

Studijski posjet Sveučilištu u Göteborgu (11. – 15. veljače 2019.)

Suradnica na projektu *Hrvatski mrežni rječnik – Mrežnik* Daria Lazić posjetila je Sveučilište u Göteborgu u Švedskoj kako bi se upoznala s njihovim leksikografskim radom te izradom jezičnih alata i digitalnih izvora za švedski jezik. Sveučilište u Göteborgu zanimljivo je iz e-leksikografske perspektive zbog dvaju znanstvenoistraživačkih centara koji ondje djeluju: centra za jezične tehnologije Språkbanken i Centra za leksikologiju i leksikografiju (u kojem se među ostalim sastavljaju dva rječnika Švedske akademije).

Språkbanken

[Språkbanken](#) je osnovan 1975. s ciljem prikupljanja i razvoja jezičnih izvora te njihovog stavljanja na raspolaganje znanstvenoj i široj javnosti. Ta se njegova funkcija održala do danas te se danas ondje prikupljaju i izrađuju jezični izvori poput velikih računalnih korpusa i raznih leksikona i razvijaju se alati za njihovu obradu. Osim toga, provode se i istraživanja na području jezičnih tehnologija. Od 2018. Språkbanken je dio Nacionalne jezične banke (*Nationella Språkbanken*), projekta čiji je cilj razvoj nacionalne e-infrastrukture za istraživanja temeljena na jezičnim podacima, te njezin koordinator. U Språkbankenu se prvenstveno radi na obradi pisanoga jezika, dok se u drugim uključenim ustanovama razvija obrada govorenoga jezika, digitalizacija dijalektnih zbirki i drugo.

Språkbanken ima 30-ak zaposlenika, koji se mogu podijeliti na stručnjake i programere. Stručnjaci su znanstvenici koji su među ostalim zaduženi za komunikaciju s korisnicima izvora i alata koji se razvijaju u Språkbankenu te za pružanje podrške i davanje informacija o mogućnostima uporabe postojećih tehnologija ili razvoja novih za potrebe određenoga istraživanja. Korisnici Språkbankenovih usluga uglavnom su lingvisti i istraživači u području jezičnih tehnologija, ali i humanističkih i društvenih znanosti općenito.

U Språkbankenu se provodi velik broj projekata, koji primjerice obuhvaćaju razvoj jezičnih tehnologija za uporabu u:

- istraživanju jezičnih varijacija (analiza povijesnih podataka, računalna morfologija, poredbeni lingvistika)
- analizi sadržaja (razlučivanje i promjena značenja, analiza sentimenta, rano otkrivanje demencije)
- učenju jezika (analiza složenosti tekstova, računalno potpomognuto učenje).

Rezultati istraživanja dostupni su na mrežnim platformama s otvorenim pristupom kojima se može pristupiti sa stranica Språkbankena. To su među ostalim sustav za pretraživanje korpusa Korp, leksička infrastruktura Karp, alat za anotiranje teksta Sparv i platforma za učenje švedskoga Lärka.

Korp

[Korp](#) je sustav za pretraživanje korpusa razvijen u Språkbankenu. Trenutačno sadržava 462 korpusa i više od 15 milijardi pojava, a korpusi obuhvaćaju korpusne suvremenoga švedskog jezika (akademski tekstovi, književni, medicinski, tekstovi s društvenih medija...), povijesne i paralelne korpusne te manji broj korpusa na drugim jezicima. Program je besplatan te ga je

