

Egy kvantitatív szemiotika felé?

Takács Boglárka

<http://www.prezzey.net>

Medizinische Universität Wien
Mikata Kft.



Így merült fel a probléma:

Hogyan lehet megállapítani egy jelsorozatról, hogy természetes nyelvi jellegű-e? (Meg lehet-e?)

Sokan azt gondolják, ez egy megoldott kérdés.

“If we apply [signal entropy] to dolphin records, we get values similar to human languages, a new fact to support the idea that these animals possess a sophisticated language” (Ballesteros, 2010)

Akár nyelvészeti cikkekben is:

“Using the technique on the Pictish symbols established that it is unlikely that they are random or sematographic (heraldic) characters, but that they exhibit the characteristics of written languages.”

(Lee et al, 2010)



(Ilyesmikről szól a második idézet...

(a)



(b)



)



Sokáig nem érdekelt senkit a téma,
de a közelmúltban több új cikk.
Elsősorban Rao és mtsai, cél:
igazolni, hogy az Indus-völgyi civilizáció által
hátrahagyott feliratok *szövegesek*
és nem csak valamilyen képnnyelvet tartalmaznak



Innen indult a vita:

Rao et al (2009). Entropic Evidence for Linguistic Structure in the Indus Script. *Science*, 324: 1165.

→ Hol is jelent meg? Miért ott? (Sproat 2010)

→ Miért is ez a címe?

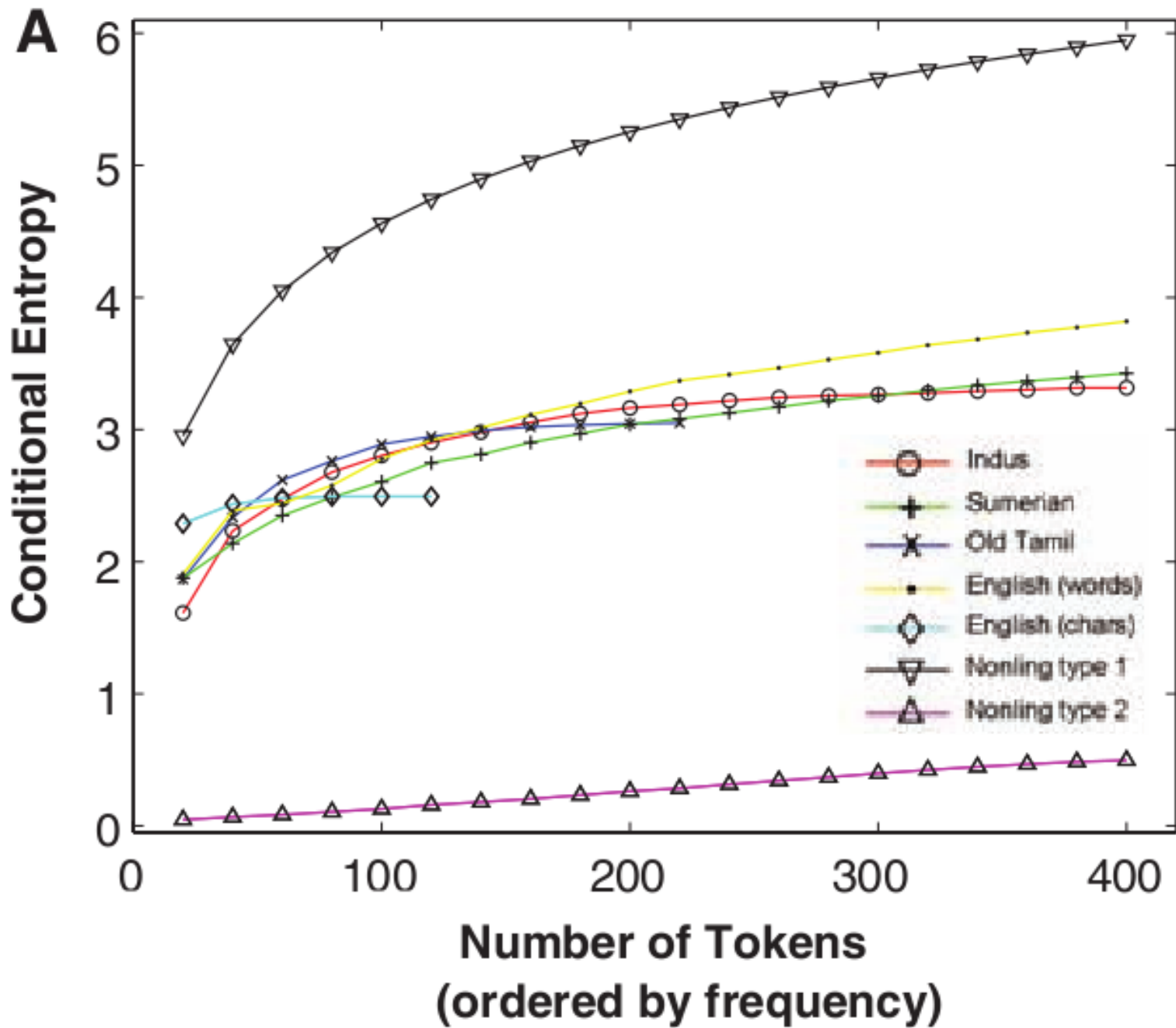
Állításuk:

