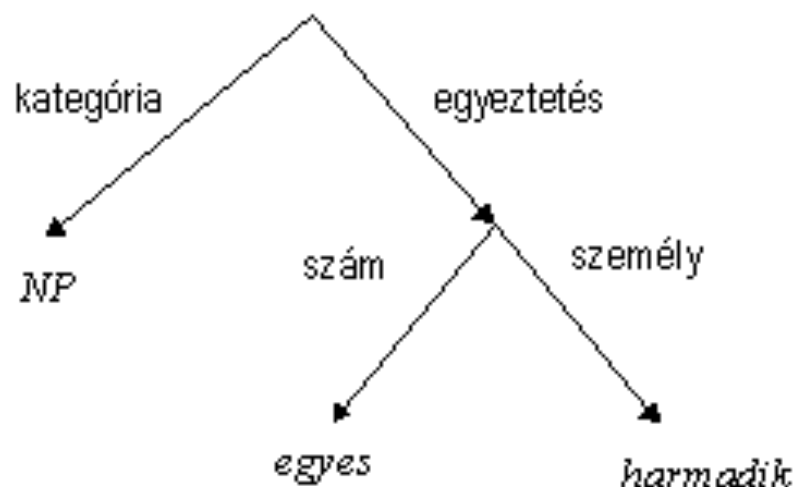


**Bevezetés a nyelvtechnológiába**  
**5.**  
**A morfológia alkalmazásai**

# Az unifikáció művelete

$$\left[ \begin{array}{l} \textit{kategória} = \textit{NP} \\ \textit{egyeztetés} = \left[ \begin{array}{l} \textit{szám} = \textit{egyes} \end{array} \right] \end{array} \right]$$
$$\left[ \begin{array}{l} \textit{kategória} = \textit{NP} \\ \textit{egyeztetés} = \left[ \begin{array}{l} \textit{személy} = \textit{harmadik} \end{array} \right] \end{array} \right]$$

A két jegyszerkezet unifikációjának eredménye:

$$\left[ \begin{array}{l} \textit{kategória} = \textit{NP} \\ \textit{egyeztetés} = \left[ \begin{array}{l} \textit{szám} = \textit{egyes} \\ \textit{személy} = \textit{harmadik} \end{array} \right] \end{array} \right]$$


# Az unifikáció közelebbről

[NUMBER SG] --> [NUMBER SG] = [NUMBER SG]

[NUMBER SG] --> [NUMBER PL] = *fail!*

[NUMBER SG] --> [NUMBER []] = [NUMBER SG]

[NUMBER SG] --> [PERSON 3] =  
                  [[NUMBER SG][PERSON 3]]

[AGREEMENT <sup>[1]</sup>][[NUMBER SG][PERSON 3]]  
                  [SUBJECT [AGREEMENT <sup>[1]</sup>] U

[SUBJECT [AGREEMENT [[PERSON 3][NUMBER SG]]]]  
= [AGREEMENT <sup>[1]</sup>][[NUMBER SG][PERSON 3]]  
  [[SUBJECT [AGREEMENT <sup>[1]</sup>] ]]

