

A „Beszédmester” szoftver adatbázis-specifikációja

SEJTES GYÖRGYI*-KOCSOR ANDRÁS×

*Szegei Tudományegyetem Juhász Gyula Tanárképző Főiskolai Kar
Gyakorló Általános és Alapfokú Művészeti Iskola

× Szegei Tudományegyetem Számítástudományi Tanszék

sejtes@hung.u-szeged.hu

kocsor@inf.u-szeged.hu

Abstract

Speech Master is a software package based on machine learning that uses graphical prompts to associate phonemes and graphemes. The program can be used in speech therapy and in teaching partially handicapped children to learn reading. The program is developed in a joint work by the Research Group on Artificial Intelligence of the University of Szeged and the Hungarian Academy of Sciences, whose staff include informaticians, linguist, surdopedagogues, teachers and developing pedagogues, and an artist. They all work together to produce a real-time phoneme recognition system. The program has its own database of annotated and segmented records drawn from a large corpus, which can be further used in sociolinguistic, phonetic, phonological research or speech technology. The goal of this study is to give a detailed account of the database in order to help linguists or informaticians in their research.

1. Bevezető „Beszélni nehéz” (Péchy, 1974)

„Olvasni nehéz”- állítják kisiskolások, nagyiskolások, felnőttek. Az állítás igazságát bizonyítják az utóbbi évek magyar és nemzetközi olvasás-szövegértési vizsgálatainak lesújtó eredményei is. A legutóbbi trauma a PISA (Programme for International Student Assessment) felmérés kapcsán ért minket. Az OECD (Organisation for Economic Cooperation and Development) szakembereit az a kérdés foglalkoztatta, hogy a tagországokban folyó oktató-nevelő munka felkészíti-e a tanulókat a jövőre, ellátja-e őket azokkal az ismeretekkel, készségekkel, amelyekre a mindennapokban szükségük lesz. A felmérésből az olvasás-szövegértés vizsgálat két aspektusára szeretnénk felhívni a figyelmet.

- A magyar tanulók negyede komoly nehézséggel küzd az olvasás hatékony eszközként való alkalmazásában, ami akadályozza az újabb ismertek, készségek elsajátítását.
- Mivel az élethosszig tartó tanulás koncepciójához kötődő szövegértési képességek terén diákjaink gyengébben teljesítettek a nemzetközileg elvárható normánál, fel kell tárnunk, hogy milyen módszerek, segédanyagok elkészítésével és alkalmazásával érhető el a hatékonyabb olvasástanítás.

Az olvasás – bár alapvetően receptív jellegű – erősen hat a nyelvhasználat produktív formáira: a beszédre és az írásra; az olvasástechnika elsajátítása pedig erősen hat a személyiségfejlődésre.

Kutatócsoportunk az Oktatási Minisztérium támogatásával olyan programot dolgozott ki kisiskolások számára, amely segíti az olvasástechnika készségszintűvé válását.

A kutatás célja egy olyan közhasznú szoftver elkészítése volt, amely a fonológiai tudatosság kialakítását, fejlesztését támogatja, ezáltal alkalmazható a hallássérültek beszédtanítására, mivel vizuális síkon képes helyettesíteni a kiesett, sérült auditórikus percepciót; alkalmazható az olvasástanításban, mert elősegíti a fonéma-graféma megfeleltetést és tanulást, illetve egyes típusú diszlexiák terápiájában és a beszédterápiában, képes kiváltani a nehézkes „tükör módszert”.

Az SZTE MTA Mesterséges Intelligencia Kutatócsoportjában informatikusok, általános és alkalmazott nyelvészek, szurdopedagógusok és tanító-fejlesztőpedagógusok, valamint egy számítógépes grafikus két éve dolgoznak együtt egy valós idejű fonéma-felismerő rendszer kifejlesztésén, és annak beszédfelismerési alkalmazhatóságán.