a szimbólumsorozatok feltételes entrópiája hasonlít a természetes nyelvek feltételes entrópiájához, de nem hasonlít másfajta jelsorozatokéhoz.

Mint például a:

- * Fortran
- * DNS bázissorrend
- * Mesterséges, nem nyelvi jelsorokat modellező adatsorok





ÁLLJUNK MEG

STÁVA

EGY SZÓRA !

Hogyan definiáljuk azt, hogy “hasonlít”?
(Az nem érv, hogy “ránézünk a grafikonra”.)

Még súlyosabb probléma:

A kritikusok (Sproat, Shalizi, Liberman) rögtön előálltak olyan jelsorozatokkal, amelyeknek ugyanolyan a feltételes entrópiája, mint a nyelveké, de nem nyelvek.

(Python, R scriptek – mindenki utánuk csinálhatja)

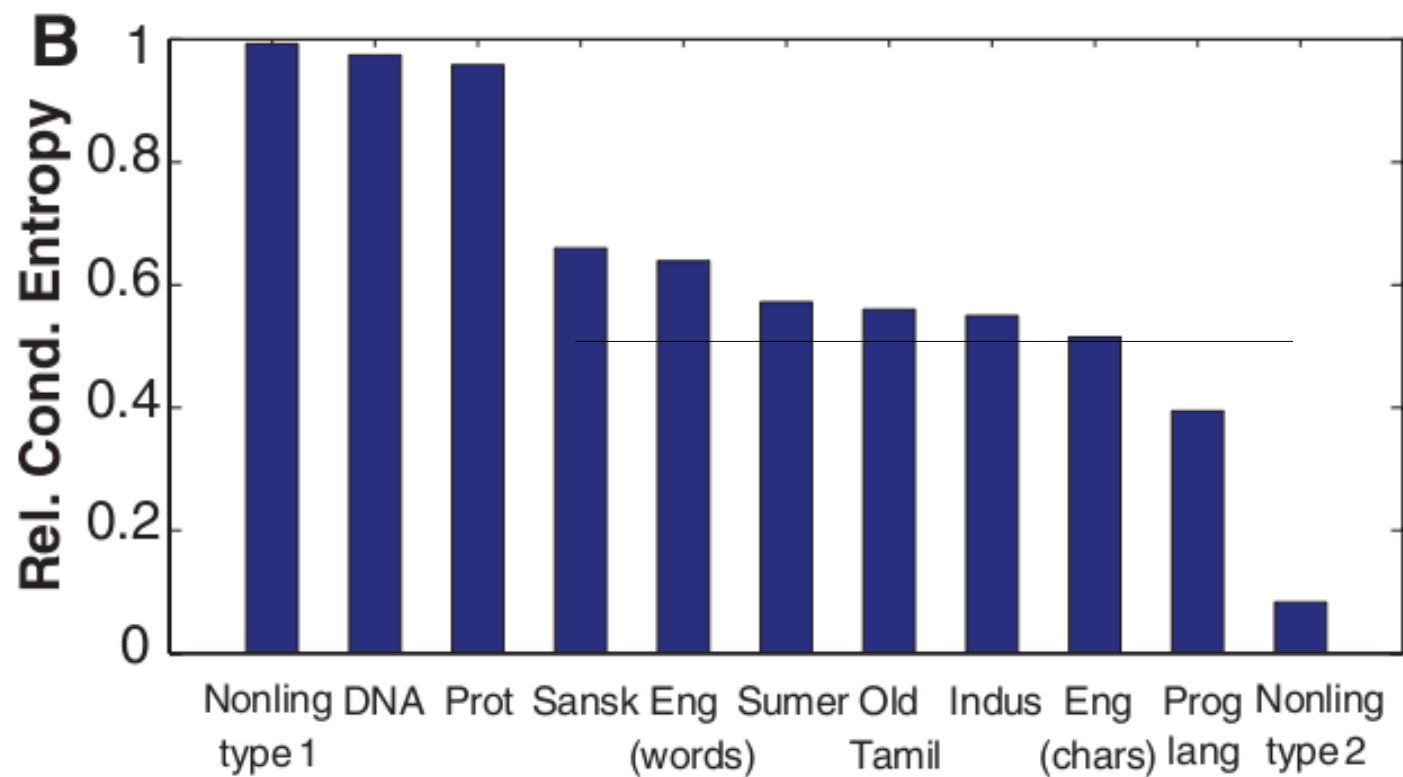
→ Tehát ez továbbra sem döntő bizonyíték az Indus-völgyi feliratokról