# Az unifikáció definíciója

**function** UNIFY(*f1*, *f2*) **returns** *fstructure* or failure

*f1-real*  $\leftarrow$  Real contents of *f1*

*f2-real*  $\leftarrow$  Real contents of *f2*

**if** *f1-real* is null **then**

*f1.pointer*  $\leftarrow$  *f2*

**return** *f2*

**else if** *f2-real* is null **then**

*f2.pointer*  $\leftarrow$  *f1*

**return** *f1*

**else if** *f1-real* and *f2-real* are identical **then**

*f1.pointer*  $\leftarrow$  *f2*

**return** *f2*

**else if** both *f1-real* and *f2-real* are complex feature structures **then**

*f2.pointer*  $\leftarrow$  *f1*

**for each** *feature* **in** *f2-real* **do**

*other-feature*  $\leftarrow$  Find or create

            a feature corresponding to *feature* in *f1-real*

**if** UNIFY(*feature.value*, *other-feature.value*) **returns** failure **then**

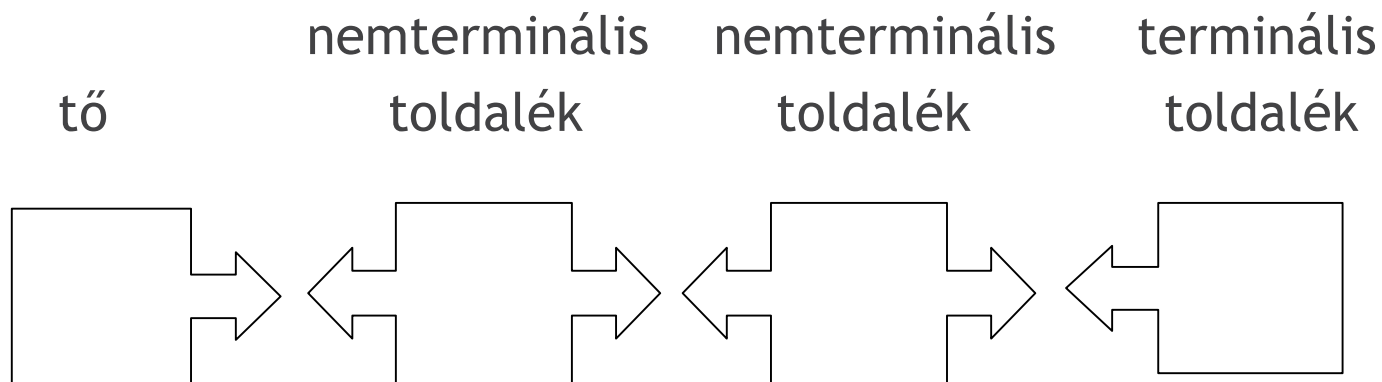
**return** failure

**return** *f1*

# HUMOR

- High-speed Unification Morphology
- jegy-érték párok
- unifikáció
- ld. Prolog, de nem rögzített aritás
- unifikáció vs. unifikálhatóság
- minden tulajdonság jegyként
- folytatási osztályok
- nincs más „valós” művelet, csak az unifikálhatóság-ellenőrzés

