

Technologie ve službách překladaatele – prezentace knihy

Miroslav Pošta

TranslatoBlog.cz

Knihy-Apostrof.cz

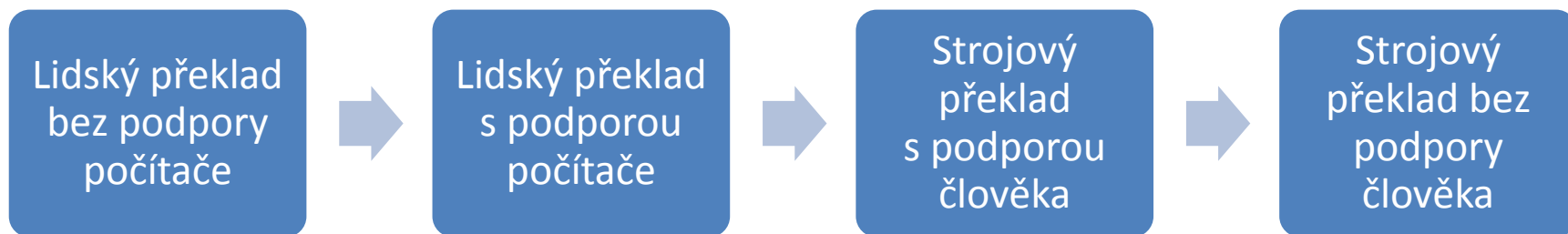
Technologie ve službách překladatele



Tři hlavní okruhy

- Jazykové korpusy
 - Počítačem podporovaný překlad (CAT)
 - Strojový překlad (MT)
-
- Další podpůrné nástroje
 - Zamyšlení nad budoucností

Metody překladu – historické fáze



Jazykové korpusy

- „hromada textů, které máte uložené v počítači“
- 1. korpus v 60. letech 1 milion slov, dnes např. SYN2015 (ČNK): 100 mil. slov
- Různé velikosti; témata (např. prezidentské projevy, čeština cizinců); míra zpracování (lemmatizace, značkování)
- Jednojazyčné vs. paralelní

Hlavní výstupy z korpusů

- **Frekvenční seznamy** – pro překladatele příliš zajímavé nejsou (spíš pro jiné profese, studium/výuku jazyka)
- **Konkordance** + z nich získané další informace (kolokace aj.)

Konkordance (Český národní korpus)

KonText SyD Morfio KWords Treq Wiki Podpora Biblio Miroslav Pošta Odhlášení English

kon text Dotaz Korpusy Uložit Konkordance Filtr Frekvence Kolokace Zobrazení Nápvěda

Korpus: syn2015 | Dotaz: pošramocení (149 výskytů) ▶ Promíchat: ✓

Výskytů: 149 | i.p.m.: 1,23 (vztaženo k celému "omezení/syn2015") | ARF: 86,26 | Výsledek je promíchán 1 / 3

Výběr řádků: základní | Atributy:

<input type="checkbox"/>	Nejen trh	regulace a výkonu vlastnických vztahů a tím i k obnovení	pošramocené	důvěry občanů v obě tyto instituce :</p><p>“ Až příliš
<input type="checkbox"/>	Regionální týdeník	s Moravskou Slavií 5.3 a postupem na čtvrté místo napravuje	pošramocenou	reputaci z úvodních kol.</p><p>Před ním na třetí příče
<input type="checkbox"/>	Právo	znamenal určitě uklidnění,“ prohlásil Psootka.</p><p>S nápravou	pošramocené	reputace podle něj Hanáci začali už ve středu, když
<input type="checkbox"/>	Lidové noviny	v rodině střední třídy, z něhož vyjdou tři sestry	pošramocené	tak, že se z toho nemohou vzpamatovat ani v
<input type="checkbox"/>	Den jako stvořený pro váno...	kolej a že se z toho dostane jen s mírně	pošramocenou	figurou. Upřímně jsem mu to přál, protože i
<input type="checkbox"/>	Hospodářské noviny	.</p><p>Podobnou cestu navíc ODS zvolila napříč republikou. Stranu	pošramocenou	skandály mají ve volbách zachránit často nové tváře, zatímco
<input type="checkbox"/>	Evropské noviny	si totiž náš národ bude moci v Evropě napravit svoji	pošramocenou	reputaci. Podářili se toto, jsem přesvědčen, že
<input type="checkbox"/>	Moravští Šternberkové	procesům. Když otec Ješek zemřel, převzal Albrecht dost	pošramocené	dědictví, jehož skutečný rozsah lze jen odhadovat: hošteljské
<input type="checkbox"/>	Gravitace	pomálu, každý z přistěhovalců v České Lužici byl nějak	pošramocený	válkou, chtěl začít nový život a nezkoumal hříchy minulých
<input type="checkbox"/>	Extra PC	tyto starosti trápit. Kdo však má z tohoto důvodu	pošramocenou	reputaci, je samotný Grooveshark. Na toho se valí
<input type="checkbox"/>	Blesk	Z. se zpracovává od počátku května a plánuje zlepšit	pošramocenou	pověst stacionáře. „ Podávala jsem trestní oznámení na bývalou
<input type="checkbox"/>	Nepokojná léta	Přestože mnohé lodě byly zasaženy a odplahočily se z boje	pošramocené	a nahnuté, nikdy žádná za bílého dne nespustila vlnku
<input type="checkbox"/>	Lidové noviny	„ Je to přirozené, to jen náš trh byl	pošramocený	a teď se narovnává. Doufám, že časem se
<input type="checkbox"/>	Almanach českých novinářů	Už bychom se nedočkali jediného zázraku na tom i tak	pošramoceném	světe. Pán Bůh by totiž neměl čas. A
<input type="checkbox"/>	Moje vítězství nad autismem	naivita v jejich očích, čistota jejich tělíček, ještě	nepošramocených	žádnými nemocemi nebo úrazy, to všechno v nás vyvolávalo
<input type="checkbox"/>	Ikárie	s čistým štítem od nuly je vždycky jednodušší než napravovat	pošramocenou	pověst. A špatná kniha vám ji mezi nakladateli pošramotí
<input type="checkbox"/>	Jihlavské listy	, odcházají obě instituce z celé „ lapálie “ s	pošramocenou	pověstí. Jihlava to hned po listopadovém zveřejnění negativních výsledků
<input type="checkbox"/>	Rekviem za Východ	jsem Šacha a měl jsem přitom zvláštní pocit, že	pošramocená	mozaika, kterou tvořil Vinnerův život, se stává mým
<input type="checkbox"/>	Nepokojná léta	se omylem pustily do sebe navzájem a podobně. Zle	pošramocené	zbytky dánského vojska se poklusem vydaly směrem k domovu v
<input type="checkbox"/>	Regionální týdeník	i vlastní zdroj pitné a nezávadné vody.</p><p>Osudný den</p><p>	Pošramocená	pověst turistické ubytovny na Harusově kopci snad mohla být znovu
<input type="checkbox"/>	Sport	vzhůru k nebi a rozhodně je muselo následovat i Španělovo	pošramocené	sebevědomí. Vždyť titul vyhrál poprvé od French Open 2011
<input type="checkbox"/>	Národní očista	zhoršuje jeho výchozí pozici pro poválečné vyjednávání. Aby napravili	pošramocenou	národní pověst, byli z ciziny do protektorátu vysláni parašutisté
<input type="checkbox"/>	Říše světa	průmyslovou špiónáž nemohly způsobit nic jiného, než ještě zhoršit	pošramocené	vztahy s GE. Věk také mnozí, včetně historika
<input type="checkbox"/>	Mladá fronta Dnes	„ Její ekonomika, ale i veřejné mínění jsou dost	pošramocené	. Hlavním cílem je obnovit její důvěru. A to
<input type="checkbox"/>	Lidové noviny	, rozpraženými rukama a vyčítavým pohledem k rozhodčímu.</p><p>Na	pošramocené	Barošově imagi se podepsala řada skandálů, které nesouvisí s
<input type="checkbox"/>	Mapa času	svém nakladatelství.</p><p>Povzbuzen Henleho houževnatostí se Wells rozhodl své	pošramocené	dílo důstojně dokončit. Ale jak už začínalo být zvykem

