

ЈЕЗИК КАО САМООРГАНИЗУЈУЋИ СИСТЕМ

Александар Костић

Српска академија наука и уметности

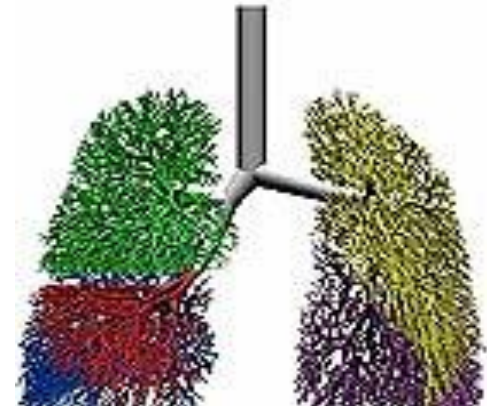
Структурација природних система

- Облици (структуре) у природи нису *sui generis* већ су **post hoc**, генерисани бројним параметрима и њиховом интеракцијом.
- Облици (структуре) природних система су решења близу оптимума за дати скуп ограничења.
- Да би се разумела логика морфологије природних система неопходно је идентификовати релевантне параметре који генеришу њихове структуре.

Морфологија природних система

Пример приципа из кога се генеришу различите структуре

- При расту сваког система долази до неједнаког прираштаја површине и запремине (чињеница геометрије).
- Неједнак раст површине и запремине може да доведе до распада система.
- Решење које срећемо у природи је максимизација површине и минимизације запремине (принцип минимума-максимума).