# Szóalaktani alapséma



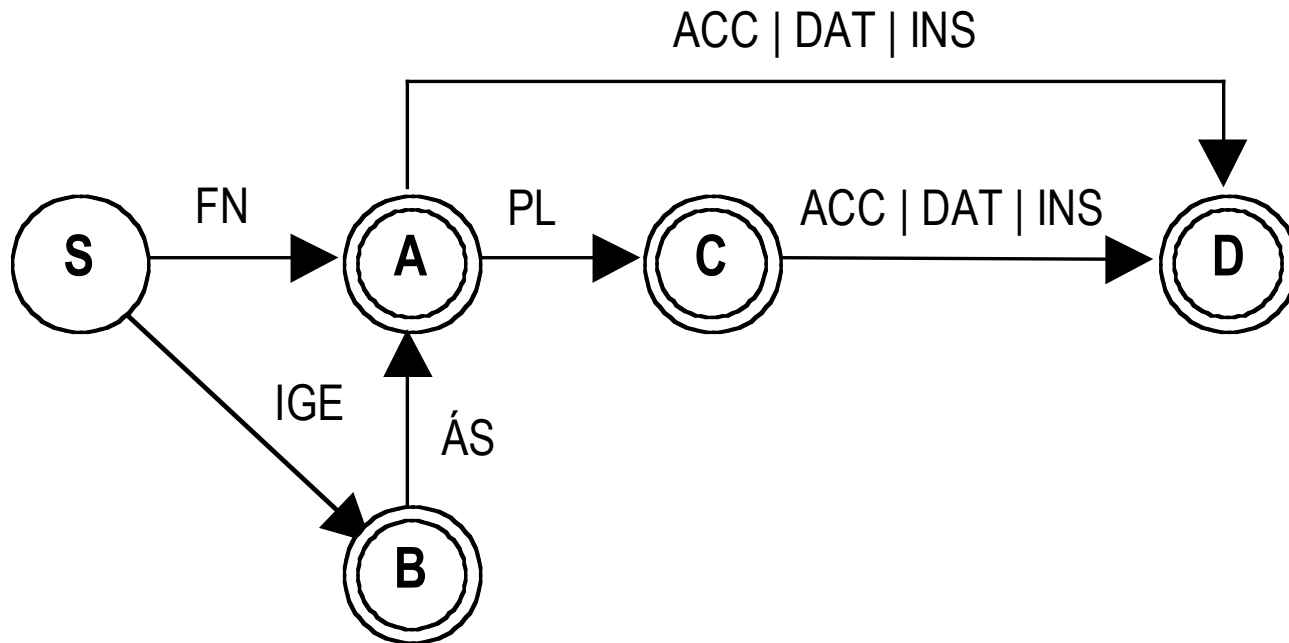
(relatív) tő / relatív toldalék

relatív tő / relatív toldalék

relatív tő / (relatív) toldalék

# Folytatási osztályok

Leegyszerűsített magyar névszói toldalékolás:



# Bináris kérdések a magyar morfo-fonológiáról

		$\alpha = +$	$\alpha = -$
1	$\alpha$ névszó	névszó	ige
2	$\alpha$ fn	főnév	melléknév, számnév
3	$\alpha$ szótári	szótári alapalak	nem szótári alapalak
4	$\alpha$ elöl	elöl képzett	hátul képzett
5	$\alpha$ kerek	ajakkerekítéses	nem ajakkerekítéses
6	$\alpha$ PL	többes szám	nem állhat többes számban
7	$\alpha$ PLkötő	PL kötőhanggal	PL nem kötőhanggal
8	$\alpha$ PERS	birt. szem.ragos	nem kap birt. szem.ragot
9	$\alpha$ ACC	van tárgyesete	nem tárgyesetes
10	$\alpha$ ACCkötő	ACC kötőhanggal	ACC nem kötőhanggal
11	$\alpha$ DAT	van részesesete	nincs részesesete
12	$\alpha$ INS:β	van eszk.h.esete	nincs eszk.h.esete
13	$\alpha$ ÁS	-ás/-és képzős	nem kap -ás/-és képzőt



# Szótövek tára

**szó** []  
[+névszó +fn +szótári –elől –kerek –PL  
–PERS +ACC –ACCkötő +DAT +INS:V]

**szav** []  
[+névszó +fn –szótári –elől –kerek +PL  
+PLkötő +PERS –ACC +DAT –INS]

**képez** []  
[–névszó +szótári +elől –kerek –ÁS]

**képz** []  
[–névszó –szótári +elől –kerek +ÁS]

# Toldalékok tára

<b>ás</b>	$[-\text{névszó} -\text{elöl} +\text{ÁS}]$ $[+\text{névszó} +\text{fn} +\text{szótári} -\text{elöl} -\text{kerek} +\text{PL}$ $+ \text{PLkötő} +\text{ACC} -\text{ACCkötő} +\text{DAT} +\text{INS:S}]$
<b>és</b>	$[-\text{névszó} +\text{elöl} +\text{ÁS}]$ $[+\text{névszó} +\text{fn} +\text{szótári} +\text{elöl} -\text{kerek} +\text{PL}$ $+ \text{PLkötő} +\text{ACC} -\text{ACCkötő} +\text{DAT} +\text{INS:S}]$
<b>ak</b>	$[+\text{névszó} -\text{elöl} -\text{kerek} +\text{PL} +\text{PLkötő}]$ $[+\text{névszó} -\text{elöl} -\text{kerek} -\text{PL} -\text{PERS} +\text{ACC}$ $+ \text{ACCkötő} +\text{DAT} +\text{INS:K}]$
<b>ek</b>	$[+\text{névszó} +\text{elöl} -\text{kerek} +\text{PL} +\text{PLkötő}]$ $[+\text{névszó} +\text{elöl} -\text{kerek} -\text{PL} -\text{PERS} +\text{ACC}$ $+ \text{ACCkötő} +\text{DAT} +\text{INS:K}]$
<b>nak</b>	$[+\text{névszó} -\text{elöl} +\text{DAT} ]$ $[]$
<b>nek</b>	$[+\text{névszó} +\text{elöl} +\text{DAT}]$ $[]$

