

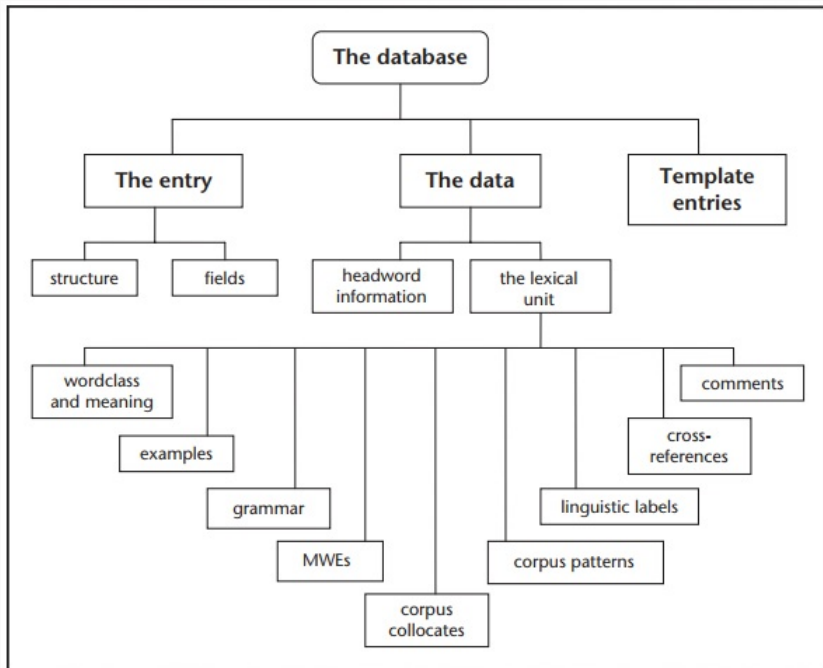
Building the database (2):

the lexical unit

Bognár Ivett

2019. 04. 25.

- Ebben a fejezetben: egynyelvű adatbázis bejegyzés készítése.
- (Angol példákkal fogunk dolgozni.)
- Minden szótár adatbázis.
- A lexikai egységre (lexical unit: LU) koncentrálunk.
- Az adatbázis bejegyzés szerkezetéről lesz szó: milyen típusú információt tartalmaz.
- Cél: egy olyan adatbázis létrehozása, ami nem csak egy szótárnak, de más könyveknek is a forrása lehet.



A bejegyzés

- A különböző mezők (field) elnevezése az adatbázisban: milyen információt tárolnak.
- Némely mező (pl. morfológia információk) automatikusan kitölthetők.
- A kész szótárhoz képest fontos különbség: lehet sokkal részletesebb
- Ebből kifolyólag a szerkezete és a mérete is különbözik.

A bejegyzés szerkezete

- Egy szó, vagy lemma: egy bejegyzés az adatbázisban
- Ez egy vagy több albejegyzést tartalmaz: az első a vezető bejegyzés, ami inkább a lemmára vonatkozik.(?)
- Saját albejegyzése van: a lemma minden LU-jának, és minden MWE-nek (multiword expression), ami tartalmazza a lemmát.

Mezők (The fields in a lexical entry)

- Ezek közül néhány automatikusan generált, de kézi ellenőrzést igényelhet.
- Nagyjából olyan sorrendben, ahogy szükség lehet rájuk a bejegyzés felépítésekor, de mivel van amelyiket többször is használni kell, nem lehet őket sorba rendezni.
- A bejegyzés tervezése nagyban függ a nyelvtől, és a nyelvtani szó szófajától.

- A legstrukturáltabb adattípusok - **WORDCLASS** (7), **CONSTRUCTION** (11) – előre meghatározott kódok tartalmaznak, amik jelölik az adott grammatikai kategóriát.
- Előre megadott listából kiválasztottak a **LINGUISTIC LABELS** (21) és a **MWE-TYPE** (16)
- Közvetlenül a korpuszból: **EXAMPLE** (10)
- Sok a szerkesztők saját nyelvi tudásából: **VARIANT FORM** (3), **INFLECTED FORM** (4), **NULLINST-TYPE** (13), **COLLOCATE TYPE** (19)
- Más típusú mezők pedig tartalmazhatnak a szerkesztő által írt szöveget: **MEANING** (9), **NULLINST-SEMANTICS** (14), **COLLOCATE-TYPE** (19), **CORPUS PATTERN** (20), or **COMMENT** (23)