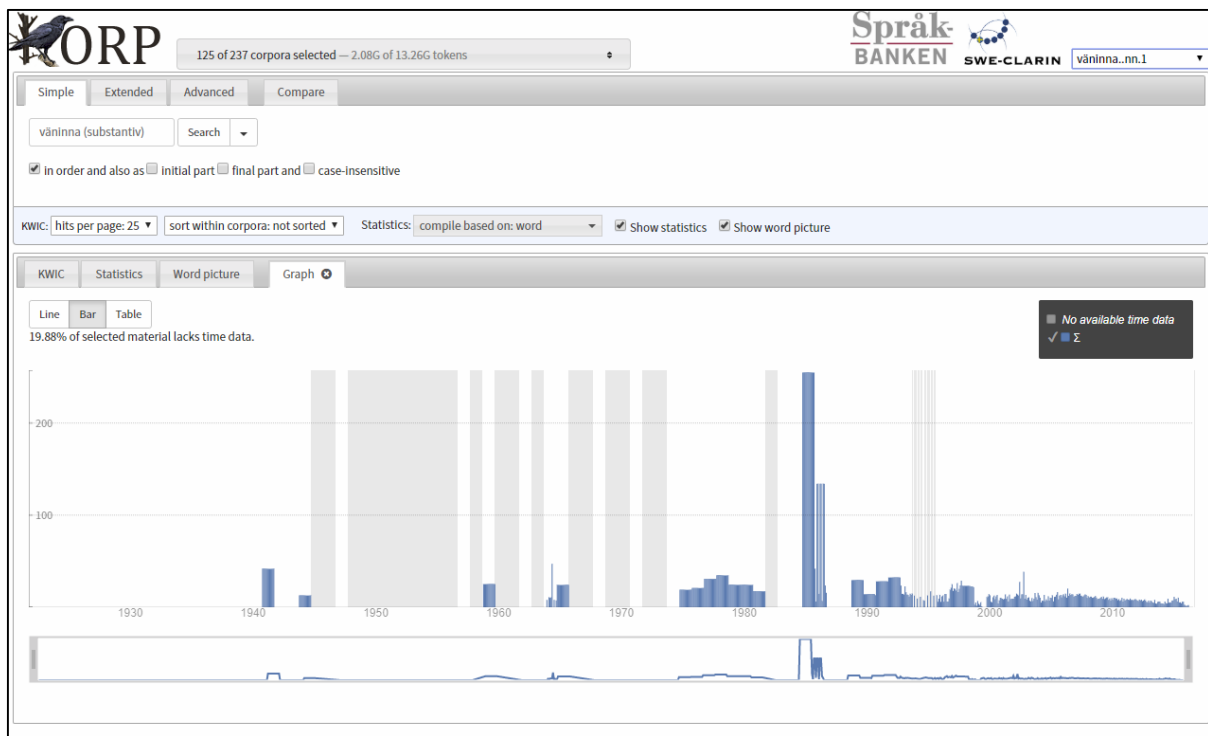
moguće preuzeti i prilagoditi za druge jezike, a u slučaju preuzimanja Språkbanken nudi besplatnu podršku.

Sustav nudi mogućnost jednostavnoga (riječ, višerječni izraz ili lema), proširenoga (prema određenim kategorijama) i naprednoga (s pomoću upitnoga jezika) pretraživanja korpusa. Zanimljivo je da je unutar proširenoga pretraživanja višeznačne riječi moguće pretraživati prema značenjima, što još nije moguće u dostupnim korpusima hrvatskoga jezika, a bilo bi korisno u leksikografskome radu. Nadalje, sustav ima nekoliko mogućnosti prikaza rezultata: KWIC (ključna riječ u kontekstu), statistika (pojavnice u pojedinim korpusima) te „slika riječi“ (*ordbild*), koja slično skicama riječi u Sketch Engineu prikazuje najčešće kolokacije podijeljene prema vrsti sintaktičkoga odnosa među sastavnicama.

