



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Ondřej Michálek

# **Biblické parafrázování**

Ústav formální a aplikované lingvistiky

Vedoucí bakalářské práce: Mgr. Rudolf Rosa, Ph.D.

Studijní program: Informatika

Studijní obor: Obecná informatika

Praha 2020

Děkuji Janovi Haluzovi za poskytnutá data a své rodině za podporu. Děkuji i vedoucímu své práce doktorovi Rudolfovi Rosovi za podnětné připomínky.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V ..... dne.....

podpis

Název práce: Biblické parafrázování

Autor: Ondřej Michálek

Katedra / Ústav: Ústav formální a aplikované lingvistiky

Vedoucí bakalářské práce: Mgr. Rudolf Rosa, Ph.D., Ústav formální a aplikované lingvistiky

Abstrakt: V této bakalářské práci se zabývám parafrázováním textu na úrovni slov, konkrétně parafrázováním biblických textů v češtině. Součástí parafráze by měla být i modernizace textu, aby se stal srozumitelnějším pro běžného čtenáře. Pro tyto účely jsem vyvinul desktopovou uživatelskou aplikaci, ve které využívám nástroje pro analýzu a generování jazyka: Word2vec a MorphoDiTu. Ve své práci porovnávám výsledky jednotlivých přístupů v parafrázování. Vyhodnocení výsledků sleduje několik kritérií a porovnávám účinnost použitých metod na verších z Bible kralické a Českého ekumenického překladu.

Klíčová slova: parafrázování, modernizace, bible, korpusy textů, Word2vec

Title: Biblical paraphrasing

Author: Ondřej Michálek

Department: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Rudolf Rosa, Ph.D., Institute of Formal and Applied Linguistics

Abstract: In this bachelor's work, I deal with the paraphrasing of text at the word level, namely the paraphrasing of biblical texts in Czech. Paraphrasing should also include modernising the text to make it easier for the average reader to understand. For this purpose, I have developed a desktop user application in which I use tools to analyze and generate language: Word2vec and MorphoDiTa. In my work, I compare the results of each approach in paraphrasing. The evaluation of the results follows several criteria and compares the effectiveness of the methods used on verses from czech translations: Bible kralická and Český ekumenický překlad.

Keywords: paraphrasing, modernization, bible, text corpuses, Word2vec

<b>PŘEDMLUVA</b>	<b>1</b>
<b>1 ÚVOD K PARAFRÁZOVÁNÍ</b>	<b>2</b>
<b>2 BIBLE</b>	<b>3</b>
2.1 Motivace pro práci s biblí	3
2.2 Biblické překlady	3
2.2.1 Rozdílné pojetí bible mezi židy a křesťany	3
2.2.2 Historie překladů	4
2.2.3 Moderní trend překladů	5
2.3 Autorství bible	5
2.3.1 Wellhausenova teorie o vzniku Starého zákona	5
2.3.2 Nový zákon	6
2.3.3 Protiklady v bibli	6
2.4 Obtíže s parafrázováním bez kontextu	7
<b>3 POUŽITÉ NÁSTROJE PRO PARAFRÁZOVÁNÍ BIBLE</b>	<b>9</b>
3.1 Word2vec	9
3.1.1 CBOW	12
3.1.2 Skip-gram	12
3.1.3 Porovnání Skip-gramu a CBOW	13
3.1.4 Kosinová vzdálenost	14
3.2 Morphodita	15
3.2.1 Vstup a výstup	15
3.2.2 Metoda tokenize	16
3.2.3 Metoda analyze	16
3.2.4 Metoda tag	16
3.2.5 Metoda generate	17
3.2.6 Morfologická značka	17
<b>4 METODY ŘEŠENÍ PROBLÉMU</b>	<b>19</b>
4.1 Nástin řešení	19
4.2 Trénování	19
4.3 Parafrázování – Word2Vec	20
4.4 Parafrázování – MorphoDiTa	20
4.5 Parametry Word2vecu	21
4.5.1 Nastavitelné parametry	21
4.5.2 Napevno zvolené parametry	22
4.6 Rešerše	23
4.6.1 Parafrázování bible	23
4.6.2 Modernizace bible	23

4.6.3	Modernizace pomocí strojového překladu	23
<b>4.7</b>	<b>Data</b>	<b>24</b>
4.7.1	Bible	24
4.7.2	Data Svědků Jehovových	24
4.7.3	Fantastika a filosofie	24
4.7.4	Titulky	24
4.7.5	Odborný text	25
4.7.6	Tabulka použitých dat	25
4.7.7	Použité modely	25
<b>5</b>	<b>IMPLEMENTACE</b>	<b>27</b>
5.1	Využití zdroje	27
5.2	Inicializace programu – DirectoryManager.cs	27
5.3	Módy programu – Menu.cs	27
5.4	Úprava zdrojových textů - TransformTXTfiles.cs	28
5.5	Vytvoření modelu – TrainNewModel.cs	29
5.6	Parafrázování textu – Inline_paraphrasing.cs	29
5.7	Úprava parafrázovaného textu – MorphoDiTa.cs	30
5.8	Unit testy	31
<b>6</b>	<b>UŽIVATELSKÁ PŘÍRUČKA</b>	<b>32</b>
6.1	Spuštění aplikace	32
6.2	Natrénování modelu	32
6.3	Parafrázování textu	33
6.4	Nápověda	33
6.5	Problém s internetovým připojením	34
6.6	Natrénované modely	34
<b>7</b>	<b>VYHODNOCENÍ VÝSLEDKŮ</b>	<b>35</b>
7.1	Výběr modelu	35
7.1.1	Zadání dotazníku	35
7.1.2	Gramatická správnost	36
7.1.3	Stejný význam	37
7.1.4	Modernizace	37
7.1.5	Vhodnost modelu	38
7.1.6	Celkové skóre	39
7.2	Parafrázování bible	40
7.2.1	Zadání dotazníku	40

7.2.2	Modernizace	40
7.2.3	Stejný význam	41
7.2.4	Gramatická správnost	42
7.2.5	Celkové skóre	42
<b>ZÁVĚR</b>		<b>44</b>
<b>SEZNAM POUŽITÉ LITERATURY</b>		<b>45</b>
<b>SEZNAM OBRÁZKŮ</b>		<b>48</b>
<b>SEZNAM TABULEK</b>		<b>49</b>
<b>PŘÍLOHA A</b>		<b>50</b>
<b>Pokyny pro hodnocení parafrází</b>		<b>50</b>
<b>Bible 21</b>		<b>51</b>
	Mystika	51
	Právo	52
	Popis pracovního postupu	57
	Podobenství	59
<b>Bible kralická</b>		<b>61</b>
	Modlitba	61
	Poezie	63
<b>PŘÍLOHA B</b>		<b>65</b>
<b>Pokyny pro hodnocení parafrází</b>		<b>65</b>
<b>Parafráze – Český ekumenický překlad</b>		<b>66</b>
<b>Parafráze – Bible kralická</b>		<b>71</b>
<b>ZOBRAZENÍ PROGRAMU</b>		<b>76</b>

## **Předmluva**

V době, kdy je na internetu dostupné obrovské množství odborných či historických textů a knih, se ukazuje, že ne všichni mluvíme stejným jazykem. Čteme-li lékařské zprávy, právní dokument, náboženský text či archeologickou zprávu, vždy se pohybujeme v nějakém uzavřeném jazykovém světě, který pro nás není vždy srozumitelný. Parafrázování textu sice nemůže plně nahradit text původní, například právo stojí na jasné formulaci, lékařská zpráva na přesném specifikování problému apod., ale nezasvěcenému člověku může alespoň přiblížit smysl textu, který se mu zdá jinak nejasný. Jako základní dokument pro parafrázování jsem zvolil biblický text. Jednak v sobě obsahuje poezii, zápisy z kronik, právní nařízení či žánr podobný pohádkám, fantasy a science fiction, jednak je otázka určitého druhu automatizace církevně ožehavým tématem, konkrétně zda a nakolik automatizace může generovat kázání či přetlumočit biblický text. Navíc, původní biblický text je dnešnímu čtenáři vzdálený i časově, tím se stává ještě nesrozumitelnějším a tím spíše by se na něm mohla parafráze uplatnit jako vhodná pomůcka k četbě.

## 1 Úvod k parafrázování

Parafrázování textu znamená extrahovat, generovat příp. rozpoznat informaci ve slově, větším celku, větě, odstavci či celém textu o stejné či případně co nejbližší výpovědní hodnotě.

(1) Pán prstenů je román napsaný Tolkienem.

(2) Tolkien napsal román Pán prstenů.

(3) Tolkien je autorem románu Pán prsten.

Věty (1), (2) i (3) mají stejnou výpovědní hodnotu, věta (3) však v parafrázi přibírá nové slovo – autor, které značí jisté pochopení věty. Pochopení věty je však ne vždy možné bez většího celku, například

(4) Doprovodil jsem dívku na hranici.

(5) Doprovodil jsem dívku na pohřební hranici.

(6) Doprovodil jsem dívku k hraničnímu přechodu.

Věta (4) není jednoznačná, neboť není jasný význam slova „hranice“ a proto se sice zdá, že (5) i (6) by mohly být správně utvořené parafráze, závisí však na kontextu, zda jsem jako vykonavatel děje inkvizitor či hraniční kontrola. Některá homonyma se dají snadno odlišit, pro některá potřebujeme větší celky textu. [3]

V této práci se hlavně věnuji parafrázování pomocí synonym či přibližných synonym.

(7) Dal jsem si víno do čutory

(8) Dal jsem si víno do flašky

Věty (7) a (8) nemají stejnou informační hodnotu, ale v případě modernizace či zjednodušení textu je to právě ten typ parafrázování, na který je zaměřena tato práce.

## 2 Bible

### 2.1 Motivace pro práci s biblií

Bible je nejčtenější a nejznámější knihou v lidské historii. Lidé, kteří se hlásili ke křesťanství v roce 2012, tvoří třetinu světové populace. Kromě křesťanů bibli užívají Svědci Jehovovi či Církev Ježíše Krista Svatých posledních dnů, známá jako mormoni. Na bibli se přímo či nepřímo odkazují mnohá další náboženství či nová náboženská hnutí, jako je například Církev sjednocení, jejíž členové jsou nazýváni moonisti. [1]

Dále je to kniha, nebo spíše soubor knih, které mají obrovskou historickou hodnotu. Její nejstarší části nám mohou říct mnohé o fungování společnosti před 7. stoletím před Kristem. Biblická archeologie potvrzuje i vyvrací mnohé, co je v této knize psáno, ale co se týče rituálních zvyků, názorů a příběhů, je bible unikátní a díky důležitosti pro židovskou víru zakonzervovanou sbírkou po tisíciletí.

Díky tomu je to však kniha, která nás spojuje s jejími autory přes téměř tři tisíciletí a je pro nás proto často nesrozumitelná. Právě díky její oblibě, množství překladů a jejímu vlivu ve světě je bible ideální knihou pro parafrázování neboli „zdnešnění“ starého hebrejsko-řeckého spisu. [2]

### 2.2 Biblické překlady

#### 2.2.1 Rozdílné pojetí bible mezi židy a křesťany

Samotné slovo bible pochází z řeckého τὰ βιβλία (ta biblia), což znamená knihy či svitky. Bible není jednou knihou, ale spíše svazkem menších knih, které mají zásadní vliv na dvě světová náboženství – křesťanství a judaismus. Někdy se hovoří o křesťanské a židovské bibli. Křesťanská bible se dělí na Starý a Nový zákon. Židovská bible, která se hebrejsky řekne תנ"ך (Tanakh), se téměř shoduje s křesťanským starým zákonem. Tanakh je spojením tří částí hebrejské bible: תורה – Torah = zákon, נביאים – Nevi'im = proroci a כתובים – Ketuvim = spisy. Oproti

křesťanskému zákonu má jiné pořadí knih a většinou i jiný počet. Katolická církev uznává 46 knih Starého zákona, ortodoxní (pravoslavné) církve jich uznávají 49. Protestantské církve a církve vzešlé z reformace jich uznávají 39 a jejich seznam knih se shoduje s knihami, které jsou obsaženy v Tanakhu, kde je však několik knih sloučených a židé tedy uznávají knih 24.

Pořadí knih je dáno také tím, že první křesťané, kteří nebyli židy, tzv. pohanokřesťané, začali používat řecký překlad Starého zákona, Septuagintu, která vzešla z alexandrijské židovské komunity do 1. století před Kristem a svojí strukturou je tedy starší než mladší židovské uspořádání knih, které bylo sestaveno na koncili v Javne na konci 1. století po Kristu. Jistá nevraživost a snaha vymezit se proti nově vznikající židovské sektě mohly přispět k tomu, že nějaké knihy byly přidány do biblického kánonu a nějaké naopak ne.

Co se Nového zákona týče, tam panuje mezi křesťany všeobecná shoda. Martin Luther, augustiniánský mnich, který je považován za předního zastávce reformace, sice označil List Jakubův za tzv. „slaměnou epištolu“ a doporučoval se jí moc nezabývat, ale přesto ji ve svých překladech Nového zákona uváděl a podoba Nového zákona se ve všech křesťanských církvích shoduje.

### 2.2.2 Historie překladů

Církevní historie prvních staletí je komplikovaná a období pronásledování se střídala s obdobím klidu. Po Ediktu milánském (r. 313), kdy bylo zrovnoprávněno křesťanství na území římské říše a po přesunutí hlavního města z Říma do Konstantinopole, se začalo na západě i na východě rozvíjet křesťanství hospodářsky, kulturně i politicky. Na konci 4. století papež Damasus I. pověřil Jeronýma, aby vytvořil nový překlad bible, neboť i novozákonní spisy i Septuaginta byly psány řecky a papež si přál, aby se četba bible rozšířila mezi latinsky mluvící lid. Jeroným byl velmi vzdělaný a přeložil sám celý Starý zákon z hebrejštiny a Nový zákon z řečtiny do latiny. Knihy, které neobsahovala hebrejská bible, přeložil ze Septuaginty. Jeho latina byla srozumitelná běžným lidem, neboť použil tzv. vulgární latiny, která se lišila od antické, používané třeba Ciceronem. Jeho překlad se nazývá

Vulgata a pro římsko-katolickou církev byl až do raného novověku závazným překladem.

Když v 16. století Martin Luther začal překládat Starý zákon do němčiny, využil hebrejské texty židovské bible, které však neobsahovaly několik knih, které obsahovala Vulgata, neboť Jeroným využíval k překladu i Septuagintu. Od té doby vzniklo několik překladů do národních jazyků více či méně zdařilých, které jsou často závislé na tom, která církev je zaštiťuje. [7]

### 2.2.3 Moderní trend překladů

Překladů bible je dnes hodně, ať už latinských, anglických nebo českých. Dříve se posuzovaly překlady podle toho, jak moc věrné byly originálu. To usnadňovalo práci s doslovným textem, ale obtížnější bylo, aby ten text zněl skutečně a neumělkovaně. S přibývajícím množstvím překladů se začalo dbát i na to, aby text zněl moderně a případně podal nový pohled. Kromě doslovného překladu se čím dál více objevuje nový typ biblických překladů, parafráze. Zmíním zde dva, které jsou v českém jazyce zásadní. Slovo na cestu a Parabible. Parabible je soubor vybraných příběhů z bible, kde Ježíšovo působení zasadí do České republiky 21. století. Ježíš se narodil v Ostravě, mluvil proti politikům a kněžím, pomáhal alkoholikům atd... Autor se zde úplně odpoutal od původního znění textu a snažil se přetlumočit myšlenku dnešnímu čtenáři. Slovo na cestu je naopak parafrázování biblického textu, které se jinými slovy snaží postihnout tu samou věc. Tento typ se ukazuje jako možná aplikace této bakalářské práce.

## 2.3 Autorství bible

### 2.3.1 Wellhausenova teorie o vzniku Starého zákona

Autorství bible je po dlouhou dobu sporným tématem. Je jisté, že bibli napsalo mnoho autorů za velký časový úsek. Podle Julia Wellhausena<sup>1</sup> jsou ve starém zákoně 4 druhy pramenů, či teologických škol. Označuje je jako P, D, J a E. J jako Jahvista a E jako Elohista jsou autoři, kteří používají odlišné jméno pro Boha a vyznačují se důrazem na příběhy. P jako Priest, neboli kněz, je označení pro texty, které vzešly

---

<sup>1</sup> Julius Wellhausen byl protestantský teolog, jeho teorie pramenů je dodnes z velké části všeobecně přijímána.

z kněžského kruhu a jsou to texty věnující se rituálům. Obsahují návody, jak přesně pracovat s hříchy, jak správně obětovat Hospodinu apod. Poslední je D jako Deuteronomista. Hlavní knihou tohoto pramene je kniha Deuteronomium, kniha zákonů, která dává důraz na vnitřní prožívání víry, oproti kněžským rituálům. Tato teorie je dnes z velké části všeobecně přijímaná. Její důsledky jsou pozoruhodné a vysvětluje časté protiklady v bibli. Jednotliví autoři či možná spíše jednotlivé teologické školy tehdy neměly problém dát dohromady několik verzí jednoho příběhu a zaobalit je svým vlastním teologickým rámcem. Díky tomu máme hned v první knize bible, Genesis, dvě zprávy o stvoření světa, z nichž díky opačnému pořadí tvoření světa a člověka jedna druhou popírá. [6]

### 2.3.2 Nový zákon

Nový zákon je mnohem mladší než Starý, ale i přesto je autorství knih častou záhadou. Je jisté, že ne všechny Pavlovy listy napsal Pavel. I zde se najdou jistá místa, která kladou důraz na něco jiného, a i zde jsou částečně místa, která si mohou protirečít. Nový a Starý zákon sdílejí spoustu témat, v mnoha ohledech však pohlíží na různé oblasti víry, života či hodnot odlišně, neboť je mezi nimi několik staletí rozdílu. Jako kdybychom chtěli dnes vydat knihu s názvem Učitelé českých zemí, kde by se jednotliví učitelé od Cyrila a Metoděje přes Komenského až po Františka Novotného vyjadřovali k jednotlivým tématům jako je morálka, podstata vzdělání apod.

