

IKTA-00056/2003

A Magyar Referencia Beszédadatbázis műszaki leírása

A konzorcium:

BME Távközlési és Médiainformatikai
Tanszék

Szerződésszám: OMFB-00257/2004

Projek- és témavezető: Dr. Vicsi Klára



SZTE Informatikai Tanszékcsoport

Szerződésszám: OMFB-00258/2004

Témavezető: Kocsor András



Bevezetés

Ez a magyar nyelvű referencia beszédadatbázis (MRBA rövidítéssel), amelyet általános felhasználói környezetben, irodákban, laboratóriumokban, lakásokban rögzítettünk, a Budapesti Műszaki Egyetem Távközlési és Médiainformatikai Tanszékén (a Beszédakusztikai Laboratóriumban) készült a Szegedi Tudományegyetem Számítástudományi Tanszékének együttműködésével. Szerkezetét az MLAP LRE-63343 SPEECHDAT (M) [1] EU projekt javaslatát figyelembe véve terveztük meg. Ez a szabvány meghatározta a fájlok elnevezését és azt, hogy milyen formában legyenek tárolva a felvételhez tartozó tulajdonságok (beszélők, és környezet jellemzői).

Az adatbázis a statisztikai feldolgozási módszereken alapuló személyfüggetlen gépi beszédfelismerők létrehozását teszi lehetővé. A szöveganyag újságcikkekből vett mondatokból és szavakból áll. Az adatbázis 332 beszélő olvasott szövegét tartalmazó hanganyag, beszélőnként 12 mondat és 12 szó speciális fonetikai elvárásoknak megfelelően összeállítva került bemondásra. A felvételek többféle mikrofonnal, hangkártyával és személyi számítógéppel készültek.

Az adatbázis szerkezeti felépítése

A beszédadatbázis könyvtárszerkezete

Az adatbázis szerkezete kis mélységben egymásba ágyazott könyvtárakat tartalmaz, folyamatos sorszámozással megkülönböztetve a felvételeket tartalmazó könyvtárakat. A könyvtárszerkezet felépítése a következő:

\<adatbázis>\<feldolgozás típusa>\<felvétel>

Ahol:

1. táblázat: Az adatbázis könyvtárszerkezete

Könyvtár	Tartalom
<adatbázis>	Az adatbázis elnevezésének rövidítése – jelen esetben MRBA (M agyar R eferencia B eszéd A datbázis)
<feldolgozás típusa>	Az adatbázis feldolgozásának típusát jelöli. Az „annotated” vagy a „segmentated” értékeket veheti fel, aszerint, hogy az adott könyvtárban levő felvételek annotálást vagy szegmentálást tartalmaznak.
<hívás>	S<sss>, ahol <sss> egy 001-től folyamatosan növekvő sorszám. Ezekben az alkönyvtárakban egy - egy bemondás során felvett hanganyag feldolgozott formája, beszédfájlok, és a hozzájuk tartozó szöveges formátumú címkefájlok találhatóak. A könyvtár elnevezése ugyanaz a háromjegyű felvételazonosító sorszám, amely minden egyes beszédfájlból (*.wav), címkefájlból (*.huo), és kézi szegmentálás fájlban (*.kez) is megtalálható.

A gyökérkönyvtár az alábbi két fájlt tartalmazza:

- UTMUTATO.DOC, az adatbázis műszaki leírása
- DISK.ID: ASCII szövegfájl, amely a kötet 11 karakteres nevét tartalmazza; megléte csupán a Unix rendszerek CD-ROM formátumánál lényeges

A beszédfájlok

A beszédfájlok formátuma

A rögzített hanganyagot 16 bites, 16 kHz-es PCM formátumban tároljuk.

A beszédfájlok elnevezése

A fájlnevek az *ISO 9660* szabvány konvencióit követik (8+3 karakter), amelyek így minden elterjedt CD-ROM formátumnak megfelelnek. A fájlnevek az alábbi minta alapján épülnek fel:

s<HHH><SS>.<KKK>

Ahol:

2. táblázat. A fájlnevek felépítése

Mező	Tartalom
HHH	Felvétel sorszáma (000-332)
SS	Szövegtest-kód (s1, s2 vagy sw), amely a szövegtest-azonosítóból és a szövegelem-azonosítóból tevődik össze.
KKK	Kiterjesztés, mely lehet wav a beszélő által bementett hanganyag esetén, vagy kétbetűs <i>ISO 639</i> nyelvkód (magyar nyelvénél HU), amit a fájl típusát meghatározó betű követ (O = ortografikus címkefájl)

Például: s027s2.WAV = a 027 sorszámú beszélő által bementett második paragrafust tartalmazó beszédfájl.