trados_32017D196...jpg trados_32017D1960.jpg Zobrazit vše ×

Co překladatel najde v jednojazyčných korpusech

- 1) **Srovnání frekvence dvou konkurentů** (slov, slovních spojení atd.)
 - Co je v autentických textech častější? *Ve třetí.**, nebo *v třetí.** (obojí ve všech rodech, číslech a pádech)? *Steva vs. Stevea*
- 2) **Hledání kolokací**
 - S jakými adjektivy se pojí substantivum *potlesk*, adjektivum *pošramocený*?
- 3) **Zjištění či ověření** gramatických jevů, úzu, významových nuancí, asociací
 - Má anglické slovo *repercussion* spíše pozitivní, nebo negativní konotace/asociace? Co v češtině *snědý* – je neutrální, nebo pejorativní?

Kolokace (ČNK) – pošramocený

Korpus: [syn v3](#) | Dotaz: [pošramocený](#) (3 914 výskytů) ▶ Promíchat: ✓

Kolokace

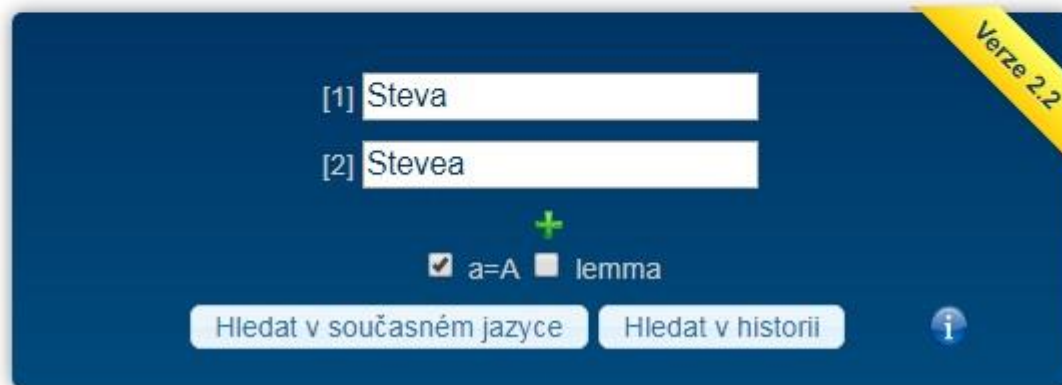
strana 1 ▶

	Filtr		Frekvence	T-score	MI	logDice
1.	p/n	reputace	493	22.203	14.962	10.121
2.	p/n	pověst	1011	31.792	12.742	8.299
3.	p/n	sebevědomí	291	17.054	11.900	7.407
4.	p/n	renomé	69	8.305	12.126	7.284
5.	p/n	image	92	9.587	10.987	6.428
6.	p/n	psychika	80	8.940	10.898	6.326
7.	p/n	ego	19	4.356	10.811	5.801
8.	p/n	sebedůvěra	21	4.579	10.280	5.472
9.	p/n	kredit	21	4.578	9.875	5.159
10.	p/n	prestiž	30	5.470	9.518	4.941
11.	p/n	důvěryhodnost	14	3.737	9.649	4.853
12.	p/n	kádrový	6	2.447	9.920	4.590
13.	p/n	kotník	31	5.558	9.117	4.588
14.	p/n	sebeúcta	4	1.999	10.581	4.522
15.	p/n	čest	32	5.642	8.621	4.135

SYD



Korpusový průzkum variant




Verze 2.2


[1]




[2]

+
 a=A lemma



Sketch Engine – word sketch

← → ↻ Zabezpečeno | https://the.sketchengine.co.uk/corpus/wsketch?corpname=preloaded%2Fczten12_9&reload=&lemma=potlesk&lpos=&minfr... 

Sketch Engine Czech Web 2012 (czTenTen12 v9)    Mr.

Home

Search

Word list


Word sketch

Thesaurus

Sketch diff

Corpus info

My jobs

User guide 

Save

Change options

Cluster

Sort by freq

Hide grammrels

More data

Less data

Sketch grammar

Menu position

potlesk (noun)

Czech Web 2012 (czTenTen12 v9) freq = 41.956 (8.27 per million)

