

MAGYAR HANGSÚLYJELÖLÉSI SZÖVEGES ADATBÁZIS FEJLESZTÉSE ÉS REFERENCIA-VIZSGÁLATA

Olaszy Gábor – Abari Kálmán – Bartalis Mátyás

Bevezetés

Hangsúlyt is jelölő magyar mondatadatbázis eddig még nem állt a beszéd-kutatók rendelkezésére. Nemzetközi viszonylatban is kevés ilyen irányú kutatás található. A szón belüli hangsúly témakörével foglalkozik Goedemans–Hulst (2009) összefoglaló leíró jellegű munkája több nyelv vonatkozásában, összehasonlító nyelvészeti céllal. Angol mondatok szótagszintű hangsúly szintjeinek gépi megállapítására is vannak kísérletek, ezek főleg a nyelvtanulást, a kiejtés segítségét célozzák (Minematsu et al. 2002; Lu 2010). Konkrét hangsúlyadatbázisról, amely mérésekre, összehasonlításokra használható, nem találtunk szakirodalmi forrást, holott hangsúlycímkék a ma már rendelkezésre álló eszközökkel, manuális munkával támogatva bármely létező beszédadatbázishoz hozzáadhatók.

A magyar nyelv esetében a szón belüli hangsúlyozási szabály egyértelmű, alapvetően, ha hangsúlyos a szó, akkor az első szótagját hangsúlyozzuk. Ettől eltérő esetek is lehetnek (Kálmán–Nádasdy 2001), de ezekkel nem foglalkozunk. A mondaton belül pedig a szavak hangsúlyozása több tényezőtől is függ (kontextus, a beszélő akarata stb.). Sok esetben az értelmezés határozza meg, hogy melyik szóra kerül hangsúly a mondatban. Ezzel kapcsolatosan a leíró jellegű munkákon (É. Kiss et al. 1998; Kálmán–Nádasdy 2001; Varga 2002) túl már olyan kutatásokat is folytattak, amelyeknek algoritmusok kifejlesztése volt célja a magyar mondat szövegének automatikus elemzésére és a szavak hangsúly-kategóriákba való besorolására (Tamm–Olaszy 2005). Ezekre alapozva kezdtünk hozzá az első magyar szöveges hangsúlyjelölési mondatadatbázis elkészítéséhez, amely egységes ábrázolással biztosítja a mondat szavainak hangsúlycímkével történő ellátását kijelentő mondatokra (más modalitással nem foglalkoztunk). A modellünk szerint az elkészült szöveges adatbázisban minden mondat minden szava hangsúlycímkét kap. A szövegbe ágyazott hangsúlyjelöléseken túl előállítottuk a mondatok hangzó változatát is (gépi beszéddel), így nemcsak vizuális, hanem akusztikai formában is elérhetővé válik a kutató számára a teljes mondat hangsúlyképe. A munka három évet vett igénybe. Ez a hangsúlyadatbázis a későbbi kutatásokat szolgálja, egyrésztől viszonyítási alap is lehet, másrésztől konkrét, újszerű vizsgálatokra is ad inspirációt: hangsúlyjelölő algoritmusok tesztelése, beszédszintézis modellezése, gépi beszédfelismerés segítése, automatikus tartal-

lomkezelés támogatása, dialógusvizsgálatok szerkezeti kutatásának kiszélesítése, mondatelemzési kutatások tárgykörének bővítése, hangsúlykutatások támogatása stb. Az adatbázis tervezésekor az egyszerű, jól átlátható, egységes szerkezet kialakítása volt a cél. Ez – mint látni fogjuk – kompromisszumokat vont maga után. Az adatbázis a <http://magyarbeszed.tmit.bme.hu/hangsuly> címen az interneten is elérhető.

Anyag és módszer

Hangsúlyozzuk, hogy ebben a munkában a hangsúly fogalom nem azonos a fonológiai hangsúly fogalommal, a hangsúlyok meghatározása kizárólagosan akusztikus paramétereken és percepciók értékelésén alapul.

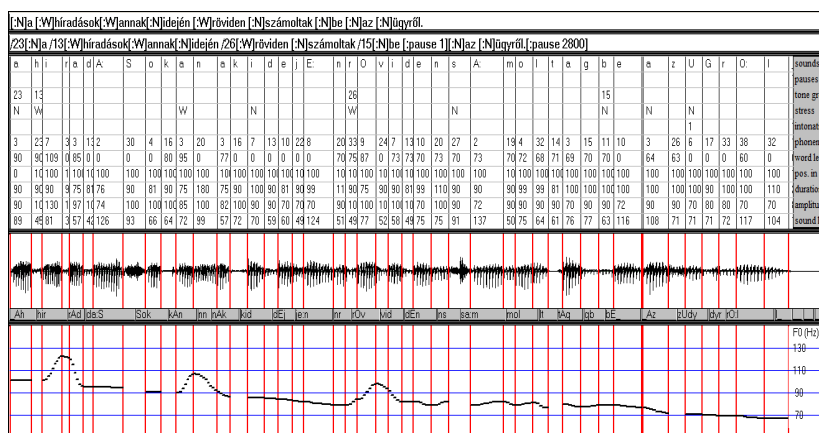
Az adatbázis nyelvi anyaga a BME TMIT-en korábban fejlesztett precíziós, párhuzamos beszédatadabázis 1866 db kijelentő mondatát tartalmazza. Ezek a mondatok irodalmi művekből származnak (Vicsi–Vígh 1998), tehát nem előre preparált szövegek, ezért a legkülönbélebb mondat szerkezetek és mondat szók fordulnak elő. A mondatok nem alkotnak egységes szöveget, az adatbázis izolált mondatok halmazának tekinthető, mindenfajta tartalmi összefüggés nélkül. A kontextus mellőzése nagyfokú szabadságot biztosított, hogy a több lehetséges hangsúlycímke-kiosztás közül melyik kerüljön be az adatbázisba. A végleges döntés, mint később látni fogjuk, három magyar anyanyelvű beszélő nyelvi intuícióján alapult. A hangsúly jelölésére bináris elvet választottunk, kétféle állapotot különböztettünk meg: a szó vagy hangsúlyos (jelölése: H) vagy nem hangsúlyos (jelölése: -). (Itt jegyezzük meg, hogy e tanulmány szövegében több helyen a példamondatokban az egyszerűbb olvashatóság végett a hangsúlyos szavakat nem a H címkével jelöljük, hanem félkövér betűstílussal emeljük ki.) Nem foglalkoztunk a hangsúlyozás módjának és különböző elméleti fokozatainak finomabb részleteivel (mondathangsúly, fő- és mellékhangsúly, szakaszhangsúly stb.), mivel nincs kialakult metodológia ezen a téren. Célunk, hogy biztosítsuk az átláthatóságot és az egyszerű, egységes szerkezetet. További szempont volt, hogy minden mondatot csak egyfajta hangsúlycímke-kiosztással tároljunk el az adatbázisban.

A tervezett hangsúlycímkek meghatározásához gépi és kézi elemzést kombináltan használtunk. Alapként a Profivox gépi szövegfeldolvasó (Olaszy et al. 2000) hangsúlyjelölő moduljával vizsgáltattuk végig a mondatokat. Ez a program felszíni szövegelemzés és egyszerű szabályok segítségével állapítja meg a hangsúlyelosztást a mondaton belül. A program automatikusan felcímkézte a szavakat a kétféle hangsúly jellel. Egy felcímkézett mondat képe tehát a következő:

(1) [-]a [H]híradások [H]annak [-]idején [H]röviden [-]számoltak [-]be [-]az [-]ügyről

Mivel a Profivox hangsúlyjelölő algoritmus csupán 70% körüli pontosságú, azaz olyan helyekre is tesz hangsúlyjelölést, amelyeknek hangsúlytalan-

nak kellene lenniük és fordítva, kétlépcsős ellenőrzést is lefolytattunk három magyar anyanyelvű személy bevonásával, hogy minél pontosabbá tegyük a mondatok szavainak a hangsúlyjeleket. Az első lépcsőben hangos olvasásos módszerrel kiejtettük a mondatokat, figyelembe véve az adott hangsúlyjeleket, és ha kellett, javítottunk a címkéken, csupán az anyanyelvi intuíció alapján (23%-ban kellett korrekciót végezni). Az esetleges hangsúlycímkejavításokat a többségi vélemény alapján fogadtuk el. Ezzel elkészült a hangsúlycímkekkel ellátott mondatok szöveges halmazának első verziója. A második ellenőrzéshez ezután minden mondatot beszéd szintetizátorral hangzó formájává alakítottunk, melynek eredményeként a megadott hangsúlycímkek szerint állt elő a mondat kiejtett formája. A szintetizált beszéd alkalmazása biztosított egyfajta egységes beszédformát és hangzást (nem köthető adott személyhez a felolvasás). A hangsúlyok előállításánál kompromisszum volt, hogy csak az alaphangfrekvenciát (f_0 -t) változtattuk, intenzitás- és hangidőtartam-változtatást nem alkalmaztunk. Az f_0 -t megemeltük az első szótag magánhangzójában, majd fokozatosan visszavezértük a második szótag magánhangzójában. Az f_0 -csúcsot tehát szuperponáltuk a kijelentő mondat alap dallamvonulatára (1. ábra).



1. ábra

A szavakon elhelyezett hangsúlycímkek a vizsgált mondat szövegébe ágyazva (fent) és az ebből beszéd szintetizálással előállított hangzó forma adatmátrixa, hullámformája és alaphangfrekvencia-menete (legalul) a Profivox rendszerben (W = hangsúlyos)

A mondatok hangzó formáját ugyanazon három személy hallgatta meg. Most is elvégeztük az esetleges címkekorrekciókat a szövegben a többségi vélemény alapján. Mindösszesen 7%-ban kellett címkét javítani. A kétszeres

laboratóriumi ellenőrzés után előállt a végleges hangsúlycímkékkel ellátott szöveges adatbázis, amely alapvetően a három magyar ellenőrző személy döntéseit tükrözi.

Az adatbázis véglegesített hangsúlyjelöléseit a mondatok szövegében korrektnak tekintjük, amit úgy kell érteni, hogy nincs címkézési hiba, vagyis ahol hangsúlyt jelöltünk, ott a hangsúlyos ejtés nem okoz megértési zavart. Vannak olyan mondatok, amelyek többféle hangsúlykiosztással is ejthetők az értelmezés, illetve a közlési szándék szerint. Ezeknél a mondatoknál az egyik helyes formát tartalmazzák a jelölések.

A nyilvános teszt

Mivel célunk egy mindenki által biztos alapnak tekinthető, tudományos igényességgel megalkotott hangsúlyadatbázis elkészítése volt, felmerült a kérdés, hogy hogyan lehet módszertanilag is igazolni, hogy a szóhangsúlycímkék jó helyen vannak a mondatok szövegében. Véleményünk szerint, akkor tekinthetjük a fenti hangsúlyadatbázisban megadott hangsúlyjeleket validátnak, ha sok ember véleménye alapján igazolni tudjuk, hogy a szavakhoz rendelt hangsúlyjelek megfelelő hangzást biztosítanak, nem okoznak zavart a prozódiai (esetleg értelmezési) megítélésben. Ezért az adatbázis referenciavizsgálatához webes nyilvános percepciók tesztet terveztünk.

A kidolgozott percepciók tesztünk a korszerű beszédminősítő rendszerekben megszokott összehasonlítási alapon működik (Fegyő 2010). Az összehasonlítási módszer alkalmazásához a következő gondolatmenet vezetett el: ha ugyanazon mondatot többféle hangsúlyozással, de – az f_0 -változtatást kivéve – teljesen azonos akusztikai tartalommal (ugyanazon hangidőtartamok, ugyanazon szünetek, azonos f_0 -alapgörbe) hozzuk létre, és ezeket páros összehasonlításban ütköztetjük, akkor az eredmények tükrözni fogják, hogy melyik a jobb hangsúlyozású. Azért vetettük el a mondatok abszolút minősítését (amely során egy mondat meghallgatását követően a minősítést végző személy egy kategóriába sorolja a mondatot), mert az ember számára nehezebb jó/rossz hangsúlyozású, vagy akár többértékű kategóriákba sorolni a mondatokat. Az egyszerűbb döntéshozatal mellett a páros összehasonlításnak másik előnye, hogy a döntési helyzetben szereplő, géppel előállított mondatok csak a hangsúlyparaméterben különböznek, minden más változó hatását ki tudjuk küszöbölni.

Ennek megfelelően az adatbázis ellenőrzött és jóváhagyott hangsúlycímkéit tartalmazó (továbbiakban A-típusú) mondataihoz készítettünk további két preparált szintetizált mondatot is, amelyek csak a hangsúlycímkékben tértek el az A-típusú mondatoktól (ezek a B- és C-típusú mondatok). Ezeket is manuálisan, a Profivox fejlesztői rendszerével (Olaszy et al. 2000) szintetizáltuk. Így hangsúlycímkével ellátott mondatok és hozzájuk kapcsolódó hangálmányok három halmaza állt elő:

A-típusú mondatok: mondat a helyesnek tartott hangsúlyozással (ezek a mondatok alkotják a hangsúlyadatbázist).

B-típusú mondatok: ugyanazon mondat, melynek minden szava hangsúlytalan címkét kapott (a hangsúlyt levettük, a dallam vonulata maradt).

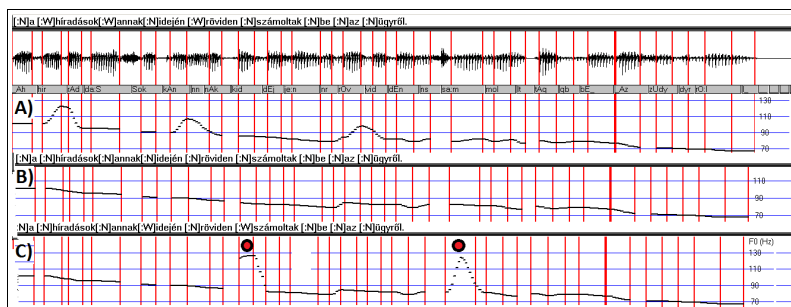
C-típusú mondatok: Kontrasztot próbáltunk létrehozni az A-típusú mondatokkal szemben, azaz ugyanazon mondatot megpróbáltuk rossz hangsúlyozással megvalósítani (a hangsúlyt többnyire egy korábban nem hangsúlyozott szóra tettük át, a dallam vonulata változatlan maradt). Itt megjegyezzük, hogy ennek a változatnak az elkészítése jelentette a legnagyobb nehézséget, mivel bizonyos mondatok tartalmához nem lehetett az A-típusú mondatról jól elkülöníthető rossz hangsúlyozást készíteni.

Az 1. táblázatban példát láthatunk, hogy a kiinduló A-típusú alakból hogyan származtattuk a B- és C-típusú mondatokat.

1. táblázat: Egy példa a B- és C-típusú mondatok származtatására

A-típus: jó hangsúlyozás	[-]a [H]híradások [H]annak [-]idején [H]röviden [-]számoltak [-]be [-]az [-]ügyről
B-típus: neutrális forma	[-]a [-]híradások [-]annak [-]idején [-]röviden [-]számoltak [-]be [-]az [-]ügyről
C-típus: rossz hangsúlyozás	[-]a [-]híradások [-]annak [H]idején [-]röviden [H]számoltak [-]be [-]az [-]ügyről

A háromféle mondat hangzó formáiban lévő különbségekre a 2. ábrán láthatunk példát.



2. ábra

A különböző hangsúlyozással előállított mondatok a szövegbe ágyazott címkéikkel és a megvalósított három alapfrekvencia-menet a Profivox rendszerben (W = hangsúlyos)

A nyilvános teszthez tehát $1866 \times 3 = 5598$ db szintetizált mondatot állítottunk elő. Minden mondat három különböző hangsúlymintázattal készült el. Az A-típusú mondatokat referenciamondatoknak tekintettük. A tesztelők számára a páros összehasonlításokhoz AB és AC mondatpárokat hoztunk létre: (i) AB

mondatpár: az A-típusú mondatot ütköztettük a B-típusú megfelelőjével, tehát az elhangzási sorrendet tekintve véletlenszerűen vagy A-B vagy B-A mondatpárt hallott a tesztelő. (ii) AC mondatpár: Az A-típusú mondatot ütköztettük a C-típusú megfelelőjével. Itt is véletlenszerű sorrendben A-C vagy C-A mondatpárt hallott a tesztelő.

Így minden A-típusú mondat összehasonlításra került a saját B- és C-típusú változatával a teszt során. Mivel sok mondatot kellett ilyen módon összehasonlítani, sok tesztelőre volt szükség ahhoz, hogy statisztikailag értékelhetőek legyenek a válaszok. Több ezer tesztelőnek küldtük ki a felhívást a teszt elvégzésére, zömében egyetemi hallgatóknak és dolgozóknak, szakmai közösségeknek az ország minden tájára. Minden tesztelőtől általános adatokat is kértünk: az életkorát, nemét és egy városnevet, ahol él, valamint azt, hogy hangszórón vagy fejhallgatón hallgatta-e a mondatokat. Egy-egy tesztelő 40 mondatpárt hallgatott meg, 20 db AB és 20 db AC mondatpárt, és mindegyikről ítéletet mondott. Egy tesztfolyamat átlagosan 15 percet vett igénybe. Minden tesztelő más-más 40 mondatpárt hallott. Az éppen hallott mondatpárt akárhányszor lejátszhatta a tesztelő, mielőtt meghozta az ítéletet. Egy-egy tesztelő többször is elvégezhetette a tesztet, ha újból bejelentkezett, minden alkalommal más-más 40 mondatos blokkot hallott.