Adatok

- A bejegyzésben szerepelő összes tény leírható az alapján, hogy milyen kapcsolatban áll a címszóval.
- Az adatbázis bejegyzés kétszer vagy háromszor akkora, mint a valós szótári bejegyzésünk: több minden lehet benne, mint amire a felhasználónak valóban szüksége van.
- A grammatikai és kollokációs információk és a példák nélkülözhetetlenek.
- Egynyelvű: kevesebb „normál” kontextusos példa is elég
- Többnyelvű: a fordítást minden lehetséges kontextusban tesztelni kell: minél több példa van, annál jobb

Az adatbázisnak tartalmaznia kell a lemmához tartozó összes LU-t, és minden LU-ra:

- szemantikai tartalma vagy alapjelentése
- semantic scope: mennyire specifikus, metaforikus jelentések
- morfológiai tulajdonságok
- grammatikai tulajdonságok (nem, szófaj, stb.)
- valency: hogyan kombinálódik szintaktikailag
- idiómák, amikben szerepel
- leggyakrabban milyen grammatikai szerkezetben szerepel: V-O, S-V, Adj-N
- fontosabb nyelvtani és szemantikai viselkedésminták a korpusz alapján
- a jelentés egyéb aspektusai (stílus, regiszter, régió, stb.)
- a korpuszban szereplő idézetek forrásai, és azok sajátosságai (médiium, dátum, szerző, műfaj, stb.)

Mi egy címszó?

Általában könnyen eldönthető (?), de vannak problémás esetek:

- Participial adjectives
surprised, broken, amazing, disturbing.
- Gerunds (nouns in -ing)
Minél ritkább, annál kevésbé címszó.
- Hyphenated combining forms
brown-haired -> hair
mud-splashed -> splash
- Compound prepositions
in spite of, according to, owing to
- Plural headwords

Headword information in the 'leader' record

Headword information in the 'leader' record

Minden bejegyzés és albejegyzés kezdete a címszó mezőben:

- kanonikus forma (főnévnél egyes szám, stb.)
- Minden bejegyzésben általában csak egyszer fordul elő, a vezető rekordban:
 - HOMOGRAPH NUMBER
 - VARIANT FORM
 - FULL FORM
 - INFLECTED FORM

Headword information in the 'leader' record

HEADWORD	weave
HOMOGRAPH NUMBER	1
INFLECTED FORM	wove (past), woven (past participle)
MEANING	interlace (threads etc.) to make fabric
HEADWORD	aluminium
REGION	GB
VARIANT FORM	aluminum
REGION	US
HEADWORD	EU
FULL FORM	European Union

Fig 9.2 Fields and data from the initial 'leader' record

Az LU szófaja és jelentése

- Néha a szabálytalan alakok egy LU-hoz vannak kötve:
- where both the strong and weak forms -> the 'move in and out' LU
- where only the strong past tense -> the LU meaning 'make a fabric by interlacing threads etc.'

1	minutes before the final whistle, he	weaved	his way infield
2	dizzy with pride and vodka, she	weaved	back to Drew's car
3	parked behind the dustbins – she	weaved	towards it, finding her vision blurring
4	As the ambulance	weaved	its way dramatically through the ...
5	A helicopter tracked the car as it	weaved	in and out of traffic near Bristol
6	he yelled something as he	weaved	from side to side across the road
7	They	wove	off through the theatre crowd
8	Herr Nordern	wove	unsteadily across the living room
9	you	wove	your way along the arcade
10	The car	wove	through the traffic on
11	They	wove	in between the cars that lined the drive
12	the demonstrators	wove	through the downtown area

Fig 9.3 Strong and weak past tense forms of *weave* in similar contexts

Minden albejegyzés az LU szófajával és egy „informális” jelentéssel kezdődik.

- Szófaj
 - adjective, adverb, conjunction, determiner, interjection, noun, preposition, pronoun, verb
 - alszófajok: noun:count, pronoun:personal, determiner:article, verb:transitive
- Jelentés
 - Elegendő legyen az LU-k megkülönböztetésére, több nem kell.
 - Csak a szerkesztők használják.

