

Využití korpusů

při výuce češtiny jako cizího jazyka

PLIN022

5.

Co se dnes dozvíte?

- základní informace o žákovských korpusech
- problémy spojené s budováním žákovských korpusů
- co je to interlanguage

Korpusová typologie

Obecné
korpusy

Specializované
korpusy

Korpusová typologie

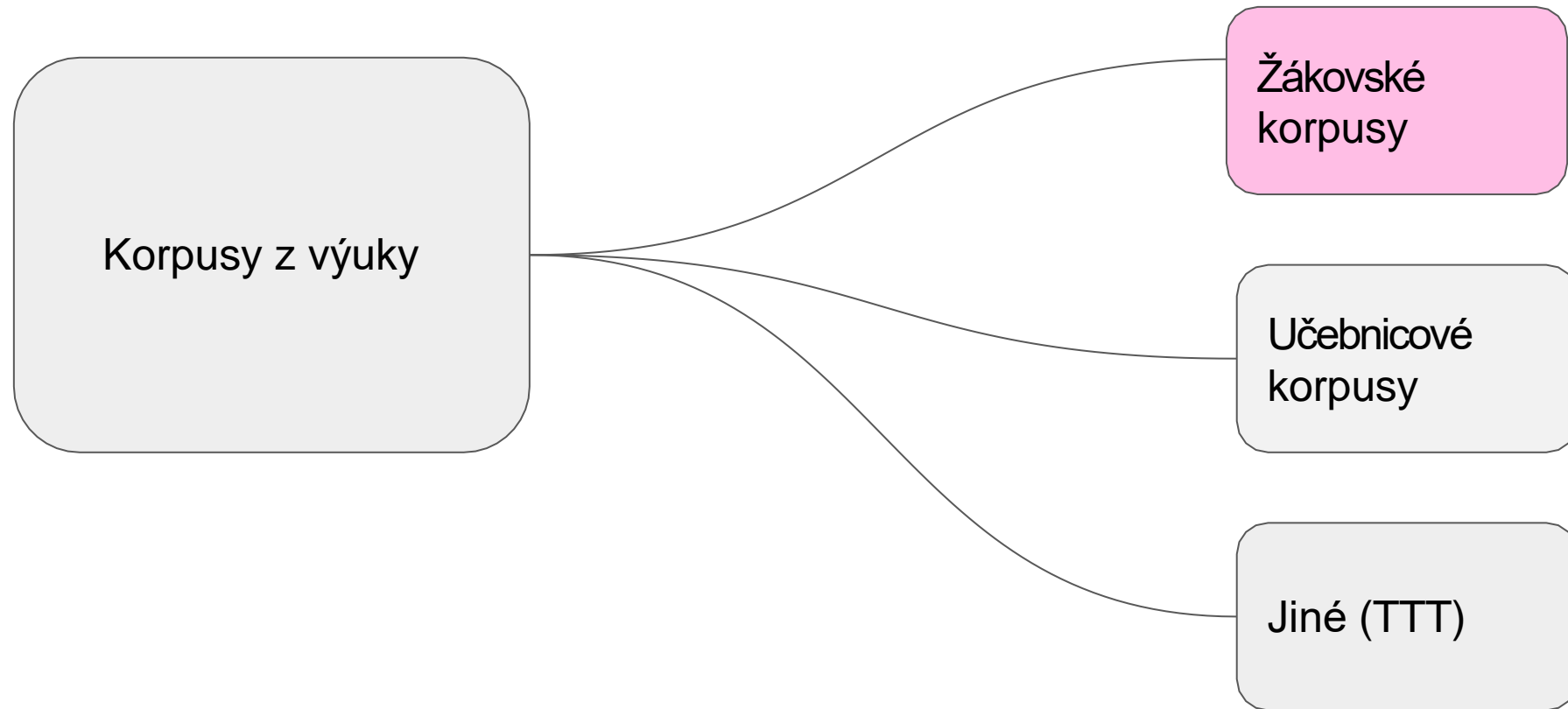
korpusy
řady SYN

KSK, Totalita,
Capek ...

Korpusová typologie: výuka

Korpusy
pro výuku

Korpusy
z výuky



Specializované korpusy pro L2

- akviziční korpusy (žákovské korpusy)
- korpusy jazyka pro specifické účely
(např. akademický jazyk, profesní mluva)
- pararelní korpusy
- jiné (např. korpus vyučovacích hodin, učebnicový korpus)

Co si o češtině myslí studenti?

