

nerada byva doma,
muž ji pravi, at' je doma,
že ji uhyně krava...“

Improving second language reading through visual attention cues to corpus-based patterns

Když dál nevěděl, obíal strom a pozoroval
kamaráda; ten se posmekl k okraji potoka, klekl a
dychtivě pil.

Kate Challis¹, Tom Drusa

¹Iowa State University, Ames, Iowa, USA
E-mail: kchallis@iastate.edu, t.drusa@gmail.com

Mužik se štětinať brade, který dělal na břízce
roztodivné prostocviky, zatřepal nohama v povětrí,
naklonil se a zašišlal polohlasitě a důvěrným tónem.

About the Researchers

Kate (Vasicek) Challis

[Czechoutyourancestors.com](https://czechoutyourancestors.com)

Technical Historical Records Linguist at FamilySearch

MA TESOL and Applied Linguistics (ISU '22)

PhD Applied Linguistics and Technology (ISU '27?)

Tom Drusa

UI Developer at Avast Software

MS Informatics and Natural Language Processing

(Masaryk University '14)



Background

- receptive vocabulary needs 95–98% text coverage for fluent reading
- L2 reading is more successful when readers attend to top-down strategies
- corpus-based data can help learners focus on important words
- visual attention cues can offload working memory burden

Czech Text

Rovněž na hukvaldském panství jsou známy případy zajetí či přímo únos civilních obyvatel vojáky, jež tímto způsobem hodlali získat peníze či uskutečnit výměnu zajatců. V lepším případě rukojmí po zaplacení výkupného vrátili zpět jeho rodině. Hůře ovšem dopadli již výše zmínění příborští měšťané Sarkandrovi, jež se svým potomkem zřejmě již neshledali („...syna mně vzali a do Turek prodali...“). [21] Při vpádu Mansfeldových oddílů do Moravské Ostravy v roce 1626 zajali vojáci (pravděpodobně na popud českých emigrantů) děkana Martina Regulina. Ten byl posléze vyměněn za nekatolické kněží vězněné kardinálem Ditrichštejnem. [22]

Czech Text

Rovněž **na** **hukvaldském** panství jsou **známy** případy **zajetí** **či** přímo **únos** **civilních** obyvatel vojáky, jež tímto způsobem hodlali získat peníze **či** uskutečnit výměnu **zajatců**. **V** lepším případě **rukojmí** **po** zaplacení **výkupného** vrátili zpět jeho rodině. **Hůře** **ovšem** dopadli již výše zmínění **příborští** **měšťané** **Sarkandrovi**, jež se svým **potomkem** zřejmě již neshledali ("...syna mně vzali **a** **do** Turek prodali..."). [21] **Při** **vpádu** **Mansfeldových** **oddílů** **do** **Moravské** **Ostravy** **v** roce **1626** **zajali** vojáci (pravděpodobně **na** **popud** českých **emigrantů**) děkana **Martina** **Regulina**. Ten byl **posléze** vyměněn **za** **nekatolické** **kněží** **vězněné** **kardinálem** **Ditrichštejnem**. [22]

red = keywords

yellow = function words

blue = lexical words

Highlighting

Rovněž **na** **hukvaldském** panství jsou **známy** případy zajetí **či** přímo **únos** **civilních** obyvatel vojáky, jež tímto způsobem hodlali získat peníze **či** uskutečnit výměnu **zajatců**. **V** lepším případě **rukojmí** **po** zaplacení **výkupného** vrátili zpět jeho rodině. **Hůře** **ovšem** dopadli již výše zmínění **příborští** **měšťané** **Sarkandroví**, jež se svým **potomkem** zřejmě již neshledali ("...syna mně vzali **a** **do** Turek prodali..."). [21] **Při** **vpádu** **Mansfeldových** **oddílů** **do** **Moravské** **Ostravy** **v** roce **1626** **zajali** vojáci (pravděpodobně **na** **popud** českých **emigrantů**) děkana **Martina** **Regulina**. Ten byl **posléze** **vyměněn** **za** **nekatolické** **kněží** **vězněné** **kardinálem** **Ditrichštejnem**. [22]

red = keywords

yellow = function words

blue = lexical words

Also **on** the **Hukvaldy** **Estate** are **known** cases of capture or outright **kidnapping** of **civilians** by soldiers who intended to obtain money **or** exchange prisoners **in this way**. **At** best, the **hostage** was returned right away to his family **after** the ransom was paid. **However**, worse befell the above-mentioned **Sarkandar** [family], **Příbor** **citizens/burghers** who were apparently never reunited with their **offspring** ("...they took my son and sold him to the **Turks...**") [21]. **During** the **invasion** of **Mansfeld's** **troops** into **Moravská** **Ostrava** in **1626**, the soldiers (probably **at** the **instigation** of **Bohemian** **emigrants**) **captured** **Deacon** **Martin** **Regulin**. He was later **exchanged** **for** **non-Catholic** **priests** **imprisoned** by **Cardinal** **Ditrichštejn** [22].

Bottom-up Processing (like a computer)

Rovněž panství jsou případy
přímo obyvatel vojáky, jež tímto
způsobem hodlali získat peníze uskutečnit výměnu
. lepším případě zaplacení
vrátili zpět jeho rodině. dopadli
již výše zmínění, jež se
svým zřejmě již ("...syna mně
vzali prodali...").
roce
vojáci (pravděpodobně českých
) . Ten byl
vyměněn .

red = keywords

yellow = function words

blue = lexical words

Also the Estate are cases of
or outright by soldiers
who intended to obtain money exchange
in this way. best, the was returned right
away to his family the ransom was paid.
, befell the above-mentioned
[family], who
were apparently their
("...they took my son and sold him to the
...") the of
into in , the soldiers
(probably of Bohemian)
exchanged . He was
exchanged .