1.1. A „Beszédmester”

(továbbiakban BeMe) olyan gépi tanuláson alapuló szoftver, amely hívóképek segítségével összekapcsolja a fonéma-grafemapárokat úgy, hogy az alkalmas legyen beszédjavítás-terápiára és olvasástanításra. Arra épül, hogy a hangos olvasás technikájának elsajátítása során a gyermek a szövegszónyi hangcsoportot hangokból, szótagokból tanulja meg előállítani úgy, hogy a szónál kisebb egységeket, a betűket megfelelteti az akusztikus jeleknek, a hangoknak. A program lehetővé teszi a részképességükben sérült, olvasási nehézségekkel küzdő gyerekek beszédjavítás-terápiáját, ill. olvasástanítását.

A motiváció adott; a gyerek akár egyedül is képes a számítógépes programmal „játszani”, ill. játszva tanulni abban a korban és annyi idő alatt, amely számára a legalkalmasabb.

2. A BeMe adatbázisa

A program speciális szempontok alapján kialakított adatbázissal működik. A nagyméretű és áttekinthetően megszerkesztett szóadatbázis az alapja a program betanító részének. A szóadatbázis 500 adatközlő 100-100 szószintű bemondását tartalmazza.

2.1. A rögzítésre kerülő hanganyag tartalmának definiálása

A BeMe olyan beszélőfüggetlen szoftver, melynek célja, hogy a beszédképzési, ill. olvasási nehézségekkel küszködő gyerekek megfelelő fejlesztésben részesüljenek. (A program része a speciális didaktika leírása is.) Ezért az első lépés egy olyan gyakorisági szótár (Bácsi, Kerekes, 2003) létrehozása volt, amely a ma Magyarországon használt, hatévesek szóanyagát tükröző 14 db első osztályos olvasókönyv szóanyagát tartalmazza.

2.2. Az adatbázis szóanyagának elkészítése

A következő feladat a 2000 leggyakrabban használt (Bácsi, *et al.*, 2004: 65) tartalmas szó (É. Kiss, *et al.*, 1998: 20) kiválasztása volt. Ebből kellett megszerkeszteni az 500 szóbemondó lapot. Minden lap három részből áll:

- Az első rész a 14 db magyar magánhangzót tartalmazza.
- A második rész 40 szóból áll. Ezek olyan tartalmas szavak, amelyek betűrendben követik egymást. Az első és második rész minden adatközlőnél azonos.
- A harmadik rész 60 szava adatközlőnként változó. A 2000 leggyakrabban használt szót osztottuk el előfordulási gyakoriságuk alapján, tehát egy-egy gyakori szó több lapon is szerepelhet.

Mivel az adatközlők kisiskolások, ügyelnünk kellett arra, hogy a bemondásokat tartalmazó szőlapon az átlagosnál nagyobb (36-os) betűméretet használjunk.

1. táblázat: Példa egy bemondólapra (kicsinyített minta)

1. rész:

a, á, e, é, i, í, o, ó, ö, ő, u, ú, ü, ű

2. rész:

alma, ág, béka, cica, csiga, dió, dzé betű, dzsungel
egér, ég, fészek, gólya, gyerek, hegy, iskola, ír
játék, kakas, lepke, lyuk, maci, nap, nyúl, oroszlán
óra, ördög, őz, pók, róka, süni, szőlő, tulipán
tyúk, uborka, út, üveg, úrhajós, virág, zokni, zsiráf

3. rész:

jó, hús, nyit, fej, kérdez, három, marad, rétes, szám, aludni
meleg, ősz, madzag, haladva, válaszoló, autó, lyukas, betű, tükör, kíváncsi
áll, lúd, lel, Füles, néz, könyvtár, galamb, szó, farkas, végez
hoz, karika, kár, kis, alkot, téli, mély, mell, köszön, Imre
Gyuri, van, hall, fut, róka, pap, nagy, zaj, olvas, fehér
mondat, érdekes, fiú, ejt, kész, fa, ész, egy, mos, fűz

2.3. Az anyaggyűjtés módszere (Kiss, 2001)

Anyaggyűjtésünk, mint a regionális vizsgálatok többsége, a korpusznyelvészethez tartozik. Célja a nyelvi adatok segítségével fonémák, ill. szavak olyan pontosságú gépi felismertetése, hogy az a BeMe programban alkalmazható legyen a beszédterápiában és az olvasástanításban.

