

JEZIKOVNI PRIROČNIKI IN NOVI MEDIJI

1. Jezikovni priročniki – vsi priročniški viri, ki so namenjeni javnosti kot vir podatkov o sodobnem slovenskem jeziku ali pri katerih nastopa slovenščina v paru z drugim jezikom.

Novi mediji – opredeljeni z negativnim izborom kot vsi mediji, ki niso tradicionalni knjižni medij jezikoslovnih priročnikov. Jezikovni priročniki v njih nastopajo v digitalizirani obliki (na računalniku ali v drugih digitaliziranih okoljih).

Delimo jih po različnih kriterijih:

- razlikovanje med priročniki, ki jih poznamo v knjižni obliki in se z novo funkcionalnostjo pojavljajo v digitalnem okolju, do popolnoma programskih virov, ki izven digitalnega okolja ne obstajajo
- digitalno okolje v katerem se pojavlja določen priročnik, bodisi da gre za računalniško okolje zaprtega tipa – aplikacija za osebni rač., strežnik, dlančnik, specializirana naprava itd. ali komunikacijski medij – internet, mobilna telefonija.

2. Slovarji

Najbolj klasični jezikovni priročniki v elektronski obliki so obstoječi knjižni slovarji, bodisi enojezični ali dvojezični, neposredno preneseni v računalniško okolje. Uporabnost slovarja v elektronski obliki je ob ustrezni programski opremi neprimerno večja kot v knjižni obliki. Slovarji se z digitalizacijo znebijo tiranične abecedne razvrstitve iztočnic. Pojavljajo se v večini računalniških okolij kot aplikacije v vseh vrstah računalnikov, na internetu, v obliki SMS sporočil v mobilni telefoniji ali v specializiranih napravah – Quicktionary.

2.1 Enojezični slovarji

Samo tri slovarje bi lahko opredelili kot enojezične slovarje v knjižni obliki, ki opisujejo sodobni slovenski jezik. To so SSKJ, SPS (2001) ter Veliki slovar tujk (VST 2002). Tem lahko dodamo BSJ in OSSJ. Prvi, ki se je pojavil v elektronski obliki je bil SSKJ (1997). Trenutno je dostopen v paketu s programom ASP32 ter prek svetovnega spleta na straneh ZRC SAZU. Dostop je brezplačen. Tudi SPS 2001 je na voljo v paketu s programom ASP32, nameravajo pa ga narediti dostopnega tudi prek spleta.

ASP32:

- možnosti tiskanja gesel in geslovnika
- možnost dodajanja opomb
- prilagajanje velikosti ali barv črk na ekranu
- sito (omogoča boljši pregled vsebine pri daljših geslih),
- je močno orodje za kompleksna iskanja: celotno bazo lahko preiskujemo po posameznih delih gesel – geslo, zaglavje, razlaga, primer, podgeslo, kvalifikator
- možnost shranjevanja in iskanja po izboru gesel

2.2 Dvojezični slovarji

Bili so poleg črkovalnikov prvi jezikovni priročniki, ki so obstajali v obliki računalniške aplikacije. Prvi, ki je bil legalno dostopen, je bil Veliki nemško-slovenski slovar (1994) v programu ASP. Trenutno so v tej obliki dostopni vsi obsežnejši in novejši dvojezični slovarji založbe DZS, osemjezični Evropski slovar Založbe Mladinska knjiga ter nekaj manjših avtorskih slovarjev podjetja Amebis.

Širjenje interneta je prineslo razmah dostopnosti dvojezičnih slovarjev – tudi brezplačno. Pojavili so se specializirani iskalniki, ki iskano besedo ali besedno zvezo iščejo po vseh brezplačno dostopnih slovarjih (www.onelook.com). Dvojezični slovarji prednjačijo pri prehodu v druge tehnologije, saj jih lahko razen na internetu in računalniku uporabljamo tudi prek digitalne telefonije s pomočjo SMS-ov ter dlančnikov.

2.3 Tezavri ali slovarji sinonimov

Za slovenski jezik v knjižni obliki ne obstaja, zato pa ga ponuja MS Office, ki ponuja tudi orodje, ki uporabniku predlaga sopomenke, nekoliko pa bi bila potrebna tudi jezikoslovna obdelava le-tega.

