlení je proto sémantické.

4.2.3 Typ ačkoli – ačkoliv, kdokoli – kdokoliv

```
p = "Typ ačkoli – ačkoliv, kdokoli – kdokoliv"
(v,0)
      chyba je na posledním znaku AND slovo končí na "koli" OR "koliv" AND
      druh != SUBS:
            >> NECHYBA „Zájmena jako např. kdokoli(v), cokoli(v) atd.,
            příslovce kdykoli(v), kamkoli(v), částice nikoli(v) a spojka
            ačkoli(v) mohou mít dvě zcela rovnocenné podoby zakončení: -
            koli a -koliv.“ (IJP)
```

Slova, která končí na *–koli*/*-koliv*, se v korpusu SYN2010 vyskytuje na 68 976 případů. S pravidlem nesouvisí pouze některé tvary substantiv, která nejsou ve vyhodnocení zahrnuta. Jde o tvary jako *sokoli* nebo *Rukoli*.

4.2.4 Vokalizace předložek (*v třetí – ve třetí*)

4.2.4.1 Vokalizace předložek *k, s, v, z* před slovem se stejnou souhláskou jako předložka

```
[0,e]
p = "Vokalizace předložek"
      předchozí znak je jeden z {k, s, v, z} AND slovo má jen 2 znaky:
            začátek následujícího slova == vzorový znak:
                  >> „Před slovem, které začíná stejnou souhláskou jako
                  předložka, vokalizujeme vždy. Např.: ke kořenům, se
                  sestrou.“ (IJP)
            (předchozí znak je „s“ AND 1. znak následujícího slova {z, ž,
            š}) OR (předchozí znak je „z“ AND 1. znak následujícího slova
            {s, š, ž}) OR (předchozí znak je „v“ 1. znak následujícího
            slova {f}) OR (předchozí znak je „k“ 1. znak následujícího
            slova {g}):
                  >> "Tendence k vokalizaci je silná, pokud předložka stojí
                  před slovem začínajícím podobnou souhláskou, tedy s před
                  z-, ž-, š-; z před s-, š-, ž-; v před f-; k před g-.
                  Např.: ke groši, se zemí, ve Finsku. U některých slov
                  však pravidlo nemusí platit (Např.: nevycházím z šoku).
                  Důležitou roli zde hraje právě faktor výslovnosti." (IJP)
```

1. Vždy když se neslabičné předložky (*k, s, v, z*) shodují se začátkem dalšího slova, dojde k vokalizaci předložky (*ke, se, ve, ze*). Z tohoto pravopisného pravidla existuje pouze jedna výjimka, a to spojení „s sebou“, kterým se zabývám v další části kódu.

2. V tomto případě jde spíše o tendenci než o pravidlo, avšak spolehlivé, protože se většinou píše jedním způsobem. Zde pokryjeme vokalizované varianty. Tendenci můžeme vidět například u spojení *ve Finsku*, které se vyskytuje v korpusu SYN2010 ve 133 případech, zatímco *v Finsku* se nevyskytuje v žádném z případů. U spojení *ze železa* jde o 109 případů, *z železa* pouze ve 4 případech.

4.2.4.2 Některé případy, ve kterých nedochází k vokalizaci předložek *k, s, v, z*

```
[e,0]
      předchozí znak je {k, s, v, z} AND slovo má jen 1 znak:
            následující slovo má 1. znak z Vokály:
                  >> „Před slovem začínajícím samohláskou se neslabičné
                  předložky nevokalizují (z "k" se nestane "ke" atp.).
                  Např.: k ústavě, s ementálem.“ (IJP)
            předchozí znak je „s“ AND následující slovo == „sebou“:
                  >> „Výjimka, jde o ustálené spojení. Vyskytuje se pouze
                  v instrumentálu. Např. brát s sebou, vzít s sebou. Sebou
                  vyjadřuje pohyb těla (mrskat sebou, …).“ (IJP)
            1. znak následujícího slova == Konsonanty AND 2. znak
            následujícího slova patří mezi Vokály:
                  >> „Před jednou souhláskou se neslabičné předložky
                  nevokalizují. Např.: k řece, s pentličkami.“ (IJP)
```

1. Zde stačí ověřit, že následující slovo začíná na vokál (porovnáním prvního znaku slova a seznamu vokálů). Seznamy vokálů a konsonantů jsou na začátku pseudokódu.

2. Druhé vysvětlení zachycuje výsadní postavení spojení s sebou, které se používá ve frázích typu vzít s sebou.

3. Pravidlo, že „*před jednou souhláskou se neslabičné předložky nevokalizují*“ (Pravdová a Svobodová, 2014, s. 234), lze ověřit tak, že první znak následujícího slova (řetězce znaků) patří mezi konsonanty a druhý znak mezi vokály.