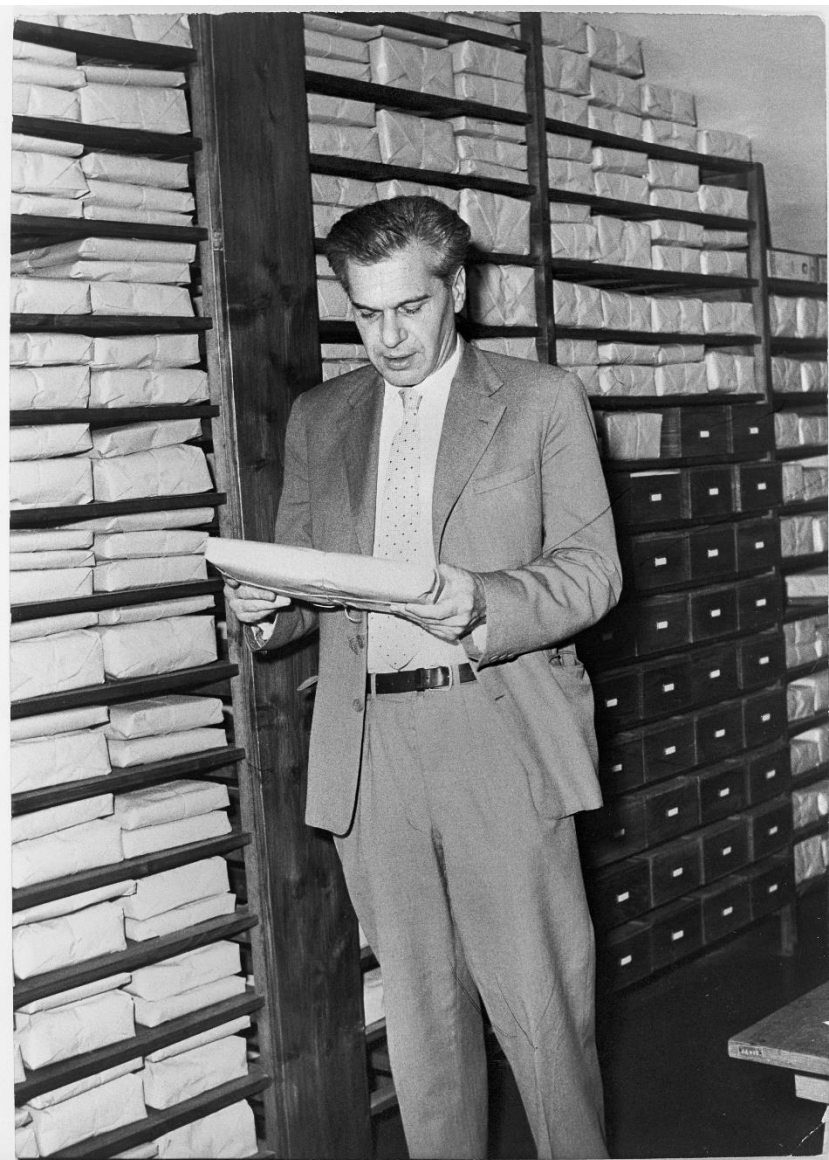
Језик као природни систем

- Лингвисти и пихолингвисти се слажу да је језик природни систем.
- Импликације оваквог става су, међутим, у великој мери занемарене.
 - Ако је језик природни систем, он је по дефиницији **самоорганизујући систем** који дели опште организационе принципе за другим природним системима.
 - Скуп ограничења која генеришу језичке структуре је **когнитивне природе**.
 - Поставља се питање која су то ограничења.

Когнитивна ограничења

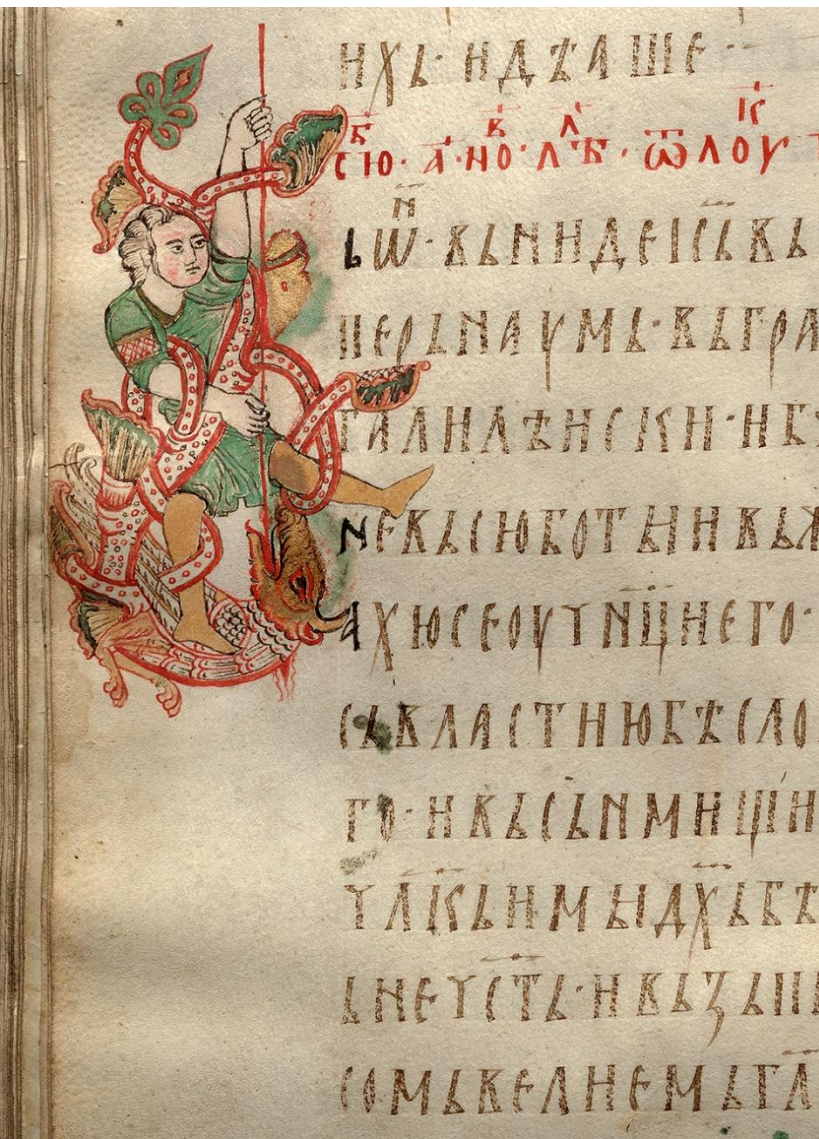
- Когнитивни систем није систем опште намене који са подједнаким успехом обавља различите когнитивне задатке.
- Постоје интринзична ограничења когнитивног система (нпр. ограничени капацитети радне меморије).
- Ова ограничења су **онтичко својство**.
- Когнитивне структуре (укључујући и језичке структуре) су решења близу оптимума за дата когнитивна ограничења.