### 2.3.3 Protiklady v bibli

Bible je tedy sbírka zákonů, příběhů, prorocství, ponaučení apod. Nechci zde vyvracet ani obhajovat bibli jako takovou, její aspoň zdánlivá rozpornost je však výhodou, pokud se budeme snažit o její parafrázi. Proto zde uvedu několik příkladů, které v sobě zdánlivou rozpornost ukazují:

Boží milosrdenství:

*Hospodin je ke všem dobrotivý, nade vším, co učinil, se slitovává.*

- žalm 145:9

*a roztříštím je jednoho o druhého, otce i syny, je výrok Hospodinův; budu bez soucitu, bez lítosti a bez slitování, až je budu ničít.*

- Jeremiáš 13:14

Vztah k hlupákovi

*Neodpovídej hlupákovi podle jeho pošetilosti, abys nebyl jako on.*

- Přísloví 26:4

*Odpověz hlupákovi podle jeho pošetilosti, aby se sám sobě nezdál moudrý.*

- Přísloví 26:5

Zde je ukázka dokonce po sobě jdoucích protikladných veršů, které autor knihy vložil přesně za sebe. Ať už mají jednotlivé protiklady teologická zdůvodnění, nebo ne, parafrázování textu to v mnohém usnadňuje.

#### **2.4 Obtíže s parafrázováním bez kontextu**

Před samotným parafrázováním je třeba se nejdříve podívat na možná úskalí. Mějme první dva verše z 1. kapitoly Janova evangelia:

*Na počátku bylo Slovo, to Slovo bylo u Boha, to Slovo byl Bůh. To bylo na počátku u Boha.*

Máme zde dvě jména – Slovo a Bůh. Velká písmena zde slouží hlavně jako ukazatel důležitosti či výlučnosti těch slov. Že Bůh není jen tak ledajaký bůh cizích národů, není jeden z bohů Egypta ani Babylonu, ale je to Bůh. A je jen jeden. Stejně tak to Slovo. Nebylo to jen tak ledajaké slovo, ale to Slovo, skrze které se tvořilo, které bylo od začátku. Toto Slovo se ztotožňuje s Kristem.

Je třeba si uvědomit, že je to pouze náš výklad, neboť biblická hebrejšтина ani řečtina nepoužívala velká a malá písmena. Co se týče víry v jednoho Boha, po většinu doby trvání starého zákona lidé věřili v existenci několika bohů, takže už jen jméno pro Boha jako Bůh je zavádějící.

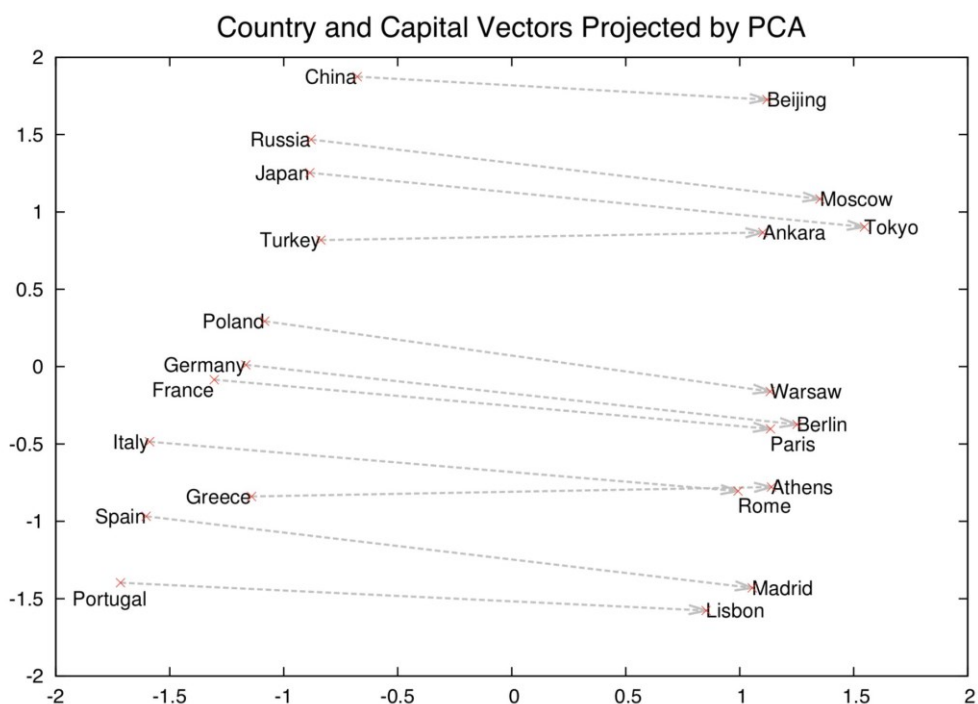
V hebrejštině navíc pro boží jméno slouží jednak tzv. tetragramaton JHWH, neboť tehdy nepoužívali v textu psané samohlásky, jednak označení Elohim či Adonai (pán). Díky tomu, že židé nezobrazují ani nevyslovují toto jméno boží (JHWH – Jahve), tak když židé na toto slovo narazí, místo JHWH přečtou Adonai Elohim – něco jako pán Bůh. V češtině se někteří překladatelé snažili přeložit Elohim jako Hospodin a JHWH jako Bůh, ale pro náš jazyk to nemá takový význam.



### 3 Použité nástroje pro parafrázování bible

#### 3.1 Word2vec

Word2vec je nástroj vyvinutý týmem v čele s Tomášem Mikolovem[4], který provádí word embeddings (vnoření slov), tzn. převádí slova na vektory. Soubor dat (textová data) se převede na několikadimenzionální prostor vektorů jednotlivých slov. Díky tomu lze se slovy provádět matematické operace. Velkou výhodou Word2vecu je, že není třeba slova nijak předem anotovat. Vychází totiž z hypotézy, že slova, která se vyskytují na podobných místech, mají podobný význam. Podobnost slov se měří pomocí tzv. kosinové vzdálenosti.



Obrázek 2 – projekce vektorů zemí a měst

Z obrázku 2 vyplývá, že Francie a Německo mají k sobě blízko, stejně tak Paříž a Berlín. Jednotlivá slova se shlukují dle jisté pravidelnosti – státy a jejich vzájemná poloha vs hlavní města a jejich vzájemná poloha.

Word2vec s vektory umí provádět následující úlohy. První z nich je výběr nejbližšího slova. Například pro slovo „kapr“ může výstup nejbližších slov vypadat jako v tabulce 1.

Slova	Kosinová podobnost vektorů
štika	0.7453
ryba	0.7294
Vánoce	0.6649

**Tabulka 1 – kosinová podobnost nejbližších slov**

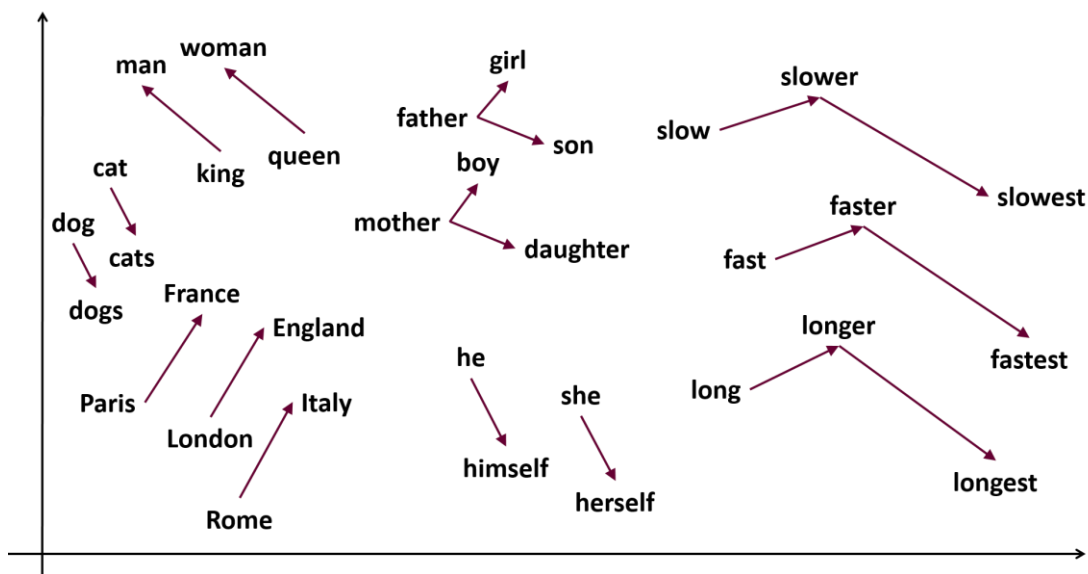
Blízkost slov ale neznamená synonymum. Podobnost mezi kaprem a Vánoce je velmi vzdálená.

Dále je možné se ptát na vztah mezi slovy a zodpovědět kupříkladu tyto otázky:

Muž se má k ženě jako král k \_\_\_\_?

$\text{vektor}(\text{král}) - \text{vektor}(\text{muž}) + \text{vektor}(\text{žena}) \approx \text{vektor}(\text{královna})$

Máme tedy nejenom možnost posoudit, které ze dvou slov je třetímu slovu bližší, můžeme též odhadnout, jak vypadá čtvrté slovo, které se má k třetímu jako druhé k prvnímu. Neplatí zde rovnost, neboť  $\text{vektor}(\text{královna})$  nemusí být přesným výsledkem aritmetické operace a nejspíš se bude jeho vektorová reprezentace trochu lišit od vektoru, který bude výsledkem operací. Word2vec navíc nemusí vždy vrátit „královnu“, ale například slovo „cestou“, tedy slovo nijak nesouvisející. Záleží proto na natrénovaném modelu. Velmi zjednodušená 2D projekce vektorového prostoru a vztahů mezi vektory slov je na obrázku 3 [21].



Obrázek 3 – projekce prostoru vektorů slov

Dále je možné vektory sčítat, což by v běžné řeči odpovídalo otázkám typu: jak spojit významově slovo  $x$  a slovo  $y$ ?

Například:

$\text{vektor}(\text{Rusko}) + \text{vektor}(\text{řeka}) \approx \text{vektor}(\text{Volha})$

$\text{vektor}(\text{Německo}) + \text{vektor}(\text{hlavní město}) \approx \text{vektor}(\text{Berlín}).^2$

Co se týče Německa a hlavního města, je  $\text{vektor}(\text{Berlín})$  jediným správným výsledkem. Co se týče Ruska, tak to má kolem 100 000 řek.[8] Pomineme-li fakt, že nejspíš bude spousta názvů duplicitních, pořad není jasné, co přesně by měla „ruská řeka“ být. Opět výsledek závisí na natrénovaných datech.

Práce s vektory někdy ukáže pozoruhodnou schopnost objevit hlubší podobnosti slov či slovních spojení, než se dá zjistit pouze morfologickou a sémantickou analýzou.

$v(\text{Montreal Canadiens}) - v(\text{Montreal}) + v(\text{Toronto}) \approx v(\text{Toronto Maple Leafs})$

V tomto kontextu je „ $v$ “ zkratka za vektor. Montreal Canadiens a Toronto Maple Leafs jsou kanadské hokejové týmy. Na nástroji Word2vec je patrné, že existuje více druhů podobnosti slov. Lexikálně by Toronto Maple Leafes mohlo být podobné

<sup>2</sup> Toto funguje pro angličtinu, kde se hlavní město řekne jedním slovem – Capital

s Montreal Maple Leafs prohozením Toronto za Montreal, ale v kontextu sportovního slovníku zde podobnost není (předpokládejme, že Montreal Maple Leafs doposud jako sportovní team neexistuje).

Nástroj používá dva modely pro převod slova na vektor – CBOW a Skip-gram. [4]

### 3.1.1 CBOW

CBOW neboli continuous bag-of-words model predikuje slovo z nějakého kontextu. Má-li věta slova  $w_1, w_2 \dots w_n$  a velikost kontextu bude deset, pak pro slovo  $w_i$  bude jeho kontext  $\{w_{i-5}, w_{i-4}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+4}, w_{i+5}\}$ . Vstupem pro model bude kontext a výstupem bude slovo “uprostřed”.

Mějme vstup “Hobiti jsou opravdu úžasná stvoření, jak jsem už říkal.” a kontext = 6.

Vstup	Výstup
(Hobiti, jsou, opravdu, stvoření, jak, jsem)	úžasná
(jsou opravdu úžasná, jak, jsem, už)	stvoření
(opravdu, úžasná, stvoření, jsem, už, říkal)	jak

**Tabulka 2 – vstup a výstup pro CBOW**

### 3.1.2 Skip-gram

Tento model je opakem CBOW, z jednoho slova se snaží až do nějaké rozumné velikosti kontextu predikovat ostatní slova pomocí tohoto pravděpodobnostního vzorce

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t)$$

kde T je velikost korpusu, c je polovina velikosti kontextu. [5]

vstup = „Hobiti jsou opravdu úžasná stvoření, jak jsem už říkal.“ a kontext = 6.

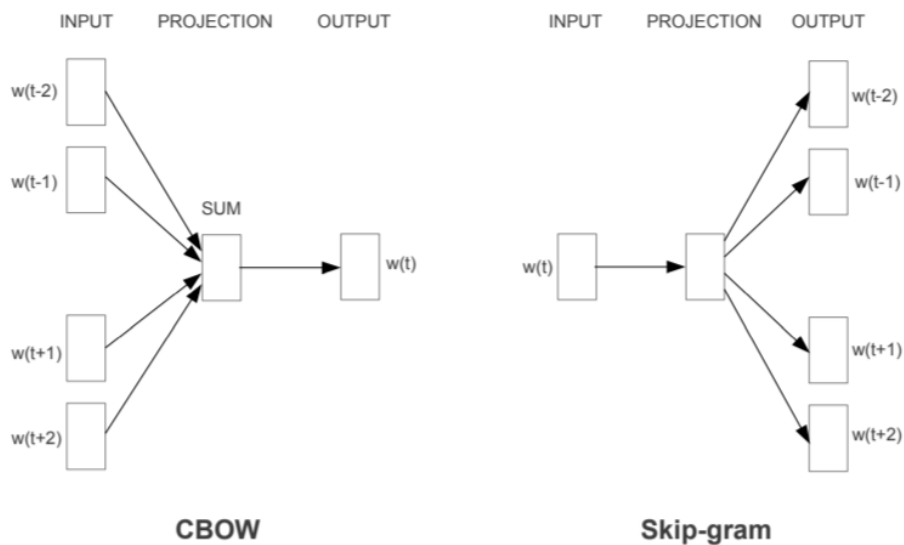
Skip-gram model se pro zadaný vstup zachová takto:

Vstup	Výstup
Hobiti	(Hobiti, jsou), (Hobiti, opravdu), (Hobiti, úžasná)
jsou	(jsou, Hobiti), (jsou, opravdu), (jsou,úžasná), (jsou, stvoření)
opravdu,	(opravdu, Hobiti), (opravdu, jsou), (opravdu,úžasná), (opravdu, stvoření), (opravdu, jak)
úžasná	(úžasná, Hobiti), (úžasná, jsou), (úžasná,opravdu), (úžasná, stvoření), (úžasná, jak), (úžasná, jsem)
stvoření	(stvoření, jsou), (stvoření,opravdu), (stvoření, úžasná), (stvoření, jak), (stvoření, jsem), (stvoření, už)
jak	(jak,opravdu), (jak, úžasná), (jak, stvoření), (jak, jsem), (jak, už), (jak, říkal)
jsem	(jsem, úžasná), (jsem, stvoření), (jsem, jak), (jsem, už), (jsem, říkal)
už	(už, stvoření), (už, jak), (už, jsem), (už, říkal)
říkal	(říkal, jak), (říkal, jsem), (říkal, už)

**Tabulka 3 – vstup a výstup pro Skip-gram**

### 3.1.3 Porovnání Skip-gramu a CBOW

Tyto dva modely mají opačný přístup ke kontextu (obr. 3) [21]. Skip-gram se používá hlavně při menších korpusech, neboť je časově náročnější pro trénování a lépe funguje pro méně frekventovaná slova než CBOW. Ten je naopak lepší pro slova frekventovaná a je také několikanásobně rychlejší pro trénování. [15] Proto jsem se rozhodl používat model CBOW.



**Obrázek 4 – porovnání CBOW a Skip-gramu**

Na obrázku 4 je patrný opačný přístup ke kontextu. CBOW z kontextu okolních slov předvídá slovo  $w(t)$ , Skip-gram místo toho předvídá okolní slova pomocí slova  $w(t)$ .

#### 3.1.4 Kosinová vzdálenost

Kosinová vzdálenost je v kontextu Word2vecu míra podobnosti vektorů. Samotná vzdálenost ale nemusí být vždy vypovídající o blízkosti významu. Pokud mají slova mezi sebou velkou vzdálenost, dá se předpokládat, že si budou významově vzdálená. Bylo by možné tedy oddělit slova vzdálená od těch blízkých tím, že se stanoví hranice blízkosti slov a pod tou hranicí by byla všechna slova považována za nevhodná. Dle Leeuwenberga [19] však zvýšení této hranice nezlepšuje přesnost (precision) parafrází. Sám navrhuje místo použití prosté kosinové vzdálenosti použít relativní kosinovou podobnost. Kosinová podobnost se počítá jako  $1 - \text{kosinová vzdálenost}$ . [20]

Relativní kosinová podobnost vychází z předpokladu, že bude-li nějaké slovo mnohem podobnější než ostatní, bude to nejspíše synonymum. Vzorec pro relativní kosinovou podobnost jednotlivých slov je tento:

$$r_{CS_n}(w_i, w_j) = \frac{\text{cosine\_similarity}(w_i, w_j)}{\sum_{w_c \in \text{TOP}_n} \text{cosine\_similarity}(w_i, w_c)}$$

Pro  $n = 10$  změřím pro každé slovo relativní kosinovou vzdálenost. Pro relativní kosinovou podobnost  $> 0.1$  je podobnost nadprůměrná, pro relativní podobnost  $< 0.1$  je podprůměrná. Zde by už dávala jistá dolní hranice podobnosti smysl, neboť Leeuwenberg tvrdí, že tato hranice precision zvyšuje.