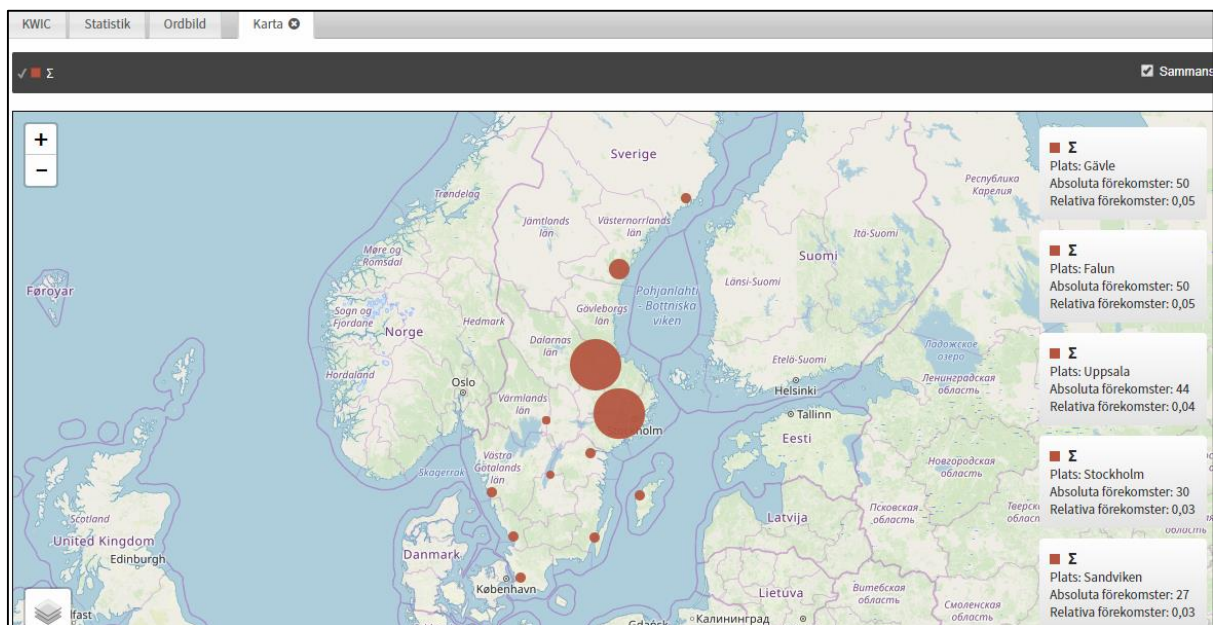
hund (substantiv)														
Preposition			Attribut			hund			Efterställt Attribut			Hund verb		
1. med	5016		1. liten	2207		1. heta	248		1. bita	227		1. skaffa	1369	
2. om	1054		2. söt	584		2. vara	472		2. skälla	166		2. ha	4781	
3. hos	242		3. röd	589		3. skälla	88		3. dö	192		3. rasta	372	
4. till	1166		4. blöt	269		4. ha	234		4. ha	774		4. rasta ²	372	
5. utan	190		5. lös ²	218		5. av ras	57		5. springa	152		5. köpa ²	596	
6. efter	377		6. lös	218		6. bli	118		6. attackera	105		6. köpa	596	
7. för	1194		7. gullig	176		7. springa	61		7. skaffa	94		7. skaffa sig	162	
8. vad för	78		8. egen	183		8. bita	45		8. bli	506		8. avliva	102	
9. mot	223		9. fin	413		9. i bil	61		9. avliva	63		9. bita	140	
10. bredvid	25		10. stackars	146		10. vilja	84		10. få	374		10. klappa ²	114	
11. i form	20		11. galen	159		11. kunna	122		11. heta	103		11. klappa	114	
12. åt	145		12. herrelös	74		12. i knä	47		12. rädda	75		12. slita	128	
13. till och med	21		13. död	154		13. skola ²	117		13. äta	87		13. lämna	244	
14. till och med	21		14. snäll	144		14. få	95		14. hitta	102		14. döda	139	
15. ihop med	19		15. ful	133		15. i koppel	28		15. bajs	29		15. se	451	

Slika riječi (*ordbild*) za riječ *hund* 'pas' u odabranim korpusima suvremenoga švedskog jezika

Osim toga, statističke podatke o pojavljivanju riječi u pojedinim korpusima moguće je grafički prikazati, osim tablično, i u obliku dijagrama trenda, koji prikazuje učestalost uporabe riječi tijekom vremenskoga razdoblja koje je obuhvaćeno odabranim korpusima, i karte koja prikazuje prostornu rasprostranjenost riječi. Prostorna rasprostranjenost riječi može se prikazati na temelju supojavljivanja te riječi i imena mjesta u kontekstu ili prema podrijetlu autora teksta za tekstove i korpusa za koje taj podatak postoji (npr. neki blogovi).



Dijagram trenda za riječ *väninna* 'prijateljica', koja se danas u švedskome rijetko upotrebljava



Prostorna rasprostranjenost oblika *tröck* (dijalektnog oblika preterita glagola *trycka* 'pritisnuti', za razliku od standardnojezičnoga *tryckte*) prema podrijetlu autora teksta

U korpusima u Korpu rečenice su zbog autorskih prava uglavnom pomiješane te je rečenica najviša razina pregleda. Trenutačno se radi na alatu [Strix](#) za analizu na razini teksta. U Strix su uključeni odabrani tekstovi iz Korpa koji nisu zaštićeni autorskim pravima.

Karp

[Karp](#) je leksička infrastruktura koja omogućava paralelno pretraživanje velikoga broja leksičkih izvora. Obuhvaća više od 40 izvora, od leksikona izrađenih u Språkbanken-u do digitaliziranih velikih povijesnih rječnika. Središnji je izvor s kojim je većina izvora povezana *Saldo*, odnosno švedski asocijativni leksikon. Saldo trenutno sadržava semantički i morfološki leksikon švedskoga jezika, a namijenjen je prvenstveno računalnoj uporabi za razvoj jezičnih alata, no pretraživ je među ostalim i u Karpu.

[Simple Search](#) [Extended Search](#) [Search History](#)

✎ Reset

Find entries [where](#) [baseform](#) [equals](#) [+ or...](#)

[+ and...](#) [+ except...](#)

Search

Compile on...

