

---

---

# **4. fejezet: Módszerek és eszközök**

**Kalivoda Ágnes**

2019. 03. 01.

---

---

# Bevezetés

## A fejezet témája:

- a korpusz szerepe a szótárírási munkában
  - 1) meglévő szótár frissítése korpuszalapon
  - 2) új szótár létrehozása korpuszalapon ←
- a lexikográfusok munkakörnyezete

## A fejezet váza:

1. A szótárírás folyamata
2. Szoftverek
3. A szerkesztési útmutató
4. Szócikk-sablonok

# **A szótárírás folyamata**

# A szótárírás folyamata

tervezés ➤ analízis ➤ (transzfer) ➤ szintézis

- célszerű először egy relációs adatbázist tervezni, és csak egy későbbi lépésben előállítani a szótárat az adatbázisból
- az adatbázisban az egyes címszavakhoz a lehető legtöbb információt rendeljük (többet, mint amit a végén megőrzünk)
- az adatbázis felgyorsítja a szerkesztési folyamatot: ha jól van összerakva, a szerkesztőknek alig kell visszanyúlni a nyers korpuszhoz
- az adatbázis újrahasznosítható: felhasználható pl. nyelvkönyvek előállításához is

# A szótárírás folyamata

tervezés ➤ **analízis** ➤ (transzfer) ➤ szintézis

- miután alaposan megterveztük az adatbázist, létre is hozzuk :)
- analízis: korpusz → adatbázis
- ebben a fázisban történnek:
  - a korpuszlekérdezések
  - a jelentések elkülönítése
  - kollokációk keresése
  - ...
- a szótárszerkesztők feladata ezután az, hogy a részletes szócikk-alapokat a célcsoport szükségletei szerint alakítsák ki

# A szótárírás folyamata

tervezés ➤ analízis ➤ **(transzfer)** ➤ szintézis

- transzfer = fordítási folyamat  
(tehát egynyelvű szótár esetén elmarad)
- nem fordítjuk le a teljes adatbázist a célnyelvre, hanem
  - minden lexikai egységhez egy-két általános fordítást adunk
  - ha adott kontextusban nem felel meg az általános fordítás, ott a minimum szükséges részt lefordítjuk

# A szótárírás folyamata

tervezés ➤ analízis ➤ (transzfer) ➤ szintézis

- szintézis: a szócikk szerkesztése
- végig szem előtt kell tartani a célcsoportot
- a legnagyobb nehézség: mit érdemes külön jelentésként feltüntetni, mit kell egyben tartani
- kétnyelvű szótár esetében:
  - a szócikk vázát a forrásnyelv anyanyelvi beszélője készíti
  - az anyagot a célnyelv anyanyelvi beszélője átnézi
  - a véglegesítést a forrásnyelvi beszélő végzi

# Szoftverek



# Szoftverek

A szótárírási folyamatot kétféle szoftver támogatja:

1. **Korpuszlekérdező rendszer 'Corpus Query System' (CQS):**  
lehetővé teszi a korpusz elemzését sokféle szempont alapján
2. **Szótáríró rendszer 'Dictionary Writing System' (DWS):**  
lehetővé teszi a szótári szöveg könnyebb szerkesztését,  
megkönnyíti a projekt menedzselését is

# A korpuszlekérdező rendszer (CQS)

- az analízis folyamatában van szerepe, ezt könnyíti meg
- tipikus lexikográfiai szükségletek:
  - döntések arról, hogy mit vegyünk bele a szótárba (pl. alakvariánsok, többszavas kifejezések)
  - különféle jelentések azonosítása, ezek sorrendje
  - frázis-szintű viselkedés leírása (pl. kollokációk, szintaktikai preferenciák)
  - címkék hozzárendelése egyes szavakhoz (pl. stílus)
- azután érdemes CQS-t választani, hogy ezekről döntöttünk

# A CQS-ek alapvető funkciói

- alap KWIC konkordancia
- egyszerű keresés
- összetett keresés, a közvetlen kontextus megadásával
- baloldali vagy jobboldali rendezés
- gyakorisági adatok
- információ a forrásszövegekről
- hosszabb részlet a forrásszövegből

# Sketch Engine

- a Sketch Engine egy piacvezető korpuszlekérdező eszköz, amely a szótárírási folyamatot is támogatja ([www.sketchengine.eu](http://www.sketchengine.eu))
- ingyenes, de minden alapfunkcióval bíró változata a NoSketchEngine (<https://nlp.fi.muni.cz/trac/noske>)
- ezt a CQS-t használja több magyar korpusz, pl. az MNSZ (<http://clara.nytud.hu/mnsz2-dev/index.html>)

# CQS – zárszó

## A szótárírók tipikus keresési stratégiája:

1. az adatok gyors áttekintése  
(konkordanciával vagy lexikális profillal)
2. a jelentések, többszavas kifejezések stb. azonosítása
3. külön keresésekkel ezek alaposabb vizsgálata

*“Knowing when to stop searching is as much a skill as knowing when a definition says enough.”*

# A szótáríró rendszer (DWS)

A teljes szótárírási folyamatot támogatja.