2.4 Terminološki enojezični ali dvojezični slovarji

So idealni priročnik za medij, kakršen je internet. Ni važno, ali je v slovarju veliko ali malo iztočnic, saj nam vsaka pride prav in nam da bolj ali manj zanesljivo informacijo. Dobimo jih na raznih spletnih straneh z zelo različnih področij. Med vsemi terminološkimi slovarji gotovo izstopa Evroterm – terminološka baza s spletnim iskalnikom Sektorja za prevajanje Službe vlade RS za evropske zadeve, ki je nastala pri prevajanju evropske zakonodaje v slovenski jezik.

3. Leksikoni besednih oblik

Informacije o besednih oblikah so bile do računalniške dobe v precej skrčeni obliki navadno dostopne v zaglavju enojezičnih ali dvojezičnih slovarjev (SSKJ, ...). Računalniški medij je za tovrstni priročnik precej bolj primeren kot knjižna oblika, kajti izpisovanje vseh oblik določene besede je prostorsko precej potratno. Leksikoni besednih oblik so navadno dostopni v dveh medijih. Na spletnih straneh ZRC SAZU in v programu ASP32. So izjemno pomembni za vse mogoče aplikacije na področju jezikovnih tehnologij.

4. Besedilni korpusi

Korpusi so obsežne zbirke besedil, po katerih lahko iščemo z računalniškim orodjem, ki se im. konkordančnik. Korpus je tako skupaj s konkordančnikom postal standardni vir podatkov o jeziku ali kar jezikovni priročnik, na katerega se lahko obrnemo ob jezikovnih zadregah. Vlogo jez. priročnika najbolje uporablja t.i. referenčni korpus – čimbolj obsežna zbirka sodobnih besedil, ki odraža stanje v določenem jeziku glede na različne jezikovne zvrsti. V slovenskem primeru, kjer slovarji temeljijo na že zastarelem gradivu, pa jih lahko do neke mere celo nadomesti. Korpus je pravzaprav vsaka zbirka besedil, ki je zbrana za določen namen. Skupaj z besedili za iskanje po korpusu potrebujemo tudi konkordančnik, računalniški program, ki omogoča hitra iskanja. Da bi izboljšali uporabnost korpusov, lahko v besedilo korpusa z različnimi postopki vnašamo tudi dodatne informacije, kot so lematizacija – osnovna oblika besede ter oblikoslovno-skladenjsko označevanje (skladenjska vloga).

4.1 Enojezični

Trenutno največji enojezični korpus slo. jezika je referenčni korpus FIDAPLUS (600 mio besed) - po korpusu iščemo s konkordančnikom ASP32. Drugi največji je korpus Nova beseda (ZRC SAZU). Brezplačno je dostopen preko interneta in obsega 95 mio besed. Konkordančnik s katerim iščemo po korpusu se im. NEVA; ni lematiziran. Tretji korpus je bistveno manjši, brezplačno dostopen na straneh Instituta Jožef Stefan. Ta korpus je postavil osnove za oblikovanje drugih korpusov, predvsem FIDE. Uporablja konkordančnik Workbench. Prednosti korpusa Nova beseda sta predvsem prosta dostopnost ter sprotno dopolnjevanje, slabosti pa izrazita neuravnoteženost in nelematiziranost. Prednosti FIDE pa so, da je korpus

dobro uravnotežen, problem pa je nerazdvoumljenost lem pri oblikah, kjer jih je možnih več, ter seveda plačljivi dostop.

4.2 Vzporedni

Pri nas obstajata dva: ELAN, ki je zbirka prevedenih večinoma pravnih in drugih besedil zbranih na Institutu Jožef Stefan in TRANS, ki je zbirka besedil s področij medicine, strojništva, zakonodaje/prava, geologije in turizma, zbranih na Oddelku za prevajalstvo FF. Vzporedni korpusi so oz. bi lahko bili pomembni za prevajalce, slovaropisce in za celotno področje jezikovnih tehnologij, ki se ukvarja z avtomatskimi prevajalnimi sistemi.