HEADWORD	mouse	HEADWORD	mouse
LU #	1	LU #	2
WORDCLASS	noun	WORDCLASS	noun
INFLECTED FORM	mice (pl)	INFLECTED FORM	mouses, mice (plural)
MEANING	small rodent	MEANING	computer device
HEADWORD	weave	HEADWORD	weave
HOMOGRAPH #	1	HOMOGRAPH #	2
WORDCLASS	verb	WORDCLASS	verb
INFLECTED FORM	wove (pt), woven (ptp)	INFLECTED FORM	wove or weaved (pt), weaved (ptp)
MEANING	interlace (threads etc.) to make fabric	MEANING	move in and out of obstacles

Fig 9.5 The first few fields in four LUs

Példák

- A példa mező szinte bárhol lehet a bejegyzésen belül.
- Egy specifikus dolgot kell illusztrálnia, ami lehet az LU jelentése, nyelvtani szerkezet, amihez köthető, stb.
- Az adatbázis példái és a szótári példák kicsit különbözhetnek:
- Az adatbázis minden belekerül, ami a korpusz alapján szükségesnek látszik.
- A példák legyenek teljes mondatok! De a korpusz mondatai lerövidíthetők.

Írányelvek a példák kiválasztása során

- Legyen rövid.
- Adjon egy informatív kontextust, ami segít a címszót megérteni.
- Ne legyenek benne olyan szavak, amiket nehezebb megérteni, mint a címszót.
- Ne legyenek benne nyelvjárási, stílusréteghez köthető, stb. szavak.
- Ne tartalmazza valós személy nevét. (Se élő, se halott!)
- Ne legyenek benne idegen szavak.

HEADWORD	mouse	HEADWORD	mouse
LU #	1	LU #	2
WORDCLASS	noun	WORDCLASS	noun
INFLECTED FORM	mice (pl)	INFLECTED FORM	mouses, mice (pl)
MEANING	small rodent	MEANING	computer device
EXAMPLE	She saw the little mouse lying dead in its corner covered in dust.	EXAMPLE	I am left-handed and find it hard to 'right click' with the mouse .
EXAMPLE	Residents of the building said it was infested with rats, mice and roaches.	EXAMPLE	These will allow you to daisy-chain devices such as scanners, mice , keyboards and more.
		EXAMPLE	That's what they use up the school – proper computers, the screens and printouts and mouses .

Nyelvtan

Az LU albejegyzésben három típusú grammatikai információ:

- Szófaj
- További nyelvtani információk a GRAMMAR mezőben
- Szintaktikai környezet a CONSTRUCTION mezőben

- Két mező kerül játékba egy ige esetén: CONSTRUCTION és EXAMPLE (11, 10)
- Az első szerkezet mindig a zéró.
- Ebben a modellben nincs szükség a tranzitív, intranszitiv megkülönböztetésre.

1	CONSTRUCTION	0 (zero)
1a	EXAMPLE	Helen stood and watched through the binoculars.
1b	EXAMPLE	People in the crowd were watching, curiously.
2	CONSTRUCTION	PP-for
2a	EXAMPLE	This is also a good place to watch for buzzards.
2b	EXAMPLE	At about 7 a.m. he walked towards his house and hid in the bushes across the road and watched for his dad to leave for work.
3	CONSTRUCTION	NP
3a	EXAMPLE	Monica watched the two men , fascinated.
3b	EXAMPLE	We all watched her , unsure of what to say.
4	CONSTRUCTION	NP Vinf
4a	EXAMPLE	He watched Joe heave his bulk out of the chair.
4b	EXAMPLE	Twenty-two athletes spend five days for the most part watching their teammates do all the work, and at the end of it all, everyone is quite happy to settle for a draw.

Problémás esetek:

- Participiumok (*deserted, distressing, confused*)
"the Style Guide for a database will usually advise you to treat these forms as full headword adjectives if they occur in the corpus modified by very or in other clearly adjectival uses."
- Fokozhatóság
 - Fokozható-e? Ha nem, feltüntetni! (*házas*)
 - Rendhagyó fokozásokat szintén! (*good, better, best*)
- Szintaktikai szerep
"Can the adjective be used both attributively (*sunny morning*) and predicatively (*the morning was sunny*), or is it attributive only (*a mere child*), or predicative only (*I want to be alone*), or is it a post-modifier (*mayor elect*)?"