Annak érdekében, hogy az adatbázis felhasználása során már a fájlnevből egyértelműen azonosítható legyen a fájl tartalma, sajátosan definiáltuk a fent említett kétkarakteres szövegtest-kódot. Ennek első karaktere egy szövegtest-azonosítót, második karaktere pedig egy szövegelem-azonosítót tartalmaz, az alábbi táblázat szerint. A beszélő a felsorolt szövegelemek mindegyikét olvasva ejti ki, hacsak nem emeljük ki külön a spontán kiejtést.

3. táblázat: Az adatbázisban alkalmazott azonosítók és a szövegtest részletes tartalma

Szövegtest-azonosító	Szövegelem-azonosító	Szövegtest tartalma	Szövegcsoport
S	1 vagy 2	5-6-7 db fonetikailag gazdag mondat úgy, hogy összesen az s1 és s2 hangfelvételek összesen 12 fonetikailag gazdag mondatot tartalmazzon.	
S	W	12 db fonetikailag gazdag szó	

A címkefájlok

A címkefájlok formátuma

A beszédfelvételekhez tartozó *.HUO címkefájlok ASCII szövegfájlok, DOS formátumú sorvégjellel (<CR><LF>). Az egyes sorok a Speechdat szabványban lefektetett SAM formátumnak megfelelően a következőképp épülnek fel:

ABC: x, y, z, ...

Ahol:

- ABC egy hárombetűs kulcsszó, amelyet kettőspont követ. Ez a mező szóközt nem tartalmazhat, így az "ABC:" karaktersorozat érvényes SAM kulcsszónak minősül.
- a kulcsszó után a számára definiált adatelemek vesszővel elválasztott felsorolása következik
- egyes adatelemek hiányozhatnak a felsorolásból; ezek helyére semmit sem kell írni a vesszők közé
- a felsorolásban a szóközök nem számítanak, a célszerű formázás ugyanakkor nagyban javítja az anyag olvashatóságát.

A címkefájl az LHD: kulcsszóval kezdődik és az ELF: kulcsszóval ér véget. Az LBD: kulcsszó határolja a címkefájl két szakaszát: a címkefájl-fejléctet (*label file header*) és a címkefájl testét (*label file body*).

Az alábbiakban felsoroljuk a címkefájlok számára definiált összes kulcsszót. Minden egyes definíció leírja az adatelemek megkívánt formátumát és az általuk felvehető értékek halmazát. Azok a kulcsszavak, amelyekhez nem tartozik adatelem, az említett felsorolás nélkül, önmagukban is szerepeltethetők a címkefájlokban (ez a javasolt eljárás), illetve teljes egészükben el is hagyhatók.

A címkefájl fejléce

Ebben a szakaszban tematikus csoportosításban bemutatjuk mindazokat a jellemzőket, amelyeket lényegesnek ítéltünk a beszédatadtbázis tartalmának részletes leírására. A kulcsszavakhoz tartozó adatelemeket vesszővel elválasztva soroljuk fel, a fentiek szerint. A felsorolást szóközök használatával tesszük olvashatóbbá.

Néhány általános szabály az adatelemek írásmódjával kapcsolatban:

- a dátumokat a kétértelműség elkerülése érdekében DD/Mmm/YYYY formában adjuk meg (pl. 25/Nov/2000). A hónapnál az angol hónapnevek ismert rövidítéseit alkalmazzuk (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec).
- az időpontokat HH:MM:SS formában jelöljük. A másodperceket megadó SS mező elhagyható vagy két nullával helyettesíthető (15:18:00).
- azokat a paramétereket, amelyek értéküket egy véges halmazból veszik, a megfelelő brit angol kifejezéssel, ill. rövidítéssel adjuk meg (pl. a beszélő nemét az M (male) vagy F (female) betűvel rögzítjük)
- a rögzített adatelemeket tartalmazó sorokat (LHD:, VOL: stb.) pontosan az alábbi táblázatokban bemutatott értékekkel kell feltölteni.

Azonosítási sorok:

Az első kulcsszó, az LHD: azonosítja a fájl formátumát ("SAM") és ennek verziószámát ("6.0"). Az ELF: kulcsszó zárja a címkefájlt. Megléteének vizsgálatával felismerhetőek az esetleges csonkolt fájlok.