# Unifikációs morfológia

<i>szó</i>	[+névszó +fn + <b>szótári</b> –elől –kerek –PL –PERS +ACC –ACCkötő +DAT +INS:V]
* <i>szav</i>	[+névszó +fn <u>–szótári</u> –elől –kerek +PL +PLkötő +PERS –ACC +DAT –INS]
<i>szó+nak</i>	[+ <b>névszó</b> +fn +szótári <u>–elől</u> –kerek –PL –PERS +ACC –ACCkötő + <b>DAT</b> +INS:V] [+névszó <u>–elől</u> + <b>DAT</b> ]
* <i>szav+nak</i>	[+ <b>névszó</b> +fn –szótári <u>–elől</u> –kerek +PL +PLkötő +PERS –ACC <u>–DAT</u> ] [+névszó <u>–elől</u> + <b>DAT</b> ]
* <i>szó+vel</i>	[+ <b>névszó</b> +fn +szótári <u>–elől</u> –kerek –PL –PERS +ACC –ACCkötő +DAT + <b>INS:V</b> ] [+névszó <u>+elől</u> + <b>INS:V</b> ]
* <i>szav</i>	[+névszó +fn <u>–szótári</u> –elől –kerek +PL +PLkötő +PERS –ACC +DAT –INS]
<i>képz+és+nek</i>	[– <b>névszó</b> –szótári + <b>elől</b> –kerek + <b>ÁS</b> ] [– <b>névszó</b> + <b>elől</b> + <b>ÁS</b> ] [+ <b>névszó</b> +fn +szótári + <b>elől</b> –kerek +PL +PLkötő +ACC –ACCkötő + <b>DAT</b> +INS:S] [+névszó + <b>elől</b> + <b>DAT</b> ]

# Morfo-fonológiai „guesser”

**kacsónak**

✓ FN

kacsóna + k

\* FN + PL

kacsón + ak

\* FN + PL

kacsó + nak

\* IGE + PL3

**kacsó** + nak

✓ FN + DAT

kacs + ó + nak

\* IGE + MNI + DAT

ka | csónak

\* FN|FN

# Hibák és normák

- a helyesírási normák, a szabályok és a gépi helyesírás-ellenőrzés
- a szabályzatok viszonya a gépi helyesírás-ellenőrzéshez
- formai és jelentéstani szempontok gépi kezelése
- a gép a „nyelvhelyességi tanácsadó” szerepében
- a nehezen értelmezhető elemzések okozta problémák

# Tipikus hibák a számítógéppel létrehozott dokumentumokban

- karakterhibák
- valódi helyesírási hibák
- nyelvhelyességi hibák
- tipográfiai hibák
- helyesírás-ellenőrzés a szavak szintjén
- a szóellenőrzés és a nyelvhelyesség-ellenőrzés viszonya
- a nyelvi programrendszer lehetséges hibái (*kör/kőr, -ít*)

# Szó szinten javítható és nem javítható hibák

## JAVÍTHATÓK

- Betűhibák és elütések
- Hibás egybeírások
- Hibás kötőjelezés
- Szóismétlések (!)

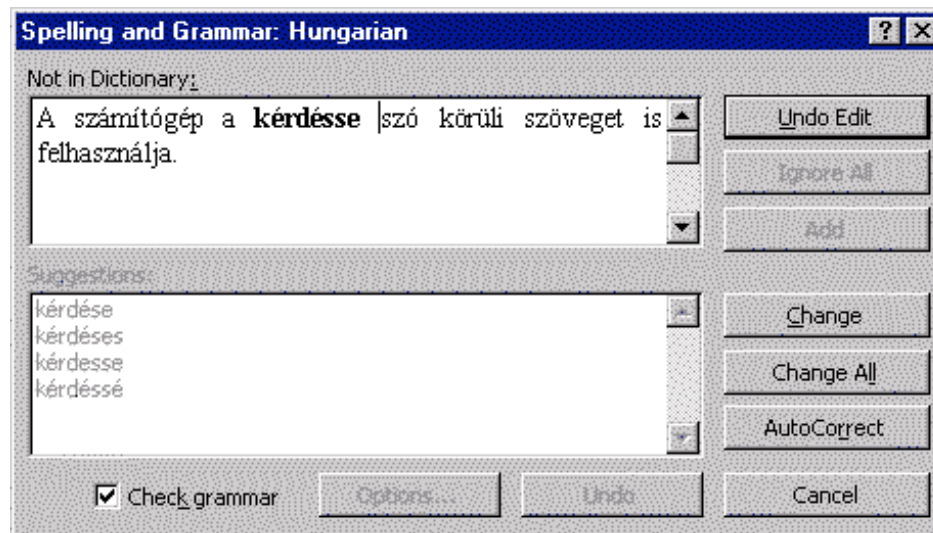
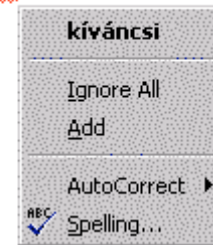
## NEM JAVÍTHATÓK