- idioms
- collocations
- phrasal verbs
- compounds
- support verb constructions

to bite someone's head off (snap at them angrily) and *to give something a clean bill of health* (report favourably on something after examination) are quite clearly 'idioms'

- Részben vagy egészében rögzült szósor
- A jelentés több/más, mint az elemek összessége (nem kompozicionális)
- Elég komplex ahhoz, hogy önmagában ne legyen értelmezhető.
- "Kilógna", ha a címszó jelentései közé akarnánk sorolni.

- A skála másik végén: szemantikailag transzparens szósorok
- Könnyű megtalálni a korpuszban.
- *better luck next time!*, *from hour to hour*
- A fentebb felsoroltak közül csak az első érvényesül
- Kétséges eseteket kollokációként vegyük fel!

- Kanonikus forma: ami a legjelöletlenebb. (például: Do you know what 'bite someone's head off' means?)
- Néha nincs kanonikus forma:
chicken and egg a korpuszban: *chicken or egg, a chicken and an egg, and the chicken or the egg*
- A MWE-TYPE és MWE mezőket használjuk!

1. Az első vagy egyetlen lexikai szó alatt vegyük fel
into the bargain -> bargain címszó; *to be hot on something* -> hot címszó; *to pull someone's leg* -> pull címszó.
2. A legritkább lexikai szónál
to open the floodgates -> floodgates.
3. Az első vagy egyetlen főnévnél
to rain cats and dogs -> cat címszó; *to pull someone's leg* -> leg címszó
4. Az első vagy egyetlen igénél
to rain cats and dogs -> rain címszó; *to twist and turn* -> twist címszó.
5. Vegyük fel önálló címszóként: *into the bargain, be hot on, pull someone's leg, rain cats and dogs*

1. Az MWE első eleméhez vegyük fel:
sky blue in the sky entry.
2. Külön LU-ként a második elemnél:
sky blue as LU subentry in the blue entry.
3. Önálló címszóként: sky blue as sky blue main entry

Phrasal verbs

1. Külön LU-ként az ige alatt
carry forward as LU subentry in the carry entry.
2. Önálló címszóként

Support verb constructions

1. Vegyük fel az MWE-t a főnév alatt:
make a decision in the decision entry.
2. Külön LU-ként a főnév alatt:
make a decision s LU subentry in the decision entry.
3. Önálló címszóként (nem javaslott)

Corpus collocates of the LU headword

"Co-occur with one another with a frequency greater than chance, and stand in a major grammatical relationship to the headword of the entry being compiled."

- főnév, ami tárgya a címszó igének.
relationship, bonds - forge
- főnév, ami alanya a címszó igének.
sun - rise, set
- főnév, ami módosítója a címszó főnévnek
maiden - speech
- melléknév, ami módosítója a címszó főnévnek
empty - promise

Három mezőt használunk:

- COLLOCATE
- COLLOCATE-TYPE
- GRAMMAR

Collocates in the database

HEADWORD	bargain
LU #	1
WORDCLASS	noun
EXAMPLE	Angelo offers her a bargain: if she will sleep with him her brother
EXAMPLE	A credit agreement could be re-opened, if the court thought just grounds that the bargain was extortionate, on the debtor's appeal to the High Court, a county court or a sheriff court.
COLLOCATE	strike
EXAMPLE	Within minutes, Sykes had struck a bargain, never stopping for a moment to ask where or how the fish had been caught.
EXAMPLE	Buyer and seller strike a bargain with each individual purchase.
COLLOCATE	hard
EXAMPLE	She says it was because the EC drove such a hard bargain on fish that public opinion turned against membership.
EXAMPLE	Don't become despondent just because it seems that your employer is determined to drive a hard bargain.

A főneveknek két olyan típusa, ami gyakran feltűnik más főnév címszavakkal együtt.