1. Hlavní problém bylo s jazykem jak jsem napsal nahoře ; s jazykem protože poprvé studuju v české skolu a take že studuju poprvé češtinu . To bylo a , to je moc **těžké** . Osobně si myslím že to je těžké protože nic v češtině vypadá jako ve Francouzštině nebo v angličtině ; Každý slovo je jiné a to se musím naučit každý den ! Učit se pady (co , koho , čeho ...) ne byl asi těžký protože už to existuje v latině (studoval jsem latinu 6 let) ale nejtěžší je psaní protože když píšu , musím se pamatovat všichni pravidla gramatiky , je respektovat a znát moc slov , to poslední neznám (to je škoda) ...

2. Jsem z Turecka , ale už šesti měsíců bydlím v Praze , protože studuju češtinu na Metropolitní Univerzitě v Praze . Čeština je těžká pro mě , ale jí naučím . Učím se češtinu , protože chci bydlet a pracovat v Praze . Tady jenom angličtina nestačí ke komunikaci .

3. Každý čtvrtek nebo pátek jím české jídlo . České pivo a bramborový salát se mi líbí ! Každý víkend chodíme se svým kamarádem na procházky po Praze a mluvíme s českými lidmi . Jsem nemám problémy s **češtinou** , protože paní Uhlířová -dobrá učitelka ! Po příjezdu do ČR největší překvapení bylo samozřejmě -český národ , protože jsou velmi dobří !

4. Když jsem jela do Prahy , myslela , že čeština je podobná ruštině , protože jsou slovanské jazyky . Na začátku studia jsem rozuměla , že **čeština** je těžká . Nemohla jsem mluvit jako Češi , a proto lidé mluvili se mnou anglicky . Teď už mluvím dobře česky a můžu úspěšně dostat na univerzitu . Čeština a ruština mají hodně společného . Některá slova v češtině vypadá jako v ruštině , protože mají společné kořeny .

5. Obvykle mluvím s kamarády a paní učitelkou z **češtiny** . Vadí mi trochu , že některá slovesa se používají s jinými pády než v ruštině , i když mají stejný význam .

6. Psát v **češtině** pro mě je snadnější , než mluvit , protože , když mluvím , vždicky začínám myslet na pravidelnost řeči , v jakém pádu musí být slovo a t. d . Proto mluvím není velmi rychle . Ačkoli každý měsíc studia cítím přírůstek znalostí a jistoty .

7. Když jsem přijela do Prahy , znala jsem jenom několik slov . Teď můžu docela dobře rozumět a trochu mluvit česky . Ale bohužel tento kurz **češtiny** nesplní mé očekávání . Měli jsme nejlepší profesorku , a také docela dobrý kurz gramatiky . Ale vůbec jsme se neměli možnost naučit hovornou češtinu

Žákovské korpusy

ang. *learner corpus*, *studijní korpus* (Čermák) = **žákovský korpus**

snaha shromáždit objektivní data pro:

- popis žákovského jazyka
- popis odchylek a rozdílů
(contrastive interlanguage analysis (S. Grangerová); computer aided error analysis)

https://www.researchgate.net/publication/277578581_Contrastive_interlanguage_analysis_A_reappraisal

- popis mimojazykových vlivů na L2 (metadata)

Jak probíhá osvojování druhého/cizího jazyka?

- L. Selinker (70. léta 20. století)
- žákovský jazyk, **mezijazyk**, interlanguage
- svébytný jazykový útvar
- proměnlivý a dynamický systém, který obsahuje řadu nestandardních prvků
 - míšení kódů
 - chyby
 - jazykový transfer: pozitivní (**kognáty**) a negativní (**false friends**)
 - naučené fráze versus improvizace

Výzkum je zaměřen na popis procesů a utváření mezijazyka či mimojazykové vlivy (např. věk).

