

The Czechoslovak Word of the Week. Re-joining Czechs and Slovaks together in an example of invisible lexicography work

Peter Malčovský¹, Michal Škrabal², Vladimír Benko¹, Jan Kocek²

¹ Slovak Academy of Sciences, E. Štúr Institute of Linguistics,
Panská 26, 811 01 Bratislava, Slovakia

² Charles University in Prague, Institute of the Czech National Corpus,
Panská 7, 110 00 Praha 1, Czech Republic

E-mail: peter.malcovsky@juls.savba.sk, michal.skrabal@ff.cuni.cz,
vladimir.benko@juls.savba.sk, jan.kocek@ff.cuni.cz

Abstract

The Czecho-Slovak Word of the Week is a joint popularization project of Czech and Slovak linguists. Throughout the year, each and every week, we are publishing a new entry on the website <https://slovo.juls.savba.sk>, written parallelly in Czech and Slovak, the central part being a language feuilleton supplemented with data drawn from language corpora and quotations from contemporary and historical monolingual and translation dictionaries. In a way, we see the website as a dictionary, with a fixed macrostructure of 52 weekly published entries, and a microstructure, determined by the order of the individual components. Thus, our project could be considered a good example of “invisible lexicography” in practice. The target audience is presented with various kinds of lexicographic information unobtrusively, covertly, and invisibly, usually not even feeling that they are “leafing through” a dictionary. At this year’s eLex, we plan to present not only the website but also the database behind it. Our solution uses modern web technologies: the *JHipster* application generator in combination with the *Vue* front-end framework and the *PostgreSQL* database. The application allows the administrator to easily enter content, including importing and formatting texts from various sources, and to use audio samples from spoken corpora as well.

Keywords: Czech; Slovak; *JHipster* application generator; *Vue* front-end framework; *Word at Glance* interface; *PostgreSQL* database

1. Introduction

The Czecho-Slovak Word of the Week is a joint year-long popularization project of the Institute of the Czech National Corpus and the E. Štúr Institute of Linguistics of the Slovak Academy of Sciences, that was inaugurated on the occasion of the 30th anniversary of the dissolution of Czechoslovakia (January 1, 1993). Throughout the year, each and every week, a new entry, written in parallel in Czech and Slovak, is published on the project website (<https://slovo.juls.savba.sk>). We intend to draw the

attention of both the Czech and the Slovak publics (especially the younger generation, for whom the former mutual intelligibility between the two languages no longer holds) to the interesting parallels, but chiefly the differences, between our two languages. We try to do so in a user-friendly and entertaining way, the central part of each entry being a language feuilleton (a very popular genre in Czechia and Slovakia), supplemented with data drawn from language corpora (SYN2015, SYN2020, and ORAL v1 for Czech; prim-10.0-public-all and s-hovor-7.0 for Slovak) and the respective entries from some older monolingual and bilingual dictionaries (Bernolák 1825, Jungmann 1835-1839, SSJČ 1960-1971, SSJ 1959-1968, KSSJ 2003, ČSS 1981, SČS 1967). In a way, we see the website as being a dictionary with a fixed macrostructure (52 entries including some multi-word units¹) and a microstructure determined by the order of the individual components (described in Škrabal & Benko, 2019: 475-476). Thus, our project could be considered a good example of “invisible lexicography” in practice. The target audience is presented with various lexicographic information – be it frequency statistics for various text types, examples from both written and spoken corpora, or quotes from older dictionaries – unobtrusively, covertly, and “invisibly”, usually without them having the feeling that they are “leafing through” a dictionary.

At this year’s eLex, we intend to present not only the website but also the database behind it within the demo section. Our solution uses modern web technologies: the *JHipster* application generator (<https://www.jhipster.tech/>) in combination with the *Vue* front-end framework (<https://vuejs.org/>) and the *PostgreSQL* database (<https://www.postgresql.org/>). The application allows the administrator to easily enter content, including importing and formatting texts from various sources (dictionary portals, Word documents, etc.), and to use audio samples from spoken corpora as well. The website itself is graphically based on the *Word at Glance* interface (Machálek, 2019, 2020), as the original layout was adapted to the needs of our project.

