

# Duplakocka, igei szerkezetek, foci

ELTE DiAGram  
2021. június 4.

Sass Bálint  
Nyelvtudományi Kutatóközpont  
sass.balint@nytud.hu

# alapgondolat

példa:

*Részt vesz az akcióban.*

vesz ACC : rész INE : akció

# alapgondolat

példa:

*Részt vesz az akcióban.*

**vesz** ACC : rész INE : akció

→ ige...

# alapgondolat

példa:

*Részt vesz az akcióban.*

vesz **ACC** : rész **INE** : akció

→ ige + helyek...

# alapgondolat

példa:

*Részt vesz az akcióban.*

vesz ACC : **rész** INE : **akció**

→ ige + helyek + kitöltők

# alapgondolat

példa:

*Részt vesz az akcióban.*

vesz ACC : rész INE : akció

→ ige + helyek + kitöltők

cél:

*'vesz részt vmiben'*

vesz ACC : rész **INE**

# igei szerkezet

*Def:*

igei szerkezet (isz) = ige + bővítmények

bővítmény = hely + kitöltő (slot + filler)

*'történik'*

→ semmi

*'néz vmit', 'bízik vmiben', 'ad vkinek vmit'*

→ kitöltetlen hely

*'von vállat'*

→ kitöltött hely

*'beleüti az orrát vmibe', 'von kétségbe vmit', 'vesz őrizetbe vkit vmi miatt'*

→ mindkettő

# valódi igei szerkezet

*Def:*

valódi igei szerkezet =

minden kitöltött hely idiomatikus és

minden kitöltetlen hely vonzat (és a vonzatkeret teljes)

*komplex ige* = v-isz, amiben van kitöltött hely

„egy egységként igeeként funkcionál”

*‘vesz részt vmiben’ ✓*

*‘olvás vmiben’ ✗*

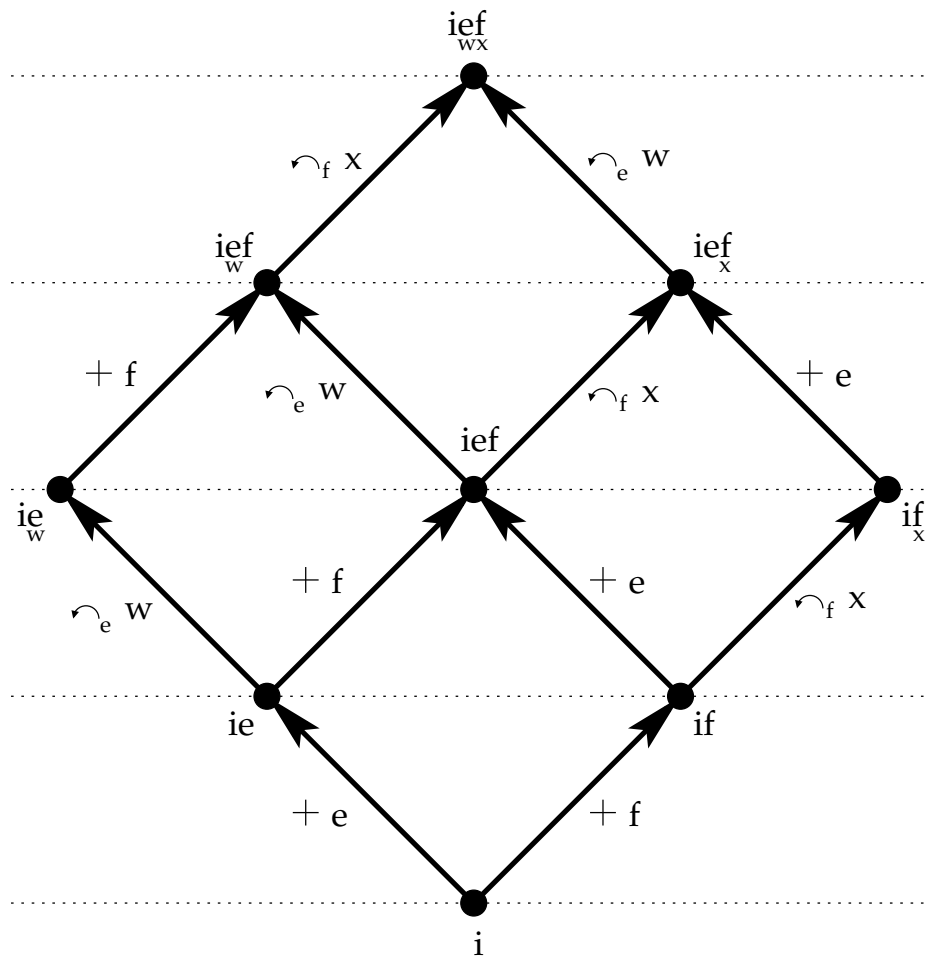
*‘vesz részt akcióban’ ✗*

*‘felhív figyelmet vmire’ ✓*

*‘süt nap’ ✓*



# duplakocka



hossz (l)