## **Előnyei:**

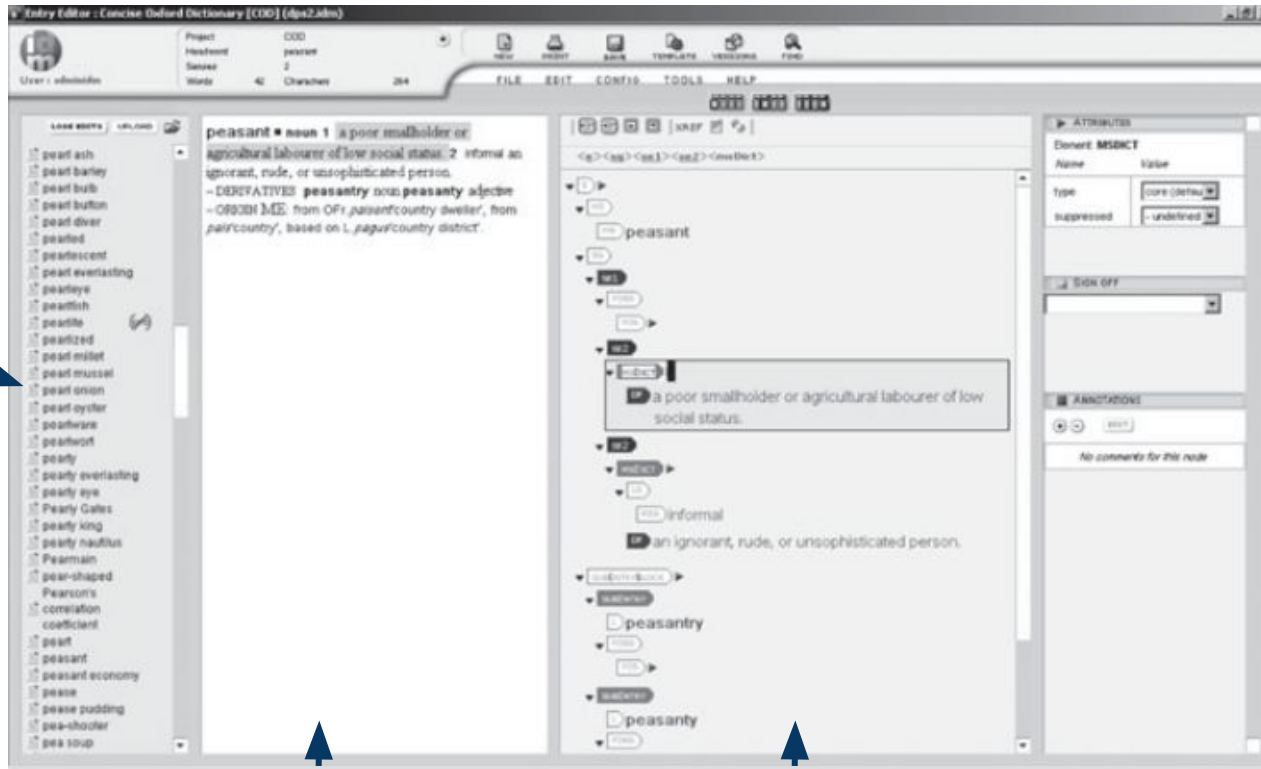
- azonos projekten fizikai helytől függetlenül lehet dolgozni
- mentesít az alapvető, de mechanikus feladatok alól
- belső konzisztenciát biztosít

## **Egy tipikus DWS három részből áll, ezek:**

- a szerkesztőfelület
- az adatbázis
- adminisztratív eszközök

# A szerkesztőfelület

a lexiko-  
gráfus  
szócikkei



adminisz-  
tratív  
funkciók

a szócikk előnézete

fa-struktúra nézet

# Az adatbázis

- ez tárolja a szerkesztőfelületen bevitt adatokat
- a lexikográfus nem érintkezik vele közvetlenül
- komplex lekérdezések futtathatók rajta, pl.
  - minden példamondat, amely egy szintaktikai mintára illeszkedik
  - minden szóvariáns, amely egy bizonyos címkével van ellátva
  - minden szócikkrészlet, amely egy bizonyos szót tartalmaz
  - ...



# Adminisztratív eszközök

A projektek menedzselését megkönnyítő, “karbantartó” eszközök.