ВЕРОВАТНОЋЕ ГРАМАТИЧКИХ ОБЛИКА (ЈЕЗИЧКИ КОРПУС)



ДИЈАХРОНИ КОРПУС СРПСКОГ ЈЕЗИКА

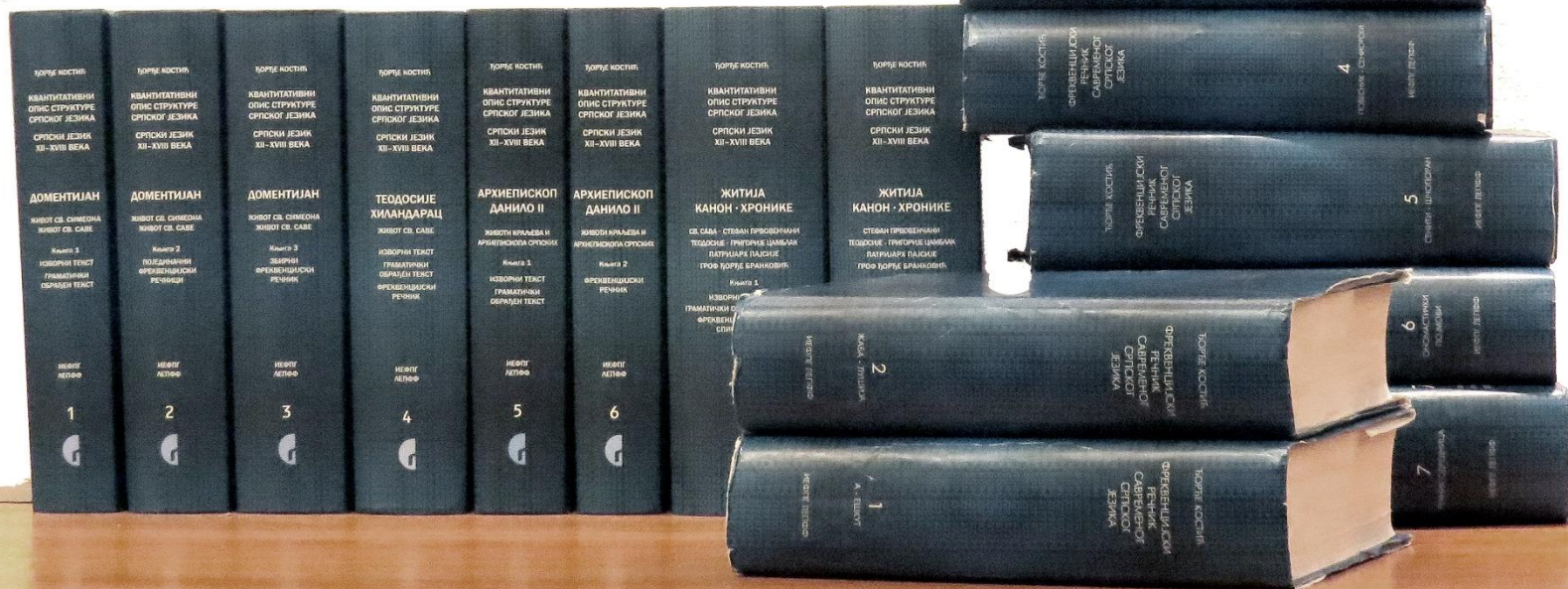
- 11.000.000 речи
- Свака реч је **ручно** граматички аотирана.
- Систем аотације обухвата око 2.500 граматичких облика.
- Узорак:
 - Српски језик од 13. до 18. века
 - Језик 18. и прве половине 19. века (славеносрпски)
 - Сабрана дела Вука Ст. Караџића
 - Језик друге половине 19. века
 - Савремени језик (три подузорка)