5

4

3

2

1

olvas Lencsi könyvet

olvas Lencsi vmit

olvas vki könyvet

olvas Lencsi

olvas vki vmit

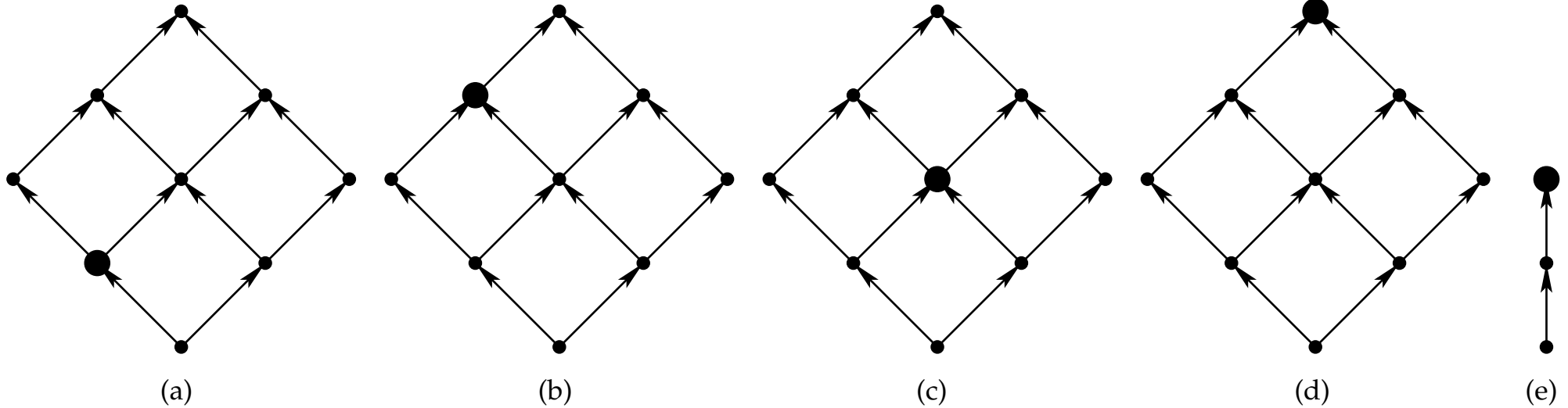
olvas könyvet

olvas vki

olvas vmit

olvas

# duplakocka – példák



Szendvicset eszik órán.

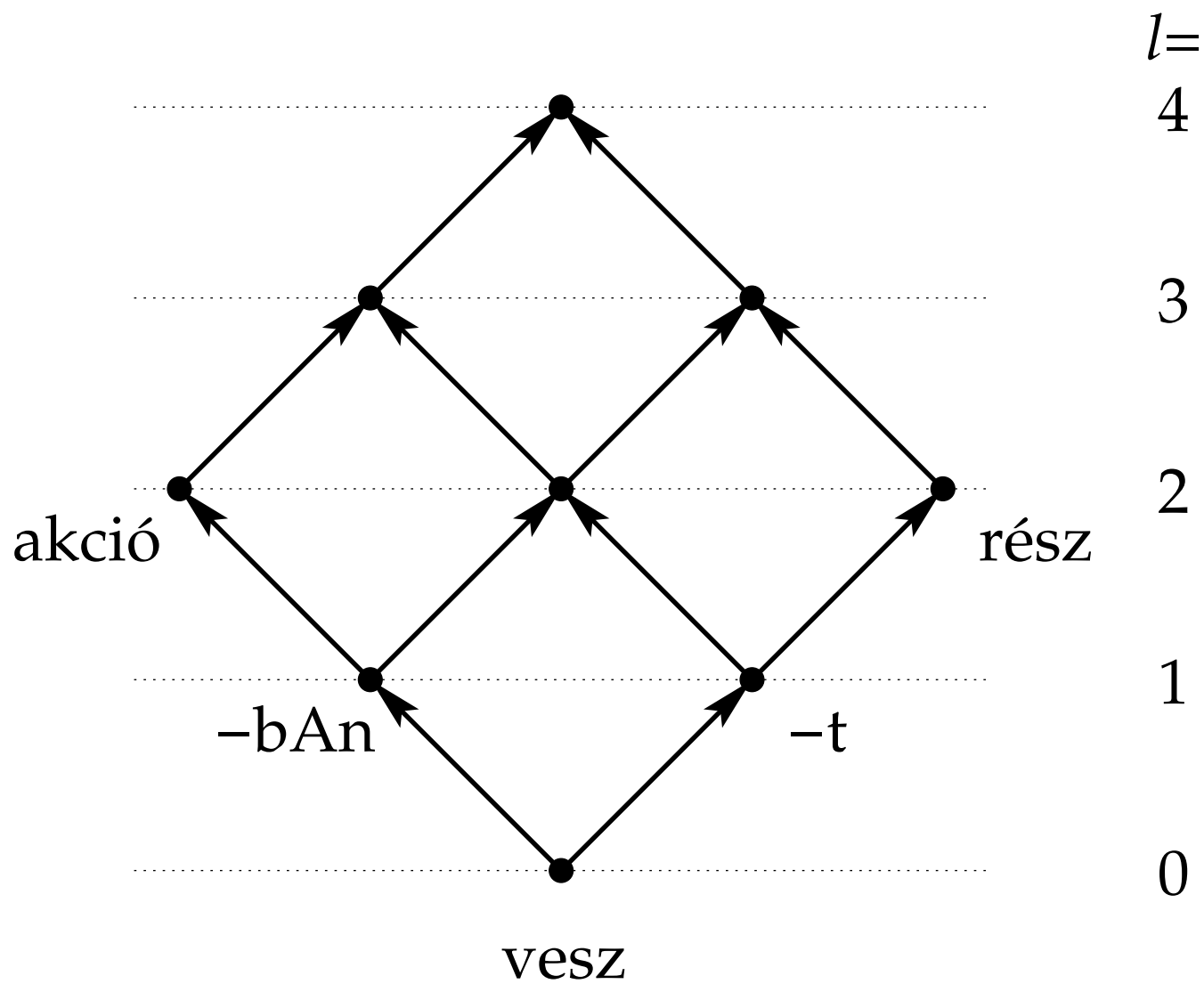
Felhívja a figyelmet a dologra.