4.2.4.3 Další způsob vokalizace předpony *k*

```
(u,0)
      předchozí vzorový znak je "k" AND toto slovo má 1 nebo 2 znaky:
            >> „Ku se dochovalo pouze v některých ustálených spojeních, a
            to jen před slovy začínajícími retnou souhláskou, zejména p
            (např. ku příkladu, ku prospěchu) a ve vyjadřování poměrnosti
            (např. pět ku sedmi).“ (IJP)
```

Slovní tvar *ku* je sice méně četný (1564 výskytů v korpusu SYN2010), avšak jeho implementace je snadno proveditelná a bezprobléмová.

4.2.4.4 Neimplementované pravopisné jevy

Tendence přiklánějící se k vokalizaci předložek, či její absenci, je slabší u slov s dvěma a třemi počátečními konsonanty. Tendence v paragrafu 4 b) a c) (Pravdová a Svobodová, 2014, s. 234) jsou málo obecné a zároveň nepřesné pro potřeby formalizace.

Předložky slabičné jsou málo frekventované. Spojení být s to, kdo s koho (71 výskytů) jsem vynechal pro jejich nízkou pravděpodobnost a zaměřil jsem se na četnější jevy.

Dělení slov není třeba v elektronicky psaném textu věnovat pozornost. Text se do určité míry formátuje sám.

4.3 Jevy, které nejsou v Akademické příručce českého jazyka přímo uvedeny

V této podkapitole se věnuji především četným jevům, které jsou v různé míře obsaženy v pravopisných procvičováních (např. pravopisne.cz, mojecestina.cz, …), popř. očekávám jejich výskyt, kvůli způsobu percepce textu, jelikož jde o diktát mluvený.

4.3.1 Zájmena typu vaši/vaší, ji/jí, ni/ní

```
p = "Zájmena typu vaši/vaší, ji/jí, ni/ní"
(i,í)
      (slovo má 2 znaky AND (1. je z {j,n})) OR (slovo má 4 znaky AND
      počáteční 3 znaky jsou ("vaš" OR "naš")):
            >> „Krátké varianty ji/ni, vaši/naši apod. jsou vázány pouze
            ke 4. pádu, ostatní pády jsou jí/ní vaší/naší. Pomůcka: Za dané
            zájmeno dosadit ukazovací zájmena té, tou a tu. Pokud se do
            věty hodí zájmeno tu, patří sem ni, ji (Nestojím o ni. ->
            Nestojím o tu dívku.). V ostatních případech (té, tou) sem
            patří jí, ní (Šli bez ní. -> Šli bez té dívky.).“ (VOJ)
```

Vybral jsem tyto čtyři tvary osobních, resp. přivlastňovacích zájmen. Neznalost skloňovacího paradigmatu těchto zájmen může nahradit uvedená pomůcka.

4.3.2 Psaní bě/bje, vě/vje, pě

```
p = "Psaní bě/bje, vě/vje, pě"
(0,j)
      následující chyba v tomto slově je (e,ě) AND jedno z {b, v, p}
      předchází chybě:
            >> „Skupiny bje a vje píšeme u těch slov, která jsou tvořena
            předponami ob- a v-, přičemž kořen slova začíná na je-. (ob-
            jezdit – objezd; v-jet – vjezd). U ostatních slov píšeme ě
            (oběd, pěnkava, větev). V žádném českém slovu se nevyskytuje
            pjě. Řídkou výjimkou jsou některé názvy, např. Skopje.“ (VOJ)
            AND další chybu v tomto slově ignorovat
```

Fungování tohoto pravidla vysvětlím na slově *objev*, které je někdy nesprávně komoleno jako *oběv*. V tomto případě dochází k více znakovým neshodám naráz, avšak jde jen o jednu chybu. Návrh algoritmu nepracuje v rámci vyhodnocení s více než jedním znakem, proto se zde uchyluji k částečné rekurzi. Chyba *oběv* je v počátku zaznamenána jako *[0,j], 3, objev*. V prvním dotazu se pravidlo ptá, jestli je v tomto slově další znaková neshoda a zda je tato neshoda ve tvaru *[ě,e]* nebo *[e,ě]*, v tomto případě jde o *[ě,e]*. V dalším kroku dochází ke kontrole pozic znaků těchto chyb, která musí být bezprostřední, aby nedošlo k zahrnutí nesouvisející chyby. Další postup algoritmu je podobný jako u jiných pravidel. V modelovém běhu nasimuluji průběh opačné záměny znaků.

Modelový běh pravidla:

Věta: *Obj**e**d je ve tři hodiny.*

1) Výpis: *[j,0]. 3, oběd*
2) Následující chyba ve slově je *[e,ě], 4, oběd* – patří do *(e,ě)*
3) Pozice předchozí chyby = 3, pozice následující chyby = 4
4) Ve vzorovém slově se nachází znak *b*, který předchází *ě*
5) Výpis vysvětlení uživateli.
6) Ignorování vyhodnocení následující chyby.