a modifier	is obj7 of	is obj4 of	is subj of	gen 1
bouřlivý + 3.646 11.07	odměnit + 2.125 11.60	sklidit + 1.914 11.61	sklidit + 658 10.84	publikum + 552 9.75
bouřlivým potleskem	odměňovat + 329 10.57	vysloužit + 527 9.72	sklidil potlesk	potlesk publika
dlouhotrvající + 843 9.77	ocenit + 840 10.43	vysloužil potlesk	sklízet + 117 8.68	obecenstvo + 133 9.61
dlouhotrvajícím potleskem	přivítat + 218 9.52	sklízet + 289 9.17	rozlehnout 68 8.67	potlesk obecnstva .
zasloužený + 1.090 9.36	vyprovodit 71 8.70	zasloužit + 423 7.82	rozlehl potlesk	tribuna + 147 9.60
zaslouženým potlesk .	vyprovodili potleskem	si zasloužil potlesk	ozvat + 394 7.92	potlesk tribun
skandovaný + 324 9.12	oceňovat 76 8.66	vyprovázet 21 7.14	se ozval potlesk	divák + 838 8.68
skandovaným potleskem	poděkovat 62 8.34	vyprovázal potlesk	vysloužit 55 7.91	potlesk diváků
naštěný + 786 8.68	poděkovali potleskem	ozvat 36 6.93	vysloužil potlesk	poslanec + 546 8.41
naštěným potlesk	vítat 71 8.31	doznít 14 6.48	doznít 36 7.89	. (Potlesk poslanců ČSSD .)
frenetický + 258 8.68	uvítat 59 8.21	dozněl potlesk	dozněl potlesk ,	dav 80 7.90
frenetickým potlesk	přerušovat 48 8.04	vychutnávat 29 6.15	rozeznít 58 7.81	potlesk davu
uznalý + 219 8.47	kvítovat 41 7.97	ozývat 18 6.09	se rozezněl potlesk	ochoz 17 7.01
uznalým potlesk	vyprovázet 39 7.92	utichnout 10 6.03	vyprovázet 26 7.74	potlesku ochozů
hlasitý + 436 7.94	vyprovázeli potleskem	svědčit 28 5.94	vyprovázal potlesk	posлуhač 52 6.33
hlasitým potleskem	šetřit 90 7.45	zaznít 31 5.89	ozývat + 130 7.43	potlesk posluchačů
spontánní + 301 7.89	nešetřili potleskem .	zazní potlesk	rozléhat 24 7.04	delegát 30 6.26
spontánním potlesk	ohodnotit 43 7.29	vychutnat 39 5.68	rozléhá potlesk	za potlesku delegátů
vřelý + 202 7.53	přerušit 42 7.19	vychutnal potlesk	odměnit 27 6.77	hlediště 12 6.05
vřelým potleskem .	přerušen potleskem	rozeznít 10 5.63	potlesk odměnil	potlesk hlediště
ohlušující + 133 7.53	rozloučit 35 7.17	zaburácet 7 5.57	utichnout 27 6.77	fanoušek 90 5.86
ohlušujícím potlesk .	děkovat 24 7.01	doprovázet 32 5.47	potlesk utichl	potlesk fanoušků
mohutný + 614 7.46	povzbuzovat 19 6.74	zasluhovat 15 5.45	zasloužit + 150 6.74	korona 6 5.83
mohutným potlesk	loučit 19 6.74	zasluhuje potlesk	zasloužil potlesk .	auditorium 6 5.79
utuchající + 139 7.37	doprovodit 23 6.68	znít 28 5.35	zaznít + 112 6.40	spolužák 14 5.73
neutuchajícím potlesk	doprovodili potleskem .	slyšet + 119 5.29	zazněl potlesk	potlesk spolužáků
hromový 80 6.95	doprovázet 75 6.58	slyšet potlesk	zaburácet 11 6.35	sál 25 5.72
hromovým potlesk	přijmout 35 6.45	. Prosím potlesk	přerušovat 13 6.21	potlesk sálu .
vlažný + 125 6.89	odvděčit 18 6.41	odměňovat 7 5.15	přerušoval potlesk	kulomet 7 5.51
vlažným potlesk	burácet 13 6.32	vyvolat + 103 5.04	utichat 11 6.19	hrstka 6 5.51
několikaminutový 76 6.85	bouřit 13 6.30	opět 36 5.04	doznívat 11 6.08	pasažér 8 5.15
několikaminutovým potlesk	bouřil potleskem	opět potlesk	vyprovodit 8 5.92	část 76 5.14
srdečný 90 6.71			ustávat 12 5.74	. (Potlesk části poslanců .)

Sketch Engine – klíčová slova

The screenshot shows the Sketch Engine interface for the 'translatoblog' corpus. The page title is 'translatoblog: Extracted keywords / terms'. The interface includes a sidebar with navigation options like 'Home', 'Advanced features', 'Manage corpus', and 'Search corpus'. The main content area displays a list of extracted keywords and terms, organized into 'Single-word' and 'Multi-word' columns. Each entry includes a checkbox, the word, its score, frequency (F), and reference count (Ref). The words are ordered by keyness score.

words: 14 % / 1,000,000 days: 63 [SL]

translatoblog: Extracted keywords / terms ?

[Change extraction options](#) Download singlewords: [TBX CSV](#). Download multiwords: [TBX CSV](#).

Singlewords and multiwords are ordered by [keyness score](#). The score and corpus frequency (leading to the respective concordance) are displayed in parentheses. Highlighted words were used as seeds in a previous WebBootCaT run within this corpus.

[<< Back to corpus files](#) [Use WebBootCaT with selected words](#)

Single-word	Score	F	RefF	Multi-word	Score	F	RefF
<input type="checkbox"/> HYPERLINK (W)	4,445.03	95	0	<input type="checkbox"/> překladová paměť (W)	983.36	21	0
<input type="checkbox"/> CAT (W)	3,696.56	79	0	<input type="checkbox"/> terminologická databáze (W)	796.25	17	0
<input type="checkbox"/> Trados (W)	2,059.29	44	0	<input type="checkbox"/> strojový překlad (W)	777.70	18	29
<input type="checkbox"/> korpus (W)	1,034.24	97	17,177	<input type="checkbox"/> národní korpus (W)	515.57	11	0
<input type="checkbox"/> SDL (W)	796.25	17	0	<input type="checkbox"/> český národní korpus (W)	515.57	11	0
<input type="checkbox"/> konkordance (W)	787.17	19	661	<input type="checkbox"/> paralelní korpus (W)	375.23	8	0
<input type="checkbox"/> překladový (W)	750.48	32	5,049	<input type="checkbox"/> uživatel nástrojů (W)	375.23	8	0
<input type="checkbox"/> ČNK (W)	702.69	15	0	<input type="checkbox"/> textový editor (W)	342.68	10	127
<input type="checkbox"/> MT (W)	702.69	15	0	<input type="checkbox"/> slovní spojení (W)	338.06	16	419
<input type="checkbox"/> našeptavač (W)	625.91	14	243	<input type="checkbox"/> původní český text (W)	328.45	7	0
<input type="checkbox"/> Slat (W)	609.13	13	0	<input type="checkbox"/> přímá řeč (W)	290.67	7	45
<input type="checkbox"/> překladatel (W)	580.81	119	43,527	<input type="checkbox"/> korpusový manažer (W)	281.68	6	0
<input type="checkbox"/> EU (W)	562.35	12	0	<input type="checkbox"/> název živočichů (W)	281.68	6	0
<input type="checkbox"/> dicend (W)	557.73	12	42	<input type="checkbox"/> tvar slova (W)	281.68	6	0
<input type="checkbox"/> lemma (W)	550.63	15	1,400	<input type="checkbox"/> určité slovo (W)	257.94	6	32
<input type="checkbox"/> AutoSuggest (W)	515.57	11	0	<input type="checkbox"/> český text (W)	257.01	9	221
<input type="checkbox"/> Alt (W)	515.57	11	0	<input type="checkbox"/> jednota tlumočnicků (W)	234.90	5	0
<input type="checkbox"/> verbum (W)	468.95	14	2,021	<input type="checkbox"/> obyčejná mezera (W)	234.90	5	0
<input type="checkbox"/> terminologický (W)	443.68	20	5,632	<input type="checkbox"/> pevná mezera (W)	234.90	5	0
<input type="checkbox"/> Memsourka (W)	422.01	9	0	<input type="checkbox"/> počítač podporovaného překladu (W)	234.90	5	0