Hits **13**

Page: 1 / 1 [Go to...](#)

SALDO ▾ 2 HITS (DISPLAYING 2)

	SENSE	LEMGRAM	PART OF SPEECH	PRIMARY	SECONDARY	CHILDREN (PRIMARY)	CHILDREN (SECONDARY)
⚙	hund	hund (noun) 🔊 2044231	noun	djur	sällskap	Karo Pompe bandhund björnhund blandrashund more...	Idefix andeögon apportera bussa³ folkiilsken more...
⚙	hund²	hund (noun) 🔊 2044231	noun	usling			

SALDO MORPHOLOGY ▾ 1 HIT (DISPLAYING 1)

	LEMGRAM	PART OF SPEECH	INHERENT	PARADIGM	INFLECTION
⚙	hund (noun) 🔊 2044231	noun	u	nn_2u_sten	hund ...

Rezultati pretraživanja riječi *hund* 'pas' u leksikonu *Saldo* s pomoću sustava Karp

Sparv

[Sparv](#) je alat za anotiranje teksta koji omogućuje leksičku, tvorbenu i sintaktičku analizu, analizu sentimenta, prepoznavanje imenskih entiteta te analizu čitljivosti unesenoga teksta.

The screenshot displays the Sparv web interface. At the top, a text input area contains a paragraph in Swedish about Ikea. Below the input, there are checkboxes for various analysis types: Lexical analysis (checked), Compound analysis (unchecked), Dependency analysis (checked), Sentiment analysis (checked), and Named entity recognition (unchecked). There is also a checkbox for Readability metrics. Below these are buttons for 'Show advanced settings', 'Reset default settings', and a green 'Run!' button. There are also buttons for 'XML' and 'Show warnings'. The results section shows the XML output for the paragraph and sentence. Below this, a dependency parse tree is visualized with arrows indicating grammatical relations between tokens. At the bottom, a table provides detailed analysis for each token.

token	msd	lemma	lex	sense	sentimentclass	deprel
Ikea	PM. NOM	Ikea	ikea_pm.1	IKEA..1, Ikea..1		SS
(	PAD					IR
namnet	NN. NEU. SIN. DEF. NOM	namn	namn_nn.1	namn..1 (0.967), namn..2 (0.020), namn..3 (0.012)	neutral	SS
är	VB. PRS. AKT	vara	vara_vb.1	vara..1	neutral	MS
bildat	VB. SUP. AKT	bilda	bilda_vb.1	bilda..1 (0.992), bilda..2 (0.008)	neutral	+F

Anotiranje teksta u Sparvu

Lärka

Språkbanken sudjeluje u projektu izrade alata za procjenu švedskoga kao stranog jezika, u sklopu kojega se prikuplja korpus švedskoga za strance te se razvijaju metode i alati za njegovu obradu. Kao dio projekta nastala je i platforma za učenje švedskoga [Lärka](#), na kojoj je dostupno nekoliko alata za primjenu u učenju i nastavi. Platforma trenutačno sadržava vježbe za lingviste, primjerice studente jezika (npr. određivanje vrsta riječi), vježbe za učenje jezika (npr. pogađanje riječi), alat za procjenu teksta za švedski kao strani jezik, alat za pronalaženje primjera te alat za anotiranje teksta.

Primjer vježbe za učenje jezika vježba je slušanja i zapisivanja riječi. Vježba se generira iz korpusa u Korpu. Odabire se jezična razina kojoj pripadaju riječi koje će se vježbati, od A1 do C2 (u sklopu projekta razvija se model računalne procjene složenosti riječi i teksta za učenje jezika) te program daje zvukovne zapise (nastale s pomoću sintetizatora govora) riječi koje je potrebno prepoznati i zapisati. Kao pomoć moguće je odabrati primjere rečenica u kojima se riječ pojavljuje (generirane iz Korpa) te kao dodatnu pomoć početno slovo riječi.

S obzirom na to da će *Mrežnikov* modul za strance sadržavati zvukovni zapis i primjere uporabe riječi, slična bi se vježba možda mogla osmisлити i na temelju rječničke građe u *Mrežniku*.

LiWriX - Listen and Write Exercise

Mode: 0 Level: A1 [Change settings \(new game\)](#) → izbor jezične razine

Correct: 1/1 zvukovni zapis Score: 0 odgovor provjera

Där kan du titta på massor av gamla tv-serier , sportprogram , b _____ och mycket annat .
 De tycker att det är fel att skämta om sådana saker i ett b _____ .
 Partiet Hamas sänder b _____ i tv-kanalen al-Aqsa .
 I Sverige är det förbud med reklam för alkohol och reklam i b _____ .
 Förslaget betyder att det kan bli reklam i b _____ och alkoholreklam i kanaler som sänds i Sverige .
 Ett b _____ för tv som vill inspirera barn till eget berättande av både enkla som allvarliga ämnen .
 I Palestina visas b _____ där självmordsbombare hyllas och det israeliska folket beskrivs som djävulen själv .
 En afrikansk gatufest med Edo Bumba Band , dansgrupper , försäljning , b _____ och fortsatt diktläsning .
 Enligt reklamen fanns figuren med i ett b _____ på en tv-kanal .
 Många föräldrar i USA protesterar mot att Apple tillåter försäljning inne i b _____ .