56% text coverage

Named entities and numbers are marked (or known)

Rovněž **hukvaldském** panství jsou **případy** **přímo** **obyvatel** vojáky, jež tímto způsobem hodlali získat peníze **uskutečnit** výměnu **lepším** případě **zaplacení** **vrátili** zpět jeho rodině. **dopadli** již výše zmínění **příborští** **Sarkandroví**, jež se svým **zřejmě** již ("...syna mně vzali **prodali...**"). [21] **Mansfeldových** **Moravské** **Ostravy** **roce** **1626** **vojáci** (pravděpodobně **českých**) **Martina** **Regulina**. Ten byl **vyměněn** **kardinálem** **Ditrichštejnem**. [22]

red = keywords

yellow = function words

blue = lexical words

Also **the Hukvaldy Estate** are **cases** of **outright** **by** soldiers who intended to obtain money **exchange** in this way. **best**, the **was** returned back to his family **the** ransom was paid. **befell** the above-mentioned **Sarkandar** [family], **Příbor** who were apparently never reunited with their **("...they** took my son and sold him to the **...")** [21]. **the** of **Mansfeld's** into **Moravská** **Ostrava** in **1626**, the soldiers (probably **of** Bohemian **Martin** **Regulin**. He was **exchanged** by **Cardinal** **Ditrichštejn** [22].

68% text coverage

Prepositions are guessable

Rovněž **na** hukvaldském panství jsou **prípady** **či** přímo **obyvatel** vojáky, jež tímto způsobem hodlali získat peníze **či** uskutečnit výměnu **v** lepším případě **po** zaplacení **vrátili** zpět jeho rodině. **dopadli** již výše zmínění **příborští** **Sarkandroví**, jež se svým **zřejmě** již ("...syna mně vzali **a do** prodali..."). [21] **Při** **Mansfeldových** **do** Moravské Ostravy **v** roce 1626 **vojáci** (pravděpodobně **na** českých) **Martina** Regulina. Ten byl vyměněn **za** **kardinálem** Ditrichštejnem. [22]

red = keywords

yellow = function words

blue = lexical words

Also **on** the Hukvaldy Estate are **cases** of **outright** **by** soldiers who intended to obtain money **or** exchange **in** this way. **At** best, the **was** returned back to his family **after** the ransom was paid. **befell** the above-mentioned **Sarkandar** [family], **Příbor** who were apparently never reunited with their **("...they took my son and sold him to the ...")** [21]. **During** the **of** Mansfeld's **into** Moravská Ostrava in 1626, the soldiers (probably **at** **of** Bohemian **Martin** Regulin. He was **exchanged** **for** **by** Cardinal Ditrichštejn [22].

79% text coverage

Cognates are guessable

Rovněž **na** hukvaldském panství jsou případy
či přímo **civilních** obyvatel vojáky, jež tímto
způsobem hodlali získat peníze **či** uskutečnit výměnu
. V lepším případě **po** zaplacení
vrátili zpět jeho rodině. **dopadli**
již výše zmínění **příborští** **Sarkandroví**, jež se
svým **zřejmě** již ("...syna mně
vzali **a do** prodali..."). [21] **Při**
Mansfeldových **do** **Moravské Ostravy** **v** roce
1626 vojáci (pravděpodobně **na** českých
emigrantů) **Martina** **Regulina**. Ten byl
vyměněn **za** **kardinálem**
Ditrichštejnem. [22]

red = keywords

yellow = function words

blue = lexical words

Also **on** the **Hukvaldy** Estate are cases of
or outright of civilians by soldiers
who intended to obtain money **or** exchange
in this way.. **At** best, the was returned back
to his family **after** the ransom was paid.
befell the above-mentioned **Sarkandar** [family],
Příbor who were apparently never
reunited with their ("...they took my son and
sold him to the ...") [21]. **During** the of
Mansfeld's into **Moravská Ostrava** in **1626**, the
soldiers (probably **at** of Bohemian
emigrants **Martin** **Regulin**. He was
exchanged **for**
by Cardinal Ditrichštejn [22].

81% text coverage

Also, I may already know some of these words

Rovněž **na** hukvaldském panství jsou **prípady** **či** přímo **civilních** obyvatel vojáky, jež tímto způsobem hodlali získat peníze **či** uskutečnit výměnu **vrátili**. **V** lepším případě **po** zaplacení **vrátili** zpět jeho rodině. **dopadli** již výše zmínění **příborští měšťané Sarkandroví**, jež se svým **zřejmě** již neshledali ("...syna mně vzali **a do** prodali..."). [21] **Při** **Mansfeldových oddílů do Moravské Ostravy v roce 1626** vojáci (pravděpodobně **na** českých emigrantů) **Martina Regulina**. Ten byl vyměněn **za** nekatolické **vězněné kardinálem Ditrichštejnem**. [22]

red = keywords

yellow = function words

blue = lexical words

Also **on** the Hukvaldy Estate are **cases** of **outright** **of** civilians by soldiers who intended to obtain money **or** exchange **was** returned back to his family **after** the ransom was paid. **befell** the above-mentioned **Sarkandar** [family], **Příbor citizens/burghers** who were apparently never reunited with their **("...they took my son and sold him to the ...")** [21]. **During** the **of** Mansfeld's **into Moravská Ostrava** in 1626, the soldiers (probably **at** **of** Bohemian emigrants **Martin Regulin**. He was **exchanged for** non-Catholic **imprisoned** by Cardinal Ditrichštejn [22].