Odaadja a haverjának a radírját.

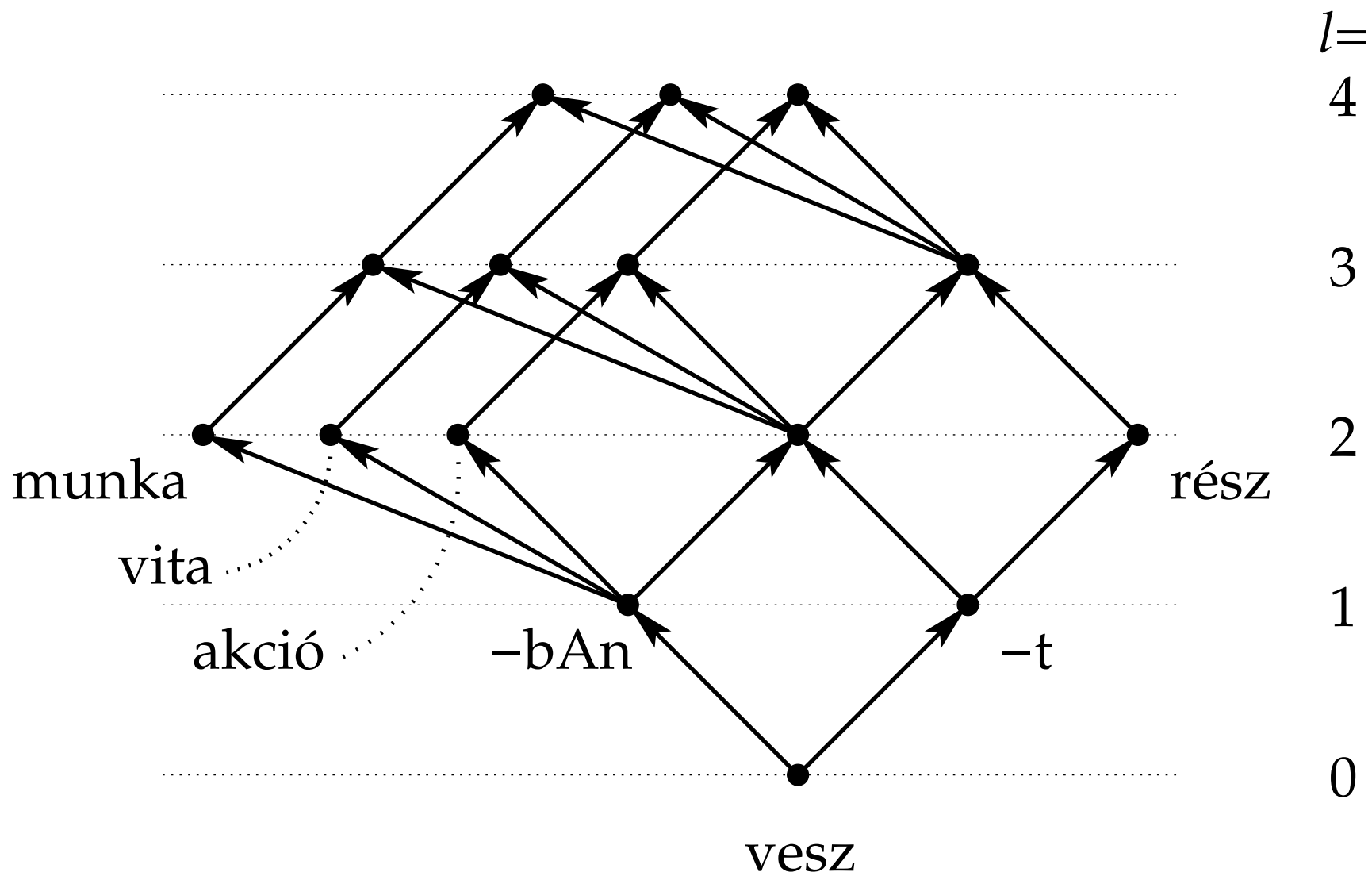
Veri az ördög a feleségét.