4. táblázat: Azonosítási sorok

Kulcsszó	Adatformátum	Példa	Megjegyzés
LHD:	"%s, %d.%02d"	SAM, 6.0	formátum neve és verziója
ELF:	""		címkefájl vége

Hívással kapcsolatos sorok:

A kötelezően megadandó jellemzők:

- az adatbázis neve
- a hívást azonosító sorszám, amelyet az adatbázis gyűjtése során osztunk ki az egyes telefonhívásokhoz. Ez egy négyjegyű, egyesével emelkedő sorszám, amely 0000-tól indul és 9999-ig terjedhet. Értéke megegyezik a címkefájl nevében található hasonló sorszámmal.

Két további opcionális sor, a szöveglap-sorszám és a szöveglap-fájlnév használható a hívás során használt szövegbemondó lap azonosítására. Ha a beszélők egyediségét nem vizsgáljuk és a beszélőket tartalmazó információs fájl indexként (és nem táblázatként) szerepel, a szöveglap-sorszámot tartalmazó sor használata mindenképpen javasolt.

5. táblázat: Hívással kapcsolatos sorok

Kulcsszó	Adatformátum	Példa	Megjegyzés
DBN:	"%.75s"	MRBA_Hungarian_Reference_Datab ase	adatbázis neve
SES:	"%04d"	273	felvétel sorszáma

Felvételi körülményeket leíró sorok:

A telefonos beszédadatbázisok változó felvételi viszonyok mellett készülnek, mivel a hívások különböző helyekről és környezetből érkeznek. Ha már a címkefájlok is tartalmaznak néhány fontosabb információt, a felhasználók könnyen kikereshetik az adatbázisból számukra érdekes anyagokat.

A felvételi viszonyokat három paraméterrel jellemezzük:

- a felvétel környezete
- a felvétel rögzítésekor használt mikrofon és hangkártya típusa.

A felvétel környezetének leírására három különböző kategóriát hoztunk létre: OFFICE, BOOTH és ETC (munkahelyen, telefonfülkében/közterületen, vagy egyéb körülmények között rögzített felvételek).

6. táblázat: A felvételi körülményeket leíró sorok

Kulcsszó	Adatformátum	Példa	Megjegyzés
ENV:	"%.75s"	OFFICE	a felvétel környezete
MIC:	"%s"	Genius Kondenzátor	a felvételkor használt

			mikrofon típusa
CRD:	"%s"	INTEL AC'97	a felvételkor használt hangkártya típusa

Beszélőt leíró sorok:

Itt a legfontosabb adatelem a beszélő egyedi kódja. Ez alapján kereshetünk a beszélő adatait leíró fájlban, de a keresés meggyorsítására néhány adat a címkefájlban is elhelyezhető:

- a beszélő neve (M vagy F)
- a beszélő életkora
- a beszélő születési helye
- a beszélő lakhelye

7. táblázat: A beszélőt leíró sorok

Kulcsszó	Adatformátum	Példa	Megjegyzés
SEX:	"%c"	M	a beszélő neve
AGE:	"%d"	23	a beszélő életkora
BPL:	"%s"	Budapest	a beszélő születési helye
RES:	"%s"	Budapest	a beszélő lakhelye

Beszédfájllal kapcsolatos sorok:

A beszédfájlok jellemzésére az alábbi paramétereket vezettük be:

- a beszédfájl elérési útja. Az elérési út DOS formátumú, első karaktere egy "backslash" (\), amely az adatbázis gyökérkönyvtárát jelenti.
- a beszédfájl neve
- szövegtest-kód. Amint fentebb már említettük, ez a kód, amely a beszédfájl nevében is megtalálható, egy kétkarakteres alfanumerikus sztring. Értékét az egy-egy karakterből álló szövegtest-azonosító és szövegelem-azonosító egyszerű összefűzésével kapjuk (lásd a 3. táblázatot).
- a beszéd kezdete a fájlban belül, átlépve az esetleges fejléceket; SAM formátumú adatbázisok esetén értéke rendszerint nulla.
- a beszéd vége a fájlban belül. Az MRBA adatbázis esetén ez általában 1-gyel kevesebb a beszédfájl teljes hosszánál, mivel a hangminták 1 byte méretűek. Ettől különböző értékek felhasználhatóak például a fájl elején vagy végén lévő felesleges hanganyag átlépésére.