## 3.2 Morphodita

MorphoDiTa [14] je nástroj, který umožňuje analyzování, tokenizování, generování a tagování textu. V této práci jsem se rozhodl pro použití REST API tohoto nástroje.

### 3.2.1 Vstup a výstup

MorphoDiTě posílám url dotazy a ona mi vrací textový výstup, implicitně v JSON formátu. Zde je ukázka pro dotaz na analyzování věty: Děti pojedou k babičce.

<http://lindat.mff.cuni.cz/services/morphodita/api/analyze?data=Děti%20pojedou%20k%20babičce.&output=vertical>

```
{
  "model": "czech-morfflex-pdt-161115",
  "acknowledgements": [
    "http://ufal.mff.cuni.cz/morphodita#morphodita_acknowledgements",
    "http://ufal.mff.cuni.cz/morphodita/users-manual#czech-morfflex-pdt_acknowledgements"
  ],
  "result":      "Děti\tdítě\tNNFP1-----A----\tdítě\tNNFP4-----A-----\tdítě\tNNFP5-----A----\npojedou\ttjet-1_^(pohybovat_se,_ne_však_chůzí)\tVB-P---3F-AA---\nk\tk-1\tRR--3-----\tk-3_^(označení_pomocí_písmene)\tNNNX-----A----\tk-4`kůň:_B_^(jednotka_výkonu)\tNNMXX-----A---8\tk-8:_B_^(ost._zkratka)\tXX-----8\tkomanditní:_B_^(jen_komanditní_společnost)\tAAXX-----1A---
```

```

8\tkoncernový_:B\tAAXXX-----1A---8\tkuo-
1_:B_,t_^(stará_jednotka_výkonu)\tNNNXX-----A---
8\nbabičce\tbabička\tNNFS3-----A-----\tbabička\tNNFS6-----A-----
\n.\t.\tZ:-----\n\n"
}

```

Tato část url adresy `analyze?data=` mě odkazuje na metodu analyzování. Pokud `analyze` vyměním za `generate`, `tag` či `tokenize`, zavolám v REST API metody pro generování, tagování či tokenizaci. Dále následuje uspořádaná množina slov, mezi kterými vkládám znak `%20` místo mezery. Na závěr dotazu mohu připojit různé specifikace pomocí znaku `&`. Já využívám tuto část pro změnu výstupu. Místo JSONu pro `result` dostávám jen prostý text.

### 3.2.2 Metoda tokenize

`Tokenize` je metoda, která mi vrátí větu, ve které jsou označena jednotlivá slova pomocí tagů `<token>`. Pro účely této práce jsem ji nevyužil. Tokenizuji sám pomocí metody `TransormString` ve třídě `TransformTXT.cs`, kde každé slovo oddělím od interpunkce a token je ohraničen mezerami.

### 3.2.3 Metoda analyze

Výstup metody `analyze` je zobrazen výše. Pro všechna slova ve větě určí všechny možné tvary. Nevýhoda této metody pro moji práci je ta, že pro jedno slovo vrátí více výsledků. Nepraktické je to proto, že pokud nebudu brát první vrácený výsledek, potřeboval bych hlubší znalost věty, abych se rozhodl, který z výsledků je správný. Místo `analyze` tedy používám pro zanalyzování věty metodu `tag`.

### 3.2.4 Metoda tag

Tato metoda je podobná `analyze`. Místo nalezení všech možných tvarů se metoda snaží „uhádnout“ tu morfologickou značku, která je správná v tom daném kontextu. Tento postup má svá úskalí. Pro větu: „Ženu holí stroj“ se metoda sama musí rozhodnout pro kontext. Výstupem, který já používám, je:

```

původní slovo, slovo v základním tvaru, morfologická značka
původního slova

```

Např:

Trolové, trol, NNMP1-----A-----

### 3.2.5 Metoda generate

Metoda `generate` pro každé slovo, pro které známe základní tvar, vygeneruje pomocí tohoto tvaru všechny jeho možné další tvary. Kvůli potřebě základního tvaru slova používám vždy tuto metodu v kombinaci s metodou `tag`, která mi vrátí základní tvar slova. Výhoda metody `generate` je, že pokud mám seznam slov, která mi má vygenerovat (v základním tvaru), pak mi na jedno zavolání metoda vrátí vygenerované všechny možné tvary pro mnou zadaná slova.

### 3.2.6 Morfologická značka

Morfologická značka (`tag`) slouží k přesnému určení tvaru slova. MorphoDiTa využívá morfologické značky o délce patnácti znaků. Každá pozice odpovídá jedné morfologické kategorii.

Jedna z možných morfologických značek pro slovo “děti” může být:

NNFP1-----A-----.

V tabulce 4 jsou vysvětleny kategorie jednotlivých pozic. [9]

No.	Jméno	Popis
1	POS	Slovní druh
2	SUBPOS	Detailní určení slovního druhu
3	GENDER	Jmenný rod
4	NUMBER	Číslo
5	CASE	Pád
6	POSSGENDER	Přivlastňovací rod
7	POSSNUMBER	Přivlastňovací číslo
8	PERSON	Osoba
9	TENSE	Čas
10	GRADE	Stupeň
11	NEGATION	Negace

12	VOICE	Aktivum/pasivum
13	RESERVE1	Nevyužito
14	RESERVE2	Nevyužito
15	VAR	Varianta (stylový příznak)

**Tabulka 4 – morfologická značka**

Rozbor značky pro slovo „děti“: 1–N značí podstatné jméno, 2–N značí všeobecné podstatné jméno, 3–F značí rod ženský, 4–P značí množné číslo, 5–1 značí první pád, další pomlčky značí, že toto slovo neobsahuje příslušné kategorie, což kupříkladu znamená, že podstatné jméno nemá žádnou osobu, tato kategorie je vyhrazena pro slovesa. 11–A značí, že slovo není negativní.

Samozřejmě děti mohou být kromě nominativu i v akuzativu. Pro jedno slovo může být několik morfologických značek, které se liší v některých kategoriích.

## 4 Metody řešení problému

### 4.1 Nástin řešení

Problém, který tato práce řeší, je parafrázování a modernizaci textu v rámci nějakého tematického okruhu, konkrétně biblických textů, na úrovni slov. Na to si vytvořím vlastní slovník. Z tohoto slovníku pak budu hledat vhodného kandidáta k parafrázi. Může se stát, že kandidát nebude vhodný, nebo že bude ve špatném tvaru. Proto sekundárně zanalyzuji jednotlivá slova v původní větě a zabráním nevhodným parafrázím. Jelikož se tato práce snaží o modernizaci biblických textů, slovníky chci mít naplněné slovy z „moderního jazyka“. Mým předpokladem je, že se mi podaří nahradit archaická slova modernějšími výrazy pomocí pouhého zaměnění slov a využiji vlastnosti Word2vecu, že slova blízká svým významem mají blízké vektorové reprezentace.

Pro získání samotného slovníku využiji nástroj Word2vec, který vytvoří slovník  $\langle \text{slovo}, \text{vektor} \rangle$ , kde slovo je klíčem a kde vektor dimenze  $n$  je uspořádaná  $n$ -tice desetinných čísel. Pomocí tohoto slovníku budu vyhledávat nejbližší slovo. V ideálním případě nahradím slovo za jeho synonymum. Nyní buď s parafrází skončím, nebo výstup ještě upravím pomocí dalšího nástroje, konkrétně MorphoDiTy, která mi „oseká“ špatné výstupy z Word2vecu v rámci několika pravidel. Pro lepší výsledky provedu nejdřív analýzu věty pomocí MorphoDiTy, poté využiji Word2vec, a nakonec ještě dopravím text MorphoDiTou.

### 4.2 Trénování

Při trénování vyberu vhodný text, které slouží jako podklad pro můj model a nejdříve ho ztokenizuji tak, že oddělím slova od interpunkce mezerou. Dále provádím tzv. truecasing, neboli převádím obecná jména na malá písmena. Pokud je slovo kupříkladu slovo „Dům“ na začátku věty a v textu se někde vyskytuje s malým písmenem uprostřed věty, opravím velké D na malé. Poté Word2Vec projde tento

upravený soubor a k jednotlivým slovům přiřadí vektory čísel. Výsledkem fáze trénování je soubor ve složce Models.

### 4.3 Parafrázování – Word2Vec

Samotné parafrázování vypadá tak, že pro každé slovo, které jsem mu označil jako vhodné pro parafrázi, vrátí Word2vec seznam nejbližších slov. Pokud používám pouze Word2Vec, nahradím prvním kandidátem, který není vlastním jménem, abych nezaměňoval jména za jména. Může se však stát, že se vlastní jméno nahradí za jméno obecné. Kupříkladu ve větě „David válčil s Goliášem“ se „Goliáš“ může nahradit za „zápasníka“. Zvláště pro starozákonní postavy to může přiblížit smysl textu, neboť symbolika hebrejských či fénických jmen je mnohem důležitější než v češtině.

Používám pro parafrázování dvě varianty výběru vhodných slov. Pokud parafrázuji pouze pomocí Word2vecu, vyberu z věty slova, která jsou delší než tři znaky. Je to z toho důvodu, že Word2vecu nepředchází žádná morfologická analýza. Tímto omezením odstráním z parafrázování většinu předložek, spojek a částic.

Druhá varianta je pomocí MorphoDiTy, kde žádné omezení na délku slov není a vybírám slova dle slovních druhů.

### 4.4 Parafrázování – MorphoDiTa

Samotné nejbližší slovo nemusí vždy být synonymum. Word2vec může vrátit kupříkladu antonymum. Aby se nestávalo, že ke slovu „jsem“ dostanu parafrázi „nejsem“, použiji další nástroj – MorphoDiTu.

Tento nástroj používám kvůli porovnání parafrází ve dvou verzích. Ta první využívá stejný výběr slov pro parafrázování jako Word2vec. Přijme pro každé slovo jeho nejbližšího souseda dle kosinové vzdálenosti. Použiji metodu tag na původní větu a pro každé slovo navržené Word2vecem jako vhodnou parafrázi získám základní tvar a vygeneruji všechny jeho tvary. Pokud se originální a navržené slovo liší ve slovním

druhu, k náhradě nedojde. Jinak ze všech tvarů vyberu k parafrázi ten, který se nejvíce podobá originálnímu tvaru, tj. má shodu na největším počtu míst v morfologické značce.

Druhá verze parafráze nejdříve otaguje původní větu, čímž získám slovní druhy jednotlivých slov. Z věty vyberu všechna podstatná jména, přídavná jména a slovesa. Nedochozí tedy k náhradě slov jako „okolo“ či „sama“.

Oproti Word2vecu, který používám skrze připojené .dll soubory, využívám webové REST API MorphoDiTy, kterému pro každou parafrázi používám tři dotazy: `tag` pro původní větu, `tag` pro každé slovo nahrazené pomocí Word2vecu a `generate` pro základní tvary nahrazených slov.

## 4.5 Parametry Word2vecu

### 4.5.1 Nastavitelné parametry

Word2vec je nástroj, který dostává velké množství parametrů. Kvůli uživatelské jednoduchosti jsem umožnil v programu měnit pouze tři parametry: dimenze vektoru, počet iterací a minimální výskyt slov.

#### Dimenze vektoru

Dimenze vektoru netriviálním způsobem ovlivňuje velikost modelu. Implicitně je nastavena dimenze 100. Dle výzkumu, který provedl Kevin Patel a Pushpak Bhattacharyya [16], se běžně dimenze vektoru pohybuje od 50 do 300 a výsledná dimenze vektoru je spoustou uživatelů stanovena „od oka“. Z jejich výzkumu vyplývá, že větší dimenze vektoru zlepšuje vlastnosti Word2vecu jen po nějakou dolní mez, poté už výsledky zůstávají téměř neovlivněny. Navíc přidávají algoritmus, který nalezne dolní mez pro vhodnou dimenzi vektoru. Potíž algoritmu je, že pro  $n$  slov vytváří tabulku o velikosti  $n*n$  a hledá minimální počet vektorů, pro které je ještě každé slovo reprezentované vektorem unikátní. To je převoditelné na hledání největší nezávislé množiny v grafu, což je NP-úplný problém. Porovnal jsem tedy velikosti svých modelů s velikostmi, které se normálně používají pro Word2vec a stanovil jsem svoji hranici na 100, díky menší velikosti mých modelů.

### Minimální počet výskytů

Minimální počet výskytů jednotlivých slov je parametr, který zajišťuje, že ojedinělá slova se do modelu nepřidají. Může se totiž stát, že náš specifický jazyk bude osekán o právě ta specifická slova. Kupříkladu budu-li mít jedno slovo v celém textu, které je pro text ale důležité, ztrácím informaci. Tento parametr slouží hlavně k snížení velikosti modelu a je tedy vhodné s ním pracovat pouze pokud máme nedostatek paměti či příliš velký model. Implicitně jsem nastavil tento parametr na 5.

### Počet iterací

Počet iterací průchodu textem je naopak parametr, který je vhodné optimalizovat. Má vliv pouze na trénování modelu a je tedy přímo úměrný době čekání. Na druhou stranu od nějakého počtu opakování se vektory téměř nemění a je tedy zbytečné čekat na další iterace. Implicitně jsem parametr nastavil na 5.

## 4.5.2 Napevno zvolené parametry

### Threads – počet vláken

V tomto parametru je nastaveno, kolika vláken dostane Word2vec k využití. Na radu autora implementace Word2vecu v C#, trénovací části, jsem nechal 12 vláken.

### Použití Skip-gramu vs CBOW

Implicitně je nastaveno použití CBOW, důvody uvádím v sekci 3.1.3.

### Negative sampling

Každý vzorek při trénování totiž trochu upraví váhy všech vektorů v modelu. Negative sampling upravuje váhy pouze na malém kontextu, čímž dramaticky zvýší rychlost trénování. Já nastavuji velikost 5.

### Hierarchical Softmax

Dle článku Tomáše Mikolova [20] negative sampling obecně překonává hierarchical softmax, zvláště pro méně častá slova, proto jsem hierarchical softmax pro trénování nevyužil a implicitně nastavuji 0 - nepoužito.

## 4.6 Rešerše

### 4.6.1 Parafrázování bible

Parafrázování bible není rozhodně nová úloha. Už ve středověku vznikl ve studentských kruzích pokus o latinské parafráze alespoň část bible. Erasmus Rotterdamský napsal dílo *Parafráze*, které parafrázuje jen určité části Nového zákona: všechna evangelia a část epištol. [17] V anglicky mluvících zemích je velmi oblíbená *The Living Bible*, parafráze Kennetha N. Taylora, vydaná v roce 1971. [18] V českém prostředí se nejvíce uchytilo *Slovo na cestu* a *Parabible*. *Parabible* však parafrázuje jen malou část Nového zákona a neobsahuje vůbec Starý zákon.

### 4.6.2 Modernizace bible

Moderní biblická parafráze by mohla být „Parabible“, avšak zde byla „modernost“ prvoplánově použita už při vytváření příběhů: Ježíš se narodil v Ostravě apod. Moderní biblický jazyk, byť je to přímo překlad, byl cílem Bible21, která se snaží o co nejmodernější, a přitom přesný překlad.

### 4.6.3 Modernizace pomocí strojového překladu

V článku *Historical Documents Modernization* [21] se autoři snaží o modernizaci holandských textů z 17. do 21. století. Využívají SMT metody (Static Machine Translation), kdy se snaží najít nejlepší překlad  $\hat{y}$  z věty  $x$ :

$$\hat{y} = \arg \max_y \Pr(y | x)$$

Jako zdrojové texty využili jednak některé knihy bible v holandštině z let 1637, 1657, 1888 a 2010, jednak holandskou literaturu ze 17. století. Slova, která neznají, zkouší rozsekat na podslova a ta „zmodernit“.

## 4.7 Data

### 4.7.1 Bible

Tato práce pro trénování modelů využívá pěti biblických překladů: Bible kralická (znění z roku 1613), Český ekumenický překlad, Bible21, Český studijní překlad a Nová bible kralická (pouze Nový zákon).

### 4.7.2 Data Svědků Jehovových

Tato data byla získána z korpusu JW300, ze stránky <http://opus.nlpl.eu/JW300.php>, kde jsem z bilingválních cs-en dat odstranil anglickou část a získal tím pouze českou část.

### 4.7.3 Fantastika a filosofie

Tato data jsou v některých ohledech tematicky podobná biblickým textům. Užívají podobné názvy nástrojů, válečné události apod. Obsahují plné znění těchto knih:

Abercombie, J - První zákon, 1-3. díl, Adams, D - Stopařův průvodce po galaxii, 1-5. díl, Asprin, R - MYTH 1-16. díl, Červeňák, J - Bohatýr, 1-2. díl; Černokněžník, 1-4. díl, Descartes, R - Rozprava o metodě, Eddings, D - Belgariad 1-5. díl; Malloreon 1-5. díl, Erikson, S - Malaszká kniha padlých 1-6. díl, Feist, R E - Hadí sága 1-8. díl, Heitz, M - Trpaslíci 1-4. díl, Kulháněk, J - Divocí a zlí 1-4. díl, Martin, G R R - Píseň ledu a ohně 1-4. díl, Pratchett, T - Úžasná zeměplocha 1-13. díl, Rowlingová, J K - Harry Potter 1-7. díl, Tolkien, J R R - Hobit; Pán prstenů 1-3. díl, Žamboch, M - JFK 1-20.

### 4.7.4 Titulky

Dalším zdrojem byly titulky k filmům. Tato data mi slouží jako zdroj moderní češtiny a získal jsem je z portálu <http://opus.nlpl.eu/OpenSubtitles-v1.php>, kde jsem

využil bilingvální data en-cs. Z nich jsem vyextrahoval pouze české věty. Tyto titulky z tohoto portálu jsou získané z webu [www.opensubtitles.org](http://www.opensubtitles.org).