85% text coverage

Words to learn

Rovněž **na** hukvaldském panství jsou **prípady** **či** přímo **civilních** obyvatel vojáky, jež tímto způsobem hodlali získat peníze **či** uskutečnit výměnu **vrátili**. **V** lepším případě **po** zaplacení **vrátili** zpět jeho rodině. **dopadli** již výše zmínění příborští měšťané Sarkandrovi, jež se svým **zřejmě** již neshledali ("...syna mně vzali **a do** prodali..."). [21] **Při** Mansfeldových oddílů **do** Moravské Ostravy **v** roce 1626 **vojáci** (pravděpodobně **na** českých emigrantů) **Martina** Regulina. Ten byl **vyměněn** **za** nekatolické **vězněné** kardinálem Ditrichštejnem. [22]

red = keywords

yellow = function words

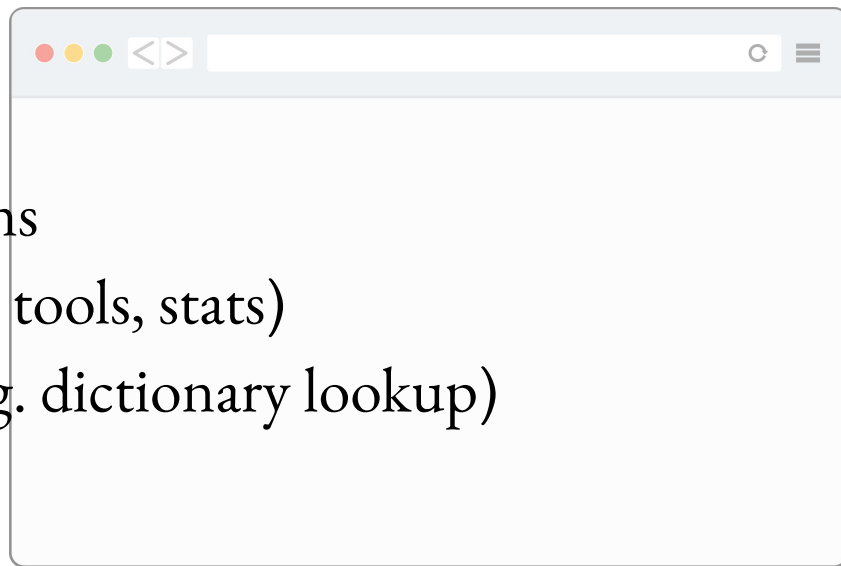
blue = lexical words

Also **on** the Hukvaldy Estate are **cases** of **or** outright **of** civilians by soldiers who intended to obtain money **or** exchange **in** this way. **At** best, the **was** returned back to his family **after** the ransom was paid. **,** **befell** the above-mentioned Sarkandar [family], Příbor citizens/burghers who were apparently never reunited with their **("...they took my son and sold him to the ...")** [21]. **During** the **of** Mansfeld's **into** Moravská Ostrava in 1626, the soldiers (probably **at** **of** Bohemian emigrants) **Martin** Regulin. He was **exchanged** **for** non-Catholic **imprisoned** by Cardinal Ditrichštejn [22].

Záměr

Build a browser extension to help with L2 reading

- user choice of L2 texts
- text authenticity
- visualization of linguistic patterns
- customization (e.g. filtering, UI, tools, stats)
- minimization of distractions (e.g. dictionary lookup)
- exportable word lists



Steps

- 1) Obtain corpus data
- 2) Build the tool
- 3) Use the tool
- 4) Improve the tool, go to 3
- 5) Share at eLex



Step 1: Obtain Corpus Data

actually... Step 0: Build CGSL

- Czech General Service List

- (Challis, 2022)

- Source Corpora

- SYN2020

(Křen et al., 2018)

- csTenTen17

(Suchomel, 2018)

- Koditex

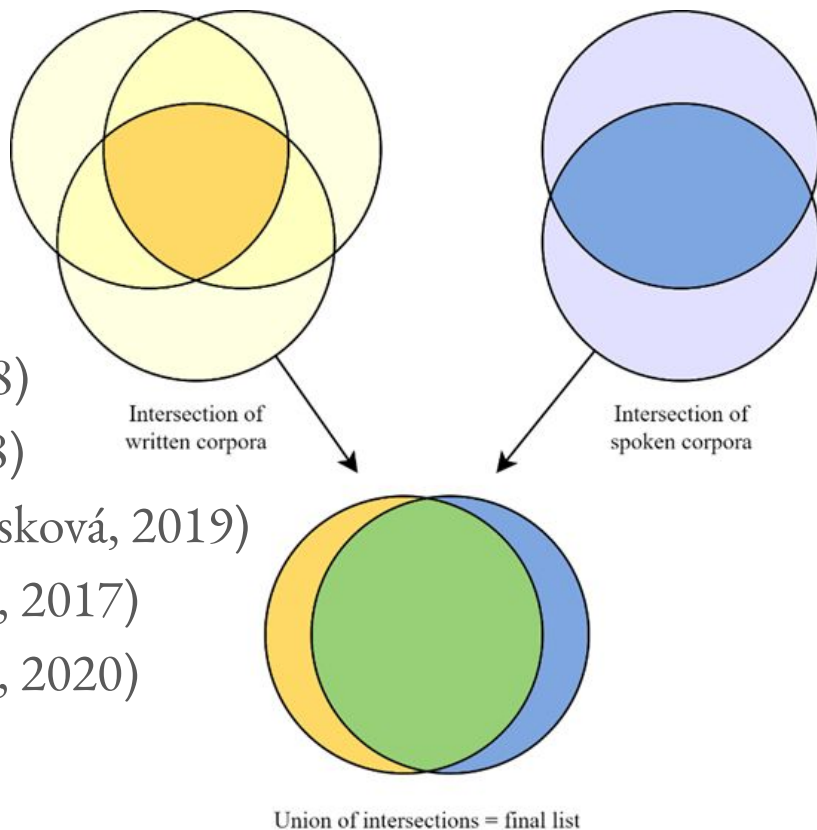
(Zasina & Komrsková, 2019)

