

Középszótáras folyamatos beszédfelismerőrendszer fejlesztési tapasztalatai

Vicsi Klára, Velkei Szabolcs, Szaszák György, Borostyán Gábor, Teleki Csaba,
Tóth Szabolcs Levente, Gordos Géza

BME Távközlési és Médiainformatikai Tanszék, Beszédakusztikai Laboratórium
1117 Budapest, Magyar tudósok körútja 2.
vicsi@tmit.bme.hu

Kivonat: A Beszédakusztikai Laboratóriumban kifejlesztésre került egy Windows XP alatt működő, statisztikai elvi alapokra épülő, folyamatos beszédfelismerő fejlesztői környezet (MKBF 1.0), amely alkalmas különböző középszótáras 1000-10 000 szavas szövegek betanítására és felismerésére. Új megoldásokat dolgoztunk ki az akusztikai előfeldolgozásban, a statisztikai modellépítésben valamint fonetikai, fonológiai és morféma nyelvi szinteket vonunk be a felismerési folyamatba. A felismerő a statisztikai alapon működő HMM akusztikai fonémamodellekkel valamint a statisztikai alapú bigram nyelvi modellekkel működik, nem lineáris simítást használva. Vizsgálataink során változtattuk a betanító anyagokat, a szótárkészletet. Kétfajta bigram alappal dolgoztunk: először a hagyományos ragozott szóalakokból építettük fel a bigram mezőket, majd a szóalakokat morfémákra bontottuk, és ezekből a morfémákból építkeztünk. A cikkben a tesztelés eredményeiről, a továbbfejlesztéshez nyert tapasztalatainkról számolunk be. A perplexitási vizsgálatok eredményeinek felhasználásával a felismerési biztonságot 70%-ról 91% fölé tudtuk vinni.

1 Bevezetés

A BME Beszédakusztikai Laboratóriumban kifejlesztett folyamatos beszédfelismerő (MKBF 1.0) optimális működését az akusztikai, fonetikai [4] és nyelvi modellek változtatásával állítottuk be. Természetesen a két szint szétválasztása nem mindig lehetