



Üzleti szemlélet és a magyar nyelv támogatása a többnyelvű világban

Prószéky Gábor

MorphoLogic & PPKE ITK

www.morphologic.hu & www.itk.ppke.hu

A magyar nyelv helyzete a digitális korban - MTA, 2013. január 18.



Hazai tervek a többnyelvűség gépi támogatásban (A Nyelv- és Beszédtechnológiai Platform megvalósítási tervéből)

1. Intelligens szótárak építése és intelligens szótárépítés
2. A magyar nyelvet is érintő szótári tartalmak gépi fejlesztése
3. Fordítómemóriák intelligensebbé tétele
4. Gépi fordítás a magyar nyelv szolgálatában



Hazai tervek a többnyelvűség gépi támogatásban 1.

- Intelligens szótárak építése és intelligens szótárépítés
- Üzleti közelítés?
22 éve fejleszt ilyeneket egy vállalkozás (MorphoLogic), ám nem ingyenes az eszköz (MoBiMouse), így nem tud versenyezni az adott esetben nála gyengébb, de ingyenes szótárprogramokkal
- Ezek a fejlesztések tehát NEM történhetnek kizárólag üzleti alapon!**



Hazai tervek a többnyelvűség gépi támogatásban 2.

- ❑ A magyar nyelvet is érintő szótári tartalmak fejlesztése
- ❑ Üzleti közelítés?
Napjainkban NEMIGEN ÉRI MEG a szótárkiadóknak új X-magyar és magyar-X szótárakat készíteni (talán a Grimm kivételével), különösen, ha X=bármely környező ország nyelve (vagy egyszerűbben: nem az angol)
- ❑ **Ezek a fejlesztések tehát NEM történhetnek kizárólag üzleti alapon!**



Ami tehát kellene...

- ❑ A „szótárlétrehozásnak” szigorúan tudományos alapokon kell történnie: további amatőr megoldásoknak helye nincs
- ❑ A mai kornak megfelelően kizárólag professzionális számítógépes nyelvtechnológusokkal: a technológiai eszközök magas szintű használata teszi az eredményeket a napi életben is, az interneten is versenyképessé
- ❑ Egy NEM kutatással, hanem „géppel támogatott gyártással” foglalkozó, lexikográfiai és terminológiai tevékenységet végző, szótári tartalmakat létrehozó, karbantartó és a meglévő tartalmakat modern eszközökkel közkinccsé tevő intézmény
- ❑ Ez nem akadémiai kutatóhely volna!
- ❑ Így az internet által fokozatosan ingyenessé tett világban a jó minőségű lexikográfiai tartalom alanyi jogon járhatna például minden magyar anyanyelvű állampolgár számára!



Hazai tervek a többnyelvűség gépi támogatásban 3.

- ❑ Fordítómemóriák intelligensebbé tétele
- ❑ Üzleti közelítés: van magyar csúcsmegoldás (MemoQ), ám amíg nincs nyelv(pár-)függő része a fordítómemóriának, addig az egész világon egyaránt eladható; ha viszont van nyelv(pár-)függő része, amely esetében a minőség javul, akkor csökken a piac!
- ❑ **A magyar-specifikus fejlesztések tehát NEM történhetnek kizárólag üzleti alapon!**



Hazai tervek a többnyelvűség gépi támogatásban 4.

- Gépi fordítás a magyar nyelv szolgálatában
- Üzleti közelítés: egy magyar kisvállalkozás (MorphoLogic) előbb volt meg (hazai forrásokból), mint a Google; nemzetközi értékelések szerint jobb is annál, de mindenki az ingyenes Google-t használja (jóllehet a webforditas.hu is ingyenes), de az ingyenesség nem teszi lehetővé az igényes továbblépést
- A magyar-specifikus fejlesztések tehát NEM történhetnek kizárólag üzleti alapon!**



Gépi fordítás: már csak ingyenesen?

- ❑ Kb. 15 éve megjelennek az első statisztikai alapú beszédfelismerési eredmények, és kb. 10 év óta az ingyenes statisztikai alapú fordítóalkalmazások
- ❑ A világ elhitte, hogy a statisztikai jellegű megoldások egymagukban megoldják a fordítási problémát (ellenpélda: „a postás megharapta a kutyát” 😊)
- ❑ Időközben átalakult a nyelvtechnológia tipikus üzleti modellje: a felhasználó az egyre nagyobb számban elérhető, de nem feltétlenül jól kidolgozott internetes eszközökért nem szívesen fizet, a finanszírozás áthárul másra
- ❑ Ingyenességbe kényszerítik a korábban nagy ráfordítással, nyelvészek bevonásával készült fordítórendszereket
- ❑ Az erre szakosodott vállalkozásoknak így üzleti modellt kell váltaniuk, ami sokszor sikertelen, hiszen az ingyenességgel csak ingyenes megoldások vetélkedhetnek



A nyelvtechnológiában mégis számít a nyelv!

- ❑ Az ingyenes internetes szótár- és fordítószolgáltatások általában nem használhatatlanul rosszak (magasan van a „baseline”)
- ❑ Ismert az „általános internetüzenet”: nem baj, ha valami nem a legjobb, csak ingyen legyen (mert a legtöbb felhasználónak egy-egy ilyen alkalmazás „annyira” azért nem kell)
- ❑ Ám a ritka nyelvek statisztikai modellezése esetében hosszú távon is marad a gyenge minőség, és jó darabig nem is lesz minőségi megoldás!
- ❑ Az általános célú nyelvtechnológiai eszközök esetén nagy nyelvekre van elég adat, ha viszont a tematikát szakirányok mentén szűkítjük, a statisztikai eszközök azoknál a nyelveknél is könnyen „sparse data problem”-ba ütközhetnek, ahol az általános területen nem
- ❑ Ha szűkül a szakterület, kevesebb a többértelműség, egyértelműbb a szakkifejezések használata: megjelent a kétféle közelítés előnyeire építő hibrid rendszerek gondolata - ami nem is annyira olcsó, mint a pusztán statisztikai rendszereké, és a fejlesztés sem olyan gyors...



Többnyelvűségi nyelvtechnológiai központ

- ❑ Olyan döntéshozói szándék szükséges ehhez, amelyik akarja és támogatja egy olyan intézmény létrejöttét, amelyben
 1. a magyar és más nyelvek közötti modern nyelvtechnológiai tartalmak (jó minőségű és ingyenes gépi szótárak, fordító-rendszerek) létrejönnek, és ha megvannak, publikussá válnak
 2. a közvetlen haszonszerzés kötelezősége nem terheli a többnyelvű nyelvtechnológia világát
- ❑ Ennek a központnak a tevékenységét nem a piac szabályainak, hanem a regionális kulturális igényeknek kell meghatároznia
- ❑ Az erre a feladatra alkalmas stáb már létre is jött: kérdés, hogy a felvázolt működéshez szükséges háttér mikor teszi lehetővé, hogy megjelenjenek azok a minőségi termékek, amelyek elnyomhatnák a kevésbé jó, de a végfelhasználók számára hasonlóképpen ingyenes, ám napjainkban még mindenütt jelenlévő, alacsonyabb igényű próbálkozásokat



Köszönöm a figyelmet!