- ORALv1

(Kopřivová et al., 2017)

- ORTOFONv2

(Kopřivová et al., 2020)



Step 1: Obtain Corpus Data

- CGSL (Challis, 2022)
 - ranks, lemmas, POS tags, modality
- Majka (Šmerk, 2007)
 - word forms
- Eufonometr (Plecháč, 2017)
 - IPA
- Wiktionary (wiktionary.org) ← work in progress
 - definitions, etymology, translations

Step 1: Obtain Corpus Data

CA. Příkazový řádek - majka -f majka.l-wt

```
D:\>chcp 1250
Active code page: 1250

D:\>majka -f majka.l-wt

učení
učení:k2eAgInSc1d1
učení:k2eAgInSc4d1
učení:k2eAgInSc5d1
učení:k2eAgMnSc1d1
učení:k2eAgMnSc5d1
učeních:k2eAgFnPc2d1
učeních:k2eAgFnPc6d1
učeních:k2eAgInPc2d1
učeních:k2eAgInPc6d1
učeních:k2eAgMnPc2d1
učeních:k2eAgMnPc6d1
učeních:k2eAgNnPc2d1
učeních:k2eAgNnPc6d1
učeníým:k2eAgFnPc3d1
učeníým:k2eAgInPc3d1
učeníým:k2eAgInSc7d1
```

Step 1: Obtain Corpus Data

Euphonometer carries out a phonetic transcription of the input text and computes the value of its euphony by means of the [method](#) proposed by Gabriel Altmann. [\[more » \]](#)

Žlutoučkový kůň pěl ďábelské ódy

eufonometr (1).txt – Poznámkový blok

Soubor Úpravy Formát Zobrazení Nápověda
| žlučoučki: ku:n pjeɪ ʒa:bɛɪlskɛ: o:di

Send

Analysis Type:

Output Format:

Transcription:

Step 1: Obtain Corpus Data

Czech [\[edit \]](#)

Pronunciation [\[edit \]](#)

- IPA^(key): [ˈlɪŋɡvɪsta]

Noun [\[edit \]](#)

lingvista *m anim* (feminine **lingvistka**)

1. linguist (scientist) [\[synonym ▲ \]](#)

Synonym: jazykovědec

Declension [\[edit \]](#)

Declension of <i>lingvista</i> (masculine animate in -a) [hide ▲]		
	singular	plural
nominative	lingvista	lingvisté, lingvisti
genitive	lingvisty	lingvistů
dative	lingvistovi	lingvistům
accusative	lingvistu	lingvisty
vocative	lingvisto	lingvisté, lingvisti
locative	lingvistovi	lingvistech
instrumental	lingvistou	lingvisty

Related terms [\[edit \]](#)

- lingvistický
- lingvistika

Further reading [\[edit \]](#)

- [lingvista](#) in *Kartotéka Novočeského lexikálního archivu*
 - [lingvista](#) in *Slovník spisovného jazyka českého, 1960–1971, 1989*
-

Step 2: Build the Extension

Google Chrome

- most used browser worldwide
- easy to port to other browsers

JavaScript, HTML, CSS



Extension icon and button

Settings page

UI overlay

- pop-up bubble
- side panel for lists
- word detail view
- statistics view

Step 3: Use the Tool

The image shows a browser window with the Lex-See extension settings page. The settings are organized into several sections:

- Highlighting:** A checkbox for automatic highlighting is unchecked. A text input field contains "být" and a slider is positioned between "být" and "antikoncepce".
- Highlighting rules:** A list contains one rule named "CGSL Rank" with edit, delete, and refresh icons. A link "add another rule" is below.
- Info Display:** A dropdown menu is set to "bubble".
- External services:** Checkboxes for "Příručka ÚJČ AV P", "Seznam Slovník", "Czech Wiktionary", and "Forvo" are shown, with the first two checked.
- Statistics & Export:** Checkboxes for "Mark ambiguous words* in statistics & export" and "Ignore case for Word FORMS" are checked.

Overlaid on the right is a preview of the "CGSL Rank" tool. It displays a list of words that meet specific criteria:

- are in CGSL, and
- have part-of-speech tag for noun, adjective or verb, and
- have 'written' modality marker, and
- only the 50% the most common of the selected

Below the list, there are controls for highlighting:

- A link "add another condition".
- Text: "shall be highlighted using the following" followed by a "color scale" dropdown.
- Text: "from this color for the most common words" followed by a color box and a slider from "transparent" to "opaque".
- Text: "to this color for the least common words" followed by a color box and a slider from "transparent" to "opaque".
- Text: "using" followed by a "logarithmic" dropdown, "scaling relative to" followed by a "CGSL" dropdown, and a period.