De minket most nem is elsősorban azok érdekelnek, hanem a kontroll minták...



Honnan tudjuk, hogy a mesterséges jelsorok tényleg “nem nyelvi szimbólumrendszereket” modelleznek-e? (Milyenek azok?)

→ Amikor valós adatokat vesznek, kisebb a különbség:



. o O (akkor most vagy a szanszkrit nem nyelv, vagy a Fortran is az?)



Hogy okosabbak lehessünk, többet kellene tudni a *nem nyelvi jellegű, de attól még strukturált szimbólumrendszerekről.*

Nem az a valós ellentét, hogy:
természetes nyelv ↔ véletlen sorozatok

(ezt a kettőt azért elég jól el tudjuk különíteni...)

Hanem egy kontinuumról van szó:

(minden esetben a biztos megoldás ;))

Term. nyelv - nyelvyszerű dolgok - nagyon nem nyelv

Minket most (kivételesen?) a középső rész érdekel.

- Milyen tekintetben nyelvyszerűek (több dimenzió?)
- Hogyan viselkednek → leíró
- Tudunk-e előrejelzéseket tenni róluk
a nyelvészet eszközeivel → magyarázó



Azért van néhány érdekes összefüggés.

Zipf-törvény (1935!):

Minél gyakoribb egy elem, annál rövidebb

Természetes nyelvekre igaz szokott lenni, de sok más emberi tevékenységre is.

Pl. számokra (Benford-törvény):
az első számjegy a leggyakoribb

Alkalmazása:

pl. pénzügyi csalások felderítése! (Varian, 1972)

Mi lenne egy Univerzális Szemiotikában
- egyáltalán lehetséges ilyen?



Van-e már ilyesmi?

Egyesek szerint a mesterséges intelligencia tulajdonképpen ugyanez lenne.

Sokminden, ami magát kvantitatív, komputációs, stb szemiotikának nevezi, tulajdonképpen szemantika vagy nyelvi alapú narratívakutatás + van egy ergonómiai irányzat is, amit így hívnak.

A maradék nagyon sokféle, nincs egységes paradigma.

Lesz-e a jövőben?

Sproat tervez gyűjteni *korpuszokat* kiindulásul
A többi rajtunk áll!



Köszönöm a figyelmet!

Ha valakit érdekel az Indus-völgy vita:
<http://www.prezzey.net/irasok-magyarul/>



Mi az a feltételes entrópia? *(nagyon köznyelvi)*

Információelméleti **entrópia** (Shannon): szövegekre alkalmazva akkor maximális, ha minden jel egyforma valószínűséggel fordulhat elő, és akkor minimális, ha mindig ugyanaz a jel jelenik meg.

Feltételes entrópia: ha ismerjük az első jelet (például egy A betűt), mit tudunk mondani a következőről. Ha az A betű után mindig W betű fordul elő és hasonlóképpen minden jel után egyértelmű, hogy mi jön, akkor a feltételes entrópia minimális. Maximális pedig akkor, ha az első betű ismeretében továbbra is akármilyen következhet. A nyelvek feltételes entrópiája a két szélsőség között helyezkedik el