In this paper, we want to focus mainly on the technical background of the whole project. In the following chapter, both the backend and the frontend will be described as well as the specific work with the database, i.e., the way to add a new entry to it. Other aspects of the project (project team setup², workflow, promotion, etc.) are left aside on purpose, as we plan to devote a separate article to them after the project is finished (December 2023).

2. Technical overview

We had several options for the technical implementation of the planned project website.

¹ The choice of the list of entries was more or less random and influenced by our personal preferences.

² Our team is largely made up of external writers of feuilletons, mostly linguists. Their texts are edited, proofread, and supplemented with information from corpora and dictionaries, for each entry has the same microstructure. In addition, a programmer and a graphic designer were necessary for the successful implementation of the project.

We rejected the simplest solution: static HTML pages, mainly because the content could be filled and changed only by a technician. A reasonable option was also to modify the existing *Word at Glance* website (<https://www.korpus.cz/slovo-v-kostce/>), which is visually based on a similar tile system. Alternatively, an established content management system, such as *WordPress*, could be used too. Considering various factors (e.g., technical limitations of our infrastructure with respect to *WordPress*, the original estimate of the scope – a word for each day, thus, up to 365 episodes of the series which would require massive automation³), we decided to base our own solution on the technology with which we already had experience. Naturally, this approach has its pros (control over every aspect of the website, no need to limit to existing templates, pay for external hosting, etc.) and cons (more overall work, inability to get “free” features for the blog that are common in established systems). An important requirement for our website – after taking the broad target audience, promotion via social networks, and overall trends into account – was to be mobile-friendly (see also Fig. 11 below).

2.1 *JHipster* application generator

Our web application was generated by *JHipster* generator. *JHipster* is a development platform that can quickly generate, develop, and deploy modern web applications and microservice architectures. The generator has been around since 2013 and is well established and popular. It generates *Spring Boot* based Java web server along with web frontend application (*Angular*, *React* or *Vue* based). Generating an application skeleton using *JHipster* is simple and straightforward, requiring only the answering of a dozen questions (application name, monolith or microservices application, database type and brand, etc.)

³ This original idea was abandoned during the preparations, as it turned out that the daily frequency would require disproportionately more time and energy, as well as a larger team, which was not possible due to the limited budget. Furthermore, we supposed that a daily periodicity would not have been beneficial in maintaining the interest of the audience; in fact, it might have had quite the opposite effect.

```
Administrator: Windows PowerShell
PS C:\work\csl_slovo> jhipster
INFO! Using JHipster version installed globally

JHIPSTER

https://www.jhipster.tech

Welcome to JHipster v7.4.1
Application files will be generated in folder: C:\work\csl_slovo

Documentation for creating an application is at https://www.jhipster.tech/creating-an-app/
If you find JHipster useful, consider sponsoring the project at https://opencollective.com/generator-jhipster

> Which *type* of application would you like to create? Monolithic application (recommended for simple projects)
> What is the base name of your application? csl_slovo
> Do you want to make it reactive with Spring WebFlux? No
> What is your default Java package name? com.juls.cs1slovo
> Which *type* of authentication would you like to use? JWT authentication (stateless, with a token)
> Which *type* of database would you like to use? SQL (H2, PostgreSQL, MySQL, MariaDB, Oracle, MSSQL)
> Which *production* database would you like to use? PostgreSQL
> Which *development* database would you like to use? PostgreSQL
> Which cache do you want to use? (Spring cache abstraction) Ehcache (local cache, for a single node)
> Do you want to use Hibernate 2nd level cache? Yes
> Would you like to use Maven or Gradle for building the backend? Maven
> Do you want to use the JHipster Registry to configure, monitor and scale your application? No
> Which other technologies would you like to use?
> Which *Framework* would you like to use for the client? Vue
> Do you want to generate the admin UI? Yes
> Would you like to use a Bootswatch theme (https://bootswatch.com/)? Cerulean
> Choose a Bootswatch variant navbar theme (https://bootswatch.com/)? Primary
> Would you like to enable internationalization support? Yes
> Please choose the native language of the application English
> Please choose additional languages to install Slovak
> Besides JUnit and Jest, which testing frameworks would you like to use?
> Would you like to install other generators from the JHipster Marketplace? No
```

Figure 1: *JHipster* questionnaire

Immediately after answering these questions, the generator creates the first version of an application, both backend server and frontend web. The application is already executable at this point, obviously, with no business logic yet.