A tesztben feltett kérdés és a rá adható négy válasz (melyből egyet kellett bejelölnie a tesztalanyoknak) a következő volt:

„Kérjük, ítélje meg az elhangzott mondatok hangsúlyozását.

1. Az első hangsúlyozása jobb, természetesebb, mint a másodiké.
2. A második hangsúlyozása jobb, természetesebb, mint az elsőé.
3. Mindkét mondat hangsúlyozása jó.
4. Egyik mondat hangsúlyozása sem jó.”

A kérdés és a rá adható válaszok megfogalmazásánál ügyeltünk arra, hogy a szubjektív beszédminősítés páros összehasonlítások során alkalmazott szokásos értékelése elvégezhető legyen. A teszt eredményeinek kiértékeléséhez a beszédtechnológiában elterjedt CMOS (Comparison Mean Opinion Score) (ITU-T 1996: 23–25) alapú értékelés egyszerűbb változatát használtuk. A tesztre adott 4 válasz egy 3 fokozatú értékelési skála kialakítására adott lehetőséget. Ha a meghallgatás során a tesztalanyok az A-típusú mondatot ítélték jobb hangsúlyozásúnak, akkor 1 pontot ért a válasz, ha a B- vagy C-típusú mondatot, akkor –1 pontot. A 3. és 4. válasz esetén 0 pontot kapott a válasz.

Az értékelés során a pontszámok átlagát, a CMOS-értékeket használtuk fel. Egy példán keresztül mutatjuk be, hogy egyetlen mondatra vonatkoztatva a pontszámok átlaga milyen jelentéssel bír (2. táblázat). A példamondat: *Ne rólam, hanem a jelenségről beszéljünk.*

2. táblázat: Az adatbázis egyetlen mondatának CMOS-értékei 11 tesztelő értékelésében (zárójelben a tesztelők száma található)

CMOS	CMOS (AB mondatpár)	CMOS (AC mondatpár)
0,91 (N = 11)	0,8 (N = 5)	1 (N = 6)

A 2. táblázat szerint 5 tesztelő AB-típusú mondatpárt, 6 tesztelő pedig AC-típusú hallgatott meg. Az AB mondatpár oszlop 0,8-as értéke azt jelenti, hogy az 5 tesztelő többsége az A-típust jelölte meg az összehasonlításból jobbnak (ha mindannyian az A-t jelölték volna meg, akkor 1 lenne az érték). Négy tesztalany az A-típusú mondatot részesítette előnyben, egy pedig egyformának ítélte a B-típusú mondattal. Azaz $0,8 = (1 + 1 + 1 + 1 + 0)/5$. Az ugyanezen AC mondatpár oszlopában az 1-es érték azt jelenti, hogy mind a hat tesztalany az A-típusú mondatot preferálta. A 2. táblázat CMOS oszlopában az összesített, az AB és AC mondatokra egyaránt vonatkozó ítéletek átlaga szerepel (0,91).

A CMOS-értékek statisztikai elemzésére egy- és kétmintás t -próbát, valamint kétszemponos varianciaelemzést használunk. A válaszkategóriák eloszlását gyakorisági táblázatokkal írjuk le, és χ^2 -próbával elemezzük.

Az elkészült hangsúlyadatbázis szerkezete és jellemzői

Az adatbázis 1866 kijelentő mondata a megadott hangsúlycímkékkel egyetlen TXT-fájlban található, és minden mondat saját azonosítóval van el látva (3. ábra). A mondatok szintetizált változatát WAV formátumban tároljuk 22 kHz-es mintavételi frekvenciával és 16 bites kvantálással ugyanazon azonosítóval, mint ami a szöveges mondatához tartozik.

[H]apámnak [H]negyvenöt [-]tavaszára [H]sikertült [H]visszasietnie
 [H]erőt, [H]egészséget, [H]szolgáljuk [-]a [H]rendőrséget
 [H]feltétlenül [-]elgőzösítik
 [H]megcsodálhattuk [H]netanjahu [-]és [H]arafat [H]partnerszellemét
 [-]pillanatnyilag [H]megfeledezett [H]bertold [H]árnyalakjáról
 [H]benjamin [H]netanjahu [H]arnold [H]kollerrel [-]találkozott
 [H]jizzadságkészlete [H]kimeríthetetlenek [-]látszott
 [H]nem [-]jó [H]mindent [H]megrohasztani, [H]meggyengíteni, [H]aláaknázni
 [H]mikszáthnak [H]kezdőként [H]nem [-]volt [H]semmije

3. ábra

Mondatpéldák a hangsúlyjelölési adatbázisból

A hangsúlyadatbázis mondatait jellemző szavak száma és hangsúlyok száma változók keresztábráját a 3. táblázat tartalmazza.

3. táblázat: A hangsúlyok számának és a szavak számának kétdimenziós gyakorisági táblázata az adatbázis 1866 mondatában

		Hangsúlyok száma								Σ
		1	2	3	4	5	6	7	8	
Szavak száma	2	1	0	0	0	0	0	0	0	1
	3	3	5	0	0	0	0	0	0	8
	4	20	23	6	4	0	0	0	0	53
	5	54	98	32	5	0	0	0	0	189
	6	60	155	80	19	2	0	0	0	316
	7	56	161	145	35	7	1	0	0	405
	8	37	145	119	54	9	0	0	0	364
	9	14	83	98	51	5	1	0	0	252
	10	4	38	60	45	3	0	3	0	153
	11	1	8	21	25	10	0	0	1	66
	12	0	7	12	11	7	0	0	0	37
	13	0	1	6	4	6	0	1	0	18
	14	0	0	1	2	0	1	0	0	4
	Σ	250	724	580	255	49	3	4	1	1866

A 3. táblázatból kiolvasható, hogy a legtöbb mondat 6–8 szót tartalmaz, és a kis vagy nagy szószámok esetén egyre csökken a hangsúlyos szavak gyakorisága. Az adatbázisban 2–3 hangsúlyszám a leggyakoribb, azonos mértékben tartalmaznak 1 és 4 hangsúlyt a mondatok. A hangsúlyszám növekedésével, egészen a maximális 8 hangsúlyig a mondatok gyakorisága jelentősen csökken. Mindösszesen 4 mondat van, amelyekben minden szó hangsúlyjellel van ellátva, például: *Minden lecsillapodott, elsimult, elcsitult*. A legtöbb hangsúlyjel 8 darab, ez egyetlen mondatban fordul elő: *Bejön a mama, foga nincs, angolul nem tud, de szépen mosolyog*. Hét hangsúlyjel 4 mondatban, hat 3-ban, öt 49-ben fordul elő. Mindössze egyetlen hangsúlyos szó (a többi hangsúlytalan) 250 mondatban található. Meg is fordíthatjuk a vizsgálatot, és a [-] címkéket is vizsgálhatjuk. Egyetlen hangsúlytalan jelölés 20 mondatban van, kettő már 85-ben. Azon mondatok száma, amelyekben több a hangsúlyos szó, mint a hangsúlytalan: 136. Mindezek a számadatok azt mutatják, hogy számos szempont szerint kaphatunk adatokat az adatbázis szerkezeti elemeiről. A hangsúlyadatbázis sok szempontú vizsgálatára webes lekérdező-felületet fejlesztettünk (Abari–Olaszy 2014). A fenti adatok ezen alkalmazás használatával is kinyerhetők.

Hangsúlymintázatok a mondathossz függvényében

A mondat szavaira tett hangsúlyjelek sorozata adja a hangsúlymintázatot. A hangsúlymintázat annyi jelet tartalmaz, ahány szó van a mondatban. Az

egyetlen kétszavas mondat hangsúlymintája H-, ami azt jelenti, hogy az első szó hangsúlyos, a második hangsúlytalan. A mondat: **Feltétlenül** *elgőzösítik*.

A három és néhány többszavas mondatban előforduló hangsúlymintázatok a 4. táblázatban láthatók.

4. táblázat: Hangsúlymintázatok a 3–6 szavas mondatokban. A számok az adatbázisbeli előfordulások számát jelzik. A zárójelben lévő százalékok az adott szószámú mondatokon belüli arányt fejezik ki

3 szavas mondatok			
H--	HH-	-HH	H-H
3 (37,5%)	3 (37,5%)	1 (12,5%)	1 (12,5%)
4 szavas mondatok (csak a 10% feletti gyakoriságúak)			
H---	HH--	-H--	-HH-
13 (24,5%)	8 (15,1%)	6 (11,3%)	6 (11,3%)
5 szavas mondatok (az első 4 legnagyobb gyakoriságú)			
H----	-H---	-HH--	-H-H-
26 (13,8%)	22 (11,6%)	22 (11,6%)	18 (9,5%)
6 szavas mondatok (az első 4 legnagyobb gyakoriságú)			
-H-H-	H-H--	-H----	H-----
36 (11,4%)	24 (8,0%)	22 (7,0%)	18 (6,0%)

A hangsúlymintázat vizsgálatával a szövegtartalom és a hangsúlyozás közötti összefüggéseket lehet vizsgálni, például hogy az azonos mintázathoz tartozó szövegekben van-e mondatszerkezeti hasonlóság.

A nyilvános percepciók teszt eredményei

Hipotéziseink a páros összehasonlításokra vonatkozó válaszokkal kapcsolatosan a következők voltak.

1. hipotézis: Az A-típusú mondatokat (hiszen ez képviseli a korrekt hangsúlyozást) többen választják jó hangsúlyozásúnak, mint a B- és C-típusúakat.

2. hipotézis: Az AB mondatpárok esetén gyakrabban kapunk 3. választ, mint akkor, ha a tesztalany AC mondatpárt hall. Ezt a hipotézist arra alapoztuk, hogy az A- és B-típusú mondatok között kisebb az akusztikai különbség, mint az A- és C-típusúak között (a C-típusú mondatokban elrontott hangsúlyozást igyekeztünk megvalósítani). Azoknál az AB mondatpároknál, amelyikek A-típusú mondatában mindössze egy szón van hangsúly, az A- és B-típus közötti akusztikai különbség szinte alig hallható. Ezért nagy valószínűséggel kaphatjuk a 3. választ is (egyformán jó mindkét mondat hangsúlyozása).

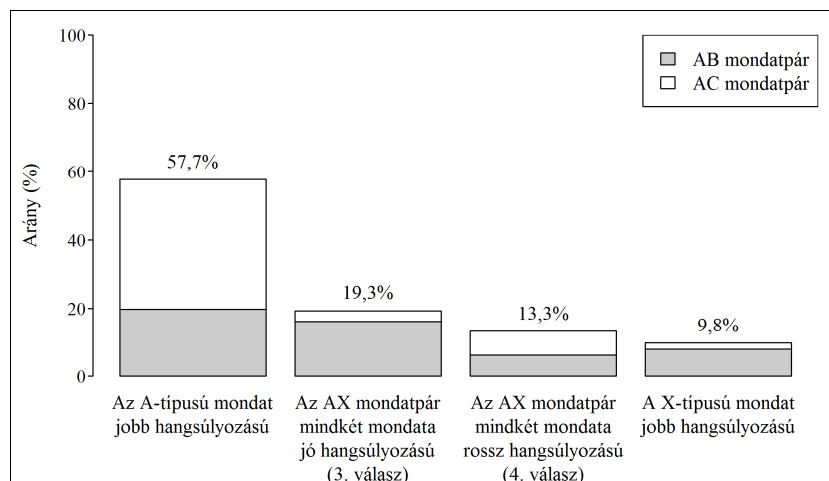
3. hipotézis: Ha a tesztelő az AC mondatpárt hallja, akkor gyakrabban fog az A-típusú mondatok javára dönteni, mint az AB mondatpároknál. Ezt a hipotézist arra alapozzuk, hogy az A- és C-típusúak között nagyobb az akusztikai különbség, mint az A- és B-típusúak között.