Step 3: Use the Tool

The screenshot shows a web browser window displaying the Wikisource page for the Universal Declaration of Human Rights in Czech. The page title is "Všeobecná deklaráce lidských práv". The browser's address bar shows the URL "cs.wikisource.org/wiki/Všeobecná_deklaráce_lidských_práv". The page content includes a navigation menu on the left, a main text area with the title "Všeobecná deklaráce lidských práv" and a sub-section "Úvod", and a metadata table on the right. A tooltip for the Lex-See tool is visible, indicating its function to highlight page content.

Všeobecná deklaráce lidských práv

Úvod

U vědomí toho,

že uznání přirozené důstojnosti a rovných a nezcizitelných práv členů lidské rodiny je základem svobody, spravedlnosti a míru ve světě,

že zneuznání lidských práv a pohrdání jimi vedlo k barbarským činům, urážejícím svědomí lidstva, a že vybudování světa, ve kterém lidé, zbavení strachu a nouze, se budou těšit svobodě projevu a přesvědčení, bylo prohlášeno za nejvyšší cíl lidu,

že je nutné, aby lidská práva byla chráněna zákonem, nemá-li být člověk donucen uchýlovat se, když vše ostatní selhalo, k odboji proti tyranii a útaku,

že je nutné podporovat rozvoj přátelských vztahů mezi národy,

že lid Spojených národů zdůraznil v Chartě znovu svou víru v základní lidská práva, v důstojnost a hodnotu lidské osobnosti, v rovná práva mužů i žen a že se rozhodl podporovat sociální pokrok a vytvořit lepší životní podmínky ve větší svobodě,

že členské státy převzaly závazek zajistit ve spolupráci s Organizací spojených národů všeobecné uznávání a zachovávání lidských práv a základních svobod a

že stejné chápání těchto práv a svobod má nesmírný význam pro dokonale splnění tohoto závazku.

Valné shromáždění

vyhlašuje tuto

Všeobecnou deklaraci lidských práv

jakožto společný cíl pro všechny národy a všechny státy za tím účelem, aby se každý jednotlivec a každý orgán společnosti, máje tuto deklaraci stále na mysli, snažil vyučováním a výchovou rozšířit úctu k těmto právům a svobodám a zajistit postupnými opatřeními vnitrostátními i mezinárodními jejich všeobecné a účinné uznávání a zachovávání jak mezi lidem členských států samých, tak i mezi lidem území, jež jsou pod jejich pravomocí.

Článek 1

Všichni lidé rodí se svobodní a sobě rovni co do důstojnosti a práv. Jsou nadáni rozumem a svědomím a mají spolu jednat v duchu bratrství.

Článek 2

Každý má všechna práva a všechny svobody, stanovené touto deklarací, bez jakéhokoli rozlišování, zejména podle rasy, barvy,

Step 3: Use the Tool

Všeobecná deklaráce lidských práv

Úvod

Ú vědomí toho, že uznání přirozené důstojnosti a rovných a nezcizitelných práv členů lidské rodiny je základem svobody, spravedlnosti a míru ve světě.

že zneuznání lidských práv a pohrdání jimi vedlo k barbarským činům, urážejícím svědomí lidstva, a že vybudování sví kterém lidé, zbavení strachu a nouze, se budou těšit svobodě projevu a přesvědčení, bylo prohlášeno za nejvyšší cíl lid že je nutné, aby lidská práva byla chráněna zákonem, nemá-li být člověk donucen uchylovat se, když vše ostatní selha odboji proti tyranii a útisku.

že je nutné podporovat rozvoj přátelských vztahů mezi národy.

že lid Spojených národů zdůraznil v Chartě znovní svou víru v základní lidská práva, v důstojnost a hodnotu lidské osobnosti, v rovná práva mužů i žen a že se rozhodl podporovat sociální pokrok a vytvořit lepší životní podmínky ve větší svobodě.

že členské státy převzaly závazek zajistit ve spolupráci s Organizací spojených národů všeobecné uznávání a zachovávání lidských práv a základních svobod a že stejně chápání těchto práv a svobod má nesmírný význam pro dokonale splnění tohoto závazku.

Vázné shromáždění vyhlásuje tuto Všeobecnou deklaraci lidských práv jakožto společný cíl pro všechny národy a všechny státy za tím účelem, aby se každý jednotlivec a každý orgán společnosti, máje tuto deklaraci stále na mysli, snažil vřučováním a výchovou rozšířit účtu k těmto právům a svobodám a zajistit postupnými opatřeními vnitrostátními i mezinárodními jejich všeobecné a účinné uznávání a zachovávání jak mezi lidem členských států, samých tak i mezi lidem území, jež jsou pod jejich pravomocí.

Článek 1 Všechni lidé rodí se svobodní a sobě rovni co do důstojnosti a práv. Usou nadání rozumem a svědomím a mají spolu jednat v duchu bratrství.

Článek 2 Všechna práva a všechny svobody, stanovené touto deklarací, bez jakéhokoliv rozlišování, zejména podle rasy, barvy,...