- **Itemizer:** Egy részére utal egy mass noun-nak.
spoonful and *slice* and *speck*
- Ezeket az adott itemizer címszava alatt érdemes feltüntetni.
- **Collective:** Egy többes számú főnévvel áll, ami a példányokra mint csoportra utal.
- Lehet idiomatikus: *a pride of lions*
- De általánosságban is segítünk a felhasználónak, ha a kollektív főnév címszava alatt megadjuk, hogy mivel szokott állni (*horde*)

no single	speck	of dust dared show itself
there isn't a	speck	of dust anywhere
he brushed a	speck	of dust from his sleeve
a	speck	of dirt never hurt anyone
A	speck	of grit danced over the lower
he came off the field without a	speck	of mud on his shorts
a rosette of small pink petals with a	speck	of green at its heart
not a	speck	of talcum powder on any surface
the bishop wiped a	speck	of sauce from his chin
I climbed high, to a tiny blue	speck	of water , unfished for decades
there was a	speck	of yellow on the horizon

Fig 9.39 The noun *speck* used as an itemizer

HEADWORD	speck
LU #	1
WORDCLASS	noun
MEANING	tiny spot, small particle
GRAMMAR	itemizer
CONSTRUCTION	PP-of NP
COLLOCATE	dust
EXAMPLE	He brushed a speck of dust from his sleeve as he waited for the silence he required.
EXAMPLE	She hung it over the line outside where she beat every speck of dust from it before replacing it over her freshly scrubbed quarry tiles.
COLLOCATE	dirt
EXAMPLE	His sympathy was probably worth about as much as the dirt beneath his fingernails – if there had been a speck of dirt present beneath those immaculately manicured items.
EXAMPLE	There was not a crumb or a speck of dirt to be found anywhere on the scrubbed and polished floors.

Fig 9.40 Part of the entry for *speck* showing itemizer uses

Korpusz mintázatok

- Intiutíve nem megismerhető (?), de a korpusz adataiból kinyerhető információ.
- Az alapján, hogy az emberek hogy használják az anyanyelvüket, következtetések vonhatók le az adott szóról.
- Milyen alakban (pl dispute gyakran passzív)
- Hol szokott szerepelni (MWE into the bargain mindig mondat végén)
- Milyen konnotációi vannak (speck: főleg negatív kontextusokban)
- Ez utóbbi: 'semantic prosody'

HEADWORD	consign
LU #	1
WORDCLASS	verb
MEANING	send, deliver, put into the care of someone
CONSTRUCTION	NP PP-to
CORPUS-PATTERN	semantic prosody: object of prep <i>to</i>; always negative, implies getting rid of something unpleasant; things are consigned to nasty places
EXAMPLE	The old adage that the Irish are at their most dangerous when they are at their lowest ebb has been consigned to the dustbin as a forgotten tradition.
EXAMPLE	Most of the people responsible for Labour's policies acknowledge that those which may have been relevant in the 1980s were disastrous and should be consigned to the rubbish heap .

Fig 9.45 Corpus patterning in the entry for *consign*

Linguistic labels,
Cross-references, Comments

Linguistic labels

HEADWORD	ape
LU #	1
WORDCLASS	noun
DOMAIN LABEL	zoology
MEANING	large monkey
EXAMPLE	It was her research that showed how close we are in evolutionary terms to the apes: how they communicate with each other, use tools and so on.
EXAMPLE	His paper on the orang-utan is especially welcome for collating so much biological information on this fascinating ape.

Linguistic labels

HEADWORD	ape
LU #	2
WORDCLASS	noun
REGISTER LABEL	informal
ATTITUDE LABEL	pejorative
MEANING	rude or stupid person (or both!)
EXAMPLE	I chained his bicycle to an ornamental fence outside the apartment block, and persuaded the uniformed ape on security that I was not a terrorist.
EXAMPLE	What on earth are you doing, grinning like an ape with your eyes shut?

HEADWORD	ape
LU #	3
MWE	to go ape
MWE-TYPE	idiom
REGISTER LABEL	very informal
MEANING	to lose or almost lose control (from excitement, anger etc.)
EXAMPLE	He kissed her on the lips, and the crowd went ape, especially when Cicely draped a hand over the dapper old gent's shoulder and kissed him back.
EXAMPLE	He'd gone ape, lost his temper completely, just about assaulted a pupil.

Hozzákapcsolja a címszót más megjelenéseihez:

- MWE-ben
the entries for cat and dog might contain a cross-reference to the rain LU
- Összetételben
- hyphenated compounds

- Pontos az, aminek hangzik. :)
- Belső jegyzetek
- Felhívni a figyelmet megoldatlan problémákra a szókikben, stb.

Köszönöm a figyelmet!