😊

pomoć:
 • rečenice
 • početno slovo

Vježba za učenje švedskoga: slušanje i zapisivanje riječi

Vježba pogađanja riječi funkcionira na principu vješala: riječ se pogađa na temelju definicije i zadanoga broja slova. Kao pomoć može se prikazati odgovarajuća definicija na nekome od niza drugih jezika, među kojima je i hrvatski. Kao podloga za vježbu služe natuknice i pripadajuće definicije iz mrežnoga rječnika [Lexin](#) namijenjenoga osobama koje uče švedski, a koji osim švedske osnove sadržava i prijevode na brojne jezike.

Word guess

Tries: 0/7

Definition:

lämnar ifrån sig mot betalning

Hjälp:

prodati/prodavati

S Ä L J 

A B C D E F G H I K M N O P Q R T U V W X Y Z Ö Å É

Vježba za učenje švedskoga: pogađanje riječi

Alat za procjenu složenosti teksta označava riječi unesenoga teksta s obzirom na jezičnu razinu (od A1 do C2) na kojoj se može očekivati njezino pojavljivanje u pasivnoj i aktivnoj uporabi jezika.



Language Acquisition Reusing **Korp**

Imorgon ska jag på bio med min bästa vän .

What do you want to assess? 

Learner essay

Text readability

Show all words of the following CEFR level(s) 

☒ A1 

☒ A2 

☒ B1 

☒ B2 

☒ C1 

Additional options 

☐ Mark all potentially incorrect words

☐ Use Spellchecker

Edit text

Reset

Evaluation

Suggested overall level: A2

Procjena složenosti teksta za aktivnu uporabu jezika



Language Acquisition Reusing **Korp**

Imorgon ska jag på bio med min bästa vän .

What do you want to assess? ?

Learner essay **Text readability**

Show all words of the following CEFR level(s) ?

☒ A1 ☐

☒ A2 ☐

☒ B1 ☐

☒ B2 ☐

☒ C1 ☐

Additional options ?

☐ Mark all potentially incorrect words

☐ Use Spellchecker

Edit text **Reset**

Evaluation

Suggested overall level: A2

Procjena složenosti teksta za pasivnu uporabu jezika

Centar za leksikologiju i leksikografiju

[Centar za leksikologiju i leksikografiju](#) (*Lexikaliska institutet*) zadužen je za istraživanja u tim područjima. Njihova se leksikografska djelatnost može podijeliti na teorijsku leksikografiju i izradu rječnika i baza podataka. Centar surađuje sa Švedskom akademijom, najvećim naručiteljem rječnika u Švedskoj, te se ondje sastavljaju dva rječnika Švedske akademije: *Svenska Akademiens ordlista* i *Svensk ordbok*. Osim toga, ondje se nalazi i švedsko uredništvo islandsko-skandinavskoga mrežnog rječnika *Íslex*.

Rječnici Švedske akademije

Na portalu svenska.se slobodno su dostupna tri rječnika Švedske akademije: osim dvaju rječnika koji se sastavljaju na Sveučilištu u Göteborgu ondje se nalazi i povijesni švedski rječnik *Svenska Akademiens ordbok*.



Portal svenska.se na kojem su dostupni rječnici Švedske akademije

Svenska Akademiens ordlista (SAOL) sadržava popis riječi s osnovnim podacima kao što su vrsta riječi, jednostavna definicija i gramatički oblici. Pristup je pri njezinu sastavljanju preskriptivan te ona služi kao pravopisni i morfološki normativni priručnik. Na mreži je trenutno dostupno posljednje izdanje iz 2015. Iako je riječ o preskriptivnome pristupu, podatci se ipak posuvremenjuju na temelju morfološke baze koja se također izrađuje u Centru za leksikologiju i leksikografiju, a osim popisa morfoloških tipova i riječi koje im pripadaju sadržava i podatke o zastupljenosti pojedinih oblika u korpusu.

