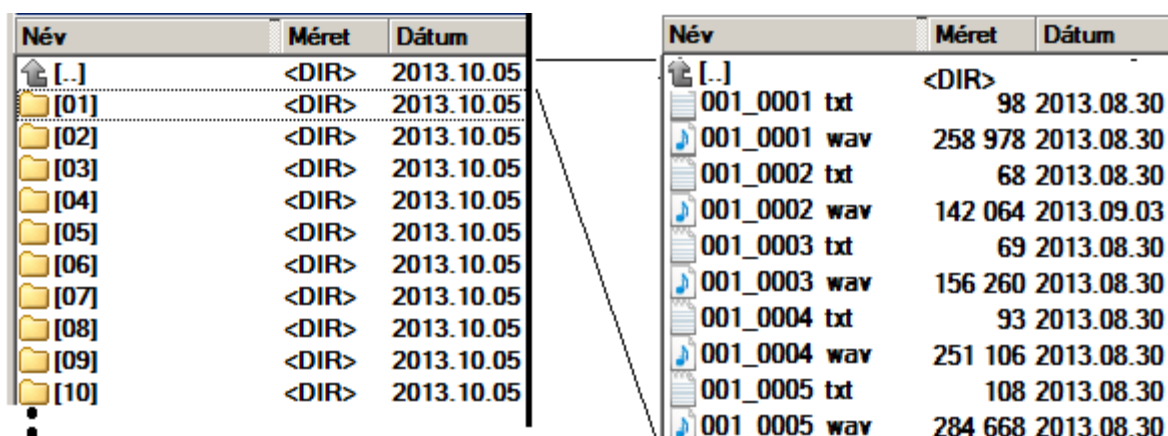


# A magyar Hangsúly adatbázis felépítése

Az adatbázis 1866 kijelentő mondatot tartalmaz szépirodalmi művekből 56 könyvtárban. A könyvtárakban jellemzően 36-36 mondat szerepel meghallgatható formában (wav) és szövegesen (txt) a hangsúlyozási jelekkel ugyanazzal az azonosítóval, csak a fájlok kiterjesztése más.

Például az első mondat a 1. könyvtárban a következő fájlneveket van ellátva:

001\_0001.txt és 001\_0001.wav



Név	Méret	Dátum
[.]	<DIR>	2013.10.05
[01]	<DIR>	2013.10.05
[02]	<DIR>	2013.10.05
[03]	<DIR>	2013.10.05
[04]	<DIR>	2013.10.05
[05]	<DIR>	2013.10.05
[06]	<DIR>	2013.10.05
[07]	<DIR>	2013.10.05
[08]	<DIR>	2013.10.05
[09]	<DIR>	2013.10.05
[10]	<DIR>	2013.10.05

Név	Méret	Dátum
[.]	<DIR>	-
001_0001.txt	98	2013.08.30
001_0001.wav	258 978	2013.08.30
001_0002.txt	68	2013.08.30
001_0002.wav	142 064	2013.09.03
001_0003.txt	69	2013.08.30
001_0003.wav	156 260	2013.08.30
001_0004.txt	93	2013.08.30
001_0004.wav	251 106	2013.08.30
001_0005.txt	108	2013.08.30
001_0005.wav	284 668	2013.08.30

## Szöveges rész (txt fájlok)

Minden mondat minden szava hangsúly jellel van ellátva. Kétféle hangsúly jelet használunk: a szó hangsúlyos: F /nem hangsúlyos: N. (A hangsúlyozás finomabb fokozatai nincsenek jelölve mondatokon belül).

Figyelem! A mondatok szövegében csak kis betűket használunk a megjelenítéseknél. Tehát itt nem érvényesítjük a helyesírás szabályait!

Például az 1. könyvtár első mondata a következő szöveges tartalommal látható:

[ :N]a [ :F]világosság [ :N]felé [ :N]fordult, [ :N]és [ :F]belebámult [ :N]az [ :F]üveg [ :N]papírnyomóba.

## Hangzó formák (wav fájlok)

Az adatbázis mondatait a ProfiVox-triad férfi hangú beszéd szintetizátorral állítottuk elő hangzó formában a hangsúlyozási jeleknek megfelelő akusztikai megvalósításban. A hangsúlyozást csak az alaphangfrekvencia változtatásával hoztuk létre. A szó első szótagi magánhangzóján erőteljes alaphangfrekvencia (F0) emelést hajtottunk végre, majd a második szótagi magánhangzó elejéig vissza vezértük az alaphangfrekvenciát mondat dallamvonulatát meghatározó alapra.

## Az adatbázis referencia vizsgálata

Az adatbázis hangsúly jelöléseit referenciaként is lehet használni. Ezt úgy kell érteni, hogy nincs benne címkézési hiba, vagyis ahol hangsúlyt jelöltünk az adott szóra, ott a hangsúlyos ejtés nem okoz megértési zavart, furcsa ejtést. Vannak olyan mondatok, amelyek többféle hangsúly kiosztással is ejthetők az értelmezés, illetve a közlési szándék szerint. Ezeknél a mondatoknál az egyik helyes formát adják a jelölések. Az adatbázis korrektségét referencia vizsgálattal bizonyítottuk 2013-ban. Lásd a szócikket.