Co překladatel najde v paralelních korpusech

Hledání obvyklých i méně častých **ekvivalentů** v druhém jazyce

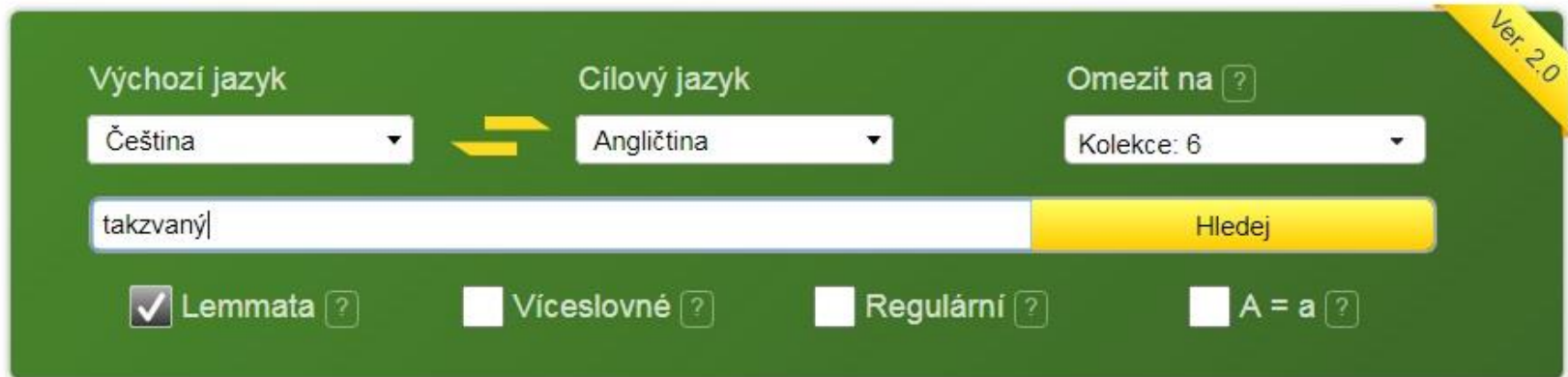
- Jaké anglické ekvivalenty má české slovo *takzvaný*? A skutečně má české *sice* v angličtině „nulový“ ekvivalent?

- 1) V konkordancích
- 2) V zarovnání slov (Treq)

Treq (ČNK)

Treq

DATABÁZE
PŘEKLADOVÝCH
EKVIVALENTŮ



Ver. 2.0

Výchozí jazyk: Čeština

Cílový jazyk: Angličtina

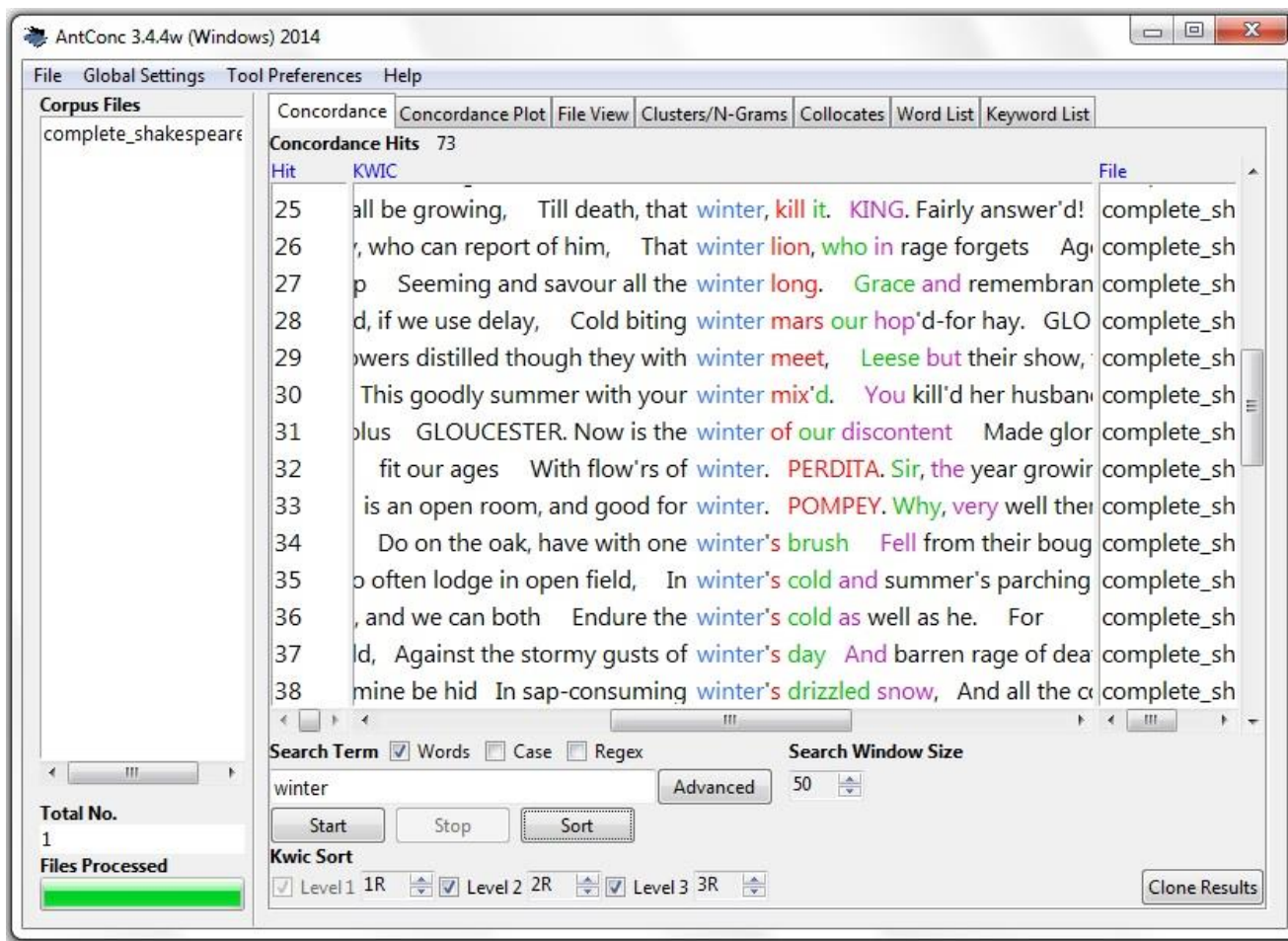
Omezit na: Kolekce: 6

takzvaný


Hledej

Lemmata Víceslovné Regulární A = a

Vlastní korpus v počítači – AntConc



Vlastní korpus Sketch Engine – konkordance (J. Levý: Umění překladu)

Query **překlad** 908 (7,402.52 per million) 