- Hibás különírások
- Központozási hibák
- Egyeztetési hibák
- Szórendhibák
- Vonatkozó névmások helytelen használata
- Stílushibák, helytelen szóhasználat
- Szóközök többszörözése vagy elhagyása
- „Értelmes” szóhibák

# A szóellenőrző moduljai

- alapszótár ill. morfológia
- ajánlómodul és adatbázisai
- időleges saját szótárak
- kiegészítő szótár
- kizáró szótár
- a ragozó saját-szótárak ötlete

kiváncsi





# Book

[illegible]

# Damerau-Levenshtein-távolság

A Damerau-Levenshtein távolság Hamming-távolság egy másik általánosítása: a minimális számú beszúrás, törlés, betűcsere és helycsere segítségével kiszámított távolság két fűzér között.

A Levenshtein-számításból úgy kapható, hogy a fő ciklust kiegészítjük ezzel:

```
if (i > 1 and  
    j > 1 and  
    str1[i-1] = str2[j-2] and  
    str1[i-2] = str2[j-1])  
then d[i, j] := minimum( d[i, j], d[i-2, j-2] + cost) // transzpozíció
```

Damerau annak idején (1964) azt is állította, hogy az emberi elütések 80%-a ezekkel korrigálható.

Ugyanakkor eredetileg ő csak egyetlen karakteres elütésekkel számol, szemben a sima Levenshtein-távolsággal.

# A szóellenőrzés menete

## (1) Morfológiai elemzés

kérdése ➡ <nincs ilyen szó a magyarban>

## (2) Ajánlás

törlés:

érdesse, krdesse, kérésse, kédesse, kérdése, kérdéss

helycsere:

ékrdesse, krédesse, kérédsse, ..., kérdéses

nyelvspecifikus csere:

kérdéssé, kérdesse, ...

...

## (3) Ellenőrzés morfológiai elemzéssel

# Szóellenőrzés morfológiával

## kérdése

kérdés[FN]+e[PSe3]

főnévi

kérd[IGE]+és[IF]+e[PSe3]

főnévi

## kérdéses

kérdéses[MN]

melléknévi

kérdés[FN]+es[SKEP]

melléknévi

kérd[IGE]+és[IF]+es[SKEP]

melléknévi

## kérdesse

kérd[IGE]+es[MUV]+se[TPe3]

igei

## kérdéssé

kérdés[FN]+sé[FAC]

főnévi

kérd[IGE]+és[IF]+sé[FAC]

főnévi

# Nyelvhelyesség-ellenőrzés a szóhatáron túl

- lehetséges-e mondat szintű helyesírás-ellenőrzés?
- „grammar checker” ?
- parciális elemzések
- hiba-nyelvtan vs. nyelvten
- hibaelemzések, a hibák súlyozása
- a hiba és a nem-hiba határának elmosódása
- a nyelvi vagy a formai természetű hibák szűrésének preferálása
- stílusellenőrzés számítógéppel

# A mondatszinten felismerhető jelenségek

- Névelőegyeztetés: **Az kutya ugat.**
- Vesszőhiány: **Nem látok vak vagyok.**
- Algoritmikus szóösszetétel: **hat lábú**
- Szemantikus szóösszetétel: **kecske béka**
- Hibás szóösszetétel: **azután a nő után**
- Hiányos szerkezetek: **Megy hasú ebédelni.**
- Téves szóhasználat: **Egyenlőre nincs mit tenni.**
- Idegen szavak szűrése: **Elromlott a printer**
- Terjengős kifejezések: **büntetést eszközöl**
- Trágár szavak szűrése: **Le van \*\*\*va.**
- Szóközhiány -felesleg: **Nem én nem akarom.**

# Mondatellenőrzés

Aki másnak vermet és maga esik bele.