Az anyaggyűjtés professzionális, mert a terepmunkában (a felvételek készítésénél) szakemberek – nyelvészek és fejlesztő pedagógusok – vettek részt.

A gyűjtés aktív módon történt, így csak azokat a nyelvi adatokat vettük fel, amelyekre szükségünk volt. Minden gyermeknek a saját szólapján lévő 100 szót kellett a számítógép mikrofonjába olvasni.

2.4. Az adatközlők kiválasztása

Adatközlőink 6-7 éves kisiskolások az ország 14 gyakorlóiskolájából. Az 500 gyerek (250 lány, 250 fiú) kiválasztásánál fő szempontunk az volt, hogy minél közelebb legyenek az olvasástanulás kezdeti szakaszához, mivel a BeMe majdani felhasználói ebbe a korosztályba tartoznak.

2.5. A hangfelvételek helyszínének kiválasztása

Elvi kiindulásunk az volt, hogy úgy jelöljük ki a felvételek helyszínét, hogy az minél jobban lefedje a magyar nyelvjárási régiókat (Kiss, 2001). Tettük ezt azért, hogy a BeMe szoftvert az ország bármely területén élő gyerek eredményesen tudja használni.

Számolnunk kellett azonban az aktív módszer azon hátrányával, hogy a felvételek során az interjúhelyzet megnöveli a gyerekek nyelvi önkontrollját, s a formális beszédhelyzet, a magyar köznyelv kívánalmi szerint olvassák a szavakat.

A gyakorlóiskolákat úgy választottuk ki, hogy legyen köztük kis-, ill. nagyvárosi iskola. Készítettünk felvételeket lakótelepi, ill. belvárosi iskolában is.

A regionális nyelvhasználat és az életkor azon összefüggését is szem előtt kell tartanunk, hogy az idősebbek jobban őrzik a nyelvjárási vonásokat, mint a hat év körüli kisiskolások.

2. táblázat: A felvételek helyszínei

Város	Az iskola neve	Nyelvjárási régió	Régió belüli csoport
Szeged	Tabán Ált. Isk.	Dél-alföldi r.	Szeged környéki cs.
Szeged	SZTE JGYTFK Gyak. Ált. Isk.	Dél-alföldi r.	Szeged környéki cs.
Győr	A. Cs. J. Tank. Főisk. Gyak. Ált. Isk.	Közép-dunántúli-kisalföldi r.	Észak-dunántúli cs.
Pécs	Pécsi Tudegy. I.sz. Gyak. Ált. Isk.	Dél-dunántúli r.	Észak-baranyai cs.
Szekszárd	IPF Gyak. Ált. Isk.	Dél-alföldi r.	Baja környéki cs.
Kaposvár	K. E. Cs.V.M. Ped. Főisk. Gyak. Ált. Isk.és Gimn.	Dél-dunántúli r.	Közép-somogyi cs.
Baja	Eötvös J. Főiskola Gyak. Ált. Isk.	Dél-alföldi r.	Baja környéki cs.
Eger	Esterházy K. Főisk. 2.sz. Gyak. Ált. Isk.és G.	Palóc r.	Déli tömb Eger vidéki cs.
Nyíregyháza	Ny. F. Bt. és Műv. K. A. Cs. J. Gyak. Ált. Isk.	Északkeleti r.	Szabolcs-szatmári cs.
Sárospatak	M. Egy. Com. Tan. F. K. Á. J. Gyak. Ált. Isk.	Palóc r.	Keleti tömb Keleti palóc cs.
Debrecen	Kölcsey F. Ref.Tanítók. Főisk. Gyakorlóisk.	Tisza-Körös vidéki r.	Hajdú-bihari cs.
Békéscsaba	Békéscsaba Belvárosi Ált. Isk.	Dél-alföldi r.	Szeged környéki cs.
Budapest	ELTE Tan.-és Óv. F. K. Gyak. Ált. Isk.	*Budapest	-----
Kecskemét	K. Tanítók. Főisk. Kar Petőfi S. Gyak. Ált. Isk.	Dél-alföldi r.	Kiskunsági cs.