[First](#) | [Previous](#) Page of 46 [Go](#) [Next](#) | [Last](#)

file516754...	jazyků a z vlivu jazykové stylizace předlohy na	překlad	plynou překladateli některé jazykové
file516754...	některé jazykové nevýhody z faktu, že výraz	překladu	je nepůvodní, že myšlenka je ex post
file516754...	barvy, která jinde nutně vybledne. Zaorálek v	překladu	Grimmels- hausenova Simplicia nemohl
file516754...	správné, ale přece jen trochu umělé. V	překladech	vůbec a zvláště v překladech ze západních
file516754...	jen trochu umělé. V překladech vůbec a zvláště v	překladech	ze západních jazyků je nápadně mnoho vztažných
file516754...	: když už měla tváře celé umazané od piva... A tak se	překlady	z francouzštiny přímo hemží výrazy typu šel se
file516754...	desítiletí pak pracuje neměnným jazykem. Naše	překlady	z let dvacátých, třicátých a i mnohé novější se
file516754...	časech užívaných podle originálu. Proto	překlad	obyčejně rychleji jazykově zastarává než dílo
file516754...	k nevhodnému poetizování. V soutěži na	překlad	Goethovy lyriky, kterou v roce 1953 vypsalo
file516754...	mne! I tento převod, který patřil mezi vítězné	překlady	soutěže, nese stopy nepůvodní, odvozené díky
file516754...	“, sentiment a stylistické manýry v	překladu	. 2. Vnějšíková charakteristika postav: herec v
file516754...	intenzivních slov, nadbytečné zdobněliny v	překladech	dětské literatury apod. 3. „Přirozená hra“:
file516754...	. Z odvozenosti jazykového vyjádření v	překladu	vyplývá, že výraz v přeloženém díle není nutný,
file516754...	. U některých jazykových prostředků je výraz	překladu	jednoznačně určen, protože domácí jazyk má k
file516754...	, směřuje ukázněný výběr k reprodukčnímu cíli	překladu	. Ničím nespoutanou možností volby mezi
file516754...	třicet jedna slov). Obdobný poměr je mezi třemi	překlady	vůbec: v našem monologu 243 : 381 : 325 slovům.
file516754...	243 : 381 : 325 slovům. Příčinou špatných	překladů	nebývá jen povrchnost. Někdy právě naopak
file516754...	improvizovaná verze, a že další opracovávání	překlad	naopak znehodnocuje. Psycholingvistickým
file516754...	jím pak někdy brání stylisticky diferencovat	překlad	. Požadavky, které na překladatele klade jeho
file516754...	a stylistické nadání. III Estetické problémy	překladu	A. Tvůrčí reprodukce 1. Překladatelství jako

[First](#) | [Previous](#) Page of 46 [Go](#) [Next](#) | [Last](#)

Vlastí korpus Sketch Engine – kolokace (lemma_lc, -1 až 1)

Collocation candidates

Page [Next >](#)

	<u>Cooccurrence count</u>	<u>Candidate count</u>	<u>T-score</u>	<u>MI</u>	<u>logDice</u>
P N teorie	40	117	6.187	5.529	10.320
P N v	84	2,635	7.036	2.106	9.601
P N český	25	487	4.278	2.793	9.197
P N při	19	209	4.003	3.618	9.122
P N umělecký	16	146	3.729	3.887	8.958
P N umění	12	76	3.301	4.414	8.642
P N o	18	633	3.138	1.941	8.580
P N strojový	9	10	2.975	6.925	8.327
P N doslovný	9	19	2.953	5.999	8.313
P N na	17	1,000	2.327	1.199	8.189
P N pro	10	403	2.218	1.745	7.965
P N jungmannův	7	10	2.617	6.563	7.965
P N kritika	7	26	2.573	5.184	7.940
P N básnický	7	68	2.455	3.797	7.876
P N funkce	7	76	2.433	3.637	7.864
P N svůj	8	292	2.064	1.887	7.771
P N dva	8	296	2.053	1.868	7.766
P N dějiny	6	26	2.370	4.962	7.717
P N že	10	662	1.612	1.029	7.705
P N literární	6	95	2.162	3.092	7.614
P N francouzský	6	128	2.062	2.662	7.568
P N německý	6	135	2.041	2.585	7.558
P N u	7	315	1.764	1.585	7.551
P N čtenář	6	168	1.941	2.270	7.513
P N zpětný	5	6	2.216	6.814	7.485
P N saudkův	5	12	2.196	5.814	7.476
P N slovenský	5	13	2.193	5.699	7.474
P N realistický	5	20	2.169	5.077	7.463

Korpus „on demand“ (Sketch Engine)

WebBootCaT: Create corpus ?

[Get seed words from Wikipedia](#)

Corpus name

Language ▼
WebBootCaT is available for languages that can be automatically tokenised.

Input type

Seed words
 URLs
 Website

Select Seed words for finding documents using a search engine. Use URLs to download texts directly from specified locations. Switch to Website in case you need to obtain all documents from a particular web domain.

Seed words

Random tuples will be selected from the seed words to query a search engine. Input 3 to 20 words or multiword expressions. Use space as separator. Enclose multiword expressions in quotes (").

Compile corpus when finished Automatically compile corpus when WebBootCaT processing is finished.

[Show advanced options](#)

Paralelní korpus vytvořený z paměti

Create parallel corpora from multilingual data

Corpus name

Source file Soubor nevybrán

Supported formats include [Translation Memory eXchange \(TMX\)](#), [XML Localisation Interchange File Format \(XLIFF 2.0\)](#) or any spreadsheet (i.e. Excel document).
Spreadsheets must contain a header with language names in the first row and then aligned segments side by side in each row, one column per language.

Česko-baskický korpus Malý princ (nahrané TMX): lemma *beránek*

Query **beránek** 10 > Filter by aligned corpus 10 (14,184.40 per million) ⓘ

MP eu-cs, Czech

zvláštní hlásek : „ Prosím pěkně ... nakresli mi **beránka** ... ” „ Cože ? ” „ Nakresli mi beránka ... ”
nakresli mi beránka ... ” „ Cože ? ” „ Nakresli mi **beránka** ... ” Vyskočil jsem , jako by do mne hrom uhodil .
, opakoval : „ Prosím pěkně ... nakresli mi **beránka** ... ” Když stojíme před příliš velkou záhadou ,
kreslit . Odpověděl mi : „ To nevadí . Nakresli mi **beránka** . ” Poněvadž jsem nikdy nekreslil beránka ,
mi beránka . ” Poněvadž jsem nikdy nekreslil **beránka** , nakreslil jsem mu jednu z těch dvou kreseb ,
místa . U mne doma je všechno malinké . Potřebuji **beránka** . Nakresli mi beránka . ” Tak jsem tedy kreslil .
malinké . Potřebuji beránka . Nakresli mi **beránka** . ” Tak jsem tedy kreslil . Díval se pozorně a pak
, shovívavě usmál : „ Ale podívej se ... to není **beránek** , to je beran . Má rohy ... ” Tak jsem kresbu znovu
. „ Právě tak jsem to chtěl ! Myslíš , že ten **beránek** bude potřebovat hodně trávy ? ” „ Proč ? ” „
” „ Jistě to postačí . Dal jsem ti docela malého **beránka** . ” Sklonil se nad kresbu . „ No , není tak moc malý

MP eu-cs, Basque

Zera ziostan : - Marraz iezadazu arkume bat mezedez !
Marraz iezadazu arkume bat ...
Marraz iezadazu arkume bat ... arren ! ...
Berdin zait , marraz iezadazu arkume bat .
Arkumerik sekulan marraztu ez nuenez , nik ezagutzen nituen bi marrazkietarik bat berregin nuen harentzat , boa itxiarena hain xuxen .
Arkume bat behar dut nik .
Marraz iezadazu arkume bat .
Hori ez da arkumea hori aharia da ... Adarrak ba ditu .
Arkume honek belar asko beharko duela uste al duzu ?
Arkume guziz txiki bat egin dizut . Burua marrazkíra makurtu zuen :