КОРПУС СРПСКОГ ЈЕЗИКА ОД 13. ДО 18. ВЕКА

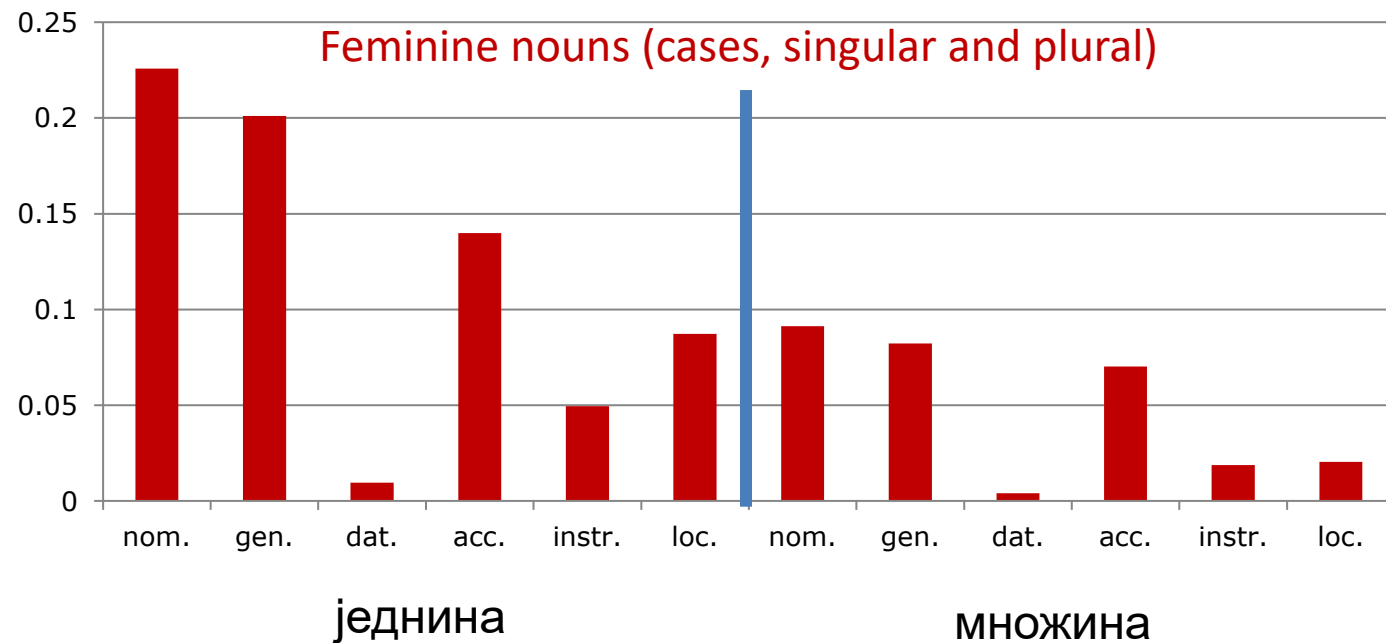
- Ручно граматички аотиран.
- Систем анотације обухвата око 3.000 граматичких облика.
- 500.000 речи
- Житија
- Старе српске повеље и писма
- Сва дела су граматички аотирана у целости.