8. táblázat: A beszédállal kapcsolatos sorok

Kulcsszó	Adatformatum	Példa	Megjegyzés
DIR:	"\\%.8s\\...\\%.8s"	\\MRBA\\S178	elérési út
SRC:	"%.8s.%.3s"	s178s1.wav	beszédállal neve
CCD:	"%.2s"	S1	szövegtest-kód
BEG:	"%lu"	0	címkézett anyag kezdete
END:	"%lu"	22527	címkézett anyag vége

Kódolási formátumot megadó sorok:

A beszédállalokban alkalmazott kódolási formátumot az alábbi adatok írják le:

- mintavételi frekvencia Hz-ben (az MRBA adatbázisnál 16000)
- adatbyte-ok száma mintánként (az MRBA adatbázisnál 1)
- értékes bitek száma az egyes mintákban (az MRBA adatbázisnál 16)
- a választott kódolási eljárás neve (az MRBA adatbázisnál *PCM* kódolás)

9. táblázat: A kódolási formátumot megadó sorok

Kulcsszó	Adatformatum	Példa	Megjegyzés
SAM:	"%d"	16000	mintavételi frekvencia (Hz)
SNB:	"%d"	1	mintánkénti byte-ok száma
SSB:	"%d"	16	értékes bitek száma mintánként
QNT:	"%.75s"	PCM	kódolási eljárás neve

A címkeállal teste

A címkeállalok teste az LBD: kulcsszóval kezdődik, ez választja el az eddig ismertettől. A címkeállal testében a beszélő által bementett szövegelemek leírása található.

A legfontosabb mezők a következők:

- az LBR: sor a rögzített időablak adatait, valamint a beszélő számára előírt bementendő szöveget tartalmazza (*LaBelling during Recording*). Az egyes adatok: a bementett beszédanyag kezdete és vége a hangmintán belül, a felvétel során használt bementési erősítés, a minimális és maximális mintaérték, illetve a bementendő szöveg ortografikus írásmódban. A bementési erősítés, a minimum és a maximum mezők opcionálisak, üresen hagyhatók.
- az LBO: sorban adjuk meg a beszélő által ténylegesen bementett szöveg ortografikus átíratát (*LaBelling Orthographic*). Az itt felsorolt adatok: a beszédanyag kezdete, közepe és vége a hangmintán belül, végül a javított ortografikus szöveg. A beszéd közepét megadó mező opcionális, üresen hagyjuk.

10. táblázat: A címkefájl testében szereplő sorok

Kulcsszó	Adatformátum	Példa	Megjegyzés
LBD:	""		címkefájl testének kezdete
LBR:	"%lu, %lu, %d, %d, %d, %s"	0, 13267, , , , Senki nem állhat az utamba.	a beszéd kezdete és vége, erősítés, minimális és maximális minta, bemondandó szöveg
LBO:	"%lu, %lu, %lu, %s"	0, , 13267, senki sem állhat az utamba	a beszéd kezdete, közepe és vége, a ténylegesen bemondott szöveg

Az ortografikus szövegek az *ISO-8859-1* (Latin-1) karakterkészlettel íródtak

A fonetikailag gazdag mondatok, illetve szavak esetében a LBO: sorban megadott szövegről egy speciális algoritmus segítségével fonotipikus fonetikai átírat készült. Ez úgy történik, hogy a nyelv fonetikai szabályainak alapján a szöveggörnyezetet is figyelembe vesszük (pl. koartikuláció, hasonulás).

Általában a különböző fonetikai átírásoknál a fonémák lejegyzésére az IPA szimbólumrendszert szokás használni. Sajnos ez a jelölésrendszer nem illeszkedik a számítógép billentyűzetéhez. Ezért nemzetközi szinten bevezetésre került egy újfajta, úgynevezett SAMPA jelölésrendszer, amely alkalmazkodik a számítógéppel kezelhető karakterkészlethez. Így a számítógépes gépelés és továbbítás egyszerűen megoldható, ellentétben a hagyományos IPA jelölésrendszerrel. A magyar fonémák IPA és SAMPA jelölésének a rendszerét a 2.2. ábra tartalmazza. A fonetikai átírást az ortografikus, SAMPA készlet segítségével végeztük.

