

HIBRID NYELVTECHNOLÓGIÁK

Ács Judit, Borbély Gábor, Makrai Márton, Nemeskey Dávid,
Recski Gábor, Kornai András

**#Istand
withCEU**

2017 november 14



MTA
SZTAKI

Kornai



Hibrid nyelvtechnológiák



M Ű E G Y E T E M 1 7 8 2

2017 november 14

1 / 13

- **Strukturalizmus** Saussure (1879) *Mémoire sur le système primitif des voyelles dans les langues indo-européennes* – Tesnière (1959) *Éléments de syntaxe structurale*
- **Generativizmus** Chomsky (1955) *The Logical Structure of Linguistic Theory*; Chomsky-Halle (1968) *The Sound Pattern of English*
- **Szintézis** Pereira (2000) *Formal grammar and information theory: Together again?* *Philosophical Transactions of the Royal Society* vol A 358 pp 1239–1253

A RACIONALISTA ÉS AZ EMPIRIKUS MEGKÖZELÍTÉS

	racionalista	empirista
kiindulópont	elmélet	mérés
apparátus	deduktív	induktív
bemenet	diszkrét	folytonos
formalizmus	algebra, logika	statisztika, valsz.
UG	csakis	talán

A MÉRLEG ÁTBILLEN (HANGTAN)

- SPE (1968): környezetfüggő szabályok, soros alkalmazás **Ez nagyon nehéz**
- Johnson (1970) **véges állapotú**
- Jelinek (1976) **rejtett állapotok**
- Koskenniemi (1983) **párhuzamos alkalmazás**
- Atal, Itakura, Pols, Mermelstein (-1975) lineáris prediktív ill. kepsztrális kódolás **ma minden mobilban**
- Prince and Smolensky (1993) **prioritásos szabályrendezés**
- Pater, Hayes, Tessier, Tesar, Zuraw (2007-) **súlyozott prioritások gépi tanulása**

A MÉRLEG ÁTBILLEN (MONDATTAN)

- Chomsky (1955,1956,1957,1965): transzformációs szabályok, soros alkalmazás **Ez nagyon nehéz**
- Bresnan (1983) **ko-deszkripció**
- Gazdar et al (1985) **környezetfüggetlen**
- Brown (1990) **véges állapotú (ngram)**
- Legendre and Smolensky (1992) megosztott harmonikus
- de Marneffe, Petrov, McDonald, Nivre (2003-) **univerzális függőségi nyelvtan**
- Kalchbrenner, Blunsom, Cho, Bengio (2008-) **ideghálós gépi fordítás**

TÖRTÉNETÜNK

- Kezdetek: Origo/Northern Light együttműködés (2002-2006)
- Szintér: BME Média Oktató és Kutató Központ (MOKK)
számítógépes társadalomtudomány első magyar műhelye:
számítógépes nyelvészet, digitális szerzői jog, kulturális termelés, a digitális tér/szociológia, formális cselekvésemélet, új média, peer to peer hálózatok, stb
- 2010 óta: MTA SZTAKI Nyelvtechnológiai Kutatócsoport
- Két évig a Nyelvtudományi Intézet fogadta be, máig erős együttműködés

<http://hlt.bme.hu>

- Az első generáció: Halácsy Péter, Pajkossy Katalin, Simon Eszter, Varga Dániel, Zséder Attila
- Frissen végzett: Recski Gábor
- Kurrens: Ács Judit, Benyeda Ivett, Borbély Gábor, Makrai Márton, Nemeskey Dávid

- Az első generáció: Webkorpusz, Hunglish, Hun* eszközlánc *HunSpell, HunMorph, HunToken, HunNER, HunPOS, HunTag*
Források: Telematikai Alap, TÁMOP (SzóSzablya, PI Szakadát István), Linguistic Data Consortium
- Generációkon átnyúlva: *4lang* (OTKA); *digitális életképesség* (Summer Institute of Linguistics), *digitális nyelvléírás* (LDC SIREN), *wikipédia konfliktuskezelés* (FuturICT/TÁMOP, PI Kertész János)
- Kurrens: morfológiai mélytanulás (OTKA); emOSA nyílt beszédarchívum (Infra2, PI Váradi Tamás); szemantikai nyelvtechnológia az e-kormányzásban (PI Vámos Tibor)

SZEMANTIKA: HONNAN TUDJUK, HOGY ÉRTI?

A kupa nem fér a barna bőröndbe mert túl kicsi. *Mi túl kicsi? A bőrönd vagy a kupa?*

A kupa nem fér a barna bőröndbe mert túl nagy. *Mi túl nagy? A bőrönd vagy a kupa?*

Johanna megköszönte Zsuzsának a sok segítséget amit adott. *Ki adta a segítséget, J vagy Zs?*

Johanna megköszönte Zsuzsának a sok segítséget amit kapott. *Ki adta a segítséget, J vagy Zs?*

Feri irigykedett amikor Vili azt mondta ő nyerte meg a versenyt. *Ki nyert, Feri vagy Vili?*

A golyó átütötte az asztalt mert hungarocellból/acélból készült. *Mi készült hungarocellból/acélból?*

HIBRID TECHNOLÓGIA

- Szimbolikus módszerek (általánosított véges automaták) és statisztikai módszerek (elosztott számítási modellek (distributed computing), Markov-láncok, mélytanulás, szövektorok) keveréke
- A nehéz (környezetfüggő ill. Turing-teljes) apparátust felváltotta a könnyű (véges automaták, lineáris módszerek)
- Magyarázó adekvatság mélytanulás által
- A racionalista és az empirista megközelítések nem mondanak ellent egymásnak, hanem kiegészítik egymást!