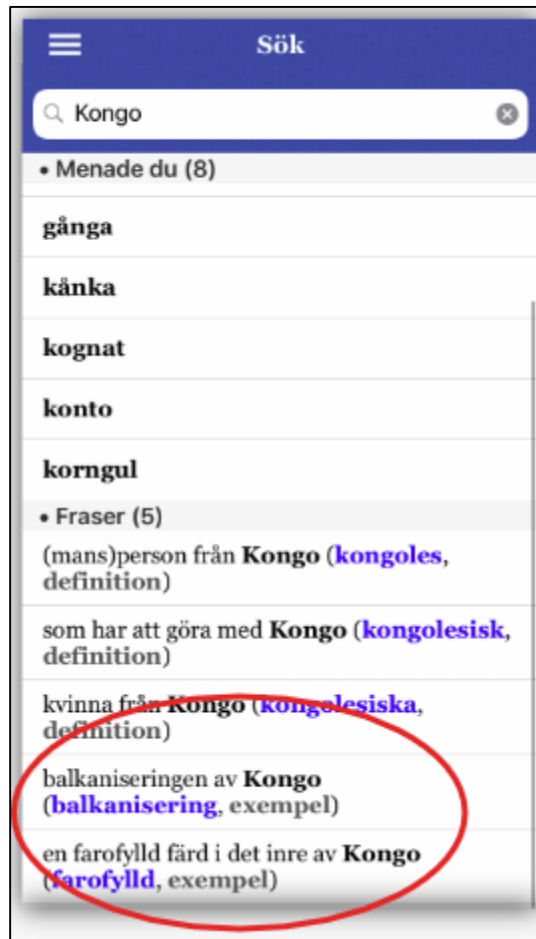
Razlika između opisa u *SAOL*-u i rječniku *Svensk ordbok (SO)*, u kojem se primjenjuje deskriptivan pristup, može se prikazati na primjeru natuknice *handsfree*. U *SAOL*-u uz natuknicu *handsfree* stoji normativna napomena da je bolje upotrebljavati pravopisno prilagođeni oblik *handfri*, dok su u *SO*-u ta dva oblika navedena kao sinonimi, bez davanja prednosti kojemu od njih. Međutim, s obzirom na to da podatci iz korpusa pokazuju da se oblik *handfri* gotovo i ne upotrebljava, u novome izdanju *SAOL*-a ta će normativna napomena biti izbačena.

SAOL	SO
tryckår: 2015	tryckår: 2009
¹hands free [hän´dsfri långt i] oböjligt substantiv • <i>en handsfree</i> – Använd hellre <i>handfri</i> , se ¹ <i>handfri</i> . Visa mer ☰	handsfree  [hän´dsfri långt i] substantiv, ingen böjning, n-genus ORDLED: hands-free • anordning i anslutning till mobiltelefon som gör att mobiltelefonen inte behöver manövreras med händerna utformad t.ex. med tanke på bilförare SYN. <i>handfri</i> Visa mer ☰
²hands free [hän´dsfri långt i] oböjligt adjektiv – Använd hellre <i>handfri</i> , se ² <i>handfri</i> . Visa mer ☰	

Rječnici Svenska Akademiens ordlista i Svensk ordbok

Posljednje izdanje rječnika *Svensk ordbok* objavljeno je 2009. u tiskanome obliku, a 2015. i u digitalnome, prvo kao mobilna aplikacija, a potom i na portalu svenska.se. Rječnik sadržava 65 000 natuknica s podacima o izgovoru, zvukovnim zapisom izgovora, gramatičkim oblicima, definicijama, primjerima uporabe, konstrukcijama i kolokacijama, poveznicama na značenjski povezane riječi poput sinonima i antonima, podacima o tvorbi i etimologiji riječi.

Trenutačno se radi na novome izdanju *SO*-a, koje bi trebalo biti objavljeno krajem godine, i to isključivo u digitalnome obliku. Naglasak je na posuvremenjavanju natuknica iako se u manjoj mjeri mijenja i njihov inventar. Posebna je pozornost posvećena sadržaju natuknica koji se preispituje iz rodne, nacionalne perspektive i slično. U opisu se nastoji postići ravnoteža između stvarne jezične uporabe kakva je, primjerice, dokumentirana u korpusima na kojima se temelji leksikografski opis i društvene osjetljivosti. Provjerava se, primjerice, kako su u rječniku prikazane pojedine zemlje i nacionalnosti, pripadnici rodni, vjerskih i drugih skupina, pojavljuju li se u primjerima samo švedska imena i tako dalje. Razmišljanje o korisnicima rječnika prisutno je i kod samoga oblikovanja rječničkoga članka: primjerice, pri oblikovanju primjera nastoje se izbjeći elementi koji bi odvlačili pozornost s riječi čija se uporaba opimjeruje. Nadalje, semantički i morfološki opis općenito se provjerava na temelju korpusa prikupljenih u Språkbanken i već spomenute vlastite morfološke baze na kojoj kontinuirano rade. Osim toga, u manjoj mjeri mijenja se i struktura rječničkoga članka, primjerice, razvija se opis rekcije, a nastavlja se i rad na prilagodbi digitalnomu mediju i razvijanju mogućnosti pretraživanja rječnika.