Výzkum žákovského jazyka



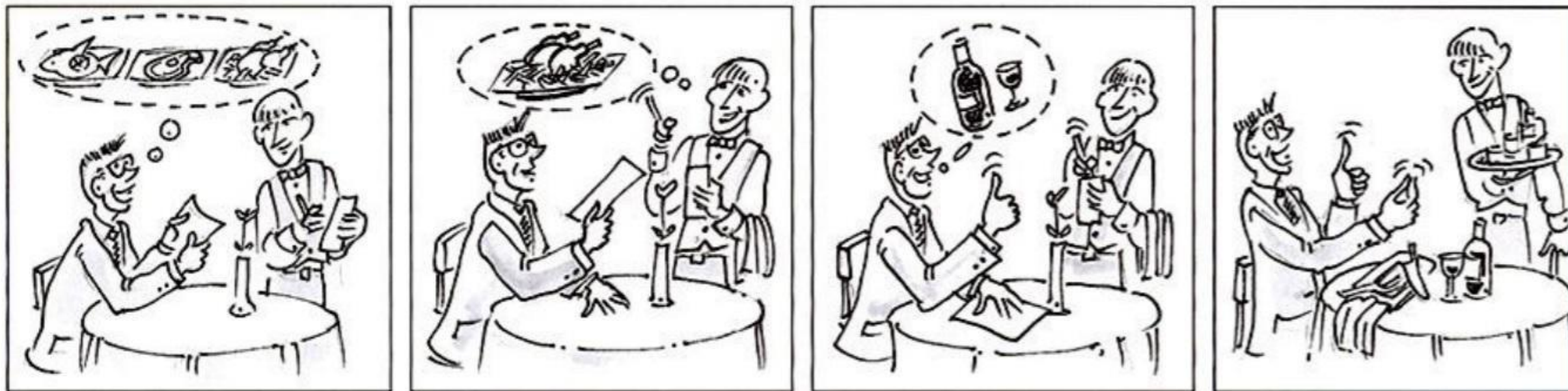
Čeština Andreje Babiše

- 2014: žádost o tlumočnicka při projevech Andreje Babiše ve sněmovně
- Karel Oliva: Babišova úroveň je C2, žádost zamítnuta
- **Vít Benešovský (2018): Analýza Babišova mezijazyka**
 - analýza přepisu rozhovorů z let 2011, 2014, 2017
 - případová studie projevu nerodilého mluvčího češtiny v českojazyčném prostředí
 - interferenční chyby (výslovnost, morfologie), slovakismy a střídání jazykového kódu
 - mírné zlepšení po roce 2014 (mimojazykové vlivy)

Jak vznikají data pro žákovské korpusy?

- elicitace (vznik projevu a jeho průběh iniciovaný a řízený druhou osobou, zprav. učitele)
- řízené projevy
- strukturované projevy
- improvizace

Jak vznikají data pro žákovské korpusy?



Obrázek 1 Příběh v obrázcích. Zdroj: Španělský mluvený korpus, online.

Guided Interview Task

The aim of this task was for learners to talk in an open ended way about past events, with reference both to themselves and to others. There were two sub-tasks, "Famous People" and "Autobiography". For "Famous People", the learners viewed a series of 8 pictures of famous people from the past, and chose four to explain what they knew about them.

Sample pictures, "Famous people" task



John Lennon



Diana, Princesa de Gales



Adolf Hitler

Zdroj: Tasks description, Spanish Learner Language Oral Corpora, online, 2016.

Jak vznikají data pro žákovské korpusy?

- certifikované zkoušky
- jazykové školy
- univerzity

!!!

= minimum přirozených (autentických) promluv

Jak vznikají data pro žákovské korpusy?

doc.t_id	UJA3_TT_001	doc.t_datum	2011-10-22
doc.t_medium	rukopis	doc.t_limit_minut	45
doc.t_pomucka	ne	doc.t_zkouska	ano prubezna
doc.t_limit_slova	20	doc.t_nazev	Popis pokoje
doc.t_tema_typ	obecne	doc.t_aktivita	
doc.t_tema_zadane	zadane	doc.t_postup_zadany	volne
doc.t_postup_prevazujici	popis	doc.t_pocet_slov	32
doc.t_rozsah_slov	-50	doc.s_id	UJA3_TT
doc.s_pohlavi	z	doc.s_vek	17
doc.s_vek_kat	16-	doc.s_jazyk1	fi
doc.s_jazyk1_skupina	nIE	doc.s_dalsi_jazyky	sv
doc.s_cj_SERR	A1	doc.s_cj_v_rodine	
doc.s_roky_v_CR		doc.s_studium_cj	VS
doc.s_studium_cj_mesice		doc.s_studium_cj_hod_tyden	15-
doc.s_ucebnice	NCSS	doc.s_bilingvni	ne
s.id	2		

Žákovské korpusy

typologie podle různých kritérií

- velikost, dostupnost
- povaha dat a povaha sběru dat
- L1, cílový jazyk
- metadata a anotace

Štindlová (2013): přehled 57 žákovských korpusů

dominují psané korpusy střední velikosti (do 1 mil. slov)

pokročilejších studentů s angličtinou jako cílovým jazykem

Learner Corpus Association (2018): 167 žákovských korpusů

Žákovské korpusy ve světě

- Komerční korpusy
 - CLC: Cambridge Learner Corpus (část přístupná přes Sketch Engine)
- Lovaňské korpusy (Katolická univerzita v Lovani, Belgie; S. Grangerová)
 - ICLE: International corpus of learner English
 - LINDSEI: Louvain International Database of Spoken English Interlanguage
 - LONGDALE: Longitudinal Database of Learner English
 - VESPA: Varieties of English for Specific Purposes Database
 - MUST: Multilingual Student Translation corpus
 - TeMa: Corpus of teaching materials of English for general purposes
- Korpusy slovanských jazyků
 - RLC: Russian Learner Corpus; CroLTeC: Croatian Learner Text Corpus