#### 4.7.5 Odborný text

Kvůli obohacením dat o nějaký odborný text, který by mohl zvýšit kvalitu parafráze, jsem zvolil texty studentů medicíny, konkrétně články z hlavního portálu pro studium klinické medicíny: [http://wiki.medik.cz/wiki/Hlavní\\_strana](http://wiki.medik.cz/wiki/Hlavní_strana).

#### 4.7.6 Tabulka použitých dat

Pro shrnutí přikládám tabulku s velikostmi jednotlivých dat.

Data	Velikost
české biblické překlady	17 MB
data jehovistů	131 MB
fantasy-filosofie	104 MB
texty studentů medicíny	31 MB
titulky k filmům	5.4 GB

**Tabulka 5 – použitá data**

#### 4.7.7 Použité modely

Jména pro své modely jsem vybíral tak, aby z nich bylo patrné, jakou mají kombinaci volitelných parametrů. První číslo je dimenze vektoru, druhé je minimální počet výskytu slova a třetí je minimální počet opakování. Pokud je model bez čísla, implicitně pro něj platí parametry: 100, 5, 5. Modely jsem se snažil sestavit tak, aby byly biblické texty, pokud jsou v modelu zahrnuty, významněji zastoupeny. Díky malé velikosti dat biblických překladu (17 MB) chci zajistit, aby případný malý výskyt určitých biblických slov nebyl Word2vecem vyřazen, proto v některých modelech biblický text několikrát zkopíruji.

**bibleDohromady\_100\_1\_5**

Tento model obsahuje pouze biblické texty. Nevýhodou je malá velikost modelu. Jeho parametry jsou dimenze 100, minimální výskyt slova 1, počet iterací 5.

### **Fantasy\_medic\_filosof**

Tento model je vytvořen z textů z wikipedie studentů medicíny [10] a z knih žánru filosofie a fantastiky.

### **Jehovisti1\_Bible12\_100\_5\_5**

Tento model obsahuje 12x všech 5 biblických překladů a 1x data Svědků Jehovových.

### **TitulkyABilbe\_100\_5\_5**

Tento model obsahuje 50x zkopírované biblické překlady a 1x titulky.

### **medickeTexty**

Tento model je obsahuje pouze texty z medik.wiki [10]

## 5 Implementace

### 5.1 Využití zdroje

V této bakalářské práci jsem využil implementaci Word2vecu v jazyce C# dvou vývojářů na GitHubu a REST API MorphoDiTy. Word2vec pro trénování modelu jsem převzal od uživatele eabdullin [11], použití natrénovaného modelu pro hledání nejbližších slov jsem převzal od uživatele GuntaButya [12]. Z jejich zdrojových kódů jsou vytvořeny .dll knihovny, které musí být spolu s .exe souborem přítomny v jedné složce při spuštění programu.

MorphoDiTa [13] je nástroj vyvinut doktorem Milanem Strakou a doktorkou Janou Strakovou z Ústavu formální a aplikované lingvistiky na Matematicko-fyzikální fakultě Univerzity Karlovy.

### 5.2 Inicializace programu – DirectoryManager.cs

Při startu aplikace je nutné, aby se v adresáři, kde se nachází .exe soubor, nacházely další složky, konkrétně `Models`, `PlainTexts`, `Results`, `SourceTXTFiles`, `Temp`. Pokud tam nejsou, například po prvním startu aplikace, program je sám vytvoří. Je pro přehlednost vhodné si tedy vytvořit nový adresář a v něm teprve spustit .exe soubor. Program se dokáže adaptovat na operační systém Linux a Windows, takže na obou OS se složky vytvoří bez problému. Neměl by být problém program spustit i na macOS, ale odladěn na něm nebyl. Pro spuštění na Linuxu je třeba mít nainstalované Mono pro C#.

Obstarávání vhodné cesty a tvorbu adresářů má na starost třída `DirectoryManager`. Kromě tvorby adresářů vloží do složky `PlainTexts` malý soubor pro testové účely.

### 5.3 Módy programu – Menu.cs

Program pracuje ve dvou módech. Trénování a následné parafráze. Proto je třída `Menu.cs` rozcestníkem v programu. Pokud by uživatel zamířil rovnou do sekce parafrázování, program ho sám varuje, že je třeba nejdříve natrénovat nějaký model. Třída pro trénování, `TrainNewModel.cs`, poskytuje rozhraní pro upravení zdrojového .txt souboru, aby vyhovoval požadavkům Word2vecu. Dále poskytuje možnost natrénovat si z několika již upravených .txt souborů vlastní model. Druhá

cesta z menu vede skrze třídu `InlineParaphrasing.cs`, která je zodpovědná za finální parafrázování textu a případné spolupráce s `MorphoDiTou`.

#### 5.4 Úprava zdrojových textů - `TransformTXTfiles.cs`

O úpravu zdrojových textů se stará třída `TransformTXTfiles.cs`. Tato třída je zvlášť zaměřena na český jazyk. Protože v UTF-8 nejsou písmena české abecedy poskládána za sebou od a-ž, mám vytvořená statická pole s českými znaky. Dále, kvůli možnému budoucímu rozšíření, mám definované pole znaků „DangerousSymbols“. To jsou symboly, které chci obalit mezerami, velmi často stojí v těsném sousedství slov, aniž by měla přímo význam. Jedná se především o různé druhy interpunkce, závorky apod.

Zde narážíme samozřejmě na problém u jmen, kdy zvlášť irská jména jako O'Brien či elfská fantasy jména typu Drizzt Do'Urden budou uměle přerušena. To se týká i některých jmen místních. Pokud budou v textu názvy jako Praha-Východ či Praha-Vysočany, tato oslova rozdělím.

Při rozdělování textu na jednotlivá slova čtu text znak po znaku a pokud narazím na interpunkci, nebo na bílý znak, uložím si slovo do slovníku a hledám vlastní podstatná jména, která začínají velkými písmeny. Všechna obecná podstatná jména potom převedu na malá písmena, takže se mi nestane, že by „Ahoj“ a „ahoj“ byla dvě slova.

S tímto řešením mohou nastat komplikace. Pokud budu kupříkladu parafrázovat texty s westernovou tematikou, kde spousta indiánů bude používat obecná jména jako vlastní. „Indián Silný Býk byl silný jako býk“ převedu na Indián silný býk byl silný jako býk (předpokládám, že celý text je pouze tato věta, jinak bych pravděpodobně převedl i indiána na malá písmena) V českém prostoru by mohl být podobný problém. U obce Měretice v okrese Chrudim jsou vesnice se jmény Studená Voda, Smrček, Louka, Hlína, Silnice a Skála.

Představme si třeba tuto větu:

Probudil jsem se ve Studené Vodě, v Smrčku jsem si chvíli poseděl, pak jsem to vzal autem přes Louku, Hlínu a Skálu, až jsem se dostal k Silnici.

Po mém parafrázování se z této věty stane:

Probudil jsem se ve studené vodě, v smrčku jsem si chvilku poseděl, pak jsem to vzal autem přes louku, hlínu a skálu, až jsem se dostal k silnici.

Dalším problémem jsou málo častá slova. Budu-li mít nějaké takové slovo na začátku věty, navíc to bude jeho jediný výskyt, opět není žádná možnost, jak zjistit, že to není jméno. Pro velké slovníky ale tento problém nastává méně často.

### 5.5 Vytvoření modelu – TrainNewModel.cs

Model vytvářím přímo ze zdrojových textů. Třída `TransformTXTFiles.cs` mi upravený text uloží do složky `SourceTXTFiles`, ve které jsou všechny doposud upravené soubory. Tato třída obstarává uživatelské rozhraní pro tvorbu vlastního modelu. Dává uživateli volnost, aby si sám mohl „namíchat“ vlastní model. Pro každý soubor, který chce vytvořit, zadává počet. Všechny txt soubory se nejdříve slíjí do jednoho jediného, který bude uložen ve složce `Temp`. Po vytvoření modelu je tento soubor smazán.

Třída poskytuje doprovodné informace o procentuálním zastoupení souborů i o jejich velikosti. Je možné zde ovlivnit parametry jako dimenze vektoru, počet iterací i minimální nutný výskyt slov.

Dalo by se uvažovat, že pro výzkumné účely by uživatel mohl nastavit více proměnných, na druhou stranu většina uživatelů nebude nastavovat žádné proměnné a ponechá program tak, jak je. Samozřejmě jsou všechny vstupy ošetřeny proti jakékoliv zákeřnosti či nepozornosti ze strany uživatele.

### 5.6 Parafrázování textu – Inline\_paraphrasing.cs

Tato třída je zodpovědná za samotné parafrázování. Okna jsou jednotlivé objekty, které jsou definovány v třídě `Window.cs`. Je možné parafrázovat jeden úryvek až pěti různými modely s/bez použití `MorphoDiTy`. Po zadání vstupního řetězce se zavolá metoda parafrázování textu. Pokud už bylo parafrázováno 5x, vyskočí dialogové okno, které nabádá buď k uložení výsledků, nebo k jejich resetování. Parafrázování samo o sobě je deterministické, takže pro stejný vstup a (ne)využití `MorphoDiTy`

vydá stejný výstup. Pro použití MorphoDiTy je potřeba webového připojení. Pro jeho ověření používám jednoduchý webrequest. Pro uložení výsledků slouží adresář `Results`, ve kterém vytvářím unikátní jméno souboru pomocí převedení data a času do integeru.

Pokud jsou všechny podmínky splněny, zavolá se metoda `ParaphraseText()`. V této metodě dochází k transformaci vstupního řetězce a výběru vhodných slov pro parafrázování, tedy všech slov větších než tři, případně všech vhodných slov otagovaných MorphoDiTou, a ta uložím do slovníku `<původní slovo, náhrada>`.

## 5.7 Úprava parafrázovaného textu – MorphoDiTa.cs

Pro MorphoDiTu jsou klíčová 3 volání, 2x `tag` a 1x `generate`. Z prvního volání `tag` pro celý vstupní text si do slovníku uložím každé slovo a jeho morfologickou značku.

Druhé volání metody `tag` slouží k nalezení základních tvaru těch slov, která mi vrátil `Word2vec`. Ty poté vložím do dotazu metodě `generate`, která mi vygeneruje všechny potřebné tvary.

Pro parsování výstupu metod `tag` a `generate` mám vlastní metody pro parsování, neboť každá z metod MorphoDiTy má lehce odlišný druh vstupu:

```
GetResultJsonVariableFromTagging(string)
```

```
GetResultJsonOutputInGenerate(string)
```

Obě metody vrací slovníky, jen s odlišnými klíči pole toho, zda chci vyhledávat podle slova, či podle morfologické značky.

2. verze MorphoDiTy se ukazuje jako vždy úspěšnější, než 1. verze. Proto je implicitně v programu možnost parafrázovat buď bez MorphoDiTy, nebo s 2. verzí. Pro stažení první verze je třeba využít zdrojových kódů z první verze programu:

<https://github.com/Tricerator/BibleGenerator>

## 5.8 Unit testy

Kvůli kontrole vstupů a výstupů jsem vytvořil několik Unit testů. Testoval jsem tři třídy: `MorphoDiTa.cs`, `TransformTXTFiles.cs` a `ManageDirectories.cs`.

Ve třídě `TransformTXTfiles.cs` jsem otestoval vstupy i pro hebrejštinu a ruštinu. Pro ruštinu funguje úprava textu pro `Word2vec` bez problémů, s hebrejštinou jsou drobné komplikace, takže jedno z možných rozšíření práce by byla vícejazyčnost i pro jazyky s opačným směrem textu. `ManageDirectories` jsem otestoval pouze pro hladký průběh vytváření složek.

Nejvíce testů je pro třídu `MorphoDiTa`, kde jsem testoval postupné odsekávání vstupu a provedení či neprovedení parafráze.

## 6 Uživatelská příručka

### 6.1 Spuštění aplikace

Pro spuštění aplikace je třeba vytvořit novou složku a do této složky umístit .exe soubor a 2 .dll soubory – `Word2vec.Net.dll` a `Word2vec.Tools.dll`. Po spuštění aplikace se vytvoří 5 složek – `Models`, `PlainTexts`, `Results`, `SourceTXTFiles` a `Temp`.

Ve složce `PlainTexts` je cvičný malý textový soubor, jen pro kontrolu, že vše funguje v pořádku. Je to také složka, kde můžete vkládat své textové soubory, aby byly na jednom místě. Ve složce `Models` budou postupně ukládány natrénované modely, pomocí kterých budete moci parafrázovat vlastní text.

Ve složce `SourceTXTFiles` už bude program sám ukládat soubory, které byly už natrénované. Do této složky žádné textové soubory nekládejte. Parafrázování bude stále fungovat, ale slova „ahoj“, „ahoj,“, „ahoj?“, „ahoj(“ a „ahoj!“ budou `Word2vecem` vnímána jako pětice různých slov.

Složka `Temp` slouží jen jako dočasné uložení souborů, ze kterých program vytváří model a které program později smaže.

Složka `Results` si v sobě ukládá výsledky jednotlivých parafrázování. Tyto výsledky budou pojmenovány podle času vytvoření.

### 6.2 Natrénování modelu

Aby byl model natrénován, musí mít nejdřív nějakou množinu textů. Prvně musíte kliknout na `Natrénovat nový model` a `Uprav soubor txt`. Potom je třeba vybrat .txt soubor. Doporučuji upravit více souborů, ať je možné udělat model zajímavější a větší. Až budete mít nějaké upravené .txt soubory, zmáčkněte `natrénuj model`.

V levé části aplikace jsou na výběr vaše upravené .txt soubory. Každý soubor můžete přidat pomocí levého kliknutí myši a stisknutí tlačítka `přidej`. Objeví se dialogové okno s počtem. Každý soubor takto můžete přidat jen jednou (s počtem třeba 10x). Pokud si počet rozmyslíte, či nastala chyba, můžete kliknout na název souboru v kategorii `model` a pak zmáčkněte `Zpět`. Sloupce počet, velikost a procenta ukazují dodatečné informace o velikosti či procentuálním zastoupení souboru v modelu. Až

budete mít všechny soubory pro trénování modelu vloženy, napište jméno modelu. Bylo by vhodné, aby zcela jasně říkalo, co je to za model. Jména jako „modell“ apod. nejsou příliš vhodná, při více modelech nebudete vědět, o který model se jedná. Dodatečné elementy jako délka vektoru, počet iterací a minimální výskyt slova můžete, nebo nemusíte měnit. Pokud se zdá, že po stisknutí tlačítka aplikace nefunguje, funguje. Jen je třeba být trpělivý. Doba trénování se prodlužuje s velikostí modelu a s počtem iterací.

### 6.3 Parafrázování textu

Pokud už máte natrénovaný model, můžete začít parafrázovat. Vraťte se do `Menu` a zvolte `In-line parafrázování`. Objeví se 6 oken, jeden `ListBox` obsahující natrénované modely a zaškrtačací tlačítko `Použít MorphoDiTu`. Nad okny máte označeno, co které okno značí. Do vstupního okna napište text. Pokud nepoužíváte `MorphoDiTu`, jste omezeni pouze velikostí `TextBoxu`, tzn. nějakými 32000 znaky. Pokud budete používat `MorphoDiTu`, jste omezeni nějakými 2500 až 3000 znaky na jedno parafrázování. Pro delší části textu je třeba text rozdělit na menší části.

Zvolte model a poté zmáčkněte `Parafrázovat`. Po pěti parafrázích vás aplikace upozorní, že byste měli výsledky buď uložit, nebo resetovat okna. Uložení probíhá automaticky a nemusíte zadávat jméno souboru. Uložit výsledky můžete kdykoliv. Tlačítko `Uložit` i tlačítko `Resetovat` smažou obsah všech oken.

U každého parafrázování si můžete zvolit, zda budete přidávat `MorphoDiTu`. Musíte však být připojeni k internetu. Pokud připojení selže, program vás zarazí.

### 6.4 Náповěda

V každé části programu je přítomné tlačítko s nápovědou, která vás provede onou částí.

## **6.5 Problém s internetovým připojením**

Pro použití MorphoDiTy bude třeba být připojen k internetu. Pokud připojení k internetu během parafrázování vypadne, program se sám resetuje. Dále se program resetuje při použití příliš dlouhého textu pro MorphoDiTu. Pokud vás program stále vrací do Menu, zkuste zkrátit text a zkontrolovat internetové připojení.

## **6.6 Natrénované modely**

Pokud chcete použít mé natrénované modely, jsou dostupné na této adrese:

<https://invite.bublup.com/001-si-97a28404-49ef-4c76-9506-142cabe5c2a1>

## 7 Vyhodnocení výsledků

### 7.1 Výběr modelu

#### 7.1.1 Zadání dotazníku

Natrénoval jsem 5 modelů a z bible vybral 6 typových žánrů. Cílem bylo vybrat vhodný model pro modernizaci a parafrázování.

V Příloze A je dotazník s parafrázovanými částmi, který respondenti bodově hodnotili dle přiložených kritérií. V tomto dotazníku jsem vybíral nejznámější biblické úseky. Z mystických textů je to první verš Janova evangelia – Jan 1,1. Z právních textů to bylo desatero, první znění – Exodus 20:2–17. Pro popis pracovního postupu jsem měl několik možností – stavbu chrámu, stavbu schrány úmluvy, stavbu oltáře apod. Zvolil jsem stavbu oltáře, protože má být na oltáři položeno několik liturgických předmětů, které by mohly ukazovat kvalitu parafráze. Z novozákonních podobenství jsem zvolil podobenství o ztracené minci – Lukáš 15:8-9, neboť je jedno z nejznámějších a je krátké. Dále jsem vybral modlitbu k parafrázování, a to nejznámější z nich, tzv. „Otčenáš“. Nakonec jsem vybral poezii z knihy Píseň písní 2:2-4. Jedna z nejznámějších pasáží, kdy si milenci vyznávají svůj obdiv.