A nyilvános tesztet 351 személy (188 férfi és 163 nő) végezte el 2013 október–november hónapjaiban. A tesztalanyok átlagéletkora 30 év, szórása 14 év. A legfiatalabb tesztkitöltő 17, a legidősebb 85 éves volt. A kitöltők kétharmada 28 évesnél fiatalabb volt. A tesztalanyok 60%-a budapesti, 15%-a debreceni lakos, a többi egyéb településekről végezte a tesztet. A teszteredmények közreadását a válaszkategóriák gyakorisági elemzésével kezdjük, majd rátérünk a CMOS-értékek vizsgálatára.

Megvizsgáltuk, hogy a 351 válaszadó a páros összehasonlítások során a 4 lehetséges választ milyen arányban választotta. Az AB és AC mondatpárokra AX mondatpárként hivatkozunk. Az X jelölés tehát az A-típusú mondatnál párban hallott B- vagy C-típusú megfelelőt jelenti. A teszt során kiválasztható 4 válasz így egyértelműen besorolható a következő 4 kategória valamelyikébe: (i) az A-típusú mondat a jobb hangsúlyozású, (ii) az X-típusú mondat a jobb hangsúlyozású, (iii) mindkét mondat (A- és X-típusú is) jó hangsúlyozású és (iv) mindkét mondat (A- és X-típusú is) rossz hangsúlyozású. A $351 \times 40 = 14\,040$ db válasz alapján az egyes kategóriák arányát a 4. ábra tartalmazza. A 4. ábráról leolvasható adatok mindhárom hipotézis beigazolódását megmutatják.

Az első hipotézis teljesülését mutatja az a tény, hogy az összes válasz 57,7%-ában az A-típusú mondatokat ítélték jobb hangsúlyozásúnak a vele párban lévő B- vagy C-típusú mondatnál szemben, míg ellenkező válasz csak az esetek 9,8%-ában született. Nagyon leegyszerűsítve azt mondhatjuk, hogy mind a négy válaszlehetőséget figyelembe véve, minden második válasz az A-típusú mondatot jelölte jobb hangsúlyozásúnak, és csak minden tizedik az X-típusút. A kimaradt esetek a 3. és 4. válasz között oszlanak el. Az AX mondatpár mindkét tagját az esetek 19,3%-ban jó hangsúlyozásúnak, 13,3%-ban pedig rossz hangsúlyozásúnak ítélték a tesztalanyok. A válaszkategóriák eloszlása nem véletlenszerű, χ^2 -próba szerint a hangsúlycímke-módosítás (az A-, B- és C-típusú mondatok kialakításának) hatása szignifikáns volt ($\chi^2 = 8244,4$, $df = 3$, $p < 0,001$) a válaszokra.

Tovább erősíthetjük az első hipotézis igazolását, ha a 4. ábra két bal oldali oszlopának részarányát összegezzük. Így azt kapjuk, hogy az A-típusú mondatokat 77%-ban nem ítélték rossz hangsúlyozásúnak a tesztelők. E kijelentés mögött a válaszkategóriák egyszerű abszolút értékelésének lehetősége húzódik meg. Ugyanis a páros összehasonlítások során implicit módon az A-típusú mondatokról abszolút döntést hoztak a tesztalanyok a következő elv alapján: az A-típusú mondat nem képvisel rossz hangsúlyozású mondatot, ha a tesztalany a 4. ábra két bal oldali oszlopában leírt kategóriából valamelyiket választotta, viszont rossz hangsúlyozású, ha a jobb oldali két kategóriát választotta. Vagyis az A-típusú mondat jó hangsúlyozású, ha jobb hangsúlyozású az X-típusú megfelelőjénél (4. ábra bal első oszlop), vagy az AX mondatpárban jó hangsúlyozásúnak ítélték (4. ábra második oszlop balról).



4. ábra

A nyilvános teszt négy válaszáinak eloszlása. A válaszok százalékos gyakoriságát megjelenítő oszlopokban külön jelöltük az AB (sötét) és AC (világos) mondatok arányát

Az ábráról leolvasható, hogy az egyes válaszkategóriákban az AB és AC mondatpárok aránya eltér (ez az eltérés is szignifikáns: $\chi^2 = 2592,8$, $df = 3$, $p < 0,001$). A második hipotézis a 3. válasz egyenetlen eloszlását feltételezi az AB és AC mondatpárok között. Ez teljesül, mivel az AX mondatpár mindkét tagját jónak minősítő válaszokban az AB és AC mondatpárok között az arány 83,7-16,3%, tehát csak minden ötödik ilyen típusú válasz származik AC mondatpárból (a 4. ábra második oszlopa). A harmadik hipotézis az A-típusú mondatokat jobbnak minősítő válaszokban (a 4. ábra első oszlopa) jósol egyenetlen eloszlást, de fordított előjellel. Az AB és AC mondatpárok aránya 34,2%-65,5%, azaz az A-típusú mondatokat előnyben részesítő válaszok kétharmadát AC mondatpár meghallgatása során kaptuk. Ez bizonyítja a harmadik hipotézis teljesülését is. A 4. ábra maradék két oszlopára is elvégezhetjük a fenti elemzéseket. Az AX mondatpár mindkét tagját egyaránt rossznak minősítő válaszok (harmadik oszlop) közel azonos arányban származnak AB és AC mondatokból (46,3%-53,7%). Az X-típusú mondatokat jobb hangsúlyozásúnak minősítő válaszok (negyedik oszlop) túlnyomó többsége, 81,5%-a AB típusú mondatpárból származik, csak a maradék 18,5% jön AC mondatpárokból.