Žákovské korpusy u nás

- projekt **AKCES** (2005–2021)

Ústav českého jazyka a teorie komunikace FF UK + další pracoviště
prof. PhDr. **Karel Šebesta**, CSc.

osvojování češtiny různými skupinami mluvčích (L1 i L2)

žákovské korpusy **CzeSL (Czech as a Second Language)**

- CzeSL–plain (bez anotace)
- CzeSL–SGT (emmendace, lemmatizace, tagování vč. chyb)
- CzeFL–LONG a CzeSL–LONG (budovaný)

Kvíz o CzeSLu

Co korpus obsahuje?

- a. pouze psané projevy b. pouze mluvené c. mluvené i psané

Kolik metadat popisuje korpusová data?

- a. 10 b. 20 c. 30

Jaký je nejčastější L1 žáků?

- a. angličtina b. ruština c. arabština d. ukrajinština

Jaká je nejčastější úroveň žáků?

- a. A1 b. A2 c. B1 d. B2

Jaká je nejčastější chyba?

- a. chyba v morfologii b. chyba lexikografická c. chyba v diakritice

Práce na CzeSL

- a. jsou už dokončeny b. stále probíhají

Kvíz o CzeSLu – řešení

Co korpus obsahuje?

- a. pouze psané projevy b. pouze mluvené c. mluvené i psané

Kolik metadat popisuje korpusová data?

- a. 10 b. 20 c. 30

Jaký je nejčastější L1 žáků?

- a. angličtina b. ruština c. arabština d. ukrajinština

Jaká je nejčastější úroveň žáků?

- a. A1 b. A2 c. B1 d. B2

Jaká je nejčastější chyba?

- a. chyba v morfologii b. chyba lexikografická c. chyba v diakritice

Práce na CzeSL

- a. jsou už dokončeny b. stále probíhají

Žákovské korpusy u nás

- mezinárodní žákovský korpus MERLIN (2012–2014) (něm.-it.-čes. projekt)

http://merlin-platform.eu/C_mcorpus.php

- databáze CHRUP (2013) (**Ch**yby **R**usů **U**krajinců a **P**oláků)

<http://chrup.ff.cuni.cz/>

Databáze jazykových chyb v češtině mluvčích s prvním jazykem slovanským

prof. Ph.Dr. Hana Gladkova, CSc., FF UK

didaktický cíl: **specifika osvojování L2, který je blízký L1**


- další menší projekty


[Using MERLIN](#)[MERLIN for research](#)[MERLIN corpus](#)[MERLIN annotations](#)[Documentation](#)[Download whole corpus](#)[Simple search](#)[Advanced search](#)[Define a subcorpus](#)[Statistics](#)

Search for words in the learner texts and display them in context (e.g., full text):

examples: *welchen*, *glí*, *rád*

 [Search for word](#)  

Search in: 

Corpus: 

No subcorpus yet? [Compile your own text collection » here.](#)

The MERLIN corpus

MERLIN provides access to 2.286 texts written by learners of Czech, Italian and German.

The learner texts stem from standardized language tests and they have been reliably related to the CEFR levels. [» read more](#)

Use MERLIN ...

... to better understand the levels of the Common European Framework of Reference (CEFR). [» read more](#)

Video tutorial ...



Korpus arabských studentů češtiny

EN | CS

Hlavní nabídka

- Domů
- Dokumenty
- Hledat
- Login

Powered by TEITOK
Maarten Janssen, 2014-

IMA_Ka_033-c-001_t

Název -

Možnosti zobrazení

Text: Přepis Opravený tvar - Ukázat: Barvy - Značky: Morfologická chyba Syntaktická chyba Lexikální chyba Chyba v pádu Poznámka