Features list includes:

- User management: frontend & backend for creating and editing users with roles (Admin, User);
- Metrics: a smart console for displaying runtime characteristics of the running server (memory, CPU load, number of server threads, number of requests and their result codes);
- Health page, Configuration page, Logs settings page: further server diagnostics and settings.

2.2 Modelling and generating entities

In this step we populated the web application with data. *JHipster* comes with a handy tool – *JDL-Studio* – where data entities can be modelled and visualized, along with relations between the entities.

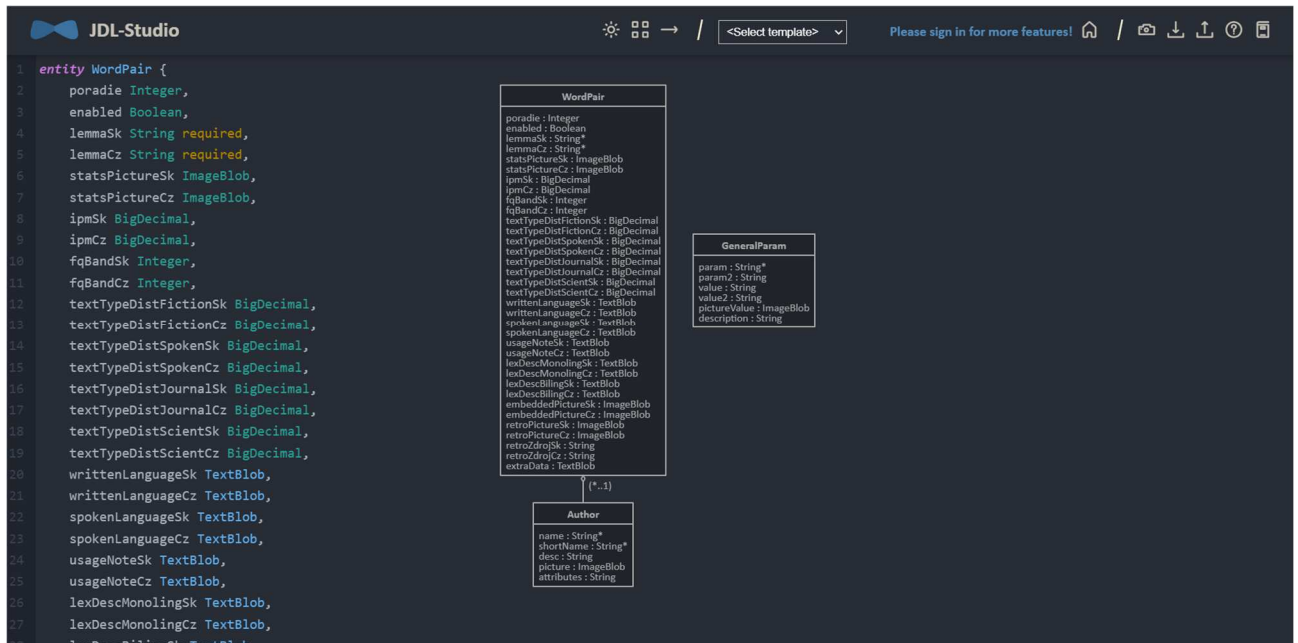


Figure 2: JDL-Studio tool

Three entities were created:

- The **WordPair** entity is a crucial entity, storing all the data necessary to display language posts. All entries have separate Czech (CZ) and Slovak (SK) items and can be mapped in a straightforward way to the user interface. Every record in the WordPair entity is used for rendering just one post (the entire “Word of the Week” article). The entity consists of text items (LemmaCZ, LemmaSK, etc.), long text items – TextBlob (UsageNoteCZ, UsageNoteSK, etc.), and image items – ImageBlob (StatsPictureCZ, StatsPictureSK), along with special items like Order (validity date for a post in numeric format), Enabled flag, ID (unique numeric value for each record), etc.
- The **Author** entity stores information data about authors of feuilletons, namely: author’s name, photograph, and short biography. Each record in the WordPair entity is linked with two records in the Author entity: AuthorCZ and AuthorSK.
- The **GeneralParam** entity stores general-purpose parameters and data for the web application, e.g., various text templates.