Výsledky prezentuji pomocí tabulek, ke kterým vždy připojím komentář. „R“ je v tabulkách označení pro respondenta, „b“ je v tabulce náhrada za bibleDohromady\_100\_1\_5, „F“ je náhradou za: Fantasy\_medic\_filosof, „J“ za Jehovisti1\_bible12\_100\_5\_5, „m“ za medickeTexty, a T je náhradou za TitulkyABible\_100\_5\_5.

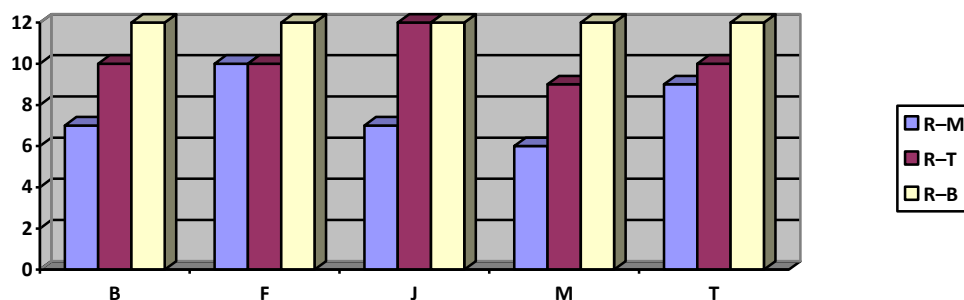
V gramatické správnosti a stejném tvrzení respondenti mohli hodnotit 0–2 body. Pro modernizaci a vhodnost modelu respondenti hodnotili 0-1 bodem. Nakonec byla možnost zanechat slovní komentář u jednotlivých parafrázi.

Parafráze probíhali pomocí Word2vecu a MorphoDiTy. Vyžadoval jsem, aby se alespoň nějaké náhrady provedli, v této části hodnotím kromě modernizace samotné schopnost modelu dobře parafrázovat.

### 7.1.2 Gramatická správnost

Respondenti	B	F	J	M	T
R-M	7	10	7	6	9
R-T	10	10	12	9	10
R-B	12	12	12	12	12

Tabulka 6 – gramatická správnost pro jednotlivé modely



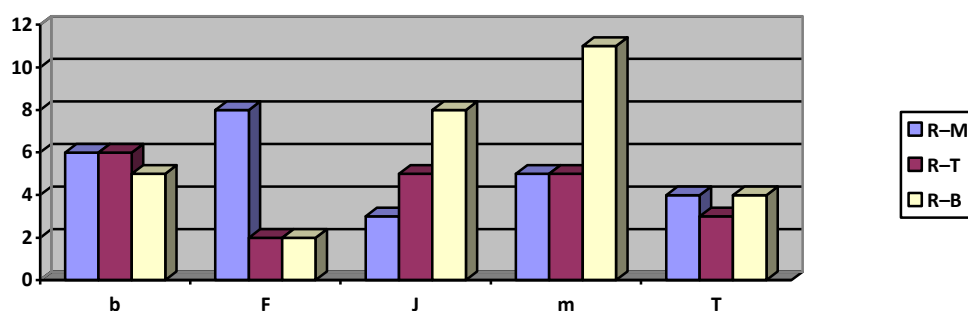
Obrázek 5 - graf tabulky 6

Maximum bodů je 12. Respondent B označil všechny parafráze za maximálně gramaticky správné. Gramaticky nejspornější byly hodnoceny parafráze pomocí modelu J, T i F.

### 7.1.3 Stejný význam

Respondenti	b	F	J	m	T
R-M	6	8	3	5	4
R-T	6	2	5	5	3
R-B	5	2	8	11	4

Tabulka 7 – významová shoda pro jednotlivé modely



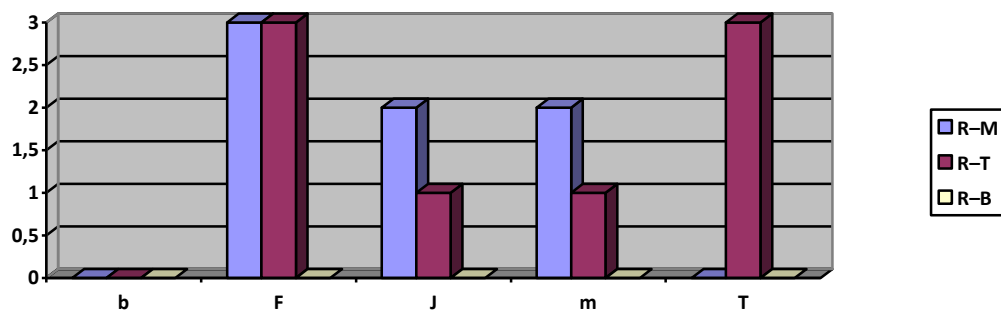
Obrázek 6 - graf tabulky 7

Maximum bodů bylo 12. Výsledky zde ukazují nejasné zadání „stejného významu“. Rozdílné hodnocení F, J i m ukazuje na problém, jak hodnotit stejný význam. Nejlépe vyšel model m, který však sám o sobě neprovedl mnoho náhrad. Druhý nejlepší model je J.

### 7.1.4 Modernizace

Respondenti	b	F	J	m	T
R-M	0	3	2	2	0
R-T	0	3	1	1	3
R-B	0	0	0	0	0

Tabulka 8 – modernizace pro jednotlivé modely



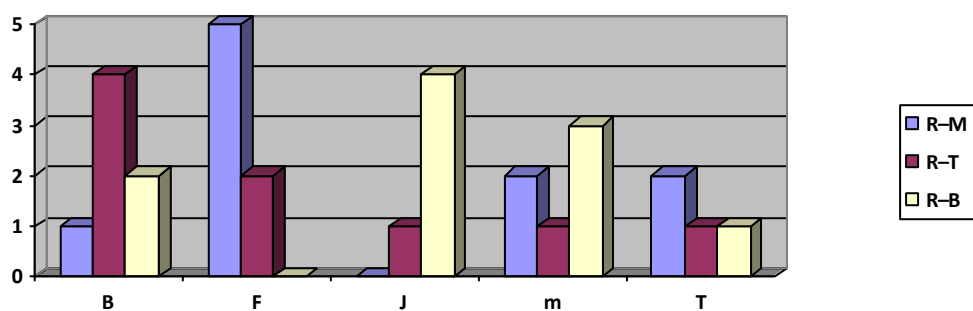
Obrázek 7 - graf tabulky 8

Maximum bodu je 6. Žádný model nepřesáhl polovinu těchto bodů. Respondent B označil všechny modely jako nevhodné pro parafrázování. Nejlepší model z hlediska modernizace je model F, za ním jsou modely J a m.

#### 7.1.5 Vhodnost modelu

Respondenti	B	F	J	M	T
R-M	1	5	0	2	2
R-T	4	2	1	1	1
R-B	2	0	4	3	1

Tabulka 9 – vhodnost modelu pro jednotlivé modely



Obrázek 8 - graf tabulky 9

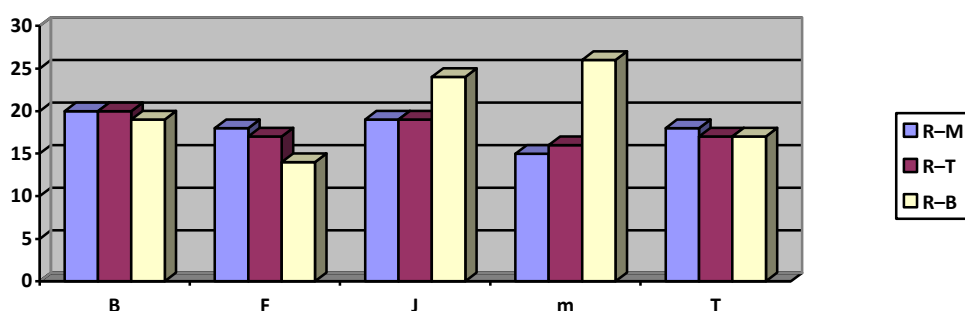
Maximum bodů je 6. Vhodnost modelu byla možnost, aby respondent vyjádřil svůj názor ohledně kvality modelů. Model F i b jsou označeny jako nejlepší. Za nimi je

model m a pak J. Očividně každý respondent upřednostňuje z nějakého důvodu jiný model.

### 7.1.6 Celkové skóre

Respondenti	B	F	J	m	T
R-M	20	18	19	15	18
R-T	20	17	19	16	17
R-B	19	14	24	26	17

Tabulka 10 – celkové ohodnocení modelů



Obrázek 9 - graf tabulky 10

Maximum bodů je 36. U respondentů M a T byl nejlepší model: bibleDohromady\_100\_1\_5, u respondenta B byl nejlepší model medickeTexty. Tyto modely však často neprovádí parafráze žádné, neboť word2vec vrátí nevhodné slovo, proto získávají maximum bodů za gramatickou správnost i za stejný smysl jako původní text. U respondenta M i T byl však model medickeTexty nejslabší. U M i T jsou rozdíly mezi jednotlivými modely téměř zanedbatelné. Respondent B navíc dal 0 u každého modelu pro modernizaci, takže je možné, že ve výsledné modernizaci se neprojeví dobře žádný model.

Díky T a B je druhým nejlepším modelem Jehovisti1\_bible12\_100\_5\_5, a proto jej použijí pro parafrázování a modernizaci samotné bible. Respondent B mi přímo napsal v poznámce, že považuje model Jehovisti1\_bible12\_100\_5\_5 jako nejlepší.

## 7.2 Parafrázování bible

Pro vyhodnocení kvality modernizování a parafráze jsem se rozhodl parafrázovat 20 náhodných veršů z bible. Nejdříve jsem si vygeneroval, z jakých knih mám vybírat verš, pak číslo kapitoly a nakonec verš. Poté jsem ještě vygeneroval, zda se jednotlivý verš vezme z Bible kralické, nebo z Českého ekumenického překladu. Přiřazoval jsem verše do dvou množin. Jakmile jedna z dvou množin dosáhla velikost deseti veršů, ubylé verše jsem přiřadil do druhé množiny. Zvolil jsem tyto dva překlady pro výsledné parafrázování proto, aby se na nich lépe hodnotila modernizace bible. Překlad pro 21. století neboli Bible21 je již přímo psán s moderním jazykem. Ekumenický překlad se vyznačuje jistou „kostrbatostí“, neboť chtěl být co nejvěrnější originálnímu znění. Generování náhodných čísel jsem prováděl pomocí unixové proměnné \$RANDOM.

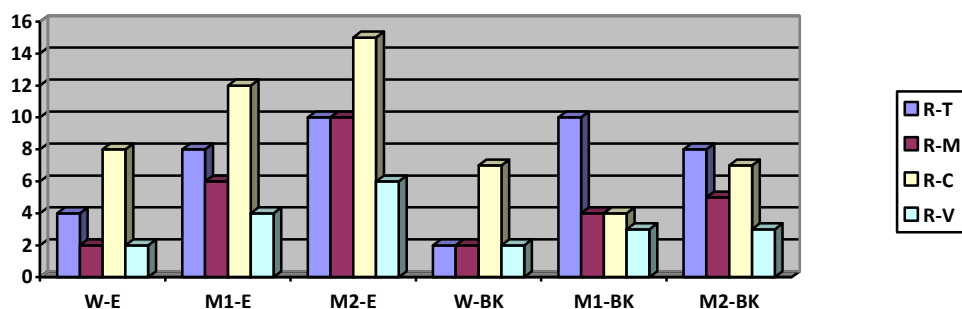
### 7.2.1 Zadání dotazníku

V příloze B je dotazník s parafrázemi, který jsem dal čtyřem respondentům, označuji je jako M, T, C a V. Měli v něm hodnotit kvalitu parafráze pomocí tří kritérií: Modernizace, stejný význam a gramatickou správnost. Jeho výsledky jsou v následujících sekcích sepsány do tabulek. „E“ znamená „Český ekumenický překlad“, „BK“ znamená „Bible kralická“, „W“ znamená parafráze pomocí Word2vecu, „M1“ znamená „MorphoDiTa - 1. verze“, neboli MorphoDiTa aplikováá na výstup parafráze z Word2vecu. „M2“ znamená „MorphoDiTa - 2. verze“, neboli otagování textu MorphoDiTou a až následné využití Word2vecu. Pro každou kategorii nejdříve uvedu tabulku s výsledky, které posléze okomentuji.

### 7.2.2 Modernizace

Respondenti	W-E	M1-E	M2-E	W-BK	M1-BK	M2-BK
R-T	4	8	10	2	10	8
R-M	2	6	10	2	4	5
R-C	8	12	15	7	4	7
R-V	2	4	6	2	3	3

**Tabulka 11 – modernizace parafrází biblických veršů**



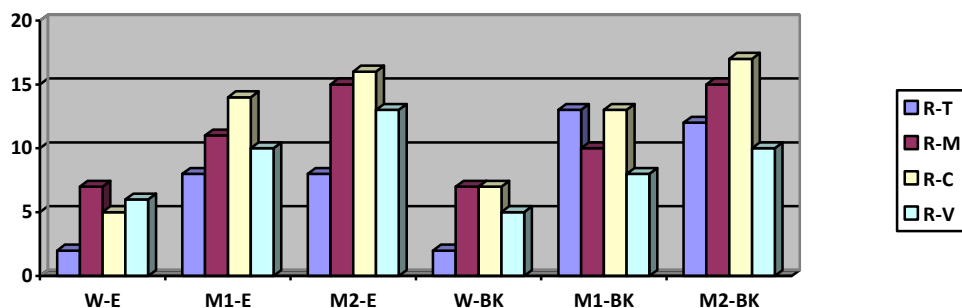
**Obrázek 10 - graf tabulky 11**

Maximální počet bodů za modernizaci je 20. Kromě respondenta C si žádný z respondentů nemyslí, že by se modernizace libovolnou metodou výrazněji zdařila. Většina si myslí, že k nějaké modernizaci došlo, ale nepovažuje ji za dostatečnou, spíše za drobné úspěšné náhrady. Bible kralická dopadla v modernizaci o poznání hůře než Český ekumenický překlad. Nejspíše dochází k tomu, že archaická slova buď nejsou nahrazena vůbec, nebo se nahradí opět za archaická.

### 7.2.3 Stejný význam

Respondenti	W-E	M1-E	M2-E	W-BK	M1-BK	M2-BK
R-T	2	8	8	2	13	12
R-M	7	11	15	7	10	15
R-C	5	14	16	7	13	17
R-V	6	10	13	5	8	10

**Tabulka 12 – shoda významu parafrází biblických veršů**



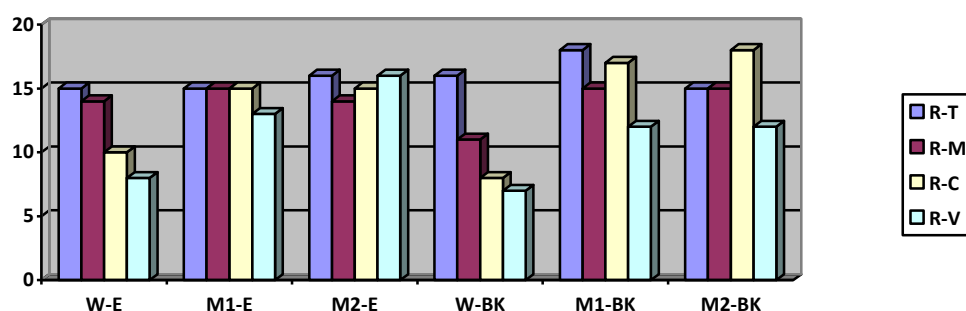
**Obrázek 11 - graf tabulky 12**

Maximální počet bodů je 20. Zde se výsledky 2. verze MorphoDiTy u E a BK podobají. Co se týče samotného Word2vecu, tabulka ukazuje, že dává často náhodné náhrady a úprava MorphoDiTou je tedy nutná.

#### 7.2.4 Gramatická správnost

Respondenti	W-E	M1-E	M2-E	W-BK	M1-BK	M2-BK
R-T	15	15	16	16	18	15
R-M	14	15	14	11	15	15
R-C	10	15	15	8	17	18
R-V	8	13	16	7	12	12

Tabulka 13 – gramatická správnost parafrází biblických veršů



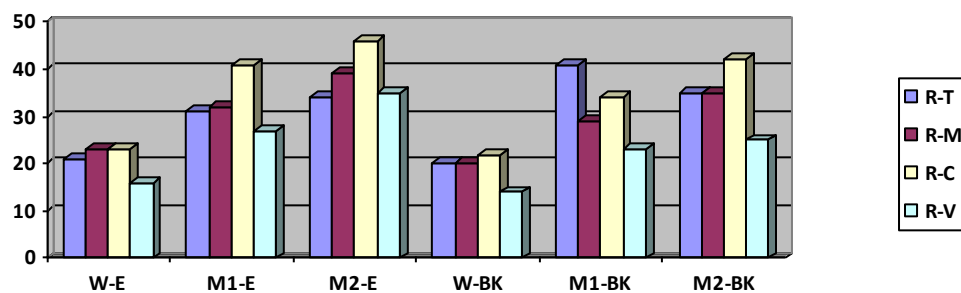
Obrázek 12 - graf tabulky 13

Z gramatické správnosti lze získat až 20 bodů. Zde se ukazuje, že word2vec funguje lépe než v jiných kategoriích. U MorphoDiTy bylo mým pravidlem, aby se shodovaly náhrady a původní slova v morfologické značce.

#### 7.2.5 Celkové skóre

Respondenti	W-E	M1-E	M2-E	W-BK	M1-BK	M2-BK
R-T	21	31	34	20	41	35
R-M	23	32	39	20	29	35
R-C	23	41	46	22	34	42
R-V	16	27	35	14	23	25

Tabulka 14 – celkové zhodnocení parafrází biblických veršů



Obrázek 13 - graf tabulky 14

Celkem je možno získat až 60 bodů. Tyto výsledky ukazují, že obecně parafrázování funguje lépe na modernějších textech než čistě archaických. Kromě respondenta T, který ohodnotil lépe parafrázování Bible kralické, se všichni shodují na preferování českého ekumenického překladu, jako vhodnějším textu pro parafrázi.