4.3.3 Souhlásky párové

```
p = " Souhlásky párové"
(b,p), (d,t), (ď,ť), (g,k), (v,f), (ž,š), (c,0)
      pokud je chyba na konci slova OR pokud je chyba (c,0), tak následující
      znak po chybě je "h":
            >> „Souhlásky rozdělujeme na znělé (b, d, ď, g, h, v, z, ž) a
            neznělé (p, t, ť, k, ch, f, s, š). Na konci slov je každá
            souhláska vyslovována nezněle. Pomůcka: Říci si slovo v jiném
            tvaru. Např. potah zní jako [potach], jiný tvar: bez potahu.“
            (VOJ)
      pokud není na konci slova:
            >> „Skupina párových souhlásek se vyslovuje celá buď zněle,
            nebo nezněle, ale označování jednotlivých hlásek se řídí podle
            toho, které hlásky jsou v jiných tvarech téhož slova nebo ve
            slovech příbuzných, např. vézt (vezu), kavka (kavek), sjezd
            (sjezdu), lehký (lehounký).“ (PČP)
```

Pravidlo je zaměřeno jen na nevhodnou substituci znělých a neznělých souhlásek. Rozhodl jsem se pro vyjmutí neshody u znaků *s* a *z*, protože slova s touto záměnou jsou často dubletní a řídí se jinými pravidly. Viz pravidla týkající se předpon *s*/*z* a psaní přejatých slov.

Chybnou spodobu znělosti u písmen *ch* a *h* vyjadřuji absencí nebo přítomností písmene *c*, algoritmus totiž pracuje s *ch* jako dvěma znaky. Např. slovo hloh zní foneticky jako *[hloch]*, což je pro některé uživatele jazyka při zápisu matoucí, výpis při neshodě je tomto případě *[c,0], 4, hloh*. Dále je nutné ověřit, zda je následující znak *h* a zda se vyskytuje na poslední pozici slova. Spodoba znělosti má totiž jinou povahu uprostřed a uvnitř slova z čehož vychází rozdílná vysvětlení.

4.3.4 Diakritika

A.

```
p = "Diakritika"
[(u,ů),ú]
      pokud je chyba na začátku slova:
            >> „Na začátku slova se dlouhé u vždy vyjadřuje jako ú.“ (VOJ)
      pokud chyba není na začátku slova AND (lemma vzoru končí na „úhelník“
      OR slovo začíná jedním z výrazů z Předpony, kterou bezprostředně
      následuje chyba):
            >> „Slovo složené z více kořenů (např. troj + úhelník) si ú
            uchovává i uvnitř slova.“ (VOJ)
      jinak:
            >> „Jde o přejaté slovo (ocún, múza), které se takto odlišuje
            od slov domácích (kůlna, důl, …). Ú se vyskytuje také u
            citoslovcí (hú, vrkú).“ (VOJ)
```

1. Na začátku slova nemůže dojít k jiné situaci. (Např. slova *úhel*, *úlovek*, …)

2. Zde je využit seznam předpon (uveden v pseudokódu), ke kterému mají přístup všechna pravidla. Pokud po jedné z těchto předpon (tedy posloupnosti znaků tvořících předponu) následuje chyba, je vysoce pravděpodobné, že se jedná o složené slovo. (Např. *spoluúčast*, *neúspěch*, …)

3. Pokud se slovo neuplatní v předchozích dvou pravidlech, bude vyhodnoceno jako cizí slovo. (Např. *súra*, *túra*, *saúdskoarabský*, …)

Pravidlo jsem zkoušel (lemma: *[a-t,v-z].*ú.**) v korpusu SYN2010 a 50 nejčetnějších lemmat, která obsahovala *ú* uprostřed slova, pravidlům odpovídala.

B.

```
[(u,ú),ů]
      >> „U nepřejatých slov (např. kůlna, důl) je dlouhé u, které není na
začátku slova, vyjádřeno jako ů.“
```

Zde není třeba ověřovat, zda jde o první znak slova, jelikož žádné takové slovo neexistuje (s *ů* na začátku).