Once the entities scripts are ready, we include them in the application, using *JHipster* command. This results in actual database tables being created on the backend (*Liquibase* scripts to create the tables). Besides, we get Java objects representing the entities and Repository and Resource Java Beans to access and manipulate the entities via *JPA* framework. Lastly, CRUD operations (Create, Read, Update, Delete) are completely implemented for all our entities, both on the back- and frontend side. REST API endpoints are created on the server so that the *Vue* frontend can access them.

Populated user interface for entities editing looks like this:

ID	PORADIE	ENABLED	LEMMA SK	LEMMA CZ	AUTHOR SK	AUTHOR CZ	
1	20230109	true	řava	velbloud	Martina Blažeková	Michal Škrabal	View Edit Delete
2	20230220	true	omrvinky	drobky	Martina Blažeková	Michal Škrabal	View Edit Delete
1101	20230123	true	vankuš	poľstár	Peter Malčovský	Gabriel Pleska	View Edit Delete
1251	20230130	true	fúrik	kolečko	Martina Blažeková	Jan Nejedlý	View Edit Delete
1351	20230515	true	zimomriavky	husi kúže	Zuzana Klučárová	Gabriel Pleska	View Edit Delete
1451	20230424	true	pery	rtý	Natália Kolenčíková	Gabriel Pleska	View Edit Delete

Figure 3: Word Pairs entity





ID	NAME	SHORT NAME	DESC	PICTURE	ATTRIBUTES	
1001	Michal Škrabal	mis	Český lexikograf a korpusev lingvista, toho času filatelista a reditel Ústavu Českého národního korpusu.	 image/jpeg, 7599 bytes	FK	View Edit Delete
1002	Martina Blažeková	martina	Československá komparatistka, doktorandka FF UK a autorka jednej básnickej zbierky.	 image/jpeg, 16694 bytes	F	View Edit Delete
1151	Gabriel Pleska	gabriel	Edituje a píše pro Peníze.cz, Finmag a Heroine. Sbírá neúžitečné informace a roste jak dříví v lese. Jen stromy přes zimu přirůstají pomaleji.	 image/jpeg, 12149 bytes	F	View Edit Delete
1201	Peter Malčovský	malco	Odborný pracovník, Jazykovedný ústav L. Štúra SAV. Moderátor galérie www.instagram.com/dedina_bnw/	 image/jpeg, 19490 bytes	FC	View Edit Delete

Figure 4: Authors entity

Česko(slovenské slovo
týdne / týždňa

Domů / Domov Archiv V médiích / médiách Info Entities Administration


Create or edit a Author

ID
1001

Name
Michal Škrabal

Short Name
mis

Desc
Český lexikograf a korpusový lingvista, toho času ředitel Ústavu Českého národního korpusu.

Picture

 image/jpeg, 7 599 bytes ×
 Choose File No file chosen

Attributes
FK

Cancel Save

Figure 5: Create/edit an author form

2.3 Creating an application homepage and other pages

With the application skeleton generated by *JHipster*, one can finalize the application by the manual creation of the homepage and other pages. Obviously, this part took most of the entirety of the development time. We have created:

- Home page – word pairs “posts” viewer with navigation (previous week, next week);
- About page – information about the project purpose and team, contacts, and credits;
- Media page – memorable promo actions for the project in various media (TV, radio, blogs);
- Archive page – a timeline with all the published posts.

2.4 Technology stack overview

2.4.1 PostgreSQL

PostgreSQL is a powerful, open-source object-relational database system with over 35 years of active development that has earned a strong reputation for reliability, feature robustness, and performance. Nowadays, the *PostgreSQL* is used even in enterprise

solutions, competing with legendary systems like *Oracle DB*. It also incorporates full search features, including stemming (Czech language stemmer is available, Slovak language stemmer not yet), removing stop words during search.