Údaje o textu	
Titulek:	Všeobecná deklaráce lidských práv
Zdroj:	osn.czif, unhcruz
Licence:	PD UN
OTRS ticket 2006091410007645	
Hlavní stránka v České Wikipedii	

#4184 spravedlnost noun Look up in Příručka ÚJČ AV

#4184 spravedlnost noun Look up in Příručka ÚJČ AV

Step 3: Use the Tool

The screenshot shows a web browser window with the URL `cs.wikisource.org/wiki/Všeobecná_deklaráce_lidských_práv`. The page title is "Všeobecná deklaráce lidských práv". The browser's address bar shows the URL. The page content is in Czech and includes a sidebar on the left with navigation links like "Hlavní strana", "Náhodný dílo", and "Wikizdroje". The main content area has a "Úvod" section with several paragraphs of text. A sidebar on the right, titled "#2089 rovny adjective", contains a search box with "equal" entered, a dropdown menu showing "Tricky for me (5)", and a list of items: "svědomí", "svoboda", "práva", "spolupráci", and "rovni". The "Tricky for me" section has a list of options: "Left mouse button click" (checked), "with Ctrl key", "with Shift key", and "with Alt key". There are also buttons for "Export word list" and "Delete word list".

Všeobecná deklaráce lidských práv

Úvod

Ú vědomí toho.

Ze uznání přirozené důstojnosti a rovných a nezcizitelných práv těmto lidské rodiny je zakladem svobody, spravedlnosti a míru ve světě.

Ze zneuznání lidských práv a pohrdání jimi vedlo k barbarským činům, urážejícím svědomí lidstva, a že vybudování světa ve kterém lidé, zbavení strachu a nouze, se budou těšit svobodě projevu a přesvědčení, bylo prohlášeno za nejvyšší cíl lidí.

Ze je nutné, aby lidská práva byla chráněna zákonem, nemá-li být člověk donucen uchýlovat se, když vše ostatní selhalo, k odboji proti tyranii a útisku.

Ze je nutné podporovat rozvoj přátelských vztahů mezi národy.

Ze lid Spojených národů zdůraznil v Chartě znovu svou víru v zakladní lidská práva, v důstojnost a hodnotu lidské osobnosti, v rovná práva mužů i žen a že se rozhodl podporovat sociální pokrok a vytvořit lepší životní podmínky ve větší svobodě.

Ze členské státy převzaly závazek zajistit ve spolupráci s Organizací spojených národů všeobecné uznávání a zachovávaní lidských práv a základních svobod a

Ze stejně chápání těchto práv a svobod má nesmírný význam pro dokonale splnění tohoto závazku.

Valné shromáždění

vyhlašuje tuto

Všeobecnou deklaraci lidských práv

jakožto společný cíl pro všechny národy a všechny státy za tím účelem, aby se každý jednotlivec a každý orgán společnosti, máje tuto deklaraci stále na mysli, snažil vylučováním a výchovou rozšířit účtu k těmto právům a svobodám a zajistit postupnými opatřeními vnitrostátními i mezinárodními jejich všeobecné a účinné uznávání a zachovávaní jak mezi lidem členských států, samých tak i mezi lidem území, jež jsou pod jejich pravomocí.

Článek 1

Všichni lidé rodi se svobodní a sobě rovni co do důstojnosti a práv. Usou nadání rozumem a svědomím a mají spolu jednat v duchu bratrství.

Článek 2

Každý má všechna práva a všechny svobody, stanovené touto deklarací, bez jakéhokoli rozlišování, zejména podle rasy, barvy,

Step 3: Use the Tool

Všeobecná deklaráce lidských práv

cs.wikisource.org/wiki/Všeobecná_deklaráce_lidských_práv

WIKIZDROJE

rovný

CGSL rank: 2089
CGSL modality: common
Part of Speech: adjective
Is not on any word list.

IPA: [rovni:]

The closest CGSL entries by phonetic distance are:

rodný	[rodní:]	distance 1
ropný	[ropní:]	distance 1
nový	[novi:]	distance 2
možný	[možni:]	distance 2
hrozný	[hrozni:]	distance 2
volný	[volni:]	distance 2
divný	[divní:]	distance 2
levný	[levní:]	distance 2
hodný	[hodní:]	distance 2
drobný	[drobni:]	distance 2
rodinný	[rodni:]	distance 2
pevný	[pevní:]	distance 2
povinný	[povíni:]	distance 2
mocný	[motsní:]	distance 2
roční	[roťní:]	distance 2
rovnat	[rovnat]	distance 2
raný	[rani:]	distance 2
nosný	[nosní:]	distance 2
hojný	[hojní:]	distance 2
rušný	[rušní:]	distance 2
rovno	[rovno]	distance 2

In the highlighted page, there are the following occurrences (5):

<i>(...) že uznání přirozené důstojnosti a (...)</i>	rovných	<i>a nezczitelných práv členů lidské (...)</i>
<i>(...) a hodnotu lidské osobnosti, v (...)</i>	rovná	<i>práva mužů i žen a (...)</i>
<i>(...) rodí se svobodní a sobě (...)</i>	rovní	<i>co do důstojnosti a práv (...)</i>
<i>(...) Každý má právo vstoupit za (...)</i>	rovných	<i>podmínek do veřejných služeb své (...)</i>
<i>(...) obdobích na základě všeobecného a (...)</i>	rovného	<i>hlasovacího práva tajným hlasováním nebo (...)</i>

In the highlighted page, the distribution of the forms of the word is as follows:

rovných (2)	
rovní (1)	
rovného (1)	
rovná (1)	
nerovným (0)	
rovnou (0)	
rovně (0)	
rovnějšími (0)	
nerovněm (0)	
nerovnějším (0)	
nejnerovnějším (0)	
nejnerovnějšíma (0)	
nejnerovnější (0)	
nerovně (0)	
nejnerovnějších (0)	

Článek 2

Každý má všechna práva a všechny svobody, stanovené touto deklarací, bez jakéhokolí rozlišování, zejména podle rasy, barvy,