Vállat von. ( $\rightarrow 3D!$ )

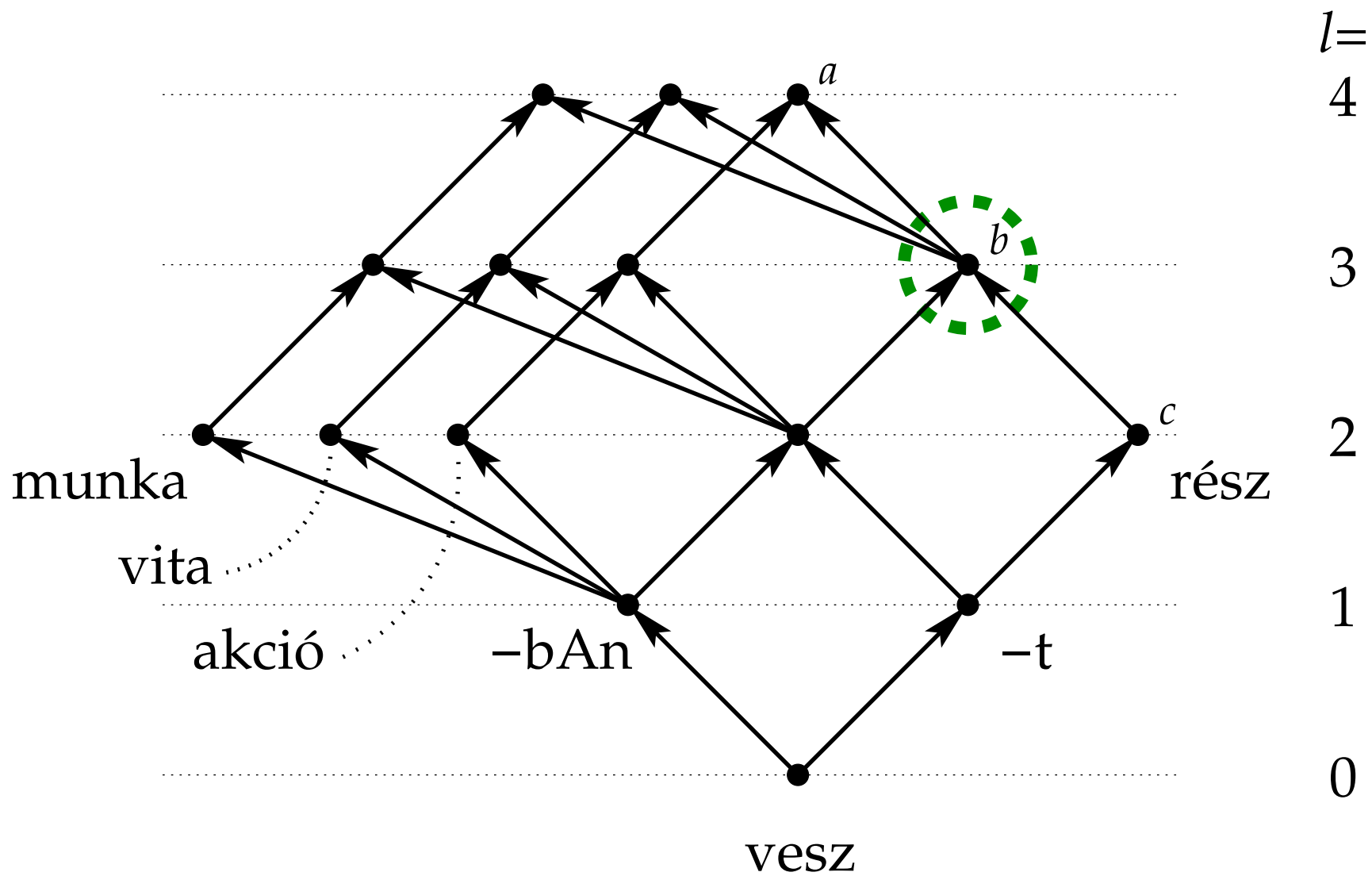
# duplakocka ...