2.4.2 JPA (Spring Data)

JPA stands for Jakarta Persistence API, a set of concepts for Java objects persistence and object-relational mapping (ORM). On our server, this layer is used for accessing the physical database. *JPA* allows various conceptual approaches to handle data in the application (Code First, Model First, Database First). On a practical level, Repository objects are created in our server for DB access purposes. Each entity has its own Repository object. The framework tries to help with DB queries as much as possible – for trivial ones like *findById*, a coder does not need to write any code. Simple queries can be written just by query method name (e.g., *findAllByNameLike*), *JPQL* database-agnostic query language, or native DB query. Combined with pagination and ordering support, it is not complicated to create backend queries for various frontend grids.

Code examples:

Query used to pick current Word of the Week record, given current system date as parameter. The *JPA* translates method name to actual query by itself:

```
// current
public List<WordPair> findTop1ByEnabledTrueAndPoradieLessThanEqualOrderByPoradieDesc(Integer
currDateNumber);
```

Figure 6: Query used to pick current Word of the Week record, given current system date as a parameter

```
//findAllTiny
@Query(
    "Select new com.peto.wotd.service.dto.WordPairTinyDTO(w.id, w.poradie, w.enabled,
w.lemmaSk, w.lemmaCz, ask.name, acz.name, ask.shortName, acz.shortName)" +
    " from WordPair w " +
    " left join Author ask on w.authorSk.id = ask.id " +
    " left join Author acz on w.authorCZ.id = acz.id "
)
public Page<WordPairTinyDTO> findAllTiny(Pageable pageable);
```

Figure 7: A more complex query for obtaining all word pairs list, using reduced DTO object for effective transfer. The query gets pagination settings from UI as a parameter (e.g., page 3, ordered by lemmaCZ).

2.4.3 *Vue* frontend framework

Vue is an approachable, performant, and versatile framework for building web user interfaces. Introduced in 2014, it has gained popularity and user base since then. Given the tremendous development rate in this web frontend area, we can look at the *Vue* as “just another web framework”. Nevertheless, the *Vue* belongs to state-of-art ones as of 2023.

From a developer’s point of view, *Vue* is similar to the *React* framework, yet improved and simplified in many ways. Unlike in *React*, *Vue* comes with handy HTML tags for if-then-else constructs, loop constructs, etc. so there is no need to combine HTML code with JavaScript code, producing a hard-to-read, hard-to-maintain mess.

Code examples:

```
<div v-if="isMobile" class="col-md-12" style="padding-left: 0; padding-right: 0">
  <h4 class="centered">
    <a @click="navigatePrev()"><font-awesome-icon icon="chevron-circle-left" size="1x" /></a>
    {{ currDate }}
    <a @click="navigateNext()"><font-awesome-icon icon="chevron-circle-right" size="1x" /></a>
  </h4>
</div>
<div v-else class="col-md-12">
  <h2 class="centered">
    <v-tooltip top>
```

Figure 8: Usage of *v-if* and *v-else* *Vue* tags to render different content for mobile and desktop web

```
<v-timeline :dense="isMobile">
  <v-timeline-item v-for="item in items" :key="item.id" color="#c7c7c7" small>
    <a v-bind:href="'/' + item.id" style="color: #0e5a9d">
      {{ getDateFormattedFromPoradie(item.poradie) }} <br />
    </a>
    <h5 class="centered" style="margin-bottom: 0px; margin-top: 0.2em">
      <span class="magenta">{{ item.lemmaCz }} / {{ item.lemmaSk }}</span>
    </h5>
```

Figure 9: *Vue* tag *v-for* in action to render all timeline items for Archive page

2.5 Visual identity

The project’s website as well as the accompanying graphic material is based on the visual identity created by Jan Kocek from the Institute of the Czech National Corpus. The homepage uses a tile system, with each tile containing a different type of content

(feuilleton, frequency statistics, dictionary data, etc.). The colour scheme is pale blue/black/purple, along with the red and blue of the Czech and Slovak flags. Combined with the modern Roboto font, this is a simple and fresh design.

The graphic designer also created an icon for the site, a logo, templates for metadata for social networks, a template for the quotes that appear at the end of the page, and a template for the side events related to the project.