C.

```
(ď,d), (t,ť), (n,ň)
      správně {d, t, n}:
            má následovat jedno z {i, í, ě}:
                  >> „Háček zaniká, i přes odlišnou výslovnost. Měkkost
                  přechází na samohlásku e (stane se z ní ě, např.: dělat)
                  nebo i (např.: dílna).“ (VOJ)
      správně {ď, ť, ň}:
            >> „U ď, ť, ň na konci slova dochází k uchování háčku (např.:
            loď, choť). Stejně tak, pokud následují samohlásky a, o, u.
            (např.: ďábel)“ (VOJ)
```

U souhlásek *ď*, *ť*, *ň* přechází háček na souhlásky *i* a *e*, ovšem nikoli na *a*, *o*, *u*, což může některé uživatele jazyka mást. U písmene *i* se háček také ztrácí, avšak už se neobjevuje na dalším znaku, a tak je nutné počítat i s jeho dlouhou variantou *í* (*dílna*, *divadlo*). Na písmeno *e* háček přechází a jeho dlouhou variantu tedy není nutné brát v potaz.

4.3.5 Psaní y/i po tvrdých, měkkých a některých obojetných souhláskách

A.

```
[ý,í],[y,i]
      předchozí písmeno je jeden z {b, f, m, l, p, s, v, z, ž, š, č, ř, j,
      d, t, n} AND lemma tohoto slova má tato 2 písmena na stejném místě
      jako chybný tvar, počítáno od začátku slova AND pokud se jedná o ADJ,
      chyba nesmí být na posledním znaku:
            předchozí písmeno je z {b, f, m, l, p, s, v, z}:
                  >> >> „Tato slova mají společné psaní i/í po souhláskách
                  (b, f, m, l, p, s, v, z), které jsou tzv. obojetné, a ke
                  každému ze zmíněných písmen existuje seznam výjimek
                  (tzv. vyjmenovaná slova), u nichž se na místo i/í píše
                  y/ý (např.: východ, pýcha, bylina). Y/ý se píše také po
                  některých přejatých slovech. (Např.: fyziognomie,
                  zygota).“ (VOJ)
            předchozí písmeno je z {ž, š, č, ř, j, d, t, n}:
                  >> „Patří mezi pravopisně měkké souhlásky (ž, š, č, ř,
                  c, j, ď, ť, ň), které jsou vždy následovány i/í.“
```

První část vyhodnocení je podobná jako u vyjmenovaných slov. Dvojitým vyhledáváním tzv. předchozích písmen zabraňuji tomu, aby se první část vyhodnocení musela opakovat celá. V druhém vyhodnocení počítám se ztrátou háčku u *ď*, *ť*, *ň*, pokud následuje *i/í*, což je splněno.

Z vyhledávání jsem vyjmul písmeno *c*, jelikož je řešené v pravidlu spjatém s tímto písmenem. Toto pravidlo je nutné vyhodnotit až mezi posledními, jelikož by bylo v rozporu např. s pravidlem 4.1.1.1.

B.

```
[í,ý],[i,y]
      předchozí písmeno je jeden z {h, k, r, d, t, n} AND lemma tohoto
      slova má tato 2 písmena na stejném místě jako chybný tvar, počítáno
      od začátku slova AND pokud se jedná o ADJ, chyba nesmí být na
      posledním znaku:
            >> "Patří mezi pravopisně tvrdé souhlásky (h ch k r d t n),
            které jsou v domácích slovech následovány y/ý." (VOJ)
```

Při kontrole předchozího písmene není nutné ověřovat, zda je souhláska ch nebo h, jelikož samotné písmeno h patří mezi tvrdé souhlásky.

4.3.6 Zájmena mně x mě a slova obsahující mě a mně

```
(n,0)
      chybě předchází "m" AND po chybě následuje "ě":
            slovo má 3 znaky nebo méně:
                  >> "Tvar mě se používá ve 2. a 4. pádu (Na mě nezbylo
                  nic). Tvar mně ve 3. a 6. pádu. (Mně se líbí Ivana. ->
                  Komu, čemu?, 3. pád.) Pomůcka: Zaměnit mě/mně za
                  tebe/tobě. Tebe nahrazuje mě (Na tebe nezbylo nic.), tobě
                  nahrazuje mně (Tobě se líbí Ivana.)." (VOJ)
            jinak:
                  >> "Souhláska n po souhlásce m označuje výslovnost
                  [ň + e]; např. měď, měkký. Tam, kde je souhláska [ň] nebo
                  [n] i v jiném tvaru slova nebo v slově příbuzném, píše
                  se mně; např. zapomněl (pomni), rozumně (rozumný)." (PČP)
```

Pravidlo ověřuje, zda je před chybným znakem přítomna souhláska *m* a za chybným znakem písmeno *ě*.

1. Pokud je splněna předchozí podmínka na přítomnost, slovo musí mít minimálně dva konkrétní znaky (*m* a *ě*) s volitelným *n*, pasují sem tedy slova *mě* a *mně*. Ve vysvětlení objasňuji rozdílné užití těchto zájmen.