A válaszok értékelését a CMOS-értékek alapján is elvégeztük. A továbbiakban ezt a mérőszámot arra használjuk, hogy különböző csoportokban az A-típusú mondatokra vonatkozó kedvező döntések számát egymással összehasonlítsuk.

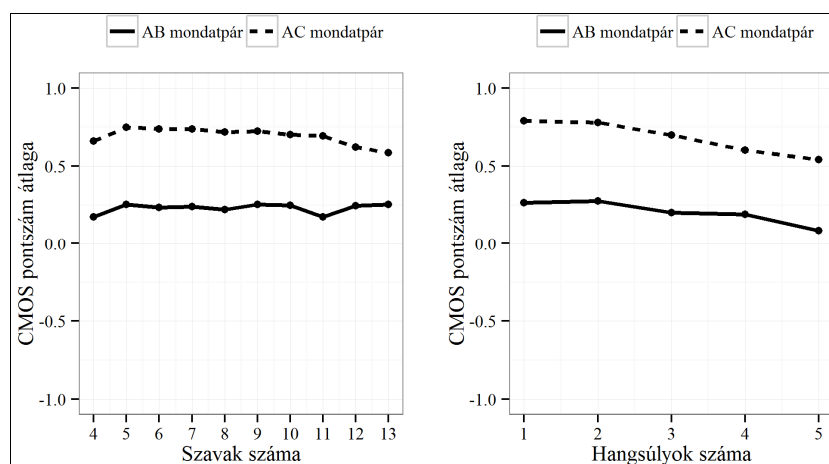
Kiszámoltuk az egész adatbázisra vonatkozó átlagos CMOS-értéket az összes AB és összes AC mondatpár esetén (5. táblázat). A pozitív értékek mindkét mondatpár esetén megmutatják, hogy szívesebben választják a tesztalanyok az A-típusú hangsúlyozással ellátott mondatokat. Az AC mondatpárok esetében ez sokkal hangsúlyosabb, mivel az akusztikai különbség nagyobb, mint az AB mondatpároknál. Mindkét várható érték szignifikánsan eltér a 0-tól (egymintás t-próbával, AB mondatpár esetében $t = 23,7$, $df = 1856$, $p < 0,001$; AC mondatpár esetében $t = 96,4$, $df = 1856$, $p < 0,001$;). Az AB és AC mondatpárookra vonatkozó eltérések is szignifikánsan eltérnek egymástól (kétmintás t-próbával, $t = -40,2$, $df = 3711$, $p < 0,001$). A CMOS-értékeken alapuló fenti három statisztikai próba eredménye egyben az első és harmadik hipotézis teljesülését is jelenti.

5. táblázat: Az átlagos CMOS-értékek az összes mondatpárra

	Átlag	Szórás	N
AB mondatpár	0,24	0,71	7020
AC mondatpár	0,72	0,52	7020

Az átlagos CMOS-értékek változását vizsgálhatjuk a tesztmondatok különböző jellemzői szerint is. A szavak száma és a hangsúlyok száma szerinti átlagos CMOS-pontszámok változását az 5. ábra tartalmazza. A 2, 3 és 14 szószámmal rendelkező mondatok a kis mintaelemszám miatt kimaradtak a vizsgálatból. Hasonló megfontolások miatt a hangsúlyszámok szerinti csoportosításban nem szerepelnek a 6–8 hangsúllyal rendelkező mondatok.

Az 5. ábra bal oldali része megmutatja, hogy a szavak száma nem befolyásolja az ítéleteket sem az AB, sem az AC mondatpárok esetén. Két szempontos varianciaelemzés segítségével egyedül a mondatpárok hatása bizonyult szignifikánsnak: $F(1, 13899) = 2135,6$; $p < 0,001$. Az 5. ábra jobb oldali részén a hangsúlyszám emelkedésével csökkenés figyelhető meg az A-típusú mondatokra vonatkozó kedvező döntések számában. Szignifikáns hatása van a mondatpár változónak: $F(1, 13915) = 2157,7$; $p < 0,001$, valamint a hangsúlyok számának is: $F(1, 13915) = 114,1$; $p < 0,001$. A CMOS-érték csökkenésének oka a hangsúlyszám növekedésével további, részletesebb vizsgálatokat kíván.



5. ábra

A CMOS-átlagok változása a szavak számának és a hangsúlyok számának függvényében

Vizsgálhatjuk az egyes mondatokra vonatkozó egyedi CMOS-értékeket is a válaszok függvényében. Ezt akár minden mondatra el lehet végezni. Itt most csak néhány érdekes eset említünk. Például megkerestük a szélsőséges CMOS-értékekkel rendelkező mondatokat. Az elemzést itt is külön végeztük az AB és AC mondatpárookra, és csak azokat a mondatokat vizsgáltuk, amelyekre legalább 5-5 válasz érkezett.

Az A-típusú mondatok számára legkedvezőtlenebb eset a -1 -es átlagos CMOS-pontszám, amely csak egyetlen mondatnál, az AB mondatpár tesztelési válaszaiból adódik. Ez azt jelenti, hogy minden tesztelő a neutrális hangsúlyozását tekintette jobbnak nem pedig az A-típusút:

(2) [-]a [H]fehér [-]köpenyes [-]ember [H]meg [-]se [H]fordult.

Az eredmény például magyarázható azzal, hogy a mondat tartalma nem kíván különösebben hangsúlyozást, anélkül sem sérül az értelmezése.

Az AC mondatpárok vizsgálatában a legkevésbé kedvező ítélet CMOS-értéke $-0,66$, ami egyetlen mondatpárra vonatkozik. Itt majdnem minden tesztelő a C-típusú mondatot tartotta jobbnak. Az A és C-típusú mondatok a következők:

(3) A) [H]arról, [-]ami [H]kísérletileg [H]nem [-]közelíthető [-]meg, [H]hallgat.