CAT ≠ MT

CAT

- Při překládání poskytuje překladateli podporu, především na základě překladové paměti (TM) a terminologické databáze
- Příklady: SDL Trados Studio, Memsource, Wordfast

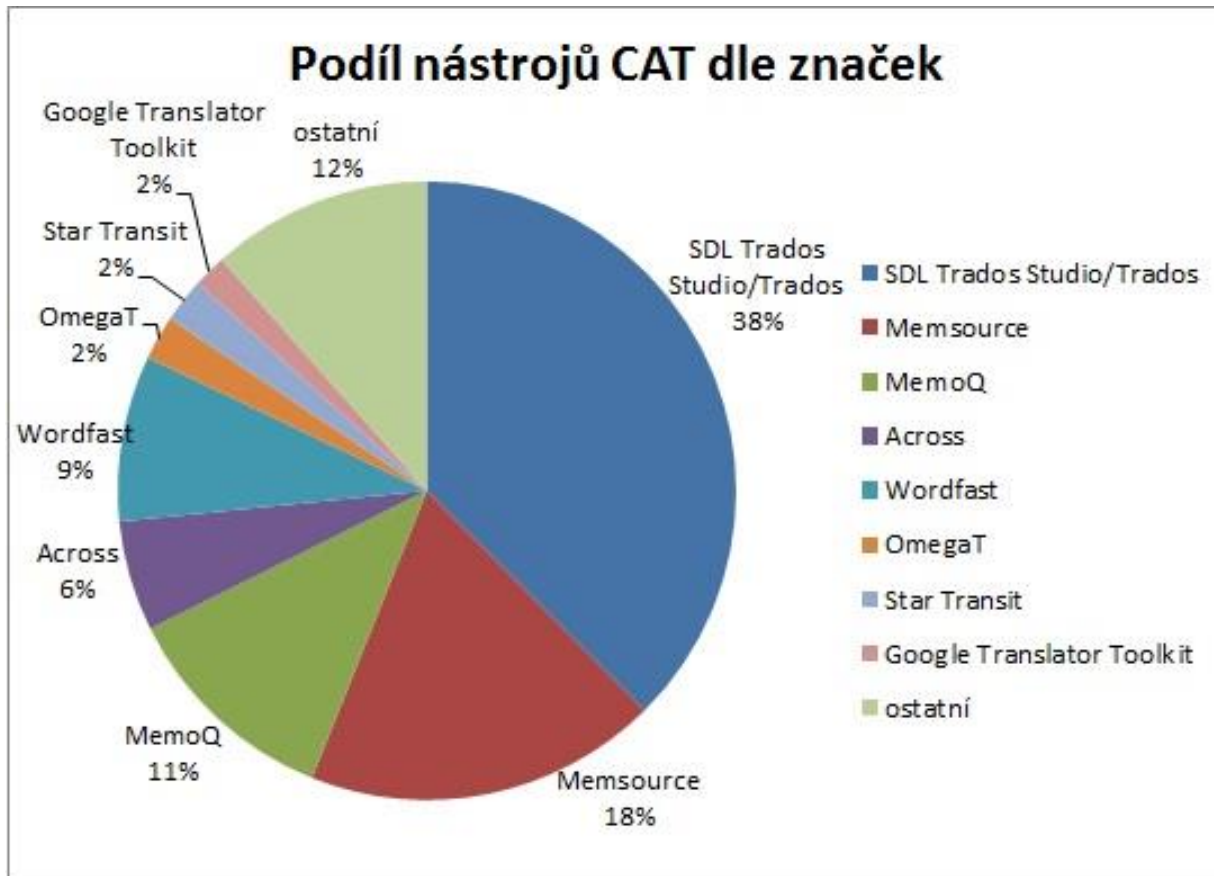
MT

- Sám překládá místo překladatele
- Příklady: Překladač Google, Microsoft Translator; Slate Desktop

CAT – hlavní funkce

- Segmentace
- Hledání shod (úplných a částečných) v paměti
- Konkordance
- Propojení s terminologickou databází
- Našeptavač
- Autopropagace
- Filtrování
- Jako textový editor: kontrola pravopisu, sledování změn, komentáře
- Kontrola kvality
- Vytvoření finálního dokumentu

Anketa: které nástroje čeští překladatelé používají?



Anketa: funkce CAT – kde jsou rezervy?

- Jen 67,78 % respondentů využívá propojení s terminologickými databázemi
- Jen 51,11 % používá našeptavač
- Jen 26,67 % používá zarovnač (aligner) – 5,56 % „Nevím“

SDL Trados Studio - našeptavač

The screenshot shows the SDL Trados Studio interface. The main window displays a table of fragment matches between English and Czech. The 'Term Recognition' window is open, showing a list of terms and their matches. The bottom of the screen shows a translation table with source and target text, along with quality scores and a glossary of terms.

Fragment ID	Source Text	Target Text
1	THE REPUBLIC OF MAURITIUS	MAURICIJSKÁ REPUBLIKA
2	Union vessels	plavidla Unie
3	Fishing zone	Rybolovná oblast
4	Fishing zone	Oblast rybolovu
5	Fishing opportunities	Rybolovná práva
6	Financial contribution	Finanční příspěvek

Segment	Source Text	Target Text	Quality
8	(1)	"1"	92%
9	On 28 January 2014, the Council adopted Decision 2014/146/EU hyperlink (1) hyperlink on the conclusion of the Fisheries Partnership Agreement between the European Union and the Republic of Mauritius ('the Agreement').	Rada dne 28. ledna 2014 přijala rozhodnutí 2014/146/EU [1] o uzavření smlouvy o rybolovu mezi Evropskou unií a Mauricijskou republikou.	90%
10	http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017D1960	http://eur-lex.europa.eu/legal-content/CZ/TXT/?uri=CELEX:32017D1960	AT
11	(2)	"2"	92%
12	The first Hyperlink to the Agreement set out, for a period of three years, the fishing opportunities granted to Union vessels in the fishing zone under the sovereignty or jurisdiction of the Republic of Mauritius ('Mauritius') and the financial contribution granted by the Union.	V prvním protokolu k dohodě o rybolovu mezi Evropskou unií a Mauricijskou republikou, který se vztahuje na období tří let, jsou poskytnuty rybolovné příležitosti pro člunům Unie v rybolovné oblasti pod suverénní nebo jurisdikcí Republiky Mauritius ('Mauritius') a finanční příspěvek poskytnutý Unií.	TC+
13	The period of application of that Protocol expired on 27 January 2017.	Období uplatnění této smlouvy vypršelo 27. ledna 2017.	TC+

Rozhodnutí Rady (EU) 2017/1960 ze dne 23. října 2017

Anketa: největší výhoda CAT?