2. Zde jsou slova, která mají více než tři znaky, a obsahují v jakékoli části *-mě-* nebo *-mně-*.

5. Vyzkoušení pravidel na autentických diktátech

V této kapitole výše uvedená pravidla aplikuji na diktáty, které obstarala Barbora Trifanová. Jsou od žáků 7. ročníku základní školy. Vypracovaných diktátů bylo 30, bez chyby jich bylo 7. Diktát byl komponován bez přihlédnutí ke schopnostem pravidel.

Text diktátu:

> Naše Národní divadlo je svým způsobem v Evropě ojedinělé. Postavil si je totiž celý národ. Lidé pořádali sbírky. Snad každý něčím přispěl. Základní kámen ke stavbě byl položen 16. května 1868. V roce 1881 postihla český národ tragédie. Národní divadlo vyhořelo. Následovaly nové sbírky a po dvou letech se břeh Vltavy mohl opět pyšnit velkolepou stavbou, na jejíž výzdobě se podíleli nejvýznamnější umělci té doby. Představení v Národním divadle byla zahájena operou Bedřicha Smetany Libuše.

Uvedena je chybná varianta v původní podobě, pokud je znak v závorce, tak chyběl. Chybný znak je podtržený, před chybou je uveden základní formát (tedy pouze dvě proměnné, nikoli pět) záznamu chyby.

Tabulka 2 Úspěšně vyhodnocené chyby

Chyba	**Četnost**	**Četnost [%]**	**Typ chyby**
[0,n] nejvýznam(n)ější	10	11,8	Zájmena mně x mě a slova obsahující mě a mně
[s,z] způsobem	4	4,7	Psaní předpon s-, z-
[z,s] zbírky	3	3,5	Psaní předpon s-, z-
[n,N] národní	3	3,5	Velká písmena
[Č,č] Český	2	2,4	Velká písmena
[e,E] evropě	2	2,4	Velká písmena
[i,y] pišnit	2	2,4	Vyjmenovaná slova
[š,ž] jejíš	1	1,2	Souhlásky párové
[b,p] zbůsobem	1	1,2	Souhlásky párové
[ý,í] sbýrky	1	1,2	i/í po měkkých a obojetných souhláskách
[c,0] břech	1	1,2	Souhlásky párové
[v,V] vltavy	1	1,2	Velká písmena
[s,S] snad	1	1,2	Velká písmena
[n,N] národním	1	1,2	Velká písmena
[k,g] trakédie	1	1,2	Souhlásky párové
[í,ý] nejvíznamnější	1	1,2	Vyjmenovaná slova

Tabulka 3 **Nevyhodnocené *chyby***

Chyba	**Četnost**	**Četnost [%]**
[i,y] následovali	7	8,2
[y,i] podíly	5	5,9
[í,ý] svím	4	4,7
[0,,] stavbou(,)	2	2,4
[í,ý] celí	2	2,4
[i,y] pořádali	2	2,4
[i,y] Vltavi	2	2,4
[0,.] Libuše(.)	1	1,2
[,,0] sbírky,	1	1,2
[,,0] výzdobě,	1	1,2
[y,i] Postavyl	1	1,2
[0,t] [0,„ “][0,s] osmnác(t)|set	1	1,2
[„ “,0] osm|desát	1	1,2
[.,0] tragédie.	1	1,2

Tabulka 4 Nepředvídatelné chyby

Chyba	**Četnost**	**Četnost [%]**
[a,á] Narodním	3	3,5
[i,í] jejiž	2	2,4
[e,é] tragedie	1	1,2
[a,á] Narodní	1	1,2
[e,é] nove	1	1,2
[r,ř] Predstavení	1	1,2
[u,ů] zpusobem	1	1,2
[e,é] Lide	1	1,2
[a,á] Zakladní	1	1,2
[0,1] (1)868	1	1,2
[0,v] (v) Evropě	1	1,2
se	1	1,2
[0,e] [0,m] způsob(em)	1	1,2
[ž,ř] požádali	1	1,2
[a,á] zahajena	1	1,2
[0,a] zahájen(a)	1	1,2

Pozn.: Při evidenci chyb jsem ignoroval chybějící tečky nad *i* a *j*, jelikož jde o typ chyby, který při psaní na počítači není možno uskutečnit.

Tabulka 5 Celkové statistiky diktátů

	Četnost	**Četnost [%]**	**Typů chyb**
Celkem chyb	85	100	46
Vyhodnoceno pravidly	35	41,2	16
Nevyhodnoceno	31	36,5	14
Nepředvídatelné	19	22,3	16

Celkově nelze z jednoho textu, ačkoli psaném 30 žáky, dělat obecně platné závěry. Jsou zde však zřetelné tendence, že velkou část chyb pokryje několik pravidel (velká písmena, párové souhlásky…), velká rozmanitost mezi nepředvídatelnými chybami a časté kontextové chyby.