2.2 ábra: Az ortografikus és fonéma karakterek közötti átírás IPA és SAMPA szimbólumokkal

magánhangzók

Betűk	Példák	IPA	SAMPA
a	há	ɔ	ɔ
á	hát	a:	A:
e	vet	ɛ	E
é	vét	e:	e:
i	hit	i	i
í	szít	i:	i:
o	sok	o	o
ó	sók	o:	o:
ö	köt	ø	2
ő	sőt	ø:	2:
u	füt	u	u
ú	kút	u:	u:
ü	süt	y	y
ű	füt	y:	y:

mássalhangzók

Betűk	Példák	IPA	SAMPA
p	pad	p	p
b	bab	b	b
t	tél	t	t
d	dél	d	d
k	kép	k	k
g	gép	g	g
c	cél	ts	ts
dz	bodza	dz	dz
cs	csó	tʃ	tʃ
dzs	dzsem	dʒ	dʒ
ty	tyúk	tʃ	tʃ
gy	gyár	dʒ	dʒ
f	fél	f	f
v	vér	v	v
sz	szép	s	s
z	zaj	z	z
s	só	ʃ	ʃ
zs	zsír	ʒ	ʒ
h	hét	h	h
r	réz	r	r
l	lép	l	l
j	jön, lyuk	j	j
m	méz	m	m
n	néz	n	n
ny	nyom	ɲ	J

A fent található átírási táblázathoz képest néhány fonéma esetében eltér az átírás, pótlólagos karakterek is be lettek vezetve.

Az egyik ilyen eltérés, hogy a felpattanó zárhangoknál, zárréshangoknál figyelembe vettük azt, hogy ezek a hangok akusztikailag továbbonthatók 3 külön egységre a következőképpen:

> – a zárképzés szakaszát a mássalhangzó SAMPA karaktere után jelöltük (pl. **t>**),
 +/- – zöngés/zöngétlen zárszakasz jelölése,

< – zárfelpattanási zörej szakasza, affrikáták esetén s spiráns zörejrel együtt jelöltük a következőképpen: a mássalhangzó SAMPA karaktere előtt egy **<** (pl. **<t**) .

Ez fontos a kettős felpattanó zárhangok felismerésében, mivel azzal különböztethetők meg a rövid zárhangoktól, hogy a szünet ,és folytatott zöngé szegmens időtartama nyúlik meg itt is jelöljük : -al (pl. **tt -> t>:-<t**), míg a csoporton belül megkülönböztető spektrális információt hordozó rész időtartama változatlan.

A SAMPA átíratot, a SAM formátum szerinti komment (CMT:) sorokban tároljuk.

Egy tipikus mintafájl

LHD: SAM, 6.0
DBN: MRBA_Hungarian_Reference_Database
SES: s025
ENV: BOOTH
MIC: Philips SBC MD 150
CRD: C-Media CM8738 (integr.)
SEX: M
AGE: 25
BPL: Szentes
RES: Szentes
DIR: \MRBA\s025
SRC: s025ws.wav
CCD: ws
BEG: 0
END: 249749
SAM: 16000
SNB: 1
SSB: 16
QNT: PCM
LBR: 0, 249749, , , , munkából, bizonyossá, mindenfélék
LBO: 0, , 249749, munkából, bizonyossá, mindenfélék
CMT: _ m u N - <k A: b> + <b o: l, _ + <b i z o J o S: A: , _ m i n + <d E F f e: l e: k> - <k
ELF:

Kézi szegmentálás fájl

A kézi szegmentálás fájl (*.KEZ) szintén egy ASCII szöveges fájl, amelyben megtalálhatjuk a fonémák szegmenshatárainak időpontját másodperc léptékben, és mellette találjuk a határ utáni akusztikai egység SAMPA kódját. 2 SAMPA kódot találunk egy-egy időpont mellett, az egyik leírás a *.HUO fájl CMT: sorában található automatikus átíratból származik. Az automatikus konvertálás nem mindig adott helyes eredményt, mivel a hasonulásokra és ejtésvariációkra csak a karakteres leiratból nem mindig lehet helyesen következtetni, ezért a szöveg meghallgatásával is járó kézi szegmentálás során javított SAMPA átírást és zaj jelöléseket is megtalálhatjuk mellette.