Step 3: Use the Tool

Všeobecná deklaráce lidských práv

cs.wikisource.org/wiki/Všeobecná_deklaráce_lidských_práv

Wikizdroje

Titul Diskuse

Číst Zobrazit zdroj

Lemma	Part of Speech	Rank	Count	Forms in Text
zavít	verb	763	2	zavít
váš	pronoun	127	1	vaše
příspěvek	noun	893	2	příspěvky
udržovat	verb	2309	1	udržují
projekt	noun	614	2	projekty, projektech
nadace	noun	7427	1	nadace
v	preposition	5	40	v, ve
chod	noun	4449	1	chodu
stáhnout	verb	1290	4	stáhnout
všeobecný	adjective	3691	7	všeobecná, všeobecné, všeobecnou, všeobecného
lidský	adjective	610	14	lidských, lidské, lidská, nelidskému, lidským
právo	noun	648	61	práv, práva, právům, právo, právech
z	preposition	14	9	Z, ze
volně	adverb	3345	3	volně
dostupný	adjective	1127	2	dostupné, dostupný
knihovna	noun	1843	1	knihovny
skočit	verb	1005	2	skočit
na	preposition	6	38	na
navigace	noun	7713	2	navigaci, navigace
vyhledávání	noun	3743	1	vyhledávání

Reset view Export as-is to TSV file Export as-is to PDF file

1547 forms of (up to) 573 items.

Článek 2

Každý má všechna práva a všechny svobody, stanovené touto deklarací, bez jakéhokoli rozlišování, zejména podle rasy, barvy,

Step 3: Use the Tool

Step 3: Use the Tool



Step 4: Improve the Tool

1) everything highlighted

práchno — troud
připučit — přimáčknout
položnica — šestinedělka
suchotnica — vřes
světadlo — buková louč

Figure 4. Screenshot from a glossary of archaisms.

Step 4: Improve the Tool

2) understand better how phonetic data helps

- pařez – Paříž [tree stump – Paris]
- chromý – chrom [lame – chromium]
- svléct – lézt [undress, take off – crawl, climb]
- huba – houba [mouth – mushroom]

Step 4: Improve the Tool

- 3) increase consistency in underlying data (e.g. POS tagging, lemmatization)
- 4) Wiktionary data
- 5) isolation of UI components from page
- 6) refine tokenization criteria
- 7) better internal data structures, real-time morphological analyzer?
- 8) add more languages?

References

- Adelman, J. S., Marquis, S. J., & Sabatos de Vito, S. G. (2010). Letters in words are read simultaneously, not left-to-right. *Psychological Science*, 21(12), pp. 1799–1801. doi: 10.1177/0956797610387442
- Agnés, E. (2015). *Vocabulary size and type goals in advanced EFL and ESL classrooms. A review of research on lexical threshold, lexical coverage, reading and listening comprehension.*
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multi-trace memory model of polysyllabic word reading. *Psychological Review*, 105, pp. 678–723. doi: 10.1037/0033-295X.105.4.678-723
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and linguistic computing*, 7(1), pp. 1–16.
- Awadh, F. H. R., Phénix, T., Antzaka, A., Lallier, M., Carreiras, M., & Valdois, S. (2016). Cross-language modulation of visual attention span: An Arabic-French-Spanish comparison in skilled adult readers. *Frontiers in Psychology*, 7, p. 307. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4779959/>. doi: 10.3389/fpsyg.2016.00307
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1).
- Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), pp. 243–257.
- Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, 104(2), pp. 198–230. doi: 10.1016/j.cognition.2006.05.009
- Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. *Journal of Research in Reading*, 32(2), pp. 230–253. doi: 10.1111/j.1467-9817.2008.01387.x
- Brantmeier, C. (2002). Second language reading strategy research at the secondary and university levels: Variations, disparities, and generalizability. *The Reading Matrix*, 2(3).
- Brezina, V., & Gablasova, D. (2015). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), pp. 1–22.
- Burch, B., Egbert, J., & Biber, D. (2016). Measuring and interpreting lexical dispersion in corpus linguistics. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 3(2), pp. 189–216.
- Čapek, K. (1935). *Dášeňka čili život štěněte.*
- Cirino, P. T., Barnes, M. A., Roberts, G., Miciak, J., & Gioia, A. (2022). Visual attention and reading: A test of their relation across paradigms. *Journal of Experimental Child Psychology*, p. 214, 105289.
- Challis, K. (2022). *Is there a core vocabulary for Czech? Introducing the Czech General Service List.* doi:10.13140/RG.2.2.17678.02889.
- Chapelle, C. (2003). *English language learning and technology.*
- Czerepowicka, M. (2021). The structure of a dictionary entry and grammatical properties of multi-word units. *Electronic lexicography in the 21st century (eLex 2021) Post-editing lexicography*, p. 79.
- Četyna, B. (1958). *Valchař se směje aneb tulaníci a pozorníci.* Krajské nakladatelství v Ostravě.
- Čermák, F., & Kren, M. (2011). *A frequency dictionary of Czech: core vocabulary for learners.* Routledge.
- Davies, A. (2005). *An introduction to applied linguistics.*
- Davies, M., & Gardner, D. (2010). *Word frequency list of American English.*
- Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and instruction*, 17(1), pp. 78–88.
- Egbert, J., Biber, D., & Gray, B. (2022). *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness.* Cambridge University Press.
- Facchetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., ... & Massetti, G. G. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive neuropsychology*, 23(6), pp. 841–855.
- Frey, A., & Bosse, M. L. (2018). Perceptual span, visual span, and visual attention span: Three potential ways to quantify limits on visual processing during reading. *Visual Cognition*, 26(6), pp. 412–429.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied linguistics*, 28(2), pp. 241–265.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied linguistics*, 35(3), pp. 305–327.
- Garnier, M., & Schmitt, N. (2015). The PhAVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6), pp. 645–666.
- Gori, S., & Facchetti, A. (2014). Perceptual learning as a possible new approach for remediation and prevention of developmental dyslexia. *Vision research*, 99, pp. 78–87.
- Gori, S., Cecchini, P., Bigoni, A., Molteni, M., & Facchetti, A. (2014). Magnocellular-dorsal pathway and sub-lexical route in developmental dyslexia. *Frontiers in human neuroscience*, 8, p. 460.
- Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A vision of reading. *Trends in Cognitive Sciences*, 20(3), pp. 171–179. doi: 10.1016/j.tics.2015.12.008
- Gray, B., & Egbert, J. (2019). Register and register variation. *Register Studies*, 1(1), pp. 1–9.
- Hajič, J., Votrúbec, J., Krbec, P., & Květoň, P. (2007, June). The best of two worlds: Cooperation of statistical and rule-based taggers for Czech. In *Proceedings of the workshop on Balto-Slavonic natural language processing*, pp. 67–74.
- Hu, M., & Nation, I.S.P. (2000) Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), pp. 403–430.