Svensk ordbok: u mobilnoj aplikaciji omogućeno je slobodno pretraživanje teksta, što može otkriti problematične primjere. Tako se otkrilo da se Kongo u primjerima spominje u negativnome kontekstu (*balkanizacija Konga, opasno putovanje u unutrašnjost Konga*).

Íslex

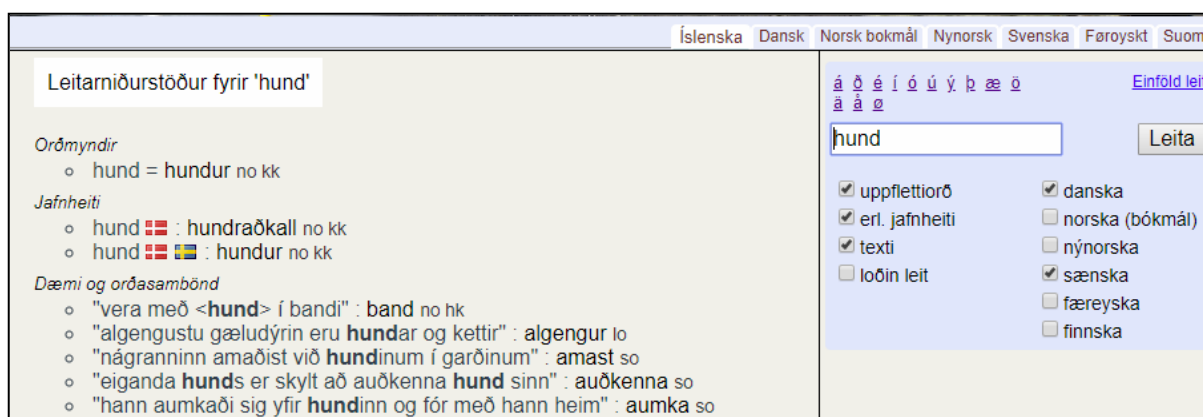
[Íslex](#) je višejezični mrežni rječnik namijenjen strancima koji uče islandski. Na Islandu je izrađen jednojezični rječnik kao temelj koji je potom preveden na danski, švedski, dva norveška standarda, ferski i finski. Iako je projekt formalno završen, islandska se baza još povremeno dopunjuje te postoji mogućnost dopunjavanja prijevoda.