◀ 17. 4. – 23. 4. 2023 ▶

Česko(-)slovenské



slovo týdne

tchyně



slovo týždňa

svokra

Pár slov o slově tchyně

Předně si slíbme: dnes žádně vtípy o tchyních, ano? To je seriózní sloupek!

A protože už jsem ve svém oboru už léta honěný a současnou kodifikací (čti: to, jak bychom měli psát podle ÚJČ) znám vcelku obstojně, už jen máloco mě překvapí. [Na rozdíl od jiných](#). Ale ruku na srdce, mluvíte-li o tchyni, vyslovujete ji krátce? Nelžete! Respektive takto: znáte vůbec někoho, kdo ji krátce vyslovuje? (Ostraváci se nepočítají!) Solva, protože takových lidí bude asi tolik jako těch, co vyslovují *matjesy* s oním [j]. A přesto tuto neděšku držíme uměle při životě. Tedy jak kde: v oficiálních psaných textech, nad nimiž bdí oko korektora, se to ještě ctí (byť i zde se občas najde dlouhá, tj. nekodifikovaná varianta). Ovšem v textech internetových, kde se plebs vyjadřuje volně a nad pravopisem si příliš hlavu neláme, je situace jiná. A proto: ještě než někoho po gramanačovsku osočíte z neznalosti normy – neřkuli celé mateřštiny, pozor! jste totiž sami v menšině... A teď mi řekněte, vážení kolegové z ÚJČ, co by vám udělalo, zrovnoprávnit dlouhou variantu (a neuvádět tak prostý lid v překvapení). Když už jste zrovnoprávnili tripletu [čurat/čúrat/čúrat](#)...

A ještě pro jednu věc mám tohleto slovo rád. Vždy jsem záviděl jiným jazykům aspiraci, neboli příděch, v první řadě romštině, na jehož fonetickém svérázu se tento jev výrazně spolupodílí. My takových (pseudo)aspirovaných slov mnoho nemáme, o to s větší rozkoší je vypouštím z úst, ať už je to [tʰi:ně], [tʰa:n] nebo [tʰoʃ].

Poche (čti [pʰe]) – a pak že prý se nedá sloupek o tchyních napsat bez vtípu o těchto. (Jen se bojím, že má sparringpartnerka tomuto lákadlu neodolá...)

Čech o slově svokra

No dobře, my máme sexy příděch, zato *svokra* zní starobyleji, vznešeněji, slavnostněji... A přitom jde o popisnost až prozaickou: 'svě krve žena'. Však jsme ji dříve taky mívali, ještě Jungmann uvádí heslo *swekra* a jeho řidiší varianty *swekry* a *swekrew*; ba co víc, rozlišuje *swekruši* (tchyně) od *swekrušny* (švagrová). Ale mamá sláva, někdejší *swekru* postupně vytlačila *tchyně*, ana se zjevila prý už za času Komenského namísto původního staročeského tvaru *tšče* (srov. mužský protějšek *test*, tchán'). Anebo snad *tchyně*...?



Michal Škrabal

Český lexikograf a korpusový lingvista, toho času filatelista a ředitel Ústavu Českého národního korpusu.

Niečo o slove svokra

O svokrách len dobre. (Hlavne ak máme nábeh čoskoro sa svokrou stať, dokonca viacnásobnou.) Slovo *svokra* (*svokrička*, *svokruša*, *svokruška*), označujúce manželovu matku, je ženským náprotivkom k mužskému pomenovaniu *svokor*, manželov otec; spoločne sa nazývajú *svokrovcami*. Na označenie matky manželky vo vzťahu k manželovi máme v slovenčine menej používané slovo *testiná* (*testinká*), s príslušným mužským pendantom *test*, spoločne *testovci*. Paralelné používanie týchto slov je však dlhodobá na ústupe, označenie *svokor*, *svokra*, *svokrovcami* sa postupne zaužívalo jednotne pre ženiných i mužových rodičov. Pri ich priamom oslovení sa však často nahrádzajú rovnakými pomenovaniami ako pre vlastných rodičov – teda *mama* a *otec*. Zatiaľ čo niekto rieši *svokrovské* problémy (najmä vo vzťahu svokra – nevesta, svokra – zať), Rudo Sloboda v románe *Krv* (1991) spomína aj iné: „*Takí zaťovia, ktorých príliš nútite prijať svokrovské móresy, odchádzajú do krčiem, aby si oddýchli.*“