* Budapestet külön régióként kezeljük a település nagysága, lakóinak száma és összetétele miatt.

2.6. A felvételek időtartama

A felvételeket négy terepmunkás készítette fél éven át heti egy alkalommal. Az anyaggyűjtés egy-egy adatközlőnél 2-10 percet vett igénybe. Az időkülönbség egyéni sajátosságokból adódott (eltérő olvasási tempó, ill. készségszint).

3. A BeMe adatbázisának felépítése

A BeMe adatbázis struktúrájának specifikációja nyilvános, a hanganyag korlátozottan elérhető. Ennek célja, hogy az adatbázis felhasználható legyen más kutatásokban pl.: a gépi beszéd felismerés más alkalmazásaiban, és nyelvészeti (fonetikai, fonológiai, szociolingvisztikai stb.) kutatásokban.

Az adatbázisban megtalálható a digitálisan tárolt hanganyag. A hullámforma mellett a hozzájuk tartozó ortografikus karaktereket is rögzítettük. A teljes hanganyagot annotáljuk és szegmentáljuk.

3.1. A hanganyag annotálása

Az annotálás azt jelenti (Vicsi, *et al.*, 2002), hogy minden hangfájl mellé egy címkefájl kerül, ami a hangfájl paramétereivel és tartalmával kapcsolatos változókat tartalmazza. A BeMe hanganyagának annotálásánál informatikai, nyelvi és társadalmi változókat veszünk figyelembe. Ezeket a SpeechDat-E (Pollak, 2000) és a The Budapest Sociolinguistic Interview (Kontra, Váradi, 1997) alapján dolgoztuk ki.

A SpeechDat (továbbiakban SD) személyfüggetlen beszéd felismerők betanítására alkalmas adatbázisok létrehozásának szabályrendszere, melyet az Európai Közösség által létrehozott szakértői bizottság állított össze (Vicsi, *et al.*, 2002).

A The Budapest Sociolinguistic Interview (Budapesti Szociolingvisztikai Interjú, a továbbiakban BUSZI) az MTA Nyelvtudományi Intézetében készült szociolingvisztikai kutatás anyaga angol nyelvű leírással.

Az általánosan elfogadott szempontokon túl specifikus mutatókat is felhasználtunk.

A BeszédMester (BeMe) adatbázisának rögzítésénél az alábbi változók relevánsak a SpeechDat (SD) és a BUSZI alapján:

3. táblázat: A BeMe változói a SD és a BUSZI alapján

A BeMe rövidítések	A BeMe rövidítések feloldása	A BeMe feloldások magyarul
SES (SD)	Session numbe	A felvétel sorszáma
ENV (SD)	Environment	Környezet (iskola, óvoda)
SEX (SD és BUSZI)	Speaker sex	Az adatközlő neme
AGE (SD és BUSZI)	Speaker age	Az adatközlő kora
ACC (SD)	Speaker accent	Az adatközlő dialektusa
REP (SD és BUSZI)	Recording place (place, city, country)	A felvétel helye
RED (SD és BUSZI)	Recording date	A felvétel dátuma
RET (SD és BUSZI)	Recording time	A felvétel időpontja
LBR (SD és BUSZI)	Labelling during recording: begin, end, gain, min, max, ortographic text prompt	Felvétel közbeni címkézés: kezdet, vég, a kinyert adat, min, max, íráskép alapú inger
LBO (SD)	Ortographic labelling: begin, centre, end, ortographic transcription text	Íráskép alapú címkézés: eleje, közepe, vége, helyesírás szerint lejegyzés
CMT (SD)	Comment row	Megjegyzések: Sampa-átirat
SNL (SD)	Subjective noise level	Szubjektív zajszint
EDU (BUSZI)	Education	Végzettség
BRP (BUSZI)	Birthplace	Születési hely
MOT (BUSZI)	Mother tongue	Anyanyelv
TIA (BUSZI)	Time abroad	Külföldön töltött idő
KFL (BUSZI)	Knowledge of foreign language	Idegennyelv-tudás
MIC (BeMe)	Microphone type	A mikrofon típusa
SPD (BeMe)	Speech defect	Beszédhiba
POC (BeMe)	Parents' occupation	A szülők foglalkozása
OCC (BUSZI)	Occupation	Foglalkozás
HLT (SD)	Health	A beszélő egészségi állapota