C) [H]arról, [-]ami [H]kísérletileg [-]nem [H]közelíthető [H]meg, [H]hallgat.

A két hangsúlyozás összehasonlításából látható, hogy az adatbázist fejlesztőknek nem sikerült megvalósítani eléggé karakteresen a rossz hangsúlyozást a C-típusú (alsó) mondatban, hiszen három esetben a hangsúlyok ugyanazonkon a szavakon vannak mindkét mondatban. A mondatok közötti akusztikai különbség tehát kicsi. A tesztlők ítéletei ezt jelzik. Szerencsére csak egy ilyen mondat van a rendszerben. Mutatunk néhány példát olyan AB mondatpárokból is, amelyekben minden tesztelő az A-típusú mondatot tartotta jobbnak (CMOS = 1).

- (4) A) [-]a [H]hogyan [-]már [H]megértette, [-]a [H]miért [-]még [-]nem.
 A) [H]nagyon [-]optimista [-]vagyok [-]a [-]jövővel [-]kapcsolatban
 A) [-]az [H]ügyvéd [-]a [H]pénzéért [-]a [H]feladatokat [-]is [-]kijelöli

A fentiek folytatásaként, ilyen részletes vizsgálatok elvégzése után, akár korrigálhatjuk is a címkéket a hangsúlyadatbázis néhány mondatában, de ezek a korrekciók a mondatok többségének megítélését az eddigi adatok szerint nem befolyásolják. Ilyen kutatást a jövőben terveznek elvégezni a cikk szerzői, de bárki előtt szabad az út a kérdéskör tetszőleges irányú vizsgálatára.

Összefoglalás

Az első magyar hangsúlyadatbázist mutattuk be, olyan szöveges mondatgyűjteményt, ahol a mondat minden szava hangsúlycímkével van ellátva. Nyilvános percepcióos teszttel igazoltuk, hogy a megadott hangsúlycímkék nem képviselnek rossz hangsúlyozást. Így az adatbázis validáltnak tekinthető az adott mondatok vonatkozásában. Az adatbázis önmagában is számos kutatási kérdés vizsgálatára ad lehetőséget. Ezekből néhányat felvillantottunk. A nagyszámú tesztelő válaszáinak részletesebb, több szempontú vizsgálata is lehetséges, erre egy külön tanulmányban térünk vissza. Tovább bővíti a kutatási teret az a tény is, hogy a hangsúlyadatbázis minden mondata 12 személy ejtésében is rendelkezésre áll a TMIT adatbázisában (Olaszy 2013), tehát összehasonlíthatók a természetes ejtések a hangsúlyadatbázis adataival. Ezzel a valós ejtés és az adatbázis közötti kapcsolat is vizsgálható. A webes lekérdezőfelülettel kiegészített adatbázis elérhető az interneten is (Abari–Olaszy 2014).

Irodalom

- Abari Kálmán – Olaszy Gábor 2014. Magyar hangsúlyadatbázis az interneten kutatáshoz, oktatáshoz. In Tanács Attila – Varga Viktor – Vincze Veronika (szerk.): *X. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Szeged, 347–356.
- É. Kiss Katalin – Kiefer Ferenc – Siptár Péter 1998. *Új magyar nyelvtan*. Osiris Kiadó, Budapest.
- Fegyő Tibor 2010. Gépi beszédminősítés távközlési rendszerekben. In: Németh Géza – Olaszy Gábor (szerk.): *A magyar beszéd*. Akadémiai kiadó, Budapest. 547–555.

- Goedemans, Rob – Hulst, Harry 2009. StressTyp: A database for word accentual patterns in the World's languages. In: Everaert, Martin – Musgrave, Simon – Dimitriadis, Alexis (eds.): *The use of databases in cross-linguistic studies*. Mouton de Gruyter, Berlin–New York. 235–282.
- ITU-T: P.800. 1996. *Methods for subjective determination of transmission quality*. <http://ip6.twbbs.org/VoIP/Measurement/p.800.pdf>.
- Lu, Jingli 2010. *CASTLE: A computer-assisted sentence stress teaching and learning environment*. PhD thesis. Massey University, Manawatu, New Zealand.
- Minematsu, Nobuaki – Kobashikawa, Satoshi – Hirose, Keikichi – Erickson, Donna 2002. Acoustic modeling of sentence stress using differential features between syllables for English rhythm learning system development. In Hansen, John H. L. – Pellom, Bryan L. (eds.): *Proceedings of Interspeech, ISCA*. 745–748.
- Kálmán László – Nádasy Ádám 2001. A hangsúly. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 393–467.
- Olaszy, Gábor – Németh, Géza – Kiss, Géza 2001. Hungarian audiovisual prosody composer and TTS development tool. In Puppel, Stanislaw – Grazina, Demenko (eds.): *Prosody*. Poznan. 167–178.
- Olaszy Gábor 2013. Precíziós, párhuzamos magyar beszédadatbázis fejlesztése és szolgáltatásai. *Beszédkutató 2013*. 261–270.
- Olaszy Gábor – Németh Géza – Olaszi Péter – Kiss Géza – Zainkó Csaba – Gordos Géza 2000. Profivox – a Hungarian TTS system for telecommunications applications. *International Journal of Speech Technology* 3–4. 201–215.
- Tamm Anne – Olaszy Gábor 2005. Kísérlet automatizált szövegelemzési módszerek kialakítására a szóhangsúlyok meghatározásához. In Alexin Zoltán – Csendes Dóra (szerk.): *III. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged. 383–393.
- Varga László 2002. *Intonation and Stress: Evidence from Hungarian*. Palgrave Macmillan, New York.
- Vicsi Klára – Vigh Attila 1998. Az első magyar nyelvű beszédadatbázis. *Beszédkutató '98*. 163–177.

A kutatást a Paelife (Grant No. AAI-08-01-2011-0001) és az EITKIC_12-1-2012-001 projektek támogatták.