- **Vyšší efektivita, vyšší rychlost a úsporu času** – („s CAT přeložím za stejný čas více textu než bez CAT“, „nerada překládám stejný text dvakrát“, „ušetří to datlování slov“)
- **konzistence**, důslednost a jednotnost,
- **jednoduché vyhledávání** v aktuálním textu i **v překladových pamětech** („možnost vyhledávání starších řešení ve vlastních pamětech místo procházení stovek jednotlivých přeložených textů“)
- propojení **s terminologickou databází**
- snazší **formátování** („nemusím řešit grafickou stránku textu“)
- **přehledné** a jednotné rozhraní („rozdělení do okýnek“), díky němuž se „nestane, že nějaký odstavec zůstane nepřeložený nebo že něco vypustím“
- **kontrola kvality**
- lepší **přehled o postupu** („vidím, kolik procent mám přeloženo“)
- Další: **ergonomie, lepší organizace práce** („klade to nižší nároky na mou systematičnost“), nástroje CAT usnadňují **týmovou spolupráci**

CAT – a co literární překladatelé?

- Nevýhody (skutečně nebo domněle) převažují nad výhodami
- Mohou ale využít jen některé funkce nebo některé malé programy (zarovnač + prohlížeč – pro vyhledávání, kontrolu jednotnosti, chybějících vět atd.)

Zamyšlení a upozornění

- Některé „parametry“ jsou pro postavení překladatele zásadní: „zdarma“ může být hodně drahé
- Kvalitní data jsou cenným aktivem, které přináší konkurenční výhodu (řešení pro budoucí překlady, jako základ CAT, ale i MT!!!) – pokud toto aktivum získává agentura a překladatel nikoliv, je to pro něj strategické/přijatelné? Odpovídá tomu sazba?
- Překladatel více nahraditelný
- „Budoucnost je v cloudu“ vs. citlivost obsahu

Strojový překlad

- Frázový strojový překlad

	0	10	20	30	40	50	60	70	80	90	100	110	120																		
1636	be	happy	only		,	budou	spokojeni	,	jen		1	0.5	0.5	0.0135791		0-1	1-2	2-4		1	2	1									
1637	be	happy	only		budou	spokojeni	,	jen		1	0.5	0.5	0.0672783		0-0	1-1	2-3		1	2	1										
1638	be	happy		,	budou	spokojeni	,		1	1	0.25	0.0135791		0-1	1-2		1	4	1												
1639	be	happy		,	budou	spokojeni		1	1	0.25	0.0672783		0-1	1-2		1	4	1													
1640	be	happy		budou	spokojeni	,		1	1	0.25	0.0672783		0-0	1-1		1	4	1													
1641	be	happy		budou	spokojeni		1	1	0.25	0.333333		0-0	1-1		1	4	1														
1642	be	it	in	shaqlawa	or	qaladze	,	he	went		ať	už	v	šaklavě	nebo	v	kaladze	,	všude		1	0.0277144	1	3.7386e-005		0-0	1-				
1643	be	it	in	shaqlawa	or	qaladze	,		ať	už	v	šaklavě	nebo	v	kaladze	,		1	0.110858	1	7.10334e-005		0-0	1-1	2-2	3-3	4-4				
1644	be	it	in	shaqlawa	or	qaladze		ať	už	v	šaklavě	nebo	v	kaladze		1	0.1875	1	8.76955e-005		0-0	1-1	2-2	3-3	4-4	5-6					
1645	be	it	in	shaqlawa	or		ať	už	v	šaklavě	nebo	v		1	0.1875	0.5	8.76955e-005		0-0	1-1	2-2	3-3	4-4		1	2	1				
1646	be	it	in	shaqlawa	or		ať	už	v	šaklavě	nebo		1	0.1875	0.5	0.00477941		0-0	1-1	2-2	3-3	4-4		1	2	1					
1647	be	it	in	shaqlawa		ať	už	v	šaklavě		1	0.25	1	0.00637255		0-0	1-1	2-2	3-3		1	1	1								
1648	be	it	in		ať	už	v		1	0.25	1	0.0254902		0-0	1-1	2-2		1	1	1											
1649	be	it		ať	už		1	0.5	1	0.0666667		0-0	1-1		1	1	1														
1650	be	safe	.		budou	v	bezpečí	.		0.25	0.240881	0.5	0.0815603		0-0	1-1	1-2	2-3		4	2	1									
1651	be	safe	.		tam	budou	v	bezpečí	.		0.25	0.240881	0.5	0.00149652		0-1	1-2	1-3	2-4		4	2	1								
1652	be	safe		budou	v	bezpečí		0.25	0.371795	0.5	0.0833333		0-0	1-1	1-2		4	2	1												
1653	be	safe		tam	budou	v	bezpečí		0.25	0.371795	0.5	0.00152905		0-1	1-2	1-3		4	2	1											
1654	be	with	my	sister	.		být	s	mou	sestrou	.		1	0.370221	1	0.00741457		0-0	1-1	2-2	3-3	4-4		1	1	1					
1655	be	with	my	sister		být	s	mou	sestrou		1	0.571429	1	0.00757576		0-0	1-1	2-2	3-3		1	1	1								
1656	be	with	my		být	s	mou		1	0.571429	1	0.00757576		0-0	1-1	2-2		1	1	1											
1657	be	with		být	s		1	0.571429	1	0.0833334		0-0	1-1		1	1	1														
1658	be		,	budou		1	1	0.2	0.0672783		0-1		1	5	1																
1659	be		ať		1	1	0.2	0.166667		0-0		1	5	1																	
1660	be		budou		0.25	1	0.2	0.333333		0-0		4	5	1																	
1661	be		být		1	1	0.2	0.166667		0-0		1	5	1																	
1662	be		tam	budou		0.25	1	0.2	0.0061162		0-1		4	5	1																

Překladač vytvořený z paměťí/paralelního korpusu

Podniky a trhy
E15.cz
GESTO DESETILETÍ: Deset projektů, které změnilo Česko

Komentáře
Komentář Bohumila Pečinky: Babišovy svázané
Bohumil Pečinka

Zprávy ▾ Volby 2017 Byznys ▾ Názory ▾ FinExpert ▾ Trendy a technologie ▾

E15.cz > E-svět > Technika > Čeští lingvisté vytvořili překladač, který pomáhá po tajfunu na Filipínách

Čeští lingvisté vytvořili překladač, který pomáhá po tajfunu na Filipínách



Tajfun na Filipínách má už několik desítek obětí (9. prosince 2014) • ZDROJ: ČTK

pav
9. prosince 2014 • 19:55

To se mi líbí 0 Sdílet Tweet

MT v CAT (Slate Desktop v SDL Trados Studiu)

The screenshot displays the SDL Trados Studio interface for a translation project. The main window is titled "Editor" and shows a list of segments with their source text, target text, and status. The segments are numbered 1 through 28. The source text for segments 1-3 is: "Small and medium-sized enterprises (SMEs) should play a crucial role in reaching the Europe 2020 Strategy objectives." The target text for these segments is: "pro malé a střední podniky (MSP) hrát podstatnou úlohu při dosahování cílů strategie Evropa 2020." The status for these segments is "AT".