## Funkciók (többek között):

- *workpack*-ek létrehozása, ezek hozzárendelése a szótárírókhoz
- az adatbázis egyes mezőinek zárolása  
(pl. a fonetikai átírást egy arra szakosodott ember csinálja, nem mindenki magának)
- a haladás ütemének nyomon követése

# **A szerkesztési útmutató**

# A szerkesztési útmutató

- a szótárírók számára készült, a szócikk szerkesztéséhez részletes instrukciókat tartalmazó könyv
- nagyon fontos, mert ez teszi lehetővé a visszatérő elemek konzisztens kezelését
- ha pl. az *informal* jelzése egyszer *infl.*, másszor *colloq.*, az nem sok bizalmat ébreszt a felhasználóban
- DE ez nem kőbe vésett → ha nem kínál megoldást egy adott helyzetre, felülírható / bővíthető

# Mit kell megadnia?

- *morfológia*: pl. az inflexió kezelése – csak a szabálytalan alakokat adjuk meg, minden alakot megadunk, vagy csak bizonyos számot és személyt?
- *alakvariánsok*: mikor nevezhető valami alakvariánsnak, kapjon-e külön címszót?
- *címkék*: milyen feltételek szabják meg, hogy valami “offenzív” vagy “elavult”?
- *példák*: mikor van szükség példákra, a korpuszpéldák módosíthatók-e?
- mi a teendő, ha egy forrásnyelvi kifejezésnek *nincsen célnyelvi megfelelője*

# Mit nem kell megadnia? (napjainkban)

- milyen sorrendben legyenek a szócikk egyes részei
- hogyan legyenek formázva a szócikk egyes részei
- rendszeresen használt elemek egységes jelölése  
(pl. címkenevek, rövidítések)

# Szócikk-sablonok

# Szócikk-sablonok

A szócikk-sablon olyan váz, amely megkönnyíti az azonos szemantikai csoporthoz tartozó szavak szócikkeinek kidolgozását.

Segít abban, hogy a lexikai tételek egységes módon legyenek kidolgozva – ne úgy, mint az itt látható példában:


**lion** *n*[C] a large strong African and Indian animal with four legs and light brown fur which eats meat and belongs to the cat family

**tiger** *n*[C] a large wild cat which has yellowish orange fur with black stripes

# Szócikk-sablonok készítési folyamata

1. adott szemantikai kategóriában válasszunk néhány (minél eltérőbb) szót (pl. ÁLLAT esetén OROSZLÁN, TEHÉN, EGÉR, EGYSZARVÚ)
2. osszuk ki a szavakat a szótárírók között
3. mindenki végez egy alapos korpuszvizsgálatot a saját szavához, és felvázol egy szócikket
4. a csoport megvitatja a szócikkeket, a fontos elemeket összegyűjtik egy szócikk-sablonba
5. ez a szócikk-sablon ellenőrző listaként használható a kategória összes többi szavánál



Query **köszön, IGE.\*, figyelem, S?PUNCT** 1,232 (0.39 per million) 

, tudja és remélheti . </s><s>	Köszönöm a figyelmüket .	( Nagy taps . ) </s><s> Köszönöm
az én munkámat is . </s><s>	Köszönöm a figyelmed !	</s><s> További szép napot
értelembe is . </s><s>	Köszönöm a figyelmüket !	</s><s> Mentsd meg emlékeid ! </s>
elé terjeszteni . </s><s>	Köszönöm a figyelmüket .	( Taps a kormánypárti
már nem maradt idő . </s><s>	Köszönöm a figyelmüket .	</s><s> Ronaldo a leggyorsabb
viszont meg szeretném	köszönni a figyelmet !	</s><s> Az iskoláról a
kormányzása alatt . </s><s>	Köszönöm a figyelmet !	</s><s> Pénzkezelési tippek II
. </s><s> A fotós hozzászólása :	Köszönöm a figyelmet .	</s><s> Ha kedveled a jó képeket
előtte mindenképp . </s><s>	Köszönöm a figyelmet !	</s><s> A következő alkalommal
képviselőtársaimnak . </s><s>	Köszönöm a figyelmüket .	( Taps a Fidesz és a KDNP
egy interjút .	Köszönjük a figyelmet .	</s><s> banyek , 2008 . június 24
leveleink érkeznek . </s><s>	Köszönjük megtisztelő figyelmét !	</s><s> Az egri vár az ország
? </s><s> Szív vagy ész ? </s><s>	Köszönöm a figyelmüket .	2008 . április 14 . hétfő
információ a honlapon . </s><s>	Köszönjük a figyelmet .	</s><s> Most kaptam az e-mailt
hozzájutni . </s><s>	Köszönöm megtisztelő figyelmüket .	( Taps az MSZP és az SZDSZ
megosztok veletek . </s><s>	Köszönöm a figyelmeteket !	</s><s> Tegnap este
lehet megbélyegző . </s><s>	Köszönöm a figyelmet !	</s><s> Kedves Ilosvai Ferenc
, rendes közgyűlés ,	köszönjük a figyelmet !	</s><s> 19 : 0 9 Katolikus öröm </s>
konzultációval . </s><s>	Köszönöm a figyelmüket .	</s><s> LÁZÁR JÁNOS