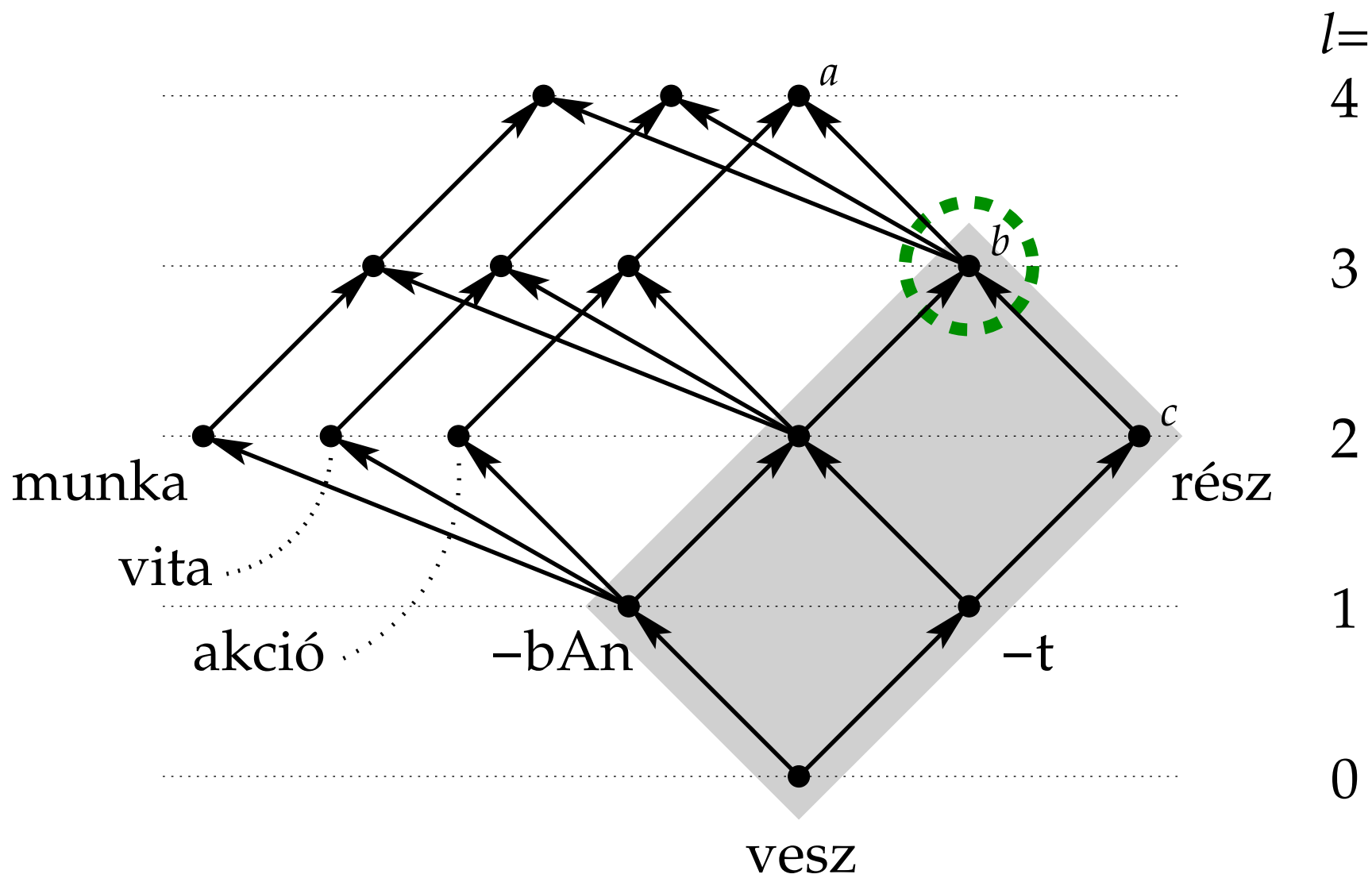
# korpuszháló



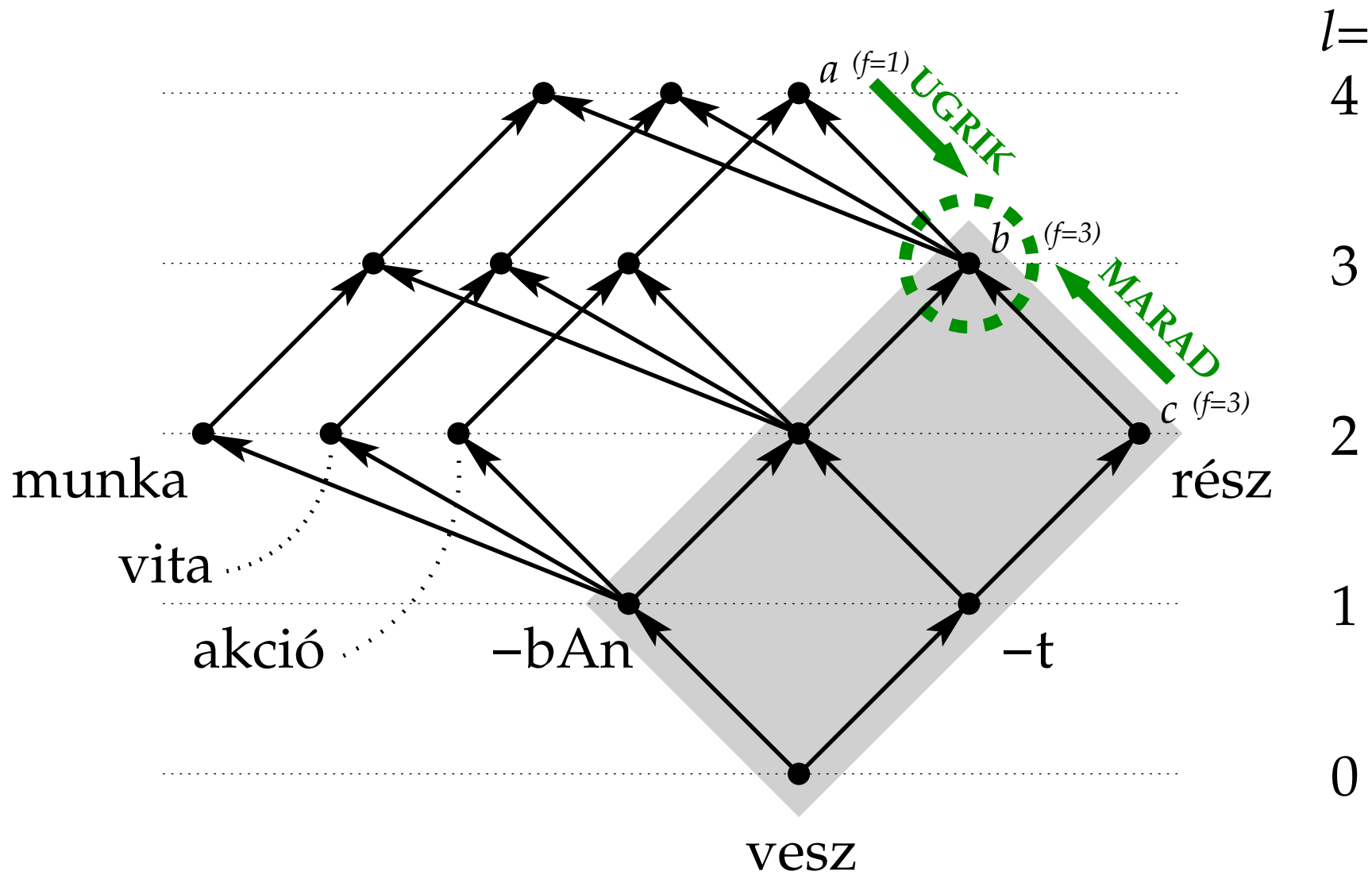
# valódi igei szerkezet

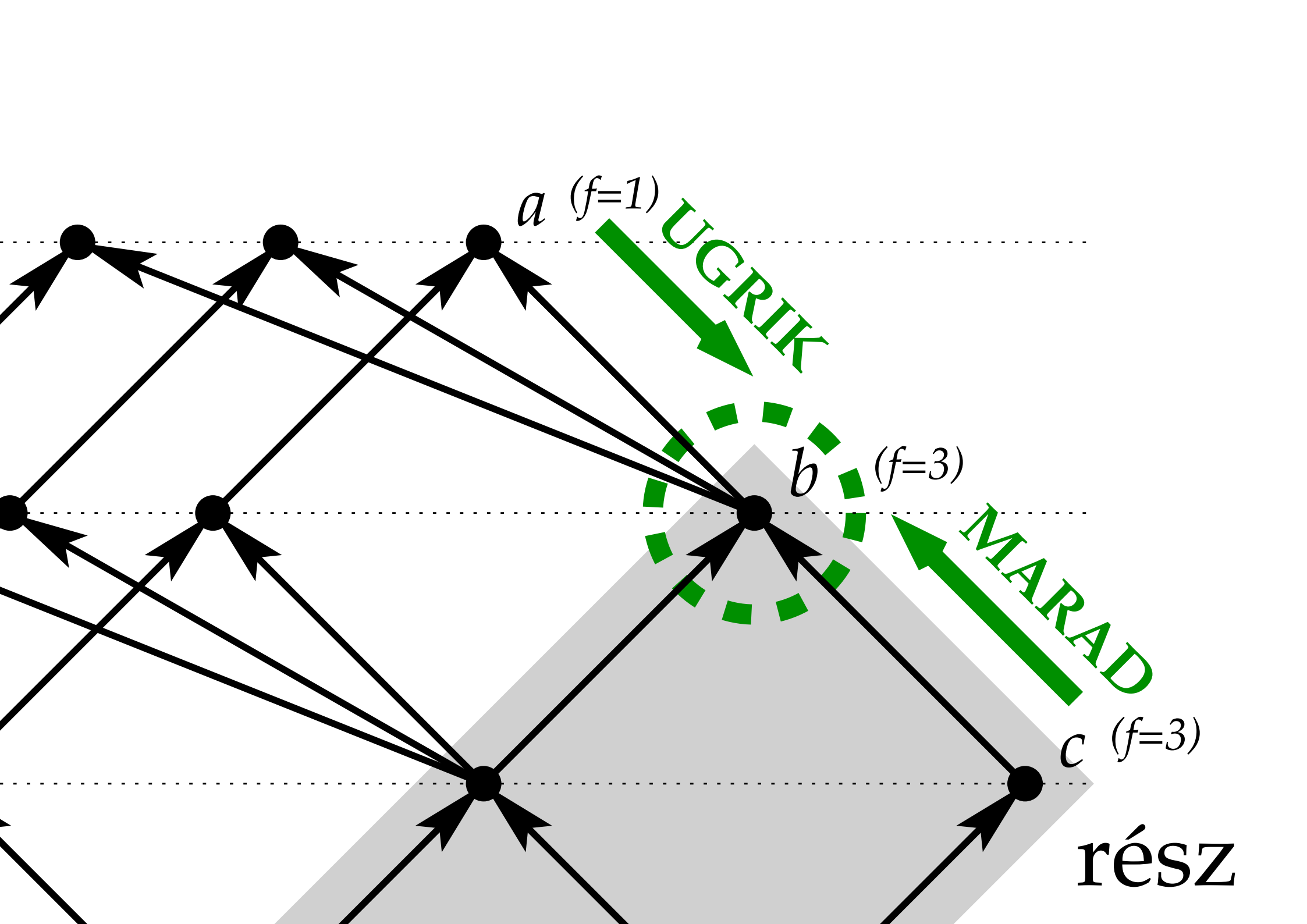


# $f$ függvény



# „ugrik és marad” módszer







# a mai újdonság

*eddig*: a mzsola adatbázison és  
különbéle nyelvű UD korpuszokon vizsgálódtam  
– ezek általános szövegek → általános isz-*ket* kapunk

*most*: a modul közvetlenül kapcsolható az e-magyar kimenetére  
függőségi elemzés szükséges! (tényleges e-magyar modul még (mindig) terv)

*azaz*: **tetszőleges magyar korpusz könnyen vizsgálható**  
speciális szövegek → **a speciális szövegeket jellemző isz-*ket* kapunk**

# korpusz beleillesztése

*kérdés:*

Hogyan feleltetjük meg a korpusz annotációját a modellünknek?

„formai” függőségek vannak = *nem stimmel*, hogy

- azonos funkciójú jelenséget mindig ugyanúgy jelöljünk,
- eltérő funkciójú jelenséget mindig eltérően jelöljünk

*cél:* formai függőségek → „**funkcionális**” függőségek