и му вспрепозодна. 19 7 154/32	и шт 19 154/32	и чловека 19 7 154/32	и му неправедна 19 7 154/33
и и 19 154/33	и му лостіва 19 7 154/33	и му избави 19 7 154/33	и му его 19 7 154/33
скрощен шал. 19 154/33	и мв.сх.а. 19 154/33	и по 19 154/33	и глаголу 19 7 154/33
и и сича 19 7 154/34	и му твоего 19 7 154/34	и му единороднаго, 19 7 154/34	и господина 19 7 154/34



ДИСТРИБУЦИЈЕ ВЕРОВАТНОЋА ГРАМАТИЧКИХ ОБЛИКА

Вероватноћа падежних облика именица женског рода у савременом српском језику



Преферентна дистрибуција вероватноћа

- Да ли постоје ограничења који условљавају дистрибуцију вероватноћа граматичких облика?
- Једно од ограничења је **КОГНИТИВНО** преферентна дистрибуција вероватноћа граматичких облика унутар дате парадигме (нпр. деκлинације).
- Претпостављамо да би критеријум когнитивно преферентне дистрибуције вероватноћа граматичких облика могла да буде њена **удаљеност од максималне ентропије**.

Ентропија и Шенонова једнакост

- Ентропија у теорији информације – просечна количина информације коју емитује неки извор.

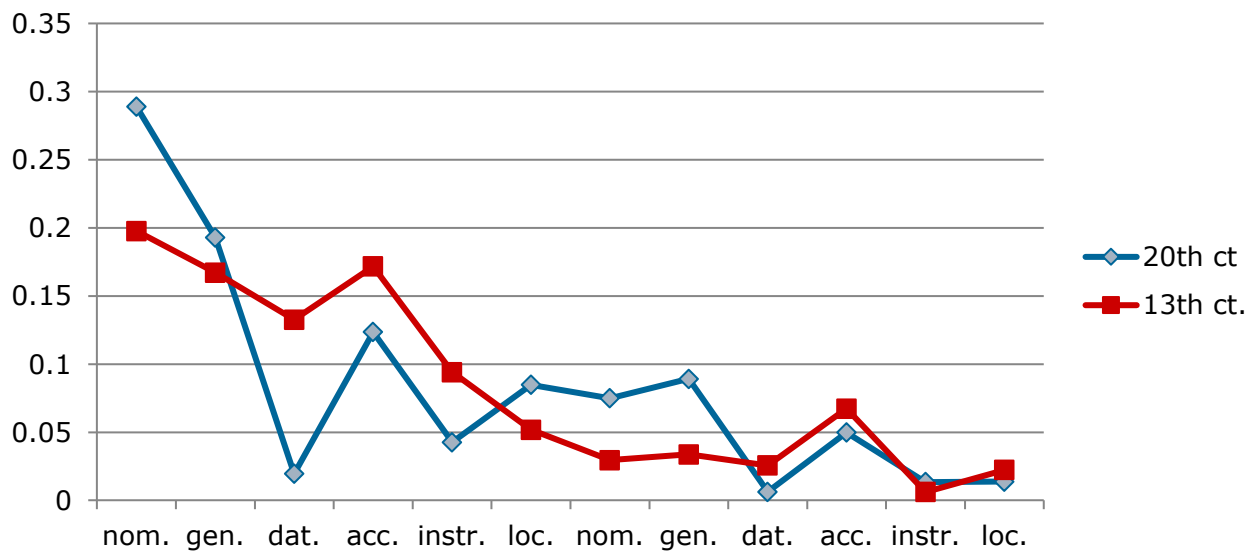
$$H_{(X)} = -\sum_i p_{(x)} \log_2 p_{(x)}$$

- Максимална ентропија – стање система у коме сви елементи имају једнаку вероватноћу.
- Поставља се питање како одредити удаљеност система од максималне ентропије.
- Дескриптор који описује ову удаљеност је **Шенонова једнакост (Shannon's equitability)** /количник добијене и максималне ентропије/.

$$E_H = H_{(x)} / H_{\max}$$

13. – 20. век (вероватноћа падежа именице)