Segment	Source Text	Target Text	Status
1	Small and medium-sized enterprises (SMEs) should play a crucial role in reaching the Europe 2020 Strategy objectives.	pro malé a střední podniky (MSP) hrát podstatnou úlohu při dosahování cílů strategie Evropa 2020.	AT
2	Small and medium-sized enterprises (SMEs) should play a crucial role in reaching the Europe 2020 Strategy objectives.	malých a středních podniků(4) hrají klíčovou roli při Evropy 2020 strategie mezinárodními.	AT
22	It also emphasises the importance of reinforcing the growth of the European economy while delivering a high level of employment, a low-carbon, resource- and energy-efficient economy and social cohesion.	dynamické a konkurenceschopnost. dále podtrhuje význam posílení hospodářského růstu evropského hospodářství, přičemž budou vysoké úrovně zaměstnanosti a nízkouhlikových technologií v energeticky resource- a hospodářství a sociální soudržnosti.	AT
23	Small and medium-sized enterprises (SMEs) should play a crucial role in reaching the Europe 2020 Strategy objectives.	pro malé a střední podniky (MSP) hrát podstatnou úlohu při dosahování cílů strategie Evropa 2020.	AT
24	Their role is reflected by the fact that SMEs are mentioned in six out of seven of the flagship initiatives of the Europe 2020 Strategy.	Tyto činnosti se odráží skutečnost, že podniky, které jsou uvedeny v šesti ze sedmi se stěžejních iniciativ strategie Evropa 2020.	AT
25	(2)	(2)	100% AT
26	In order to ensure that enterprises, particularly SMEs, play a central role in delivering economic growth in the Union, which is a top priority, the Commission adopted a Communication entitled "An Integrated industrial policy for the globalization era, putting competitiveness and sustainability at centre stage" in October 2010, which was endorsed by the Council of December 2010.	Aby bylo zajištěno, že podniky, zejména malé a střední podniky, budou hrát ústřední úlohu v oblasti hospodářského růstu v Unii, který je navíc ve sdělení Komise s názvem „Roční Integrovaná průmyslová politika pro globalizaceERA, pro konkurenceschopnost a udržitelnost ve fázi“ v říjnu 2010, který byl schválen Radou v prosinci 2010.	AT
27	This is a flagship initiative of the Europe 2020 Strategy. The Communication sets out a strategy aiming to boost growth and jobs by maintaining and supporting a strong, diversified and competitive industrial base in Europe, in particular by improving framework conditions for enterprises and strengthening several aspects of the internal market, including business-related services.	Jedná se o stěžejní iniciativě strategie Evropa 2020. Sdělení stanoví strategie zaměřené na posílení hospodářského růstu a pracovních míst, podporou a podpora silná diverzifikované a průmyslové konkurenceschopnosti Evropy, zejména v zlepšení rámcových podmínek pro podniky a rovněž různých aspektů vnitřního trhu, včetně business-related.	AT

Nebezpečí spojená s posteditací

- Důležitá oblast – proč se o ní diskutuje tak málo?
- Posteditace je náročnější než překlad: překladatel musí za pracovní den zpracovat **větší objem textu** (a informací), dá se předpokládat, že bude muset řešit hlavně ty **největší překladatelské problémy**: z překladatele řešitelem problémů
- Náročnější i psychicky
- Hrozí, že posteditor **otupí**, spokojí se s “minimaxem“, bude ztrácet cit pro jazyk a soudnost, rychleji vyhoří?

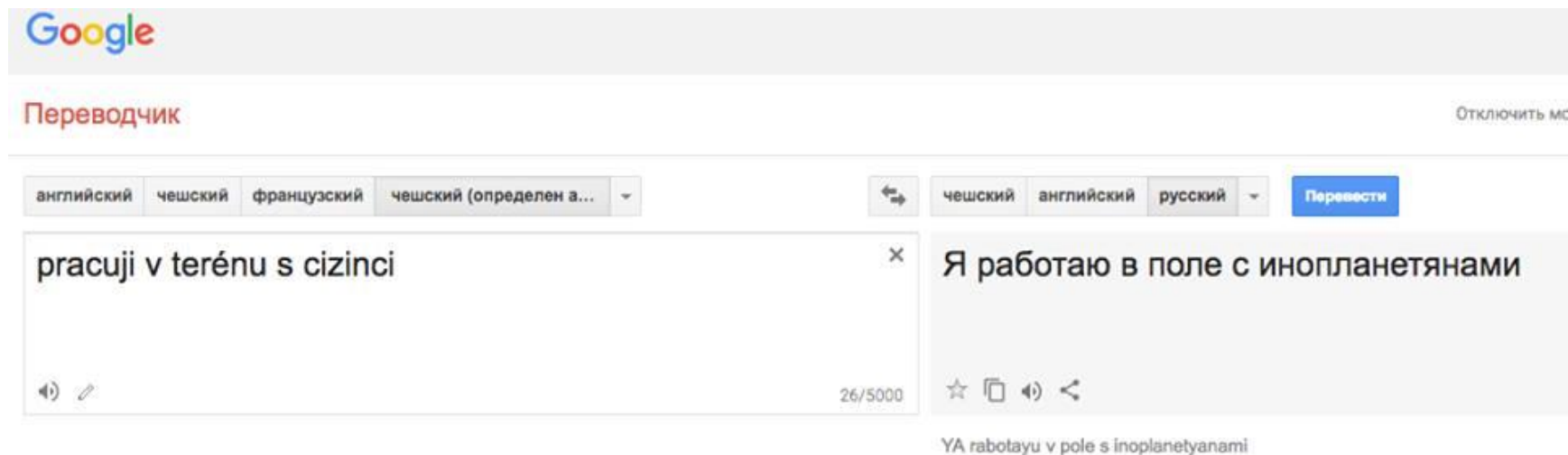
Několik závěrů

- Technologické nástroje jsou dobrý sluha, ale špatný pán
- Jazykovědci a translatologové by ve svých modelech měli začít brát v potaz i počítač (a nástroje CAT a MT)
- Dnes už překladatel nevystačí s tím, co se naučil ve škole – celoživotní vzdělávání a úloha škol, profesních organizací
- Nutnost investovat – ale také investice promítat do cen!
- Praktici by si měli uvědomovat, co je v jejich zájmu, a nepodřezávat si pod sebou větev – diskuse i v profesních organizacích (slevy za shody, co když mi nezůstane TM, posteditace aj.)

Budoucnost

- Co nám říkají šachy... - naučit se pracovat se „stroji“
- Nutnost sledovat vývoj technologií, hájit své pozice, hájit kulturu jazykového vyjadřování

Zas tak horké to nebude



The screenshot shows the Google Translate interface. At the top left is the Google logo. Below it, the word "Переводчик" (Translator) is written in red. On the right side, there is a link "Отключить MS". The main interface has a language selection bar with "английский", "чешский", "французский", and "чешский (определен а...". A swap button is in the center, and the target language is set to "русский". A blue "Перевести" button is on the right. The input text box contains "pracuji v terénu s cizinci" and has a character count of "26/5000". The output text box contains the Russian translation "Я работаю в поле с инопланетянами". Below the output, there are icons for star, copy, speaker, and back, followed by the text "YA robotayu v pole s inoplanetyanami".

TranslatoBlog.cz
Knihy-Apostrof.cz