Pro ilustraci jsem vytvořil modelový diktát, který obsahuje především jevy popsané v pseudokódu. Podtržením jsou označeny jevy, které pravidla zvládnou identifikovat a popsat.:

> Ve vlaku byl ukraden hasicí přístroj. Tento fakt odhalil pan průvodčí při ranní směně. Postupoval podle obvyklého scénáře. Ztrátu oznámil správkyni, kterou přátelsky oslovoval jako kapitánku lodi. Vedoucí prohlásila, že musí jít o jistou mánii a že musí objet místní bazary. Z kapsy si vytáhla klubko klíčů a sjela z parkoviště. Vůz jí však v půli cesty vypověděl službu, a tak musela do Litomyšle podniknout túru. Cestou přemýšlela o místním muzeu, poté si vzpomněla na svou konzervatoř. Vůni lučních bylin však narušil pluh na sousedním poli.

6. Návrhy na funkční a didaktické funkce

Funkce webové aplikace by měla jít směrem sbírání dat a vyhodnocování co největšího množství relevantních dat. Tato data o chybách, kterých se uživatelé dopustili, mohou sloužit k závěrům, jaká pravidla dále rozšiřovat. „Vhodné je vést si evidenci pravopisných nedostatků u jednotlivých žáků a podle potřeby jim pak dávat individuální úkoly.“ (Čechová a Styblík, 1998, s. 177) Vyhodnocovací systém počítá i s klasifikací a tříděním chyb, při vytvoření uživatelských profilů je tedy možné evidovat, v čem uživatel chyboval a generovat diktáty „na míru“ složené z problémových jevů. Samozřejmě je nutné položky z těchto jevů obměňovat, aby se k procvičení nevracely stejné výrazy, u nichž se uživatel mohl naučit správnou podobu. S danou chybou by bylo užitečné uložit i větu, ve které se objevila. Z dat by poté mohlo vyplynout, že např. ve větě *Děti si hrály s drahokamy*. uživatelé často chybují v koncovce u slova *drahokamy*, a pokud by aplikace vyhodnotila, že uživatel často chybuje v koncovkách substantiv, dostal by do vygenerovaného diktátu právě tuto větu. Nebylo by tedy třeba značkovat jednotlivé věty na přítomnost jednotlivých chyb.

S větší znalostí problémových slov je také snazší zjistit, o jaké výrazy seznamy dále rozšířit. Např. u slov typu *viklat*, *vikýř* je nutné si zapamatovat jejich výlučnost. Některé seznamy by bylo možné nahradit hledáním v elektronických slovnících, např. Slovníku spisovného jazyka českého na webových stránkách Ústavu pro jazyk český[8] nebo slovníkovém vyhledávači Debdict[9]. Uživatel by měl mít také možnost jednoduše dát najevo, zda mu dané vysvětlení pomohlo (vedle vysvětlení by např. mohl být palec nahoru a palec dolů). Nepopulárním vysvětlením by následně byla věnována větší pozornost.

Problémem uvedeného pseudokódu je fakt, že některá pravidla platí jen pro daná slova. Pokud jde o skloňovaný základní tvar, je možné pomocí analyzátoru Ajka nalézt základní tvar a porovnat jej se slovníkovým tvarem ve slovníku. U odvozených výrazů (určit např. že slovo *lyžařka* je odvozeno od slova *lyže*) je to možné zatím jen manuálně. Nejlepší řešením by byla komponenta, která by byla schopna rozpoznat tzv. superlemma, tedy původní neodvozený tvar slova.

U diktátů je třeba hledat vhodnou formu diktování, která bude vyhovovat maximálnímu počtu uživatelů. Příklad je možné nalézt např. u, v teoretické kapitole zmiňovaného, televizního pořadu Diktát.

[8] Dostupné z http://ssjc.ujc.cas.cz
[9] Dostupné z https://deb.fi.muni.cz/debdict/index-cs.php

7. Závěr

Cílem této práce bylo popsat nekontextové jevy v českém pravopisném systému pomocí formálních pravidel. Celkově jsem vytvořit 28 pravidel, které se větví do 51 zestručněných vysvětlení pravopisných jevů, přičemž desítky dalších vysvětlení a rozlišení vybraných slov bylo zaznamenáno v seznamech určitých pravidel. Tyto jevy jsem vyhodnotil jako nejproblematičtější či alespoň snadno formalizovatelné. Cíl práce byl tedy z větší části splněn a zbývá tak vytvoření webového rozhraní, které bylo popsáno v práci Jakuba Rumanovského, a přepis pseudokódu do programovacího jazyka.