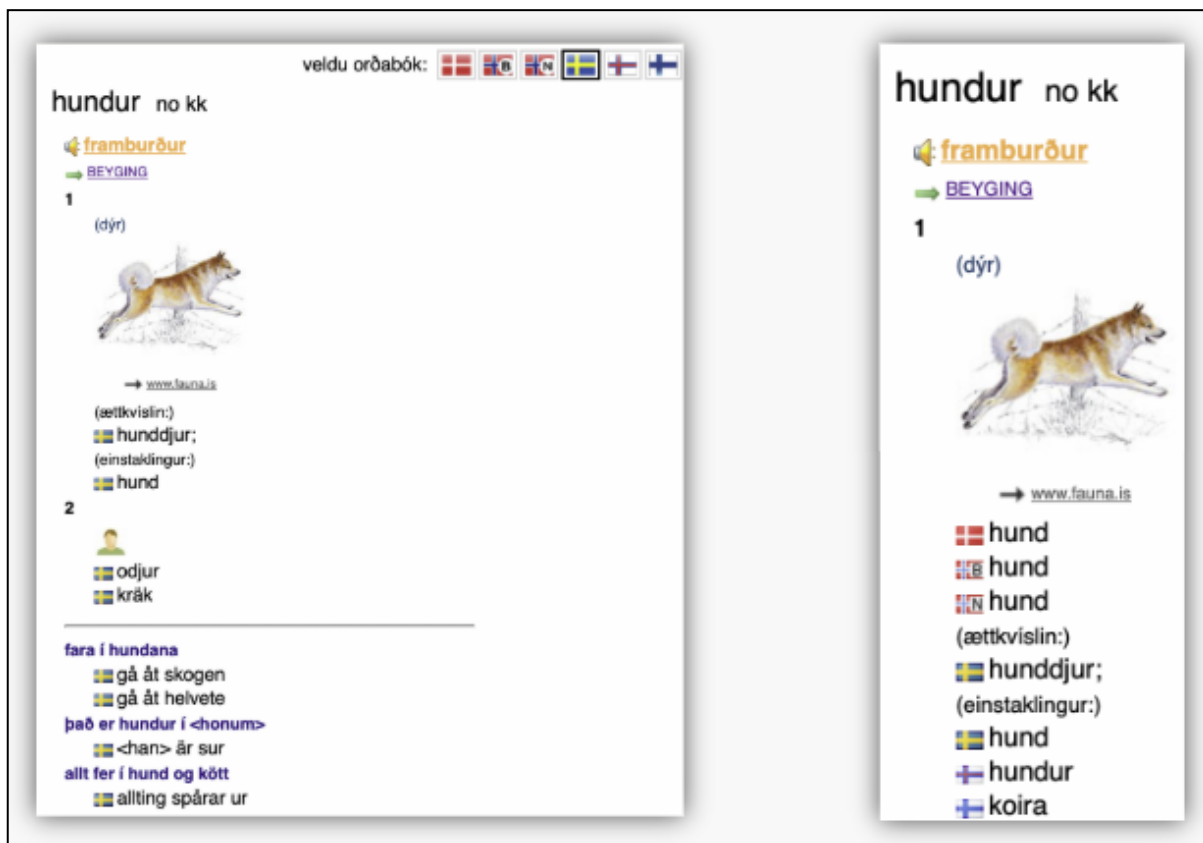
Íslex: početna stranica

Grafičko sučelje programa u kojem je rječnik pisan pregledno je te omogućuje jednostavnu komunikaciju i suradnju među uključenim stranama. Islandsko uredništvo suradnike u ostalim nordijskim zemljama obavještava o dodanim i izmijenjenim podacima te oni potom mogu dodavati i ažurirati prijevode. Program omogućuje jednostavno pregledavanje odgovarajućih podataka unesenih za druge jezike, a izmjene su odmah nakon pohrane vidljive na mreži.

Riječ je o rječniku koji je jednostavan za uporabu i u kojem su dobro iskorištene mogućnosti digitalnoga medija. Primjer su toga napredne mogućnosti pretraživanja te je za pretraživanje, primjerice, moguće odabrati natuknice na islandskome, ekvivalente na ciljnim jezicima, slobodno pretraživanje teksta, jedan ili više ciljnih jezika, proširiti ga na riječi koje se slično pišu i tako dalje.

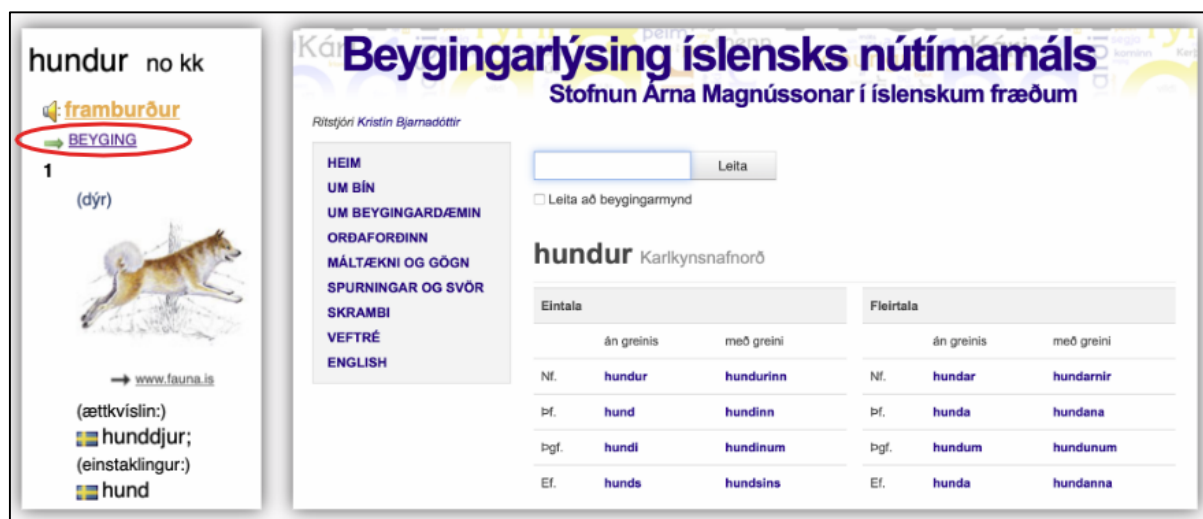


Íslex: napredne mogućnosti pretraživanja



Mogućnost odabira jednoga ili svih jezika za prikaz

Obrada natuknica uključuje zvukovne zapise izgovora, u nekim slučajevima slikovni materijal (slikovni prikazi biljaka i životinja izrađeni su za potrebe rječnika), a promjenjive riječi povezane su s vanjskom morfološkom bazom ([BÍN](#)), što je vrlo korisno s obzirom na morfološku složenost islandskoga.



Povezanost s morfološkom bazom