Co se týče úspěšnosti jednotlivých metod, opět je až na respondenta T pouhé použití Word2vecu nejméně úspěšné, poté následuje přidání MorphoDiTy, nakonec upravený algoritmus a větší využití MorphoDiTy. Z šedesáti možných bodů získalo čisté použití Word2vecu kolem třetiny bodů, druhá verze MorphoDiTy získala kolem dvou třetin.

## Závěr

Vytvořil jsem software, který umožňuje parafrázovat text pomocí podobnosti slov. Nashromáždil jsem data pro parafrázování a modernizování bible. Natrénoval jsem modely na několika typech textů a porovnal kvalitu parafráze. Nejvhodnější model jsem vybral pro parafrázování a modernizování několika náhodných biblických veršů. Použití samotného Word2vecu pro parafrázování dává v porovnání s ostatními metodami špatné výsledky. Problémem je určit, co přesně má Word2vec nahrazovat.

Rozšířením práce by mohlo být provázání z dalšími nástroji, pomocí kterých by se parafrázovala nejenom slova, ale i fráze. Bylo by třeba zjistit ve větě závislosti, například pomocí UDPipe, nástroje na <http://ufal.mff.cuni.cz/udpipe>. Nahrazení podstatného jména jiným s rozdílným rodem vyžaduje i změnu případného přídatného jména, které přímo s daným podstatným jménem souvisí. Dále by bylo dobré, pro zvýšení přesnosti parafrází s Biblí kralickou, zvětšit množinu textů českého jazyka ze 17. století.

Snížením náhrad pomocí MorphoDiTy a úprava výstupu výrazně zkvalitňuje parafrázi. Modernizovat text se podařilo spíše jen z části. Můj přístup funguje lépe pro modernější verše s archaickými slovy z Českého ekumenického překladu než pro texty ze 17. století z Bible kralické.

## Seznam použité literatury

- [1] Mind-Blowing Statistics About Christianity You Need to Know - Holyart.com Blog. Sacred art: the sale of sacred items truly unique on HOLYART.com [online]. Copyright © 2016 [cit. 22.04.2020]. Dostupné z: <https://www.holyart.com/blog/mind-blowing-statistics-christianity-need-know/>
- [2] FINKELSTEIN, Israel a Neil Asher SILBERMAN. *Objevování Bible: Svatá Písma Izraele ve světle moderní archeologie*. Vyd. 2. Přeložil Marie ČAPKOVÁ. Praha: Vyšehrad, 2010. Teologie (Vyšehrad). ISBN 978-80-7429-044-2.
- [3] I. Androutsopoulos and P. Malakasiotis, "A Survey of Paraphrasing and Textual Entailment Methods". *Journal of Artificial Intelligence Research*, 38:135-187, 2010.
- [4] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: "Efficient Estimation of Word Representations in Vector Space", 2013; [<http://arxiv.org/abs/1301.3781> arXiv:1301.3781].
- [5] NLP 101: Word2vec — Skip-gram and CBOW. *Towards Data Science* [online]. RIA KULSHRESTHA, 2019 [cit. 2020-05-21]. Dostupné z: <https://towardsdatascience.com/nlp-101-Word2vec-skip-gram-and-cbow-93512ee24314>
- [6] WELLHAUSEN, Julius. *Prolegomena zur Geschichte Israels*. 6. vydání. Berlin: Walter de Gruyter, 1927. 424 stran.
- [7] FILIPI, Pavel. *Křesťanstvo: historie, statistika, charakteristika křesťanských církví*. 4., dopl. vyd. Brno: Centrum pro studium demokracie a kultury, 2012. ISBN 978-80-7325-280-9.
- [8] Russia [online]. 2020 [cit. 2020-05-22]: <https://kids.britannica.com/students/article/Russia/276819#207538-toc>

- [9] HAJIČ, Jan. Positional Tags: Quick Reference (Czech "HM" Morphology). Ufal.mff.cuni.cz [online]. 2000 [cit. 2020-05-24].
- [10] [http://wiki.medik.cz/wiki/Hlavní\\_strana](http://wiki.medik.cz/wiki/Hlavní_strana)
- [11] <https://github.com/eabdullin/Word2Vec.Net>
- [12] <https://github.com/GuntaButya/Word2Vec.Net-CSharp>
- [13] <http://lindat.mff.cuni.cz/services/morphodita/>
- [14] Jana Straková, Milan Straka, Jan Hajič (2014): *Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition*. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 13-18, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-00-6
- [15] Tomas Mikolov, Quoc V. Le, Ilya Sutskever: "Exploiting Similarities among Languages for Machine Translation", 2013; [<https://arxiv.org/abs/1309.4168> arXiv:1309.4168v1].
- [16] Kevin Patel, Pushpak Bhattacharyya: "Towards Lower Bounds on Number of Dimensions for Word Embeddings", Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 31–36, Taipei, Taiwan, November 27 – December 1, 2017
- [17] R.A.B. Mynors, "The Publication of the Latin *Paraphrases*," in Robert Dick Sider (ed.), *New Testament Scholarship: Paraphrases on Romans and Galatians (Collected Works of Erasmus)* (University of Toronto Press, 1984: ISBN 0-8020-2510-2)
- [18] Taylor, Kenneth N., *My Life: A Guided Tour*, 1991

[19] LEEUWENBERG, Artuur, et al. A minimally supervised approach for synonym extraction with word embeddings. The Prague Bulletin of Mathematical Linguistics, 2016.

[20] COSINE, STATISTICAL ENGINEERING DIVISION, 2017,  
[www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm](http://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm).

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pages 3111–3119, 2013.

[22] Domingo, Miguel & Chinea-Rios, Mara & Casacuberta, Francisco. (2017). Historical Documents Modernization. The Prague Bulletin of Mathematical Linguistics. 108. 10.1515/pralin-2017-0028.

## Seznam obrázků

Obrázek 1 - kodex Alexandrinus.....	8
Obrázek 2 – projekce vektorů zemí a měst .....	9
Obrázek 3 – projekce prostoru vektorů slov .....	11
Obrázek 4 – porovnání CBOW a Skip-gramu .....	14
Obrázek 5 - graf tabulky 6 .....	36
Obrázek 6 - graf tabulky 7 .....	37
Obrázek 7 - graf tabulky 8 .....	38
Obrázek 8 - graf tabulky 9 .....	38
Obrázek 9 - graf tabulky 10 .....	39
Obrázek 10 - graf tabulky 11 .....	41
Obrázek 11 - graf tabulky 12 .....	41
Obrázek 12 - graf tabulky 13 .....	42
Obrázek 13 - graf tabulky 14 .....	43

## Seznam tabulek

Tabulka 1 – kosinová podobnost nejbližších slov .....	10
Tabulka 2 – vstup a výstup pro CBOW .....	12
Tabulka 3 – vstup a výstup pro Skip-gram .....	13
Tabulka 4 – morfologická značka .....	18
Tabulka 5 – použitá data .....	25
Tabulka 6 – gramatická správnost pro jednotlivé modely .....	36
Tabulka 7 – významová shoda pro jednotlivé modely.....	37
Tabulka 8 – modernizace pro jednotlivé modely .....	37
Tabulka 9 – vhodnost modelu pro jednotlivé modely.....	38
Tabulka 10 – celkové ohodnocení modelů .....	39
Tabulka 11 – modernizace parafrází biblických veršů .....	41
Tabulka 12 – shoda významu parafrází biblických veršů.....	41
Tabulka 13 – gramatická správnost parafrází biblických veršů.....	42
Tabulka 14 – celkové zhodnocení parafrází biblických veršů.....	42

## Příloha A

### Pokyny pro hodnocení parafrází

Parafráze jsou rozděleny do jednotlivých témat. Každé téma obsahuje pět modelů, které parafrázovaly zdrojovou větu. Vaše hodnocení prosím zaznamenejte do excelové tabulky, kde každému tématu odpovídá jeden list. U každé parafráze budete hodnotit tyto vlastnosti:

Gramatická správnost, stejné tvrzení, modernizace, vhodnost modelu.

**Gramatická správnost:** Pro danou parafrázi ohodnoťte, zda se vám zdá v pořádku, či obsahuje chyby. Pokud je parafráze úplně v pořádku, ohodnoťte 2, pokud obsahuje drobné chyby, ohodnoťte 1. Pokud je chyb více, ohodnoťte 0.

**Stejně tvrzení:** Pro danou parafrázi ohodnoťte, zda tvrdí to samé, co originál. Pokud ano, dejte 2 body. Pokud si myslíte, že se trochu odchyluje, ale smysl více méně zůstává stejný (třeba vymění zlato za stříbro apod.) ohodnoťte 1 bod. Pokud se smysl liší, ohodnoťte 0.

**Modernizace:** Zde ohodnoťte, zda se vám zdá, že parafráze užívá modernější výrazy než originál, případně zda vám přijde znění věty modernější. Pokud ano, ohodnoťte 1, jinak ohodnoťte 0.

**Vhodnost modelu:** V této kategorii můžete udělit 1 bod, pokud se vám parafráze líbí. Jinak nechte 0.

**Slovní komentář:** Slouží k tomu, abyste podle vás vybrali z pěti parafrází tu, která podle vás parafrázovala úryvek nejlépe, a napsali, proč tomu tak je. Může to být kvůli nejvyššímu počtu bodů, či dle jiného, subjektivního kritéria.

Pokud model nenahradí žádné slovo, nehodnoťte ho a do komentáře napište: „BEZ PARAFRÁZE“. První čtyři témata jsou převzata z Bible21. Poslední dvě jsou převzata z Bible kralické, znění 1613. Možná se vám bude zdát, že ani původní text není správně, zkuste však ohodnotit parafrázi tak, jak se vám jeví. Pokud bude výsledný text znít moderněji, měl by být kontrast výraznější než u Bible21.

## **Bible 21**

### Mystika

Původní text: Jan 1:1

Na počátku bylo slovo, to slovo bylo u Boha, to slovo byl Bůh. To bylo na počátku u Boha.

### **Model: bibleDohromady\_100\_1\_5**

Na počátku bylo poselství, to poselství bylo u bludaře, to poselství byl Bůh. To bylo na počátku u bludaře

### **Model: Fantasy\_medic\_filosof**

Na začátku zbylo pozdrav, to pozdrav zbylo u Boha, to pozdrav byl Bůh. To zbylo na začátku u Boha

### **Model: Jehovisti1\_bible12\_100\_5\_5**

Na začátku žilo poselství, to poselství žilo u Boha, to poselství byl Bůh. To žilo na začátku u Boha

### **Model: medickeTexty**

Na petechii bylo slovo, to slovo bylo u Boha, to slovo byl Bůh. To bylo na petechii u Boha.

### **Model: TitulkyABible\_100\_5\_5**

Na začátku zbylo sdělení, to sdělení zbylo u Boha, to sdělení byl Bůh. To zbylo na začátku u Boha.

## Právo

### **Původní text: Exodus 20:2–17**

Bůh promluvil všechna tato slova:

Já jsem Hospodin, tvůj Bůh, který tě vyvedl z Egypta, z domu otroctví.

Neměj žádné bohy kromě mne.

Nevytvářej si modly v podobě čehokoli nahoře na nebi, dole na zemi nebo ve vodách pod zemí. Neklaň se jim a nesluž jim, neboť já Hospodin, tvůj Bůh, jsem Bůh žárlivě milující. Trestám nepravost otců na synech do třetího i čtvrtého pokolení těch, kdo mě nenávidí, a prokazují milosrdenství tisícům pokolení těch, kdo mě milují a zachovávají má přikázání.

Neužívej jméno Hospodina, svého Boha, nadarmo, neboť toho, kdo by užil jeho jména nadarmo, Hospodin neponechá bez trestu.

Pamatuj na sobotní den, aby ti byl svatý. Šest dní budeš pracovat a dělat všechnu svou práci, ale sedmý den je dnem odpočinku, zasvěceným Hospodinu, tvému Bohu. Nebudeš dělat žádnou práci – ty, tvůj syn ani tvá dcera, tvůj otrok ani tvá děvečka, tvé dobytče ani přistěhovalec ve tvých branách. V šesti dnech totiž Hospodin učinil nebe i zemi, moře a všechno, co je v nich, ale sedmého dne odpočinul. Proto Hospodin požehnal sobotní den a posvětil jej.

Cti svého otce i matku, ať jsi dlouho živ na zemi, kterou ti dává Hospodin, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bližním.

Nedychti po domě svého bližního. Nedychti po manželce svého bližního, po jeho otroku, jeho děvečce, jeho býku ani jeho oslu – vůbec po ničem, co je tvého bližního."

### **Model: bibleDohromady\_100\_1\_5**

Bůh promluvil všechna ta řeči:

Já jsem faraón, tvůj Bůh, který tě vyvedl z područí, z domu přístřešku.

Neměj žádné modly kromě mne.

Nevytvářej si modly v zahradě čehokoli důle na nebesech, dole na zemi nebo ve vrcholcích pod královstvím. nerozervi se jim a netaj jim, nebo já faraón, tvůj Bůh, jsem Bůh žárlivě milující. Trestám nešlechtnost předků na polích do osmého i šestého kmene těchto, kdo mě nenávidí, a prokazují slitování tisícům kmene těchto, kdo mě pronásledují a dodržují má přikázání.

Neužívej símě Hospodina, svého bludaře, naprosto, nebo onoho, kdo by užil jeho jména naprosto, faraón nenechává bez trestu.

Pamatuj na sváteční den, aby ti byl zkroušený. Šest dní budeš spát a připravovat všechnu svou službu, ale osmý den je nocí odpočinku, zasvěceným bohovi, tvému usvědčování. Nebudeš připravovat žádnou službu – ty, tvůj syn ani tvá manželka, tvůj služebník ani tvá děvečka, tvé býk ani host ve tvých stopách. V pěti dnech totiž faraón učinil nebesa i zemi, řeku a všechno, co je v nich, ale osmého dne zatměl. Proto faraón žehnal sváteční den a posvětil jej.

Cti svého bratra i chůvu, ať jsi dlouho živ na zemi, kterou ti dá faraón, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bratrovi.

Nedychti po paláci svého druha. Nedychti po snaše svého druha, po jeho služebníku, jeho pršce, jeho oslu ani jeho býku – naopak po ničem, co je tvého druha. "

### **Model: Fantasy\_medic\_filosof**

Bůh pronesl veškerá tahle jména:

Já jsem Hospodin, tvůj Bůh, který tě vyvedl z mokřadu, z paláce pokladnice.

Neměj žádné duchy namísto mne.

Nevytvářej si pilíře v důsledku něčeho důle na nebesech, dole na podlaze nebo ve prostorách pod hladinou. Neklaň se jim a nesluž jim, neboť já Hospodin, tvůj Bůh, jsem Bůh úzce milující. Trestám nepravost bratrů na duších do čtvrtého i třetího tisíciletí těchtohle, kdo mě nenávidí, a prokazují požehnání tisícům tisíciletí těchtohle, kdo mě uctívají a zachovávají má povídačky.

Neužívej slovo Hospodina, svého Boha, nadarmo, neboť tohotohle, kdo by zvyknul jeho slova nadarmo, Hospodin neponechá bez činu.

Pamatuj na sobotní den, aby ti byl spravedlivý. Šest dní budeš studovat a dělat veškerou svou lekci, ale sedmý den je nadechnutím odjezdu, zasvěceným Hospodinu,

tvému Bohu. Nebudeš dělat žádnou lekci – ty, tvůj syn ani tvá manželka, tvůj válečník ani tvá děvečka, tvé dobytče ani přistěhovalec ve tvých pruhách. V třiceti dnech ale Hospodin přijal nebe i podlaze, řeku a všechno, co je v nich, ale devátého dne vychutnal. Proto Hospodin připomněl sobotní den a odtáhl jej.

Cti svého syna i sestru, ať jsi dlouho živ na podlaze, kterou ti dává Hospodin, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bližním.

Nedychti po sídle svého bližního. Nedychti po sestře svého bližního, po jeho otroku, jeho děvečce, jeho býku ani jeho řemeni – vůbec po nikom, co je tvého bližního. "

#### **Model: Jehovisti1\_bible12\_100\_5\_5**

Bůh promluvil všechna taková slova:

Já jsem bůh, tvůj Bůh, který tě přivedl z pouště, z domu zajetí.

Neměj žádné modly kromě mne.

Nevytvářej si bohy v důsledku něčeho důle na nebesech, dole na zemi nebo ve vodách pod zemí. Neklaň se jim a neklaň jim, neboť já bůh, tvůj Bůh, jsem Bůh žárlivě milující. Trestám nešlechtnost otců na králích do třetího i třetího kmene těch, kdo mě nenávidí, a prokazují milosrdenství tisícům kmene těch, kdo mě poslouchají a dodržují má přikázání.

Neužívej jméno Hospodina, svého Boha, nadarmo, neboť toho, kdo by uchovával jeho jména nadarmo, bůh neponechá bez trestu.

Pamatuj na sváteční den, aby ti byl průkopnický. Šest dní budeš pracovat a dělat veškerou svou práci, ale osmý den je dnem odpočinku, zjeveným bohovi, tvému Bohu. Nebudeš dělat žádnou práci – ty, tvůj syn ani tvá matka, tvůj služebník ani tvá vdova, tvé maso ani host ve tvých branách. V pěti dnech ale bůh učinil nebe i zemi, moře a všechno, co je v nich, ale osmého dne odpočinul. Proto bůh žehnal sváteční den a posvětil jej.

Cti svého pána i manželku, ať jsi dlouho živ na zemi, kterou ti poskytuje bůh, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bližním.

Nedychti po domě svého bližního. Nedychti po manželce svého bližního, po jeho služebníkovi, jeho děvečce, jeho pochoutce ani jeho oslíku – prostě po ničem, co je tvého bližního. "

### **Model: medickeTexty**

Bůh promluvil všechna tato slova:

Já chcípnu Hospodin, tvůj Bůh, který tě vyvedl z Egypta, z domu otroctví.