мушки род

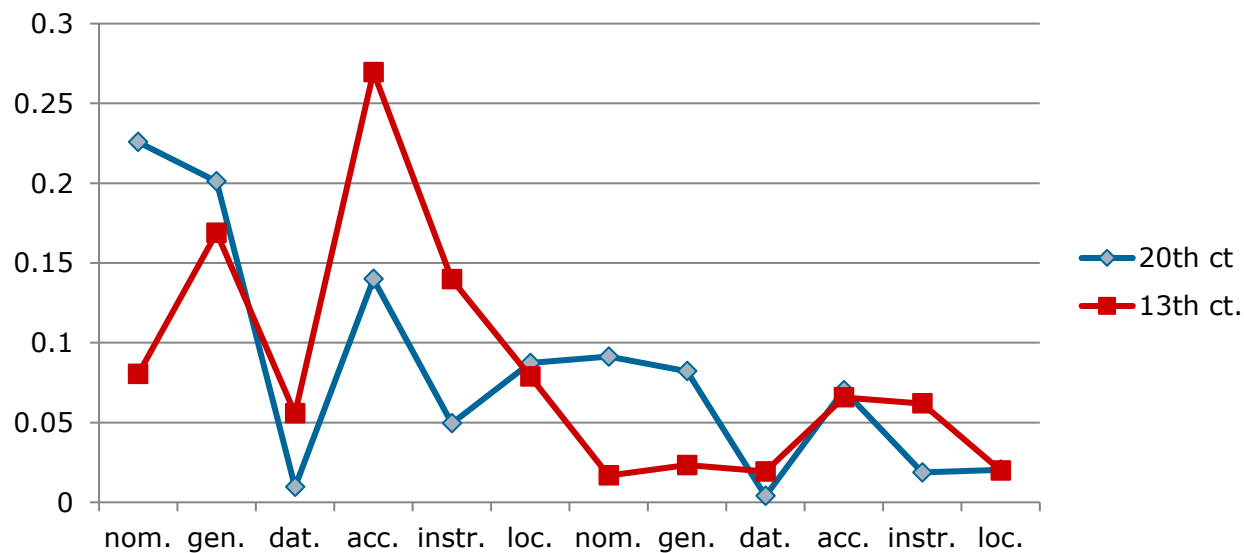


13th ct.
 $E_H=0,87$

20th ct.
 $E_H=0,83$

13. – 20. век (вероватноћа падежа именице)

женски род

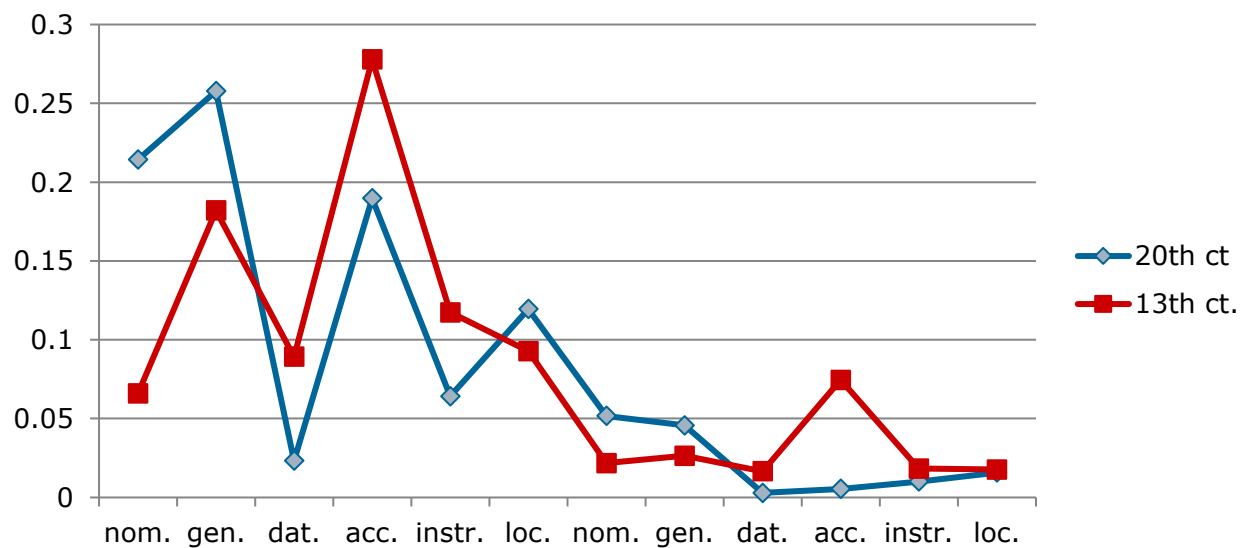


13th ct.
 $E_H=0,86$

20th ct.
 $E_H=0,85$

13. – 20. век (вероватноћа падежа именице)