Porovnal jsem webové stránky, které se věnují procvičování pravopisu, a zhodnotil jejich funkce. Zjistil jsem, že nejlépe je zpracována webová stránka pravopisne.cz, kvůli garanci rámcového vzdělávacího programu (RVP), rozsáhlému počtu cvičení, vysvětlení pravopisných jevů a přítomnosti několika zvukových diktátů. V určitých ohledech je tedy je možné brát jako vzor pro další vývoj řešeného projektu. Rámcově jsem se zmínil o aktuální podobě českého pravopisu. Stejně tak o morfologickém analyzátoru Ajka, který prošel dlouhým vývojem a považuji ho za spolehlivý pro vyhodnocení určitých typů úloh zmíněných v jednotlivých pravidlech. Uvedený korpus SYN2010 jsem používal pro zjištění relevance určitých jevů a nepřímé testování některých pravidel.

Popsal jsem základní funkce pseudokódu, ze kterých vycházejí jednotlivá pravidla v praktické části práce. Důležitý je systém pěti proměnných, který eviduje špatný znak, který napsal uživatel, správný znak, který je uveden ve vzorovém textu a liší se od špatného znaku. Další proměnnou je pozice chyby, díky ní algoritmus může snadno zjišťovat přítomnost určitých znaků v chybném slově, slovo, ve kterém se chyba stala, a celou větu. Uvedena jsou i tzv. vysvětlení, která stručně interpretují pravopisné chyby uživatele.

Ve stěžejní a nejrozsáhlejší kapitole této práce jsem zmínil poměrné zastoupení chybovosti, kterým se zabývala Barbora Trifanová ve své práci. V samotných pravidlech jsem z těchto závěrů nevycházel, ale kvůli vysoké míře kontextových chyb jsem popsal některé aspekty jejich možné formalizace. Podkapitoly „Pravopis – hláska a písmeno“, „Slovotvorba“ a „Jevy, které nejsou v Akademické příručce českého jazyka přímo uvedeny“ obsahují jednotlivá pravidla, vždy s popisem procesu vyhodnocení a ověřování. Některá složitější pravidla jsem doplnil i o průběh vyhodnocení na reálném příkladu pro lepší představu. Za nejlépe zpracované pravidlo považuji psaní *i/y* a *í/ý* po obojetných souhláskách.

Některým částem kódu jsem možná věnoval přílišnou pozornost (např. dubletním tvarům *s/z* a související shodě příznaků), ovšem při reálném užití kódu by delší pasáže neměly příliš ovlivňovat rychlost vyhodnocování. Důležitost evidence dubletních jevů však spočívá ve skutečnosti, aby uživatel nebyl upozorňován na jev, který je v pravopisné normě. Je možné a žádoucí do pseudokódu přidat vysvětlení dalších pravopisných jevů, avšak už současná databáze pravidel spolu s vyznačením všech odchylek od vzorového textu může být uživateli užitečná, a tak bych doporučil aplikaci zveřejnit i s nekompletním popisem všech pravopisných pravidel.

Pravidla jsem následně aplikoval na jeden diktát napsaný 30 žáky, přičemž zde popisovaná pravidla pokryla 41 % chyb. Z výsledků jsem vytvořil čtyři tabulky s četností různých typů chyb. Uveden je i modelový diktát, který klade důraz na pravopisné jevy řešené v předchozí kapitole.

Dále jsem navrhl možné funkce nad rámec řešených pravidel v pseudokódu. Jejich cílem je zahrnout do vyhodnocování další nástroje pro český jazyk, které by uživateli podaly více informací o chybě. Statistické metody mohou rozlišit náročnost jednotlivých jevů a slov. Uživatelský účet může vést přesnou evidenci míry zvládnutí jednotlivých jevů u uživatele.

Seznam použité literatury

CVRČEK, Václav: Kodifikační praxe In: *Sociální aspekty spisovných jazyků slovanských. In: Sborník příspěvků ze zasedání Komise pro spisovné slovanské jazyky při Mezinárodním komitétu slavistů Praha, 16. a 17. září 2005*. 1. vyd. 2006, Acta Slavica et Baltica VI. Praha: Univerzita Karlova v Praze – Filozofická fakulta, Euroslavica; s. 16–35. ISBN 80-7308-155-5.

ČECHOVÁ, Marie a Vlastimil STYBLÍK. *Čeština a její vyučování: didaktika českého jazyka pro učitele základních a středních škol a studenty učitelství*. 2. upr. vyd., v SPN-pedag. nakl. vyd. 1. Praha: SPN - pedagogické nakladatelství, 1998, 264 s. ISBN 80-85937-47-6.