References continued

- Jelínek, T. (2008). Nové značkování v Českém národním korpusu. *Naše řeč*, (1), pp. 13–20. [New Tagging in the Czech National Corpus].
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, 53(3), pp. 83–93.
- Karlík, P. (2012). *Přiruční mluvnice češtiny*. Lidové noviny. [Handbook of Czech]
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, pp. 7–36.
- Kopřivová, M., Lukeš, D., Komrsková, Z., Poukarová, P., Waclawicová, M., Benešová, L., Křen, M. (2017). *ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017*. Ústav Českého národního korpusu FF UK. Praha. Accessible at <http://www.korpus.cz> [ORAL: A Corpus of Informal Spoken Czech]
- Kopřivová, M., Komrsková, Z., Lukeš, D., & Poukarová, P. (2017). Korpus ORAL: sestavení, lematizace a morfoloické značkování. *Korpus–gramatika–axiologie*, 15, 47–67. [The ORAL Corpus: Assembly, Lemmatization and Morphological Tagging]
- Korobov, M. (2015). Morphological analyzer and generator for Russian and Ukrainian languages. In *International conference on analysis of images, social networks and texts*, pp. 320–332. Springer, Cham.
- Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., ... & Usal, K. A. (2022). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, pp. 1–35.
- Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *Tesol Quarterly*, 47(4), pp. 867–872.
- Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal*, 104(2), pp. 345–362.
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for academic purposes*, 22, pp. 42–53.
- Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, 33, pp. 70–98.
- Lobier, M., Dubois, M., & Valdois, S. (2013). The role of visual processing speed in reading speed development. *PLoS ONE*, 8, p. 4. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058097>
- Lobier, M., Peyrin, C., Le Bas, J. F., & Valdois, S. (2012). Pre-orthographic character string processing and parietal cortex: A role for visual attention in reading? *Neuropsychologia*, 50(9), pp. 2195–2204. doi: 10.1016/j.neuropsychologia.2012.05.023
- Lorusso, M. L., Facoetti, A., & Bakker, D. J. (2011). Neuropsychological treatment of dyslexia: does type of treatment matter?. *Journal of learning disabilities*, 44(2), pp. 136–149.
- Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, pp. 1–13.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of educational psychology*, 93(1), p. 187.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian modern language review*, 63(1), pp. 59–82.
- Plecháč, P. (2017). *Euphonometer 2.0*. Prague: Institute of Czech Literature, CAS. Available at: <http://versologie.cz>.
- Rowling, J. K. (1999). *Harry Potter and the Sorcerer's Stone*. Scholastic.
- Sanosi, A. B. (2018). The effect of Quizlet on vocabulary acquisition. *Asian Journal of Education and e-learning*, 6(4).
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and instruction*, 20(2), pp. 100–110.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Song, M. J. (1999). Reading strategies and second language reading ability: The magnitude of the relationship. *English Teaching*, 54(3), pp. 73–95.
- Straková, J., Straka, M., & Hajic, J. (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.
- Suchomel, V. (2018). csTenTen17, a Recent Czech Web Corpus. In *RASLAN*, pp. 111–123.
- Šmerk, P. (2007). Fast Morphological Analysis of Czech. In Petr Sojka and Aleš Horák *Proceedings of Third Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2009*. Brno: Masaryk University, 2007. pp. 13–16. ISBN 978-80-210-5048-8. Available at: <https://nlp.fi.muni.cz/ma/>
- Těšitelová, M. (1987). *O češtině v číslech*. Academia.
- Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied Linguistics*, 34(4), pp. 457–479.
- Verhallen, M. J., & Bus, A. G. (2011). Young second language learners' visual attention to illustrations in storybooks. *Journal of Early Childhood Literacy*, 11(4), pp. 480–500.
- Wiktionary: The free dictionary*. Accessed at: <https://wiktionary.org>.
- Wright, B. A. (2016). Transforming vocabulary learning with Quizlet. *Transformation in language education*. Tokyo: JALT, pp. 436–440.
- Ziková, M. (2017). Suppletivismus in Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*. Accessed at: <https://www.czechency.org/slovník/SUPLETIVISMUS>



kchallis@iastate.edu
t.drusa@gmail.com