Elment mellettem anélkül hogy köszönt volna.


A **hog**y kötőszó elé vesszőt kell írni!  
Javaslat: **mellettem**.,

Ignore Sentence

 Grammar...

Ha ez egy összetett mondat, a jelzett helyen  
vagy környékén egy vessző hiányzik.  
Ellenőrizze!  
Javaslat: **maga**.,

Ignore Sentence

 Grammar...

## Spelling and Grammar: Hungarian

Grammatical error:

**Elszeretném** adni az autómat.

Suggestions:

Az igekező nem a segédigéhez tartozik.  
Javaslat: **El szeretném**.



☒ Check grammar

Options

Undo

## Helyesírási Szabályzat

File Edit Bookmark Options Help

Contents Search Back Print

**131.c** A magyar helyesírás szabályai 11. kiadás

Az igekező külön szó marad, ha közte és igeje (vagy igeve) között más szó is van: *el ne késs, le nem tenné, újjá is építjük, közre kell bocsátani, haza szabad menni, föl sem véve* stb.

Hasonlóképpen: *be nem avatkozás, meg nem értés* stb.

Az igekezők különírása vagy egybeírása gyakran értelmi különbségre is utal: annak jelölője, hogy az igekező a közvetlenül utána álló igehez tartozik-e, vagy az ezt követőhöz. Más és más tehát: *megvan húszéves*, de: *meg van töltve*; *megfogom a lepke*t, de: *meg fogom nézni*; *leszokott a dohányzásról*, de: *le szokott utazni vidékre*; *beleszeretett a lányba*, de: *el bele szeretett volna szólni*; *ellesszek nélküle*, de: *el leszek fáradva*; *megtudták a hírt*, de: *meg tudták javítani*, stb.

# Automatikus szövegelválasztás

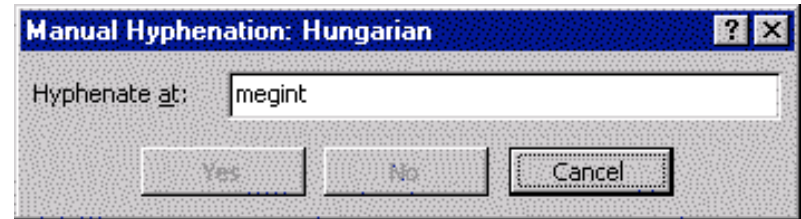
- az elválasztás alkalmazása
- automatikus és interaktív módszerek
- az elválasztás alapszabályai
- a morfológiai felülbíráltás kérdése
- alternatív elválasztások kezelése  
(többértelműség, illetve a szabályok „engedékenysége” miatt)
- tipográfiai szempontok
- különleges elválasztások (hosszú kettős mássalhangzók, mássalhangzó-háromszorozódás) helyes kezelése



# A magyar elválasztás szabályai

Alap	Elválasztva	Példa
#VV	#VV	<u>autó</u>
#VC	#VC	<u>al</u> aki
VV#	VV#	haz <u>ai</u>
VV	V-V	ba- <u>u</u> xit
VC1C2V	VC1-C2V	er- <u>ké</u> ly
VCiCiV	VCi-CiV	vet- <u>t</u> em
VCc1c2V	VC-c1c2V	mor- <u>z</u> sa
Vc1c2CV	Vc1c2-CV	as- <u>z</u> -tal
Vc11c12c21c22V	Vc11c12-c21c22V	taris- <u>sz</u> -nya
Vc1c1c2V	Vc1c2-c1c2V	össze/össz-sze

# Book



11

1

1

1



# Számítógépes szinonimaszótárak és tezauruszok

- a szinonimákról
- szinonimaszótár vagy tezaurusz?
- tárolási és keresési problémák
- a rokon értelműség definíciója
- az automatikus csere problémái
- tő-visszaállítás
- többértelműségek kezelése
- a lexikai és a szintaktikai szó különbségéből adódó nehézségek
- az összetett szavak szinonimáinak problémája
- morfológiai generálás minta alanián

# Szinonimaszótárak