ČERNÁ, Markéta. Server pravopisne.cz využijí i vysokoškoláci, říká spolutvůrce vzdělávacího portálu. In: *Munimedia.cz* [online]. 10.04.2012 [cit. 2014-05-26]. Dostupné z: http://www.munimedia.cz/prispevek/server-pravopisne-cz-vyuziji-i-vysokoskolaci-rika-spolutvurce-vzdelavaciho-portalu-3032/

Diktát. In: *Ceskatelevize.cz* [online]. 03.09.2012 [cit. 2014-04-26]. Dostupné z: http://www.ceskatelevize.cz/ivysilani/10396609755-diktat/

GREPL, Miroslav, Petr KARLÍK, Marek NEKULA a Zdenka RUSÍNOVÁ. *Příruční mluvnice češtiny*. Vyd 2., opr. [i.e. 3. vyd.]. Praha: NLN, Nakladatelství Lidové noviny, 2008, 799 s. ISBN 978-80-7106-980-5.

HLAVSA, Zdeněk. *Pravidla českého pravopisu: školní vydání včetně Dodatku*. Vyd. 2. rozš., v nakl. Fortuna 1. Praha: Fortuna, 1999, 383 s. ISBN 80-7168-679-4.

HOLUBEC, Tomáš. *Holubec – počítačové výukové programy* [online]. 2011 [cit. 2014-04-12]. Dostupné z: http://www.holubec.cz/info/index.php

KARLÍK, Petr, Marek NEKULA a Jana PLESKALOVÁ. *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, 2002, 604 s. ISBN 80-7106484-x.

KOCEK, Jan, Marie KOPŘIVOVÁ a Karel KUČERA. *Český národní korpus: úvod a příručka uživatele*. Vyd. 1. Praha: Filozofická fakulta UK, 2000, 156 s. ISBN 80-85899-94-9.

PRAVDOVÁ, Markéta a Ivana SVOBODOVÁ. *Akademická příručka českého jazyka*. Vyd. 1. Praha: Academia, 2014, 533 s. ISBN 978-80-200-2327-8.

Pravidla českého pravopisu. 2. vyd. Brno: Lingea, 2010, 895 s. ISBN 978-80-87062-88-3.

RUMANOVSKÝ, Jakub. *Systém na trénování diktátů* [online]. 2013 [cit. 2014-04-16]. Bakalářská práce. Masarykova univerzita, Fakulta informatiky. Vedoucí práce Marek Grác. Dostupné z: http://is.muni.cz/th/359581/fi_b

SKŘIVÁNKOVÁ, Jana. *Moje čeština* [online]. © 2012–2014 [cit. 2014-05-12]. Dostupné z: http://www.mojecestina.cz/

ŠMERK, Pavel a Pavel RYCHLÝ. *Majka – rychlý morfologický analyzátor Detail publikace* [online]. 2009 [cit. 2014-05-26]. Dostupné z: http://www.muni.cz/research/publications/935762

ŠMERK, Pavel. Fast Morphological Analysis of Czech. In: *Proceedings of Third Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2009*. Brno: Masarykova Univerzita, 2007. p. 13–16. ISBN 978-80-210-5048-8.

TRIFANOVÁ, Barbora. *Analýza chyb v diktátech žáků po absolvování 1. stupně ZŠ* [online]. 2014 [cit. 2014-06-03]. Masarykova univerzita, Filozofická fakulta. Vedoucí práce Marek Grác. Dostupné z: https://is.muni.cz/auth/th/382965/ff_b/bakalarka_final.pdf

TUČEK, Tomáš a Ondřej MARTINEK. O pravopisne.cz. *Pravopisne.cz* [online]. 2011 [cit. 2014-06-03]. Dostupné z: http://pravopisne.cz/o-pravopisne-cz/

ÚSTAV ČESKÉHO NÁRODNÍHO KORPUSU FF UK. *Český národní korpus – SYN2010*. Praha, 2010a. Dostupný z: http://www.korpus.cz

ÚSTAV ČESKÉHO NÁRODNÍHO KORPUSU FF UK. *Korpus SYN2010* [online]. Praha, 2010b. [cit. 2014-06-03]. Dostupný z: http://ucnk.ff.cuni.cz/syn2010.php

ÚSTAV PRO JAZYK ČESKÝ AV ČR. *Internetová jazyková příručka* [online]. © 2008–2013 [cit. 2013-11-16]. Dostupné z: http://prirucka.ujc.cas.cz

Seznam obrázků

Obrázek 1 pravopisne.cz 10
Obrázek 2 Televizní pořad Diktát.....11
Obrázek 3 Uživatelské rozhraní aplikace DictatesSystem (Rumanovský, 2013, s. 30).....14

Seznam tabulek

Tabulka 1 Statistika chyb dle pravopisných jevů (Trifanová, 2013, s. 42).....17
Tabulka 2 Úspěšně vyhodnocené chyby 43
Tabulka 3 Nevyhodnocené chyby 44
Tabulka 4 Nepředvídatelné chyby 44
Tabulka 5 Celkové statistiky diktátů 45