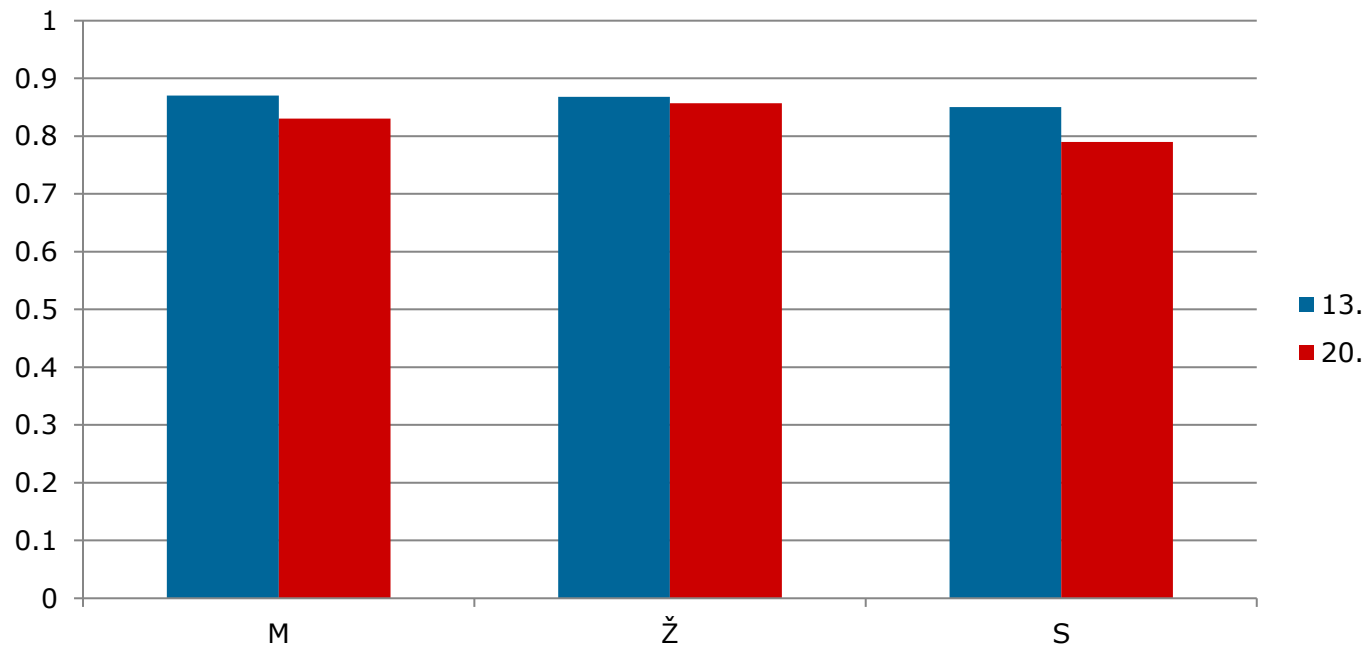
средњи род



13th ct.
 $E_H=0,85$

20th ct.
 $E_H=0,80$

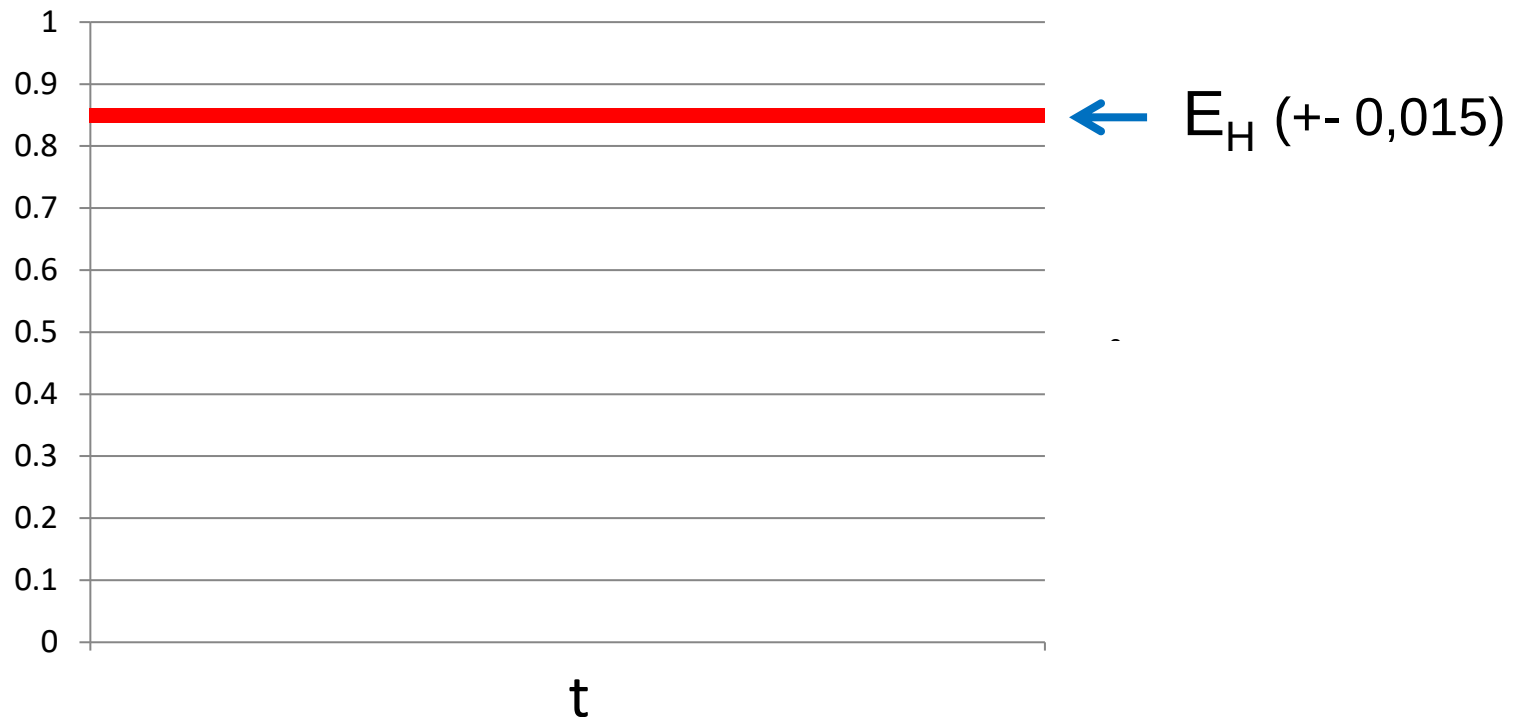
Удаљеност од максималне ентропије за два временска узорка



13th ct.
 $E_H=0,86$

20th ct.
 $E_H=0,83$

Опсег варирања E_H



- Конзервација опсега варирања E_H индикатор је **самоорганизације** система и његове стабилности у опсегу од седам векова.

Опсег варирања E_H

- Дијахроно варирање вероватноћа појединачних падежних облика именице је, локално гледано, несистематско (арбитрарно).
- Локална арбитрарност је, међутим, ограничена нивоом хомогености дистрибуције вероватноћа падежних облика унутар дате парадигме (деклинације).
- Без обзира на облик дистрибуције вероватноћа падежних облика, њена удаљеност од максималне ентропије варира у строго ограниченој маргини на мање-више фиксној удаљености од максималне ентропије.
- Вредност удаљености дистрибуције од максималне ентропије може да се схвати као **индекс стабилности система** која је когнитивно усовљена.
- Конзервација удаљености од максималне ентропије у распону од седам векова индикатор је **самоорганизације** система.