Takmer rovnako ako *svokry* sa v bežnej komunikácii skloňuje *svokrin jazyk* (po latinsky *Sansevieria trifasciata*, v češtine *tchynin jazyk* alebo *tenura*, v angličtine *mother-in-law's tongue*) – kvietok, ktorý dostal meno podľa tvaru svojich listov: sú totiž dlhé a špicaté (mečovité), nelichotivo pripomínajúce jazyk svokry. Najznámejšia a najodolnejšia izbová rastlina – *izbovka* (po česky tak pohodovo – *pokojočka*), akýsi nezmar medzi nezmarami, mnohým z nás dobre známa z parapetných dosiek československých škôl, škôlok a úradov, má dnes napriek občasnému prívlastku *socialistická* stále viac obdivovateľov. Pre svoju absolútnu nenáročnosť a zároveň dekoratívnosť a dlhovekosť je objektom čoraz väčšieho zborateľského záujmu sukulentárov. Rýchlo sa rozmnožuje a jej listy majú protizápalové liečivé účinky, podobne ako aloa pravá (*Aloe vera*). A pretože funguje ako perfektná prírodná čistička vzduchu, výborne sa hodí aj do spálne. (Tu si prímyslíme veľkého smajlika.)

Či už vnímame *svokru* ako milé, rodinne založené stvorenie, ktoré chce každému len dobre, predovšetkým svojmu už dospelému dieťaťu (a vnúčatám), alebo ako vďačný terč [vtípy](#), dokonca až čierneho humoru, každý z nás sa rád zasmieje:

– „Viete, prečo sa svokry pochovávali do pol pása?“ – „Aby sa mal kto starať o hrob!“

– „Svoju svokru volám vegeta.“ – „Prečo?“ – „Pretože sa mieša do všetkého.“

Mladomanžel príde domov podnapitý a vo dverách stretne svokru s metlou v ruke. Prižmúri oči a povie: „Mamička, zameťáte alebo odlietate?“

Slovenka o slove tchyně

Zatiaľ čo slovenské slovo *svokra* nás baví možným etymologickým východiskom v indoeurópskom **swē-* (svoj) + **ker-* (hlava/krava), malý prieskum medzi českými internetovými vtípkármi ma doviedol až k ich uletenému tvrdeniu, že slovo *tchyně* vzniklo zložením slov *tchoř* a *svině* – odpadávam a končím.



Monika Kapustová

Venuje sa jazykovede, píše, prekladá, učí. Miluje vtípy o svokrách, babkách, dedkoch, deťoch a hodinách slovenčiny.



Figure 10: Screenshot of the web page⁴



Figure 11: Screenshot of the mobile version

⁴ A short feuilletton is followed by frequency information (among various text types), similarly used words, examples from written and spoken data, and excerpts from older dictionaries.

2.6 Workflow

A logged-in user with the admin role can create and edit data in the system. A new pair of words is entered into the system as follows:

Via the menu Entities / Word Pair, we get to an overview of all word pairs (see Figure 3 above). Pressing the “Create a new word pair” button will open the “Create/edit a word pair” form. There we can enter the basic data: the publication date of the post in the YYYYMMDD form and the Czech and Slovak lemma form. Pressing the “Populate” button will fill in some of the following items using the templates defined in the General Params entity. These are templates for corpus samples and dictionary data with links. Other items can be added subsequently: frequency graphs and numeric items for frequencies, word-clouds for the “Similarly used words” tile, or screenshots from older monolingual dictionaries. We get the data from the “Word at Glance” portal, the JÚLŠ dictionary portal, and other tools.