3.2. A táblázatfájlok felosztása és tartalma

A változókat a következő fájlokba csoportosítjuk. A rövidítések mellett a lehetséges válaszokat tüntetjük fel.

3.2.1. A beszélőre vonatkozó információs fájl

Fájl: <BeMe>\TABLE\SPEAKER.TBL)

SEX: lány/fiú

AGE: 5/ 6/ 7

ACC: Dél-alföldi r. (Szeged környéki cs.) / Közép-dunántúli-kisalföldi r. (Észak-dunántúli cs.) / Dél-dunántúli r. (Észak-baranyai cs.) / Dél-alföldi r.(Baja környéki cs.) / Dél-dunántúli r. (Közép-somogyi cs.) / Palóc r. (Déli tömb Eger vidéki cs.) / Északkeleti r. (Szabolcs-szatmári cs.) / Palóc r. (Keleti tömb Keleti palóc cs.) / Tisza-Körös vidéki r. (Hajdú-bihari cs.) / Budapest / Dél-alföldi r (Kiskunsági cs.)

EDU: óvoda

TIA: az évek száma

KFL: a nyelv neve

SPD: fejlődési artikulációs zavar

POC: a szülők foglalkozásának neve

BRP: a település neve

OCC: a foglalkozás neve

HLT: ép hallású, nagyothalló

3.2.2. A felvétel körülményeire vonatkozó információs fájlok

Fájl: <BeMe>\TABLE\REC_COND.TBL

ENV: iskola

MIC: Hama Dynamic microphone IMP. 600 / Shure HW505 / Typhoon multimedia / Sennheiser / Gaia SL-8720 / Circe SL-8715

SNL: magas / közepes / alacsony

3.2.3. A felvételek sorrendjére vonatkozó fájlok

Fájl: <BeMe>\TABLE\SESSION.TBL

SES: 0001- 0500

Minden SES egy adatközlő bemondásait tartalmazza 12 fájlba elosztva. (500 fő x 12 = 6000 fájl)

A felépítés:

<BeMe>\<ses0001-ses0500>\<b01-b12>

A tartalma:

b01: a 14 db magyar magánhangzó (2.2.)

b02-b06: minden adatközlőnél azonos bemondások (2.2.), fájlanként 8 szó (8 x 5 = 40 szó)

b07-b12: az adatközlőnként változó bemondások, fájlanként 10 szó (6 x 10 = 60 szó)

RED: év, hó, nap

RET: óra, perc

3.2.4. A lexikonfájl tartalma

Fájl: <BeMe>\TABLE\LEXIKON.TBL

LBR: az adatközlőtől várt szóanyag ortografikus formában

LBO: az adatközlő által bemondott szóanyag ortografikus formában

CMT: az adatközlőtől várt szóanyag SAMPA átírata

Az elhangzott szövegek lejegyzésére a nyelvészetben általában az IPA szimbólumrendszert használják. Ez a jelölésrendszer nem illeszkedik a számítógép billentyűzetéhez, ezért a nemzetközi SAMPA jelölésrendszerrel dolgozunk (Wels, 1992).