*azaz:* annotáció → ige + helyek + kitöltők → a modellbe illeszthető

*lényeg:*

a modellünk szempontjából **egységessé** alakítjuk a reprezentációt  
= az azonos funkciójú elemek és relációk kapjanak azonos jelölést

(függőségi elemzés → tagmondatolás nem szükséges, minden ige kiindulópont)

# korpusz beleillesztése

- igekötő külön dependens: hozzákapcsoljuk
- NP bővítmény → hely = esetrag; kitöltő = fej
- rag és névutó azonos kezelése
- elöljárók, *am, to/te, sich, aufeinander*

a névutó annotációja *nehéz feladat elé állítja* a rendszert

UD:

elé case → feladat obl → állít

e-magyar:

feladat ATT → elé TO → állít

# használat

1. program letöltése:

```
git pull https://github.com/sassbalint/  
    double-cube-jump-and-stay-multilingual  
cd double-cube-jump-and-stay-multilingual
```

2. sima txt korpusz beszerzése (*min. 30000 szó*):

```
wget https://mek.oszk.hu/00900/00954/00954.htm  
lynx --dump --force-html 00954.htm > sztpe.txt
```

3. függőségi elemzés e-magyar-ral (5 perc):

```
docker pull mtaril/emtsv  
docker run --rm --name emtsv -p 10005:5000 -d mtaril/emtsv  
curl -F "file=@sztpe.txt" http://127.0.0.1:10005/  
    tok/morph/pos/conv-morph/dep/conll > sztpe.tsv
```

4. korpusz feldolgozása + isz-k kinyerése „ugrik és marad” módszerrel (2 *másodperc*):

```
mv sztpe.tsv input  
make INPUT=sztpe.tsv
```

5. az eredmény megtekintése:

```
vim result/sztpe*pVCC
```

→ ebből fogunk most szemezgetni

# példa

megbízást ad vminek a gyártására  
alkalmaz módszert  
alkalmaz eszközt  
rendszerbe állít  
hadrendbe állít vmit  
szolgálatban áll  
védelmet biztosít  
ellát szolgálatot  
tevékenységet folytat  
kilátásba helyez vmit  
rakétát hordoz  
támadást indít

# másik példa

ízelítőt ad vmiből  
a gyártó alkalmaz vmit  
maradandót alkot  
vmennyi lóerőt állít csatasorba  
forgalomba állít vmit  
rajthoz áll  
nagydíjgyőzelmet arat  
átveszi a vezetést  
hivatalosan bejelent  
a kínálat bővül  
futamban bukik  
kereket cserél