4.3 Korpusi strokovnih besedil

Prek interneta je prosto dostopen korpus prevodov zakonodaje Evropske unije. Uporablja isti konkordančnik kot slovar Evroterm. Vpisani izraz je lahko v slo. ali ang., izpis pa je lahko enojezičen ali dvojezičen in od te izbire je odvisno ali ga uporabljamo kot enojezični ali vzporedni korpus. Iskanje lahko omejimo na različna specializirana področja zakonodaje (kmetijstvo, okolje, ...).

4.4 Svetovni splet

Svetovni splet je tudi sam postal neke vrste mega korpus. Razvit je bil tudi konkordančnik, ki uporablja tudi poljubni splošni iskalnik (Google, Altavista, ...). Spletni konkordančnik deluje sorazmerno počasi, tako da je za enostavna iskanja lažje uporabljati splošne iskalnike.

5. Jezikovna orodja v pisarniških programskih paketih

5.1 Črkovalniki

So programske orodje, ki preverja pravilnost zapisa besed v različnih programih od urejevalnikov besedil, programov za razpoznavo črk itd. Pojavili so se med prvimi od orodij. En del slovensko pišočih uporablja črkovalnik MS Office, drugi del pa prosto dostopne črkovalnike aspell, ispell, myspell.

5.2 Delilniki

Delilnik je orodje za delitev besed po pravilih slovenskega pravopisa.

5.2 Preverjanje slovnice

To je orodje za preverjanje skladenjske pravilnosti stavkov. Za slovenski jezik Microsoft Office in drugi urejevalniki tega orodja še nimajo. Edino samostojno orodje za preverjanje slovnice nudi podjetje Amebis pod imenom BesAna.

5.4 Sezname samopopravkov

To je orodje, ki je na voljo v Microsoft Office kot integrirana aplikacija. Gre za spisek pogostih tipkarskih napak, ki jih delajo uporabniki pri uporabi urejevalnika besedil. Pri spisku napačnih besed se skriva tudi kakšna jezikovna past, saj med njimi najdemo tudi pare, kjer gre za besede z različnimi pomeni (tehničar-tehničarica), ...

6. Avtomatski prevajalni sistemi

Verjetno nikoli ne bodo v celoti nadomestili prevajalcev, lahko pa spravijo dokument v obliko, ko uporabnik brez znanja izhodiščnega besedila razbere osnovno sporočilo ali temo. Podjetje Amebis ponuja en prevajalnik iz slov. v ang., drugi sistem pa je prevajalnik med celo množico jezikov in obstaja tudi prek spleta. Med njimi je tudi slovenščina, ki pa še ni dosegla ravni, ki bi omogočala razumevanje brez izhodiščnega jezika.

7. Govorne tehnologije

Za slovenščino obstaja več sistemov za sintezo govora: prvi je bil razvit na IJS in se im. Govorec – namenjen gluhih in naglušnim za delo z rač., drugi pa se im. eBrskalnik, ki je na voljo na spletu.

8. Prihodnost

Prva zadrega nastane na ravni, ki še ni računalniška – spodobna dokumentiranost stanja sodobnega slovenskega jezika v obliki, ki je primerna za prenos v digitalne aplikacije, od vseh vrst slovarjev do uspešnega oblikoslovno-skladenjskega označevalnika. Drugi element je racionalnost pri uporabi razpoložljivih finančnih sredstev in človeških potencialov.

SLOVENSKA LEKSIKALNA PODATKOVNA ZBIRKA

Predlagana zasnova leksikalne podatkovne zbirke omogoča gradnjo različnih tipov slovarjev. Gre namreč za opis slo. jezika, kakršnega lahko pridobimo na podlagi referenčnega Korpusa slovenskega jezika FIDA. Nabor iztočnic je narejen na podlagi korpusne liste besed, ki je osnova za nadaljno analizo posameznih leksikalnih enot. Izhodiščno vodilo leksikalne zbirke je prikazati aktualno stanje slovenščine na leksikalni ravni: obstoj leksikalnih enot, njihovo dejansko obliko in pomen ter tipično ubesediljenje. Podatkovna zbirka nastaja v formatu XML/SGML.