Neměj žádné bohy kromě mne.

Nevytvářej si modly v chlápku či v ženském těle nahoře na nebi, dole na zemi nebo ve vodách pod zemí. Neklaň se jim a nesluž jim, neboť já Hospodin, tvůj Bůh, chcípnu Bůh žárlivě milující. Trestám nepravost otců na synech do třetího i čtvrtého pokolení těch, kdo mě nenávidí, a prokazují milosrdenství tisícům pokolení těch, kdo mě milují a zachovávají má přikázání.

Neužívej jméno Hospodina, svého Boha, nadarmo, neboť toho, kdo by užil jeho jména nadarmo, Hospodin neponechá bez trestu.

Pamatuj na sobotní den, aby ti byl svatý. Šest dní budeš pracovat a mít všechnu svou práci, ale sedmý den je dnem odpočinku, zasvěceným Hospodinu, tvému Bohu. Nebudeš mít žádnou práci – ty, tvůj syn ani tvá dcera, tvůj otrok ani tvá děvečka, tvé dobytče ani přistěhovalec ve tvých branách. V šesti dnech totiž Hospodin učinil nebe i zemi, moře a všechno, co je v nich, ale sedmého dne odpočinul. Proto Hospodin požehnal sobotní den a posvětil jej.

Cti svého otce i matku, ať jsi dlouho živ na zemi, kterou ti dává Hospodin, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bližním.

Nedychti po domě svého bližního. Nedychti po manželce svého bližního, po jeho otroku, jeho děvečce, jeho býku ani jeho oslu – vůbec po ničem, co je tvého bližního.  
"

### **Model: TitulkyABible\_100\_5\_5**

Bůh promluvil veškerá tahle slůvka:

Já jsem hospodin, tvůj Bůh, který tě vyhnal z ciziny, z bytu hříchu.

Neměj žádné smrtelníky kromě mne.

Nevytvářej si ohavnosti v atmosféře čehokoli dříve na nebesech, dole na hoři nebo ve oblastech pod oblastí. Neklaň se jim a nesluž jim, neboť já hospodin, tvůj Bůh, jsem Bůh žárlivě milující. Trestám nešlechtnost matek na vnučích do čtvrtého i třetího království těchto, kdo mě snáší, a prokazují spasení milionům království těchto, kdo mě zbožňují a dodržují má přikázání.

Neužívej jméno Hospodina, svého Boha, nadarmo, neboť tohoto, kdo by užil jeho jména nadarmo, hospodin neponechá bez rozsudku.

Pamatuj na páteční den, aby ti byl boží. Šest dní budeš makat a dělat veškerou svou praxi, ale pátý den je okamžikem truchlení, zasvěceným pánu, tvému Bohu. Nebudeš dělat žádnou praxi – ty, tvůj syn ani tvá sestra, tvůj služebník ani tvá služebná, tvé jehně ani přistěhovalec ve tvých komnatách. V pěti dnech jenže hospodin učinil nebesa i hoři, oceán a všechno, co je v nich, ale šestého dne zacvičil. Proto hospodin požehnal páteční den a posvětil jej.

Cti svého bratra i sestru, ať jsi dlouho živ na hoři, kterou ti dává hospodin, tvůj Bůh.

Nezabíjej.

Necizolož.

Nekrad'.

Nelži o svém bližním.

Nedychti po bytě svého druha. Nedychti po ženě svého druha, po jeho služebníku, jeho děvečce, jeho býku ani jeho kasinu – vůbec po nikom, co je tvého druha. "

## Popis pracovního postupu

### **Původní text: Exodus 27:1-4**

Vyrobíš také čtvercový oltář z akáciového dřeva. At' má pět loket na délku, pět loket na šířku a tři lokte na výšku. V jeho čtyřech koutech vytvaruješ rohy, jež budou vystupovat z oltáře, a obložíš jej bronzem. K oltáři vyrobíš nádoby na vynášení popela, lopaty, obětní misky, vidlice a pánvice na oharky. Veškeré jeho náčiní vyrobíš z bronzu. Zhotovíš k němu také mřížový rošt z bronzu a na jeho čtyři konce připevníš čtyři bronzové kruhy.

### **Model: bibleDohromady\_100\_1\_5**

Vyrobíš také dvojité příbytek z mřížového dřeva. At' má pět loket na šířku, pět loket na délku a tři lokte na výšku. V jeho čtyřech dírách vytvaruješ býčky, jež budou mstit z příbytku, a obložíš jej bronzem. K oltáři vyrobíš nádoby na vynášení tuku, kropenky, obětní pánvičky, pánve a pohárky na oslíky. Veškeré jeho nářadí vyrobíš z bronzu. Zhotovíš k němu také vyšitý rošt z bronzu a na jeho čtyři večery připevníš čtyři bronzové kruhy.

### **Model: Fantasy\_medic\_filosof**

Vyrobíš taky objemný oblouk z akáciového železa. At' má pět rukávů na šířku, pět rukávů na plochu a tři ramena na rychlost. V jeho třech údolích vytvaruješ korouhve, jež budou sestupovat z kašny, a obložíš jej bronzem. K mrtvole vyrobíš mísy na vynášení cihly, kostičky, zakřivené nádobky, hřívy a pánvice na oharky. Veškeré jeho hromádky vyrobíš z bronzu. Zhotovíš k němu taky mřížový rošt z bronzu a na jeho tři nynějšíky připevníš tři pomalované důlky.

### **Model: Jehovisti1\_bible12\_100\_5\_5**

zhotovíš navíc čtvercový příbytek z cypřišového dřeva. At' má pět loket na délku, pět loket na loket a tři lokte na výšku. V jeho třech úhlech vysloviš rohy, jež budou vystupovat z oltáře, a obložíš jej mědí. K oltáři zhotovíš předměty na vynášení gumy, kropenky, obětní pánvičky, lopaty a pánvičky na pánvice. Veškeré jeho vybavení

zhotovíš z mědi. Zhotovíš k němu navíc mřížový rošt z mědi a na jeho tři konce připevníš tři měděné obruby.

**Model: medickeTexty**

Vyrobíš také čtvercový oltář z akáciového dřeva. At' má pět loket na délku, pět loket na šířku a tři lokte na výšku. V jeho čtyřech koutech vytvaruješ rohy, jež budou vystupovat z oltáře, a obložíš je bronzem. K oltáři vyrobíš nádoby na vynášení popela, lopaty, obětní misky, vidlice a pánvice na oharky. Veškeré jeho náčiní vyrobíš z bronzu. Zhotovíš k němu také mřížový rošt z bronzu a na jeho čtyři konce připevníš čtyři bronzové kruhy.

**Model: TitulkyABible\_100\_5\_5**

zhotovíš též čtvercový příbytek z akáciového železa. At' má pět palců na hloubku, pět palců na délku a tři pruty na váh. V jeho třech částech vytvaruješ drápy, jež budou vystupovat z oltáře, a obložíš je mědí. K tanci zhotovíš misky na čištění hlin, hřebíky, obětní mísy, knoty a pánvičky na oharky. Veškeré jeho nářadí zhotovíš z mědi. Zhotovíš k němu též mřížový rošt z mědi a na jeho tři zítřky připevníš tři měděné otvory.

## Podobenství

### **Původní text: Lukáš 15:8-9**

Nebo má-li nějaká žena deset stříbrných mincí a ztratí jednu z nich, což nerozsvítí lampu, nevymete dům a nehledá pečlivě, dokud ji nenajde? A když ji nalezne, svolá své přítelkyně a sousedky a řekne: „Radujte se se mnou, poněvadž jsem našla peníz, který jsem ztratila.“

### **Model: bibleDohromady\_100\_1\_5**

Nebo má-li nějaká dívka pět stříbrných ovcí a ztratí jednu z nich, což nehlaholí lampu, nevymetá dům a nepřivádí pečlivě, dokud ji nenajde? A jakmile ji nalezne, láká své lásky a sousedky a praví: „Radujte se se mnou, ač jsem našla peníz, který jsem překypovala.“

### **Model: Fantasy\_medic\_filosof**

Nebo má-li nějaká dáma pět zlatých papírů a změní jednu z nich, což nerozsvítí lucernu, nevymete dům a nehledá důkladně, pokud ji nenajde? A jakmile ji nalezne, svolá své sestřenice a sousedky a řekne: „Radujte se se mnou, protože jsem vyhledala peníz, který jsem ztratila.“

### **Model: Jehovisti1\_bible12\_100\_5\_5**

Nebo má-li nějaká dívka pět stříbrných mincí a získá jednu z nich, což nelisuje lampu, nevymetá dům a neztrácí důkladně, takže ji nenajde? A jestliže ji najde, přijde své spolupracovnice a babičky a řekne: „Radujte se se mnou, jakož jsem získala denár, který jsem dostala.“

### **Model: medickeTexty**

Nebo má-li nějaká muž deset stříbrných mincí a ztratí jednu z nich, což nerozsvítí lampu, nevymete dům a nehledá pečlivě, dokud ji nenajde? A když ji nalezne, svolá své přítelkyně a sousedky a řekne: „Radujte se se mnou, poněvadž chcípnu našla peníz, který chcípnu ztratila.“

**Model: TitulkyABible\_100\_5\_5**

Nebo má-li nějaká dcera pět stříbrných kostek a ztrácí jednu z nich, což nezaslechne vázu, nevymetá dům a nerespektuje důkladně, dokud ji nenajde? A jestli ji nalezne, svolá své sestřenice a tety a poví: „Radujte se se mnou, zdali jsem našla hadřík, který jsem ztratila.“

## **Bible kralická**

### Modlitba

Pro výběr modlitby použijí znění modlitby „Otčenáš“, neboli modlitby Páně, které sice není z Bible21, ale je v českém prostředí nejznámější.

### **Původní text: Lukáš 11:2-4**

I řekl jim: Když se modlíte; říkejte: Otče náš, jenž jsi v nebesích, posvět' se jméno tvé. Přijď království tvé. Buď vůle tvá, jako v nebi tak i na zemi. Chléb náš vezdejší dávej nám každého dne. I odpusť nám hříchy naše, nebo i my odpouštíme všelikému vinníku našemu. A nevod' nás v pokušení, ale zbav nás od zlého.

### **Model: bibleDohromady\_100\_1\_5**

I pravil jim: Když se scházíte; rcete: Otče náš, jenž jsi v nebích, posvět' se símě tvé. Přijď tajemství tvé. Buď vůle tvá, jako v nebesech tak i na zemi. Chléb náš vezdejší dávej nám každého dne. I zjednej nám viny naše, nebo i my proviňujeme všelikému spolupracovníku našemu. A neuved' nás v pokušení, ale vykup nás od dobrého.

### **Model: Fantasy\_medic\_filosof**

I odpověděl jim: Když se modlíte; říkejte: Otče náš, jenž jsi v nebesích, posvět' se slovo tvé. Přijď území tvé. Buď mysl tvá, jako v nebesech tak i na podlaze. Chléb náš vezdejší dávej nám každého dne. I důvěřuj nám závazky naše, nebo i my odpouštíme všelikému vinníku našemu. A nevod' nás v pokušení, ale zbav nás od dobrého.

### **Model: Jehovisti1\_bible12\_100\_5\_5**

I řekl jim: Když se modlíte; rcete: Otče náš, jenž jsi v nebích, proklej se jméno tvé. Přijď království tvé. Buď vůle tvá, jakožto v nebesech tak i na zemi. Chléb náš vezdejší dej nám každého dne. I zjednej nám hříchy naše, nebo i my odpouštíme společnému vinníku našemu. A neuved' nás v pokušení, ale zbav nás od zlého.

### **Model: medickeTexty**

I řekl jim: Když se modlíte; říkejte: Otče náš, jenž jsi v nebesích, posvět' se jméno tvé. Přijď království tvé. Buď vůle tvá, jako v nebi tak i na zemi. Chléb náš vezdejší dávej nám každého dne. I odpusť nám hříchy naše, nebo i my odpouštíme všelikému vinníku našemu. A neuvod' nás v pokušení, ale zbav nás od zlého.

**Model: TitulkyABible\_100\_5\_5**

I řekl jim: Když se modlíte; říkejte: Otče náš, jenž jsi v nebích, posvět' se jméno tvé. Přijď panství tvé. Buď vůle tvá, jako v nebesech tak i na hoři. Chléb náš vezdejší dávej nám každého dne. I odpusť nám činy naše, nebo i my odpouštíme všelikému vinníku našemu. A neuved' nás v temnotě, ale zbav nás od strašného.

### Poezie

Bible je plná poezie, avšak nejpoetičtější kniha je rozhodně Píseň písní. Je to milostná poezie, která se do bible dostala záhadně, místy je „velmi necudná“. Jeden z možných výkladů této knihy je alegorický vztah Boha a Izraele.

#### **Původní text:**

Jako lilium mezi trnám, tak přítelkyně má mezi pannami. Jako jabloň mezi dřívím lesním, tak milý můj mezi mládenci. V stínu jeho žádostiva jsem byla seděti, a sedímť; nebo ovoce jeho sladké jest ústům mým. Uvedl mne na hody, maje za korouhev lásku ke mně. Očerstvětež mne těmi flašemi, posilňte mne těmi jablky, nebo umdlévám milostí.

#### **Model: bibleDohromady\_100\_1\_5**

Jako lilium mezi trnám, tak láska má mezi pannami. Jako jabloň mezi dřívím lesním, tak milý můj mezi služebníky. V stínu jeho žádostiva jsem zůstala seděti, a sedímť; nebo úroda jeho hořké jest cestám mým. Uvedl mne na zápaly, maje za korouhev trpělivost ke mně. flaše mne těmi flašemi, řiďte mne těmi palmami, nebo umdlévám mocí.

#### **Model: Fantasy\_medic\_filosof**

Jako lilium před ostružiním, tak sestřenice má před vránami. Jako jabloň před dřívím mírným, tak drahý můj před mládenci. V stínu jeho žádostiva jsem bývala seděti, a sedímť; nebo ovoce jeho krásné jest hrdlům mým. Uvedl mne na rituály, maje za suknicu pomstu ke mně. Očerstvětež mne těmihle flašemi, posilňte mne těmihle jablky, nebo umdlévám vlastí.

#### **Model: Jehovisti1\_bible12\_100\_5\_5**

Jako lilium s bodláčím, tak spolupracovnice má s pannami. Jako jabloň s jezerem lesním, tak milý můj s muži. V stínu jeho žádostiva jsem vzniknula procházeti, a pěstujitť; nebo ovoce jeho sladké jest dásním mým. Uvedl mne na hostiny, maje za korouhev úctu ke mně. flaše mne těmi haklíky, posilňte mne těmi jablky, nebo umdlévám vůlí.

**Model: medickeTexty**

Jako lilium mezi trnám, tak přítelkyně má mezi pannami. Jako jabloň mezi dřívím lesním, tak milý můj mezi mládenci. V stínu jeho žádostiva chcípnu byla seděti, a sedímt'; nebo ovoce jeho sladké jest ústům mým. Uvedl mne na hody, maje za korouhev lásku ke mně. Očerstvětež mne těmi flašemi, posilňte mne těmi jablky, nebo umdlévám milostí.

**Model: TitulkyABible\_100\_5\_5**

Jako jabloň mezi silničkou, tak sestřenice má mezi kráskami. Jako lilium mezi pilinou lesním, tak laskavý můj mezi chlapci. V stínu jeho žádostiva jsem bývala seděti, a pěstujit'; nebo cukroví jeho rozkošné jest nohám mým. Uvedl mne na velikonoce, maje za rámě krásu ke mně. flaše mne těmihle flašemi, posilněte mne těmihle rozinkami, nebo umdlévám vírou.

## **Příloha B**

### Pokyny pro hodnocení parafrází

Jednotlivé verše jsou po sobě 3x parafrázovány pomocí různých metod a můžete hodnotit, která parafráze z těchto tří je nejpovedenější. Hodnotí se tři kategorie: gramatická správnost, stejné tvrzení, modernizace. Odpovědi prosím zaznamenejte do přiložené excelové tabulky.

**Gramatická správnost(G):** Pro danou parafrázi ohodnoťte, zda se vám zdá v pořádku, či obsahuje chyby. Pokud je parafráze úplně v pořádku, ohodnoťte 2, pokud obsahuje drobné chyby, ohodnoťte 1. Pokud je chyb více, ohodnoťte 0.

**Stejně tvrzení(S):** Pro danou parafrázi ohodnoťte, zda tvrdí to samé, co originál. Pokud ano, dejte 2 body. Pokud si myslíte, že se trochu odchyluje, ale smysl více méně zůstává stejný (třeba vymění zlato za stříbro apod.) ohodnoťte 1 bod. Pokud se smysl liší, hodnotte 0.

**Modernizace(M):** Zde hodnotte, zda se vám zdá, že parafráze užívá modernější výrazy než originál, případně zda vám přijde znění věty modernější. Pokud ano, ohodnoťte 2, pokud se vám zdá, že se alespoň nějaká slova zaměnila za modernější, hodnotte 1, jinak hodnotte 0.

## Parafráze – Český ekumenický překlad

### **Genesis 8:17**

Vyved' s sebou všechno tvorstvo, jež je s tebou, všechnu zvěř i ptactvo a dobytek a všechnu havěť plazící se po zemi. Ať se na zemi hemží, ať se na zemi plodí a množí.

### **Word2Vec**

provějí s sebou vše pobývaly, jež je s námi, veškerou polní i ptáky a brav a veškerou zvěř hemžící se po zemi. Ať se na zemi zahemží, ať se na zemi množí a rozmnožují.

### **MorphoDiTa – 1. verze**

prověj s sebou všechno tvorstvo, jež je s mnou, veškerou polní i pták a brav a veškerou zvěř hemžící se po zemi. Ať se na zemi hemží, ať se na zemi množí a rozmnožuje.

### **MorphoDiTa – 2. verze**

pojdi s sebou všechno tvorstvo, jež je s tebou, všechnu polní i pták a brav a všechnu zvěř hemžící se po zemi. Ať se na zemi hemží, ať se na zemi množí a rozmnožuje.