3.2.5. A bemondások száma

Az 500 kisiskolás bemond 7000 magánhangzót és 50 000 szót.
Ez összesen: 57 000 bemondás.

4. A szegmentált anyag

Az adatbázis tartalmazza a korlátozottan elérhető szegmentált hanganyagot is. A szegmentálás azt jelenti, hogy a beszéd időfüggvényében bejelöljük a fizikailag megfigyelhető fonémákat és azok határait (Vicsi, *et al.*, 2002).

A szegmentált anyag a következőképp épül fel:

Pl.: ses0001 (az első adatközlő bemondásai)

s0001b01.dat (az első adatközlő első bemondásának szegmentált anyaga)

s0001b01.huo (az első adatközlő első bemondásának SAMPA átírata)

s0001b01.wav ((az első adatközlő első bemondásának hanganyaga)

5. Összegzés

Jelen cikk írásakor a rögzített hanganyagot annotáltuk és szegmentáltuk. Az adatbázis segítségével elkészült a BeMe első prototípusa. A tesztelések, majd az ezeket követő javítások után -terveink szerint- a szoftver alkalmas lesz a beszédterápiára és az olvasástanításra.

A rögzített és feldolgozott szóadatbázist azért hozzuk létre, hogy a BeMe szoftver kifejlesztéséhez szükséges beszédtechnológiai kutatásokat kiszolgálja.

Úgy gondoljuk azonban, hogy a korpusz nagysága és a feldolgozás szempontjai alapján más nyelvészeti (szociolingvisztikai, fonetikai, fonológiai stb.) és beszédtechnológiai kutatásokra is alkalmas az adatbázis, az annotált és szegmentált hanganyag. A tanulmány megírásának a célja az, hogy az adatbázis részletes bemutatásával segítsük azon nyelvészek, informatikusok munkáját, akik kutatási célokra szeretnék felhasználni munkánkat.

Irodalom:

- Bácsi J., Kerekes J.** (2003) Az első osztályos olvasókönyvek szóanyagából készült Gyakorisági szótár „Van szó”. *Módszertani Közlemények* (Szeged) 2003/2.sz. 52-57.
- Bácsi J., Kerekes J., Lódiné Szabó K., Zs. Sejtes Gy.** (2004) „BeszédMester” számítógéppel segített olvasástanítás és beszédterápia. *Módszertani Közlemények* (Szeged) 2004/2.sz. 61-69.
- É. Kiss K., Kiefer F., Siptár P.** (1998) *Új magyar nyelvtan*. Budapest: Osiris Kiadó. pp.20.
- Kiss J.** (2001) *Magyar dialektológia*. Budapest: Osiris Kiadó.
- Kontra M., Váradi T.** (1997) *The Budapest Sociolinguistic Interview: Version 3*. Budapest: Linguistics Institute, Hungarian Academy of Science.
- Péchy B.** (1974) *Beszélni nehéz!* Budapest: Magvető Kiadó.
- Pollak, P., Cernocky, J., Boudy, J., Choukri, K., Heuvel, H., Vicsi K., Virag A., Siemund, R., Majewski, W., Sadowski, J., Staroniewicz, P., Tropsch, H., Kochanina, J., Ostrouchov, A., Rusko, M., Trnka, M.** (2000) *SpeechDat(E) – Eastern European Telephone Speech Databases*. Athens: Proceeding LREC’ 2000 Satellite workshop XLDB – Very large Telephone Speech Databases, 29 May.
- Vicsi K., Tóth L., Kocsor A., Gordos G., Csirik J.** (2002) MTBA-Magyar nyelvű telefonbeszéd-adatbázis. *Híradástechnika, LVII. /8.* pp. 35-43.
- Wels, J.** et al. (1992) *Standard Computer-Compatible Transcription. Esprit Project 2589 (SAM), Doc.no. SAM-UCL-037*. London: Phonetics and Linguistics Dept., UCL.