A feldolgozást három szinten került jelölésre: fonémák szintjén, szavak szintjén és a mondatok szintjén. **Fonémaszintű szegmentálás** esetén a következő szabályrendszer került alkalmazásra:

- a szegmenshatárokat a null-átmenetekhez kerültek illesztésre (zöngés hangok esetében a pozitív meredekségű null-átmenethez). A bejelölés pontossága min. 1 ms. Zöngétlen hangoknál ez azt jelenti, hogy 1 ms pontossággal jelöltük be a zöngétlen hangok kezdetét. A zöngés hangok esetében mindig a null-átmenetekhez igazodtunk szintén 1 ms pontossággal. Zöngés hangok között a határ bejelölésének bizonytalansága alapperiódusnyi hosszú.
- zöngétlen hang után a magánhangzó kezdetét a zöngé indulásánál jelöltük.
- zárhangok, affrikáták jelölésénél ezen hangok kezdetét a megelőző hang utolsó lecsengő periódusa előtt jelöltük, vagyis a megfelelő hang utolsó periódusát a zárhangok és az affrikáták részének tekintettük.
- a magánhangzó-magánhangzó vagy magánhangzó-rezonáns mássalhangzó kapcsolatokban a határt az átmeneti rész 50%-ánál jelöltük be. A határ pontos bejelölése

esetén egy-két periódusnyi ingadozást engedünk meg. Ez itt nem hiba, hanem a feladat természetéből adódik, hiszen folyamatos mozgás által létrehozott hangtermékek kettéválasztása ekkor bizonytalan. Rendszerint két különböző szakértő a határt két különböző helyre teszi.

Allophonok esetén három különböző jelölést alkalmaztunk:

' Minden magánhangzónál létezik egy rekedt ejtés, ami igen gyakori. Jelölése a magánhangzó SAMPA karaktere előtt egy '. (Pl. 'o jelenti az o hangot rekedt ejtésben.)

h A ch betű kapcsolatot h hangnak ejtjük.***

h A h hangnak csak egy allophonját ejtjük, a zöngés h hangot.

Az aktuális kimondásnál kimaradt hangokat megjelöltük, a követő hang előtt zárójelbe tettük. Természetesen az annotálásban szereplő jelöléseket megtartottuk, azokat a kézi szegmentálás során pontosan bejelöltük. A beszéd közben tartott szünetet ~ jellel jelöltük. Szöveg betoldását vagy kihagyását a bemondó által *-al jelöltük a szegmentálás során.

A **szószintű szegmentálás** esetén a szavak határát jelöltük. A **mondatok határait** a frázisjelző írásjelek (vessző, kettőspont, pontosvessző, gondolatjel, macskaköröm, és, pont kérdőjel, felkiáltójel) adták.

Címkézésre SAMPA karaktereket használtunk. A karaktereket a határbejelölések között tüntettük fel. Azt jegyeztük le, ami ténylegesen elhangzott.

a fájl formája:

<space><space><space><szegmenshatár s- ban><space><javított SAMPA jelölés><space>

<SAMPA jelölés automatikus átiratból>

A kez fájlban a korábbi SAMPA átirathoz képest újabb jelöléseket találunk:

~ beszéd közben tartott rövid szünet

spk beszélő által keltett, beszéd közbeni, a fonémáktól eltérő jellegű zaj

* ez akkor jelenik meg a <SAMPA jelölés automatikus átiratból> oszlopában

ha a kézi szegmentálás olyan plusz szegmenst jelöl, amelyhez az automatikus fonetikus átirat alapján nem tartozik kód, vagy pedig egy, az automatikus átalakítás során jelölt kóddal jelzett szegmensnek helyes kiejtés szerint szerepelnie kellene, de a beszélő nem mondta ki, ilyenkor a <javított SAMPA jelölés> oszlopában találjuk.

A *- gal jelölt szegmens, a megfelelő oszlop szerint értelmezve a fentebb jelölt szegmenssel összevonandó.

-2 szóhatár jelölése

-3 frázishatárt jelző szünet. (frázisjelző írásjelek: vessző, kettőspont, pontosvessző, gondolatjel, macskaköröm, és, pont kérdőjel, felkiáltójel bejelölése.)

Ezzel jelöljük a felvett bemondás utolsó szegmensének végét.

egy mintafájl

-2.00000 * _
1.18000 n n
1.24000 E E
1.31000 m m
-2.00000 * _
1.39000 e: e:
1.51000 r r
-2.00000 * _
1.56000 O O
1.63000 z z
-2.00000 * _
1.69000 E E
1.74000 g> g>
1.77000 + +
1.83000 <g <g
1.86000 e: e:
1.97000 s *
-2.00000 * _
2.05000 S S:
2.17000 E E
2.26000 m: m:
2.43000 i i
2.49000 t> t>
2.51000 - -
2.57000 <t <t
-3.00000 * ,