IMA_Ka_033-c-001_t

Moje země

Moje vlast je Egypt, a proto jsem hrdá na to, hlavní město Egypt je Káhira. Egypt je považován za zemi civilizací, a proto mnoho turistů navštíví Egypt z celého světa. V Egyptě Káhiry, pyramidy, město šarm Alshaich, Luxora paláce. V památek, proto cestujou Egypt mnoho turistů ze všech států. Samozřejmě rok je rozdělen na čtyři období jaro, léto, podzim a zima. Léto je horké a suché, zima je chladné, a proto lidé musí mít teplé oblečení. V létě je suché a horké, proto lidé mají rádi roční období, protože příroda je krásnější a počasí je pěkné. Kanceláře nabídku a přilákání turistů do Egypta. Léto je období cestovního ruchu v Egyptě. Protože počasí v Egyptě je teplejší než počasí evropských zemí. Egypt usiluje o udržení své ekonomiky a prosazování svého průmyslu, aby postupoval dvěma způsoby: za prvé podporovat vývoz národního zboží do sousedních zemí a za druhé snažit se omezit dovoz zboží, výrobků a používání místního zboží jako alternativu. Egypt má takzvaný Suezský kanál.

Egypt

Opravený tvar

Egypta

Morfologická chyba

Chybná deklinace podstatného jména mužského vzoru hrad (parHrad)

Chyba v pádu

Záměna 2. pádu (gen)

říklad řeka Nil, věž
amných historických
pěkné počasí.
irzne v zimě počasí je
považováno za nejlepší
nabízejí cestovní

Žákovské korpusy – nepřímé využití

- ve slovnících a gramatických příručkách

Grammar: *information*

Do not say "an information" or "informations." Say, for example: *some information, a lot of information, or a piece of information:*

- *Do you have any information about tickets?*

- *I need a couple more pieces of information before I can finish my report.*

Zdroj: The Longman Corpus Network, 2015, online.

🕒 Spelling mistakes are among the most frequent errors made by candidates in the exam. Find and correct the spelling mistakes made by candidates in the sentences below.

1 As you can see in the advertisement, the holiday is quite cheap.

Zdroj: Brook-Hart – Haines, 2014–2015, s. 21.

in·for·ma·tion /ˌɪnfəˈmeɪʃən \$ -fer-/ ●●● S1 W1 noun [uncountable] 🔊 🔊

1 facts or details that tell you something about a situation, person, event etc

🔊 I need more information.

information that

🔊 We have received information that Grant may have left the country.

information about/on

🔊 The book contains information about a wide variety of subjects.

my/our etc information is (=used to say what you know about a situation)

🔊 My information is that Gary wants to stay with the club.

→ inside information

Register

In everyday English, in informal contexts, people often use the expression **tell someone something** rather than using the noun **information**:

I wonder if you could **give me some information** about your childhood. → I wonder if you could **tell me something** about your childhood.

GRAMMAR: Countable or uncountable?

• **Information** is an uncountable noun and is not used in the plural. You say:

I need some information about hotels.

X Don't say: I need some informations about hotels.

• **Information** is always followed by a singular verb:

The information is strictly confidential.

• When talking about one fact or detail, you say a **piece of information** or a **bit of information**:

Žákovské korpusy – nepřímé využití

- ve výukových materiálech

Příklad 1 Korekce vybraného typu chyby a reformulace pravidla užití daného jevu.

Zadání:

Přečtěte si text, který napsal žák z Ukrajiny. Čísla v textu jsou vždy za slovem, kde udělal chybu. Řekněte, jaké jsou to chyby a opravte je.

Mám moc radá (1) sport. Hrajeme (2) moc venkovní sporty. Ráno jezdím na kole do školy (3) s mym (4) kamaradem (5). Nemůžeme lyzovat (6), sáňkovat nebo běhat na lýžich (7), protože tu není sníh. Ale, v letě (8), když je teplo, plaváme (9) v bázeňe (10) (11), na řece nebo na moře. Mám ráda fotbal a obvykle jdu na fotbalové hřiště (12), ale nehraju fotbal dobře. Ještě hrajeme volejbal a basketbal. Golf a baseball nejsou na Ukrajině populární. Hrajeme taky kulečnik (13), šachy a stolní tenis. Myslím že, šach je dobrý duševní sport. Můžu (14) trochu hrát šachy, ale nemužu (15) bruslit. Na podzim nesportujeme moc, protože velmi moc prší. Pak, dívám se sport na televízi (16).

Šindelářová, J.; Škodová, S. (2012, s. 118–119)

Literatura

METODIKA PRÁCE S ŽÁKY-CIZINCI

V ZÁKLADNÍ ŠKOLE

Jaromíra Šindelářová – Svatava Škodová

MŠMT Praha 2012

TECHNICKÁ UNIVERZITA V LIBERCI

Fakulta přírodovědně-humanitní a pedagogická

Katedra českého jazyka a literatury

Karel Šebesta – Svatava Škodová
a kolektiv

**Čeština – cílový jazyk
a korpusy**

Liberec 2012