### **Leviticus 18: 14**

Neodkryješ nahotu sestry své matky. Je to blízká příbuzná tvé matky.

### **Word2vec**

Neodkryješ odkryl sestry své matky. Je to zachována družka tvé matky.

### **MorphoDiTa – 1. verze**

Neodkryješ nahotu sestry své matky. Je to blízká družka tvé matky.

### **MorphoDiTa – 2. verze**

neodhalíš nahotu sestry své matky. je to blízká družka tvé matky.

### **Leviticus 23:21**

Podáváním nabídne snopek Hospodinu, aby ve vás našel zalíbení; druhého dne po dni odpočinku jej nabídne kněz podáváním.

### **Word2vec**

Podáváním nabídl bochánek bohu, aby ve vás našla potěšení; druhého dne po dni odpočinku jej nabídl velekněz zvedáním.

### **MorphoDiTa - 1. verze**

zvedáním nabídne bochánek bohovi, aby ve vás našel potěšení; druhého dne po dni odpočinku jej nabídne velekněz zvedáním.

### **MorphoDiTa – 2. verze**

zvedáním nabídne bochánek Hospodinu, aby ve vás našel potěšení; druhého dne po dně odpočinku jej nabídne velekněz zvedáním.

### **2. Paralipomenon 21:15**

Těžce onemocníš střevní chorobou, až ti vyhřeznou střeva pro nemoc zhoršující se den ze dne.

### **Word2vec**

Těžce střevní onemocníš poruchou, až ti střeva zimmici pro nemoc rozmáhající se den ze dne.

### **MorphoDiTa – 1. verze**

Těžce onemocníš střevní poruchou, až ti vyhřeznou zimnice pro nemoc rozmáhající se den ze dne.

### **MorphoDiTa – 2. verze**

Těžce onemocníš střevní poruchou, až ti vyhřeznou zimnice pro nemoc rozmáhající se den ze dne.

### **Přísloví 11:8**

Spravedlivý bývá zachován, je-li v soužení, kdežto svévolník se dostane na jeho místo.

#### **Word2vec**

Spravedlivý je zjeven, je-li v soužení, zatímco ničema se dostává na jejich místo.

#### **MorphoDiTa – 1. verze**

Spravedlivý je zachován, je-li v soužení, kdežto ničema se dostává na jeho místo.

#### **MorphoDiTa – 2. verze**

Spravedlivý je zachován, je-li v soužení, kdežto ničema se dostává na jeho místo.

### **Přísloví 23:28**

Číhá jak zákeřník a zavíní, že přibývá mezi lidmi věrolomných.

#### **Word2vec**

Číhá jak zavile a zákeřník, že vzrůstá s lidmi přímých.

#### **MorphoDiTa – 1.verze**

Číhá jak zákeřník a zavíní, že vzrůstá s lidmi přímých.

#### **MorphoDiTa – 2. verze**

útočí jak zákeřník a zavíní, že vzrůstá mezi lidmi přímých.

### **Píseň písní 5:15**

Jeho stehna jsou sloupy z bílého mramoru, spočívající na patkách z ryzího zlata. Vzhled má jak Libanón, je ztepilý jak cedr.

#### **Word2vec**

Jeho příhbi nejsou kruhy z jasného alabastru, obsažené na podstavcích z třpytivého zlata. úprava má jak bubeník, je záhon jak ztepilý.

### **MorphoDiTa – 1. verze**

Jeho stehna jsou kruhy z jasného alabastru, spočívající na podstavcích z třpytivého zlata. úprava má jak bubeník, je ztepilý jak cedr.

### **MorphoDiTa – 2. verze**

Jeho stehna jsou kruhy z jasného alabastru, spočívající na podstavcích z třpytivého zlata. Vzhled má jak Libanón, je ztepilý jak cedr.

### **Jonáš 3:3**

Jonáš tedy vstal a šel do Ninive, jak mu Hospodin uložil. Ninive bylo veliké město před Bohem, muselo se jím procházet tři dny.

### **Word2vec**

Jonáš však vstav a šel do Ninive, jak mu bůh svěřil. Ninive nebylo velké město kravami Bohem, mělo se jím prodírat tři dny.

### **MorphoDiTa – 1. verze**

Jonáš však vstavil a šel do Ninive, jak mu bůh svěřil. Ninive bylo velké město před Bohem, mělo se jím prodírat tři dny.

### **MorphoDiTa – 2. verze**

Jonáš tedy vstal a odešel do Ninive, jak mu Hospodin svěřil. Ninive bylo velké město před pánem, mělo se jím prodírat tři roky.

### **Jan 17 2**

stejně, jako jsi učinil, když jsi mu dal moc nad všemi lidmi, aby vše, co jsi mu svěřil, dal jim: život věčný.

### **Word2vec**

podobně, jakožto jsi udělal, jestliže jsi mu dal moc nad všemi lidmi, aby vše, co jsi mu dal, dal jim: život skutečný.

### **MorphoDiTa – 1. verze**

stejně, jakožto jsi učinil, jestliže jsi mu dal moc nad všemi lidmi, aby vše, co jsi mu dal, dal jim: život skutečný.

### **MorphoDiTa – 2. verze**

stejně, jako jsi učinil, když jsi mu dal moc nad všemi lidmi, aby vše, co jsi mu dal, dal jim: život skutečný.

### **2 Tesalonickým 2:5**

Nevzpomínáte si, že jsem vám to říkal, ještě když jsem byl u vás?

### **Word2vec**

Nevzpomínáte si, že bych vám to řekl, mnohem jestliže bych byl u vás?

### **MorphoDiTa – 1. verze**

Nevzpomínáte si, že jsem vám to řekl, mnohem jestliže jsem byl u vás?

### **MorphoDiTa – 2. verze**

nevzpomenete si, že jsem vám to řekl, ještě když jsem byl u vás?

## Parafráze – Bible kralická

### Leviticus 17:1

I mluvil Hospodin Mojžíšovi, řka:

#### **Word2vec**

I mluví bůh izraelitům, řka:

#### **MorphoDiTa – 1. verze**

I mluvil bůh izraelitovi, řka:

#### **MorphoDiTa - 2. verze**

I mluvil Hospodin Mojžíšovi, řka:

### Numeri 8:10

Postavíš Levíty před Hospodinem, a vloží synové Izraelští ruce své na Levíty.

#### **Word2vec**

Postavíš Levíty kravami bohem, a položí potomci Izraelští ruce své na Levíty.

#### **MorphoDiTa – 1. verze**

Postavíš Levíty před bohem, a vloží potomci Izraelští ruce své na Levíty.

#### **MorphoDiTa – 2. verze**

Postavíš levity před Hospodinem, a vloží potomci judští ruce své na levity

### Soudci 20:32

A řekli synové Beniamin: Padají před námi jako i prvě. Synové pak Izraelští řekli byli: Utíkejme, abychom je odtrhli od města až k stezkám.

#### **Word2vec**

A říkali potomci Beniamin: Padají kravami nimi jakožto i neposlední. zobrazení pak Izraelští říkali nebyli: odlákejme, abyste je oddělíte od města až k nenastražíš.

#### **MorphoDiTa – 1. verze**

A řekli potomci Beniamin: Padají před jimi jakožto i poslední. potomci pak Izraelští řekli byli: odlákejme, abychom je oddělili od města až k stezkám

#### **MorphoDiTa – 2. verze**

A řekli potomci Beniamin: padnou před námi jako i poslední. potomci pak judští řekli byli: Utíkejme, abychom je oddělili od města až k stezkám.

#### **Žalmy 12:5**

Kteříž říkají: Jazykem svým přemůžeme, máme' ústa svá s sebou, kdo jest pánem naším?

#### **Word2vec**

Kteříž říkali: Jazykem mým napadni, odměříš rty svá s sebou, kdo bude' otcem tímto?

#### **MorphoDiTa – 1. verze**

Kteříž říkají: Jazykem svým přemůžeme, odměříme' rty svá s sebou, kdo jest otcem naším?

#### **MorphoDiTa – 2. verze**

Kteříž říkají: Jazykem svým přemůžeme, odměříme' rty svá s sebou, kdo jest otcem naším?

#### **Izajáš 7:4**

A díš jemu: Šeť se, abys se nekormoutil. Neboj se, a srdce tvé nechať se neděsí dvou ostatků hlavní těch kouřících se, před rozpáleným hněvem Rezinovým s Syrskými, a syna Romeliášova,

### **Word2vec**

A díš mu: Šeť se, abych se vystříhal. přikážu se, a srdce tvé necht' se neznepokojuje tří syrských ústřední těch rozpáleným se, kravami zbouřením zlobně rozpáleným s rozpáleným, a otce otěhotníš,

### **MorphoDiTa – 1. verze**

A díš jemu: Šeť se, abys se nevystříhal. Neboj se, a srdce tvé necht' se neznepokojuje tří ostatků hlavní těch kouřících se, před rozpáleným hněvem rozpáleným s rozpálenými, a otky Romeliášova,

### **MorphoDiTa – 2. verze**

A díš jemu: Šeť se, abys se nekormoutil. Neboj se, a srdce tvé nechať se neděsí dvou ostatků hlavní těch kouřících se, před rozpáleným hněvem Rezinovým s Syrskými, a syna Romeliášova,

### **Joel 1:12**

Vinný kmen usechl, a fík usvadl, strom zrnatých jablek, též i palma i jabloň, všecko dříví polní poschlo, a že odňato potěšení od synů lidských.

### **Word2vec**

Vinný rod neurodí, a fík nachýlený, strom jablek granátových, též i kokosová i lesními, koli dřeva zvěř dočasné, a že vzato zalíbení od synů božích.

### **MorphoDiTa – 1. verze**

Vinný kmen usechl, a fík usvadl, strom zrnatých jablek, též i palma i jabloň, všecko dříví polní poschlo, a že odňato potěšení od synů lidských.

### **MorphoDiTa – 2. verze**

štípený rod urodil, a fík usvadl, strom zrnatých jablek, též i palma i jabloň, všecko dřevo polní poschlo, a že odňato zalíbení od synů božích.

### **Ámos 4:4**

Jdětež do Bethel, a buďtež poběhlci Galgala, rozmnožujte převrácenost, a přinášejte každého jitra oběti své, třetího roku desátky své.

### **Word2vec**

Jdětež do Bethel, a buďte šepci Galgala, množte převrácenosti, a přinesete jednoho večera oběti své, onoho roce desátek své.

### **MorphoDiTa – 1. verze**

Jdětež do Bethel, a buďtež poběhlci Galgala, množte převrácenost, a přineste každého večeru oběti své, třetího roku desátky své.

### **MorphoDiTa – 2. verze**

Jdětež do Bethel, a buďtež poběhlci Galgala, množte převrácenost, a přineste každého večeru oběti své, třetího roku desátky své.

### **Jakub 3:5**

Tak i jazyk malý úd jest, avšak veliké věci provodí. Aj, maličký oheň, kterak veliký les zapálí!

### **Word2vec**

Tak i jazyk krátký úd budeť, nicméně velké věci pášice. Aj, nepatrný plamen, jakž velký les spálí!

### **MorphoDiTa – 1. verze**

Tak i jazyk krátký úd jest, nicméně velké věci provodí. Aj, patrný plamen, jakž velký les zapálí!

### **MorphoDiTa – 2.verze**

Tak i jazyk krátký úd jest, avšak velké věci provodí. Aj, patrný plamen, kterak velký oheň zapálí!

### **1. Petrův 1:9**

Docházejíce konce víry vaší, spasení duší vašich,

#### **Word2vec**

Docházejíce konce víry naší, spásy duší mých,

#### **MorphoDiTa – 1. verze**

Docházejíce konce víry vaší, spása duší vašich,

#### **MorphoDiTa – 2. verze**

Docházejíce konce víry vaší, spása duší vašich,

### **1. Janův 4:1**

Nejmilejší, ne každému duchu věřte, ale zkušujte duchů, jsou-li z Boha; nebo mnozí falešní proroci vyšli na svět.

#### **Word2vec**

Nejmilejší, ne komukoli duchu způsobíte, ale válčíte andělů, nejsou-li z Boha; či někteří proroci proroci šli na svět.

#### **MorphoDiTa – 1. verze**

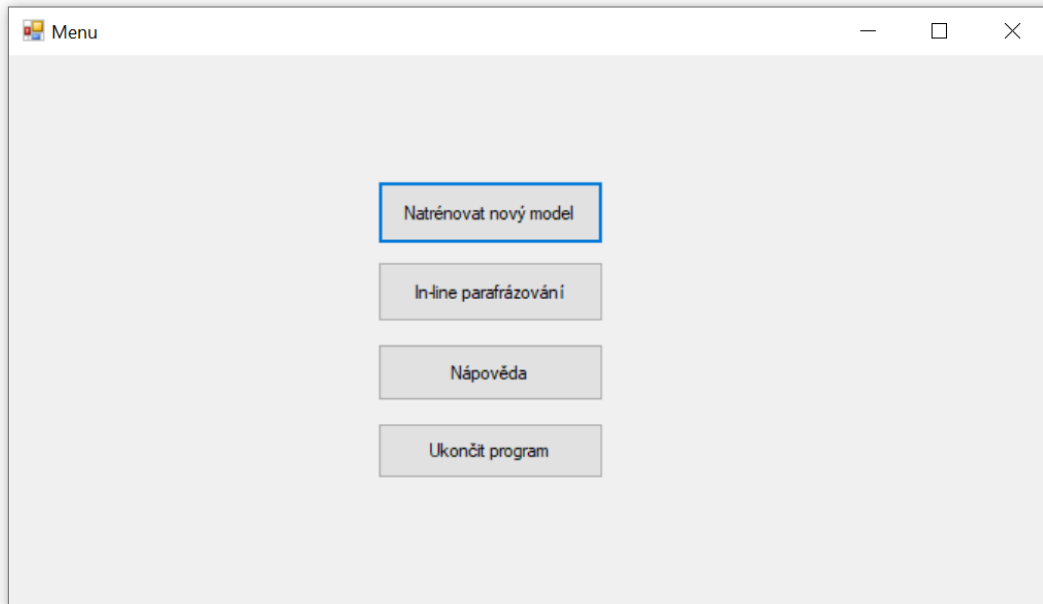
Nejmilejší, ne každému duchu způsobte, ale zkušujte andělů, jsou- li z Boha; či mnozí falešní proroci šli na svět.

#### **MorphoDiTa – 2. verze**

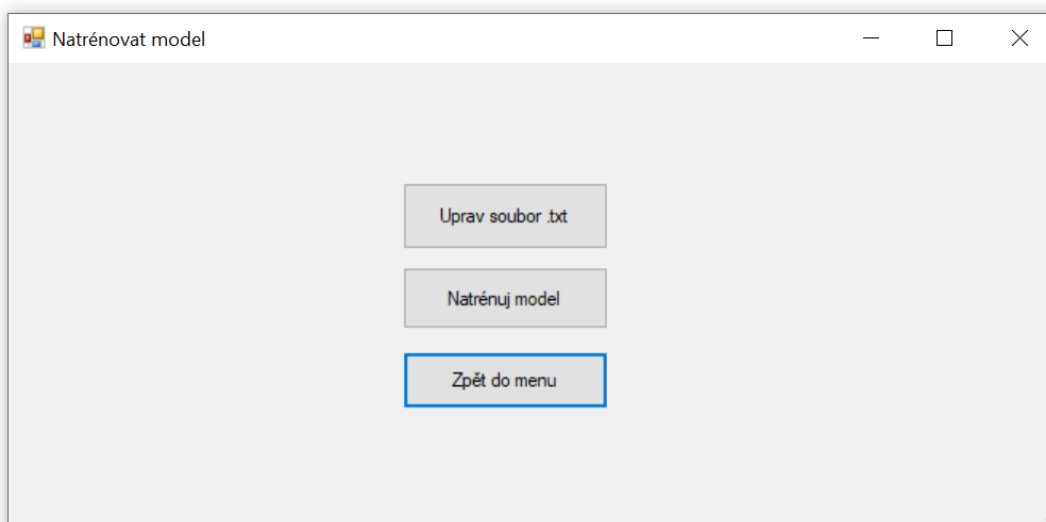
nejmilovanější, ne každému duchu způsobte, ale zkušujte andělů, jsou- li z stvořitele; nebo mnozí falešní proroci šli na svět.

## Zobrazení programu

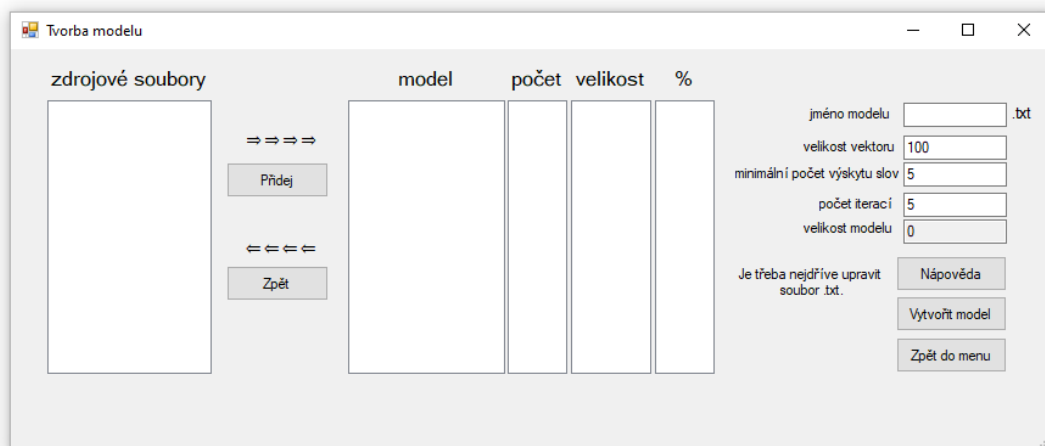
### Úvodní menu



### Natrénovat nový model



## Natrénuj model



## In-line parafrázování