# másik példa

- ✗ ízelítőt ad vmiből
- ✗ a gyártó alkalmaz vmit maradandót alkot
- ✓ vmennyi lóerőt állít csatasorba
- ✗ forgalomba állít vmit
- ✓ rajthoz áll
- ✓ nagydíjgyőzelmet arat
- ✓ átveszi a vezetést hivatalosan bejelent
- ✗ a kínálat bővül
- ✓ futamban bukik
- ✓ kereket **cserél**

# harmadik példa

lendületet ad a csapatnak  
címért harcban áll  
győzelmet arat idegenben  
átveszi a vezetést  
keményen dolgozik  
eldönti a mérkőzést  
elkezdi a felkészülést  
él a lehetőséggel  
elveszíti a mérkőzést  
magasra emeli a trófeát  
sokat fejlődik  
folytatja a pályafutását



# az igei szerkezetek *jellemzik* a témát

- *állít ↔ rendszerbe állít*
- *cserél ↔ kereket cserél*
- *győzelmet arat ↔ győzelmet arat idegenben*

csak kitöltetlen helyek esetén is:

- *'állomásozik v miben'*
- *'gyárt v mit', 'gyorsul'*
- *'megnyer v mit', 'vezet v mivel'*

# konfigurálhatóság

1. főnévi igenév bővítmény  
ilyenkor 2 db isz lesz belőle

*Egyszerűen meg kell verni minden ellenfelet.*

- *'kell megverni'*
- *'megver ellenfelet egyszerűen'*

2. tagadás

*'nem kap gólt', 'nem jelent gondot', ...*

*'sikerül' ↔ 'nem sikerül'*

*'tud' ↔ 'nem tud'*

*de: 'megnyer' ↔ 'elveszít'*

~ ha van ellentétes értelmű szó, akkor alig van tagadás!

# konfigurálhatóság

3. jelző → többszintű igei szerkezet

egy szinttel lejjebb megyünk az elemzésben

egyszintűre visszavezetve: Acc és Acc+ATT

→

*'piros lapot kap'*

*'harcban áll a bajnoki címért'*

*'meghatározó labdarúgó érkezik'*

*'jó úton jár'*

(kényelmes konfigurálhatóság kialakítás alatt)

# szerkezetek mint alapegységek

igék osztályozása:

{ fut, szalad, rohan... }, { megfog, megragad, megüt, megérint... }

↔

**igei szerkezetek** osztályozása:

*terv*: különböző hosszúságú szerkezetek megfeleltetése egymásnak  
(különböző hosszúságú, de ugyanannyi kitöltetlen helyet tartalmazó)

{ vesz górcső alá ACC, megvizsgál ACC }

{ jut eszébe NOM, kitalál ACC }

{ talál kapuba, bevesz kaput, lő gólt, rúg gólt, szerez gólt }

{ bemutat védést, véd }

{ birtokol labdát NOM, van labda ADE }

{ nem fér kétség ALL, van biztos NOM }

{ birtokol ACC<sub>1</sub> NOM<sub>2</sub>, van NOM<sub>1</sub> ADE<sub>2</sub> } (!)

# mire jó

- korpuszra jellemző igei szerkezetek kinyerése
- valamiféle továbbfejlesztett topic modelling  
*cserél* ↔ *'kereket cserél'*
- eszköz az igei szerkezetek osztályozásához
- lexikográfia
- ...

# köszönöm a figyelmet

*Korábban 'sérüléssel bajlódott', de 'bekerült a csapatba'. A mérkőzés 'végére elfáradt', sokszor 'eladta a labdát', úgy érezte, hogy 'képtelen értékesíteni a büntetőt'. De aztán 'megszakadt a sorozat'. Mikor 'bejutottak a fináléba', és 'egymásnak feszült' a két csapat, nagyon 'jó teljesítményt nyújtott', 'hátrányból álltak fel' és végül '0-ra nyerték meg a mérkőzést.' Ő 'emelte magasra a trófeát'.*

Sass Bálint — [sass.balint@nytud.hu](mailto:sass.balint@nytud.hu)

# cikkek

Sass Bálint. A duplakocka modell és az igei szerkezeteket kinyerő „ugrik és marad” módszer nyelvfüggetlensége, valamint néhány megjegyzés az UD annotáció univerzalitásáról. In: MSZNY 2020, 399-407.

Sass Bálint. The „Jump and Stay” Method to Discover Proper Verb Centered Constructions in Corpus Lattices. In: RANLP 2019, 1076-1084.

Sass Bálint. A Lattice Based Algebraic Model for Verb Centered Constructions. In: Sojka, P. et al. (Eds.): TSD 2018, LNCS 11107, 231-238.

Sass Bálint. Az igei szerkezetek algebrai struktúrája, avagy a duplakocka modell. In: Argumentum. 2018(14): 12-44. Debreceni Egyetemi Kiadó.

Sass Bálint. Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása. Ph.D. dolgozat, PPKE ITK, 2011.