Create or edit a WordPair

ID

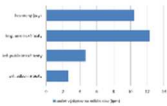
Poradie

Enabled

Lemma Sk

Lemma Cz

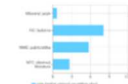
Stats Picture Sk



image/png, 3 412 bytes ✕

 No file chosen

Stats Picture Cz



image/png, 9 732 bytes ✕

 No file chosen

Figure 11: Create/edit a WordPair form

We save the document with a feuilleton in HTML format. Having opened the HTML file in a text editor, we copy-paste it into the “Usage Note Sk” and “Usage Note Cz” entries. We press the button below the item to clear the text and manually edit the section headings. Corpus entries “Written Language Sk”, “Written Language Cz”, “Spoken Language Sk”, “Spoken Language Cz” supplement the template with apposite examples from the corpora. We highlight the keyword pressing the relevant button. We equip the dictionary entries “Lex Desc Monoling Sk” and “Lex Desc Monoling Cz” by taking the formatted dictionary text from the dictionary portals page code (via the browser function “View page source”). After the copy-paste we press the button below the entry. The rich dictionary formatting will be preserved. The links in the “Additional Resources” section are already filled in. We add the audio recordings for the spoken language samples, in JSON format, to the “Extra Data” entry. We get this audio material from spoken corpora, e.g.:

```
"audioSK" : "https://korpus.juls.savba.sk/hovor-7.0-web/2008-07-26-Briock/Briock_00573.691.ogg#t=5,10"
```

Finally, at the bottom of the form, we enter the name of the writers of the Czech and Slovak feuilleton according to the authors’ list and save the form via the “Save” button. After returning to the overview list of word pairs, we can check the new entry via the “View” button.

3. Future work

The database described above is fine-tuned now: it appears to be both robust and flexible enough for further use. At least three possible uses can be imagined: 1) another, follow-up project created by the users themselves (user-generated content supervised by professional editors); 2) other language pairs (Czech-German/Polish, Slovak-Hungarian/Polish); 3) adding another language(s) (e.g., those of the Visegrad area: Czech, Slovak, Polish, Hungarian). The last option is certainly the most implementation-intensive, but even that seems relatively straightforward, adding entries for the new language(s) to the Word Pair data entity and modifying the main website to display the language data in 3(+) columns instead of the current two. After adding the new features to the data model, the *JHipster* generator can be re-run to re-generate the code for the entire updated system. Some caution is necessary here, however, as the manual edits we made to the code may be lost in this process.

4. Acknowledgements

This work has been, in part, funded by the Slovak VEGA Grant Agency, Project No. 2/0016/21. It was also supported by the European Regional Development Fund-Project “Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World” (No. CZ.02.1.01/0.0/0.0/16_019/0000734). During its creation we used the tools developed within the Czech National Corpus project (LM2015044)

funded by the Ministry of Education, Youth and Sports of the Czech Republic within the framework of Large Research, Development and Innovation Infrastructures.

5. References

- Bernolák, Antoninus (1825). *Slowár Slowenský Čěško-Latínsko-Ňemecko-Uherský seu Lexicon Slavicum Bohemico-Latino-Germanico-Ungaricum*. Budae: Typis et Sumtibus Typogr. Reg. Univers. Hungaricae.
- [ČSS] *Čěsko-slovenský slovník* (1981). Bratislava: Veda.
- Jungmann, Josef (1835-1839). *Slownjk čěsko-ňemecký*. Praha: Knížecí arcibiskupská knihtiskárna.
- [KSSJ] *Krátký slovník slovenského jazyka*. Bratislava: Veda.
- Machálek, T. (2019). Slovo v kostce – agregátor slovních profilů. Praha: FF UK. Accessed at: <http://korpus.cz/slovo-v-kostce/>. (21 April 2023)
- Machálek, T. (2020). Word at a Glance: Modular Word Profile Aggregator. In N. Calzolari et al. (eds.) *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille: ELRA, pp. 7011-7016.
- [SČS] Gašparíková, Želmíra & Kamiš, Adolf (1967). *Slovensko-čěský slovník*. Praha: SPN 1967.
- [SSJ] *Slovník slovenského jazyka* (1959-1968). Bratislava: Vydavateľstvo SAV.
- [SSJČ] *Slovník spisovného jazyka čěského* (1960-1971). Praha: ČSAV.
- Škrabal, M. & Benko, V. (2019). Make my (Czechoslovak word of the) day. In I. Kosem et al. (eds.). *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal*. Brno: Lexical Computing CZ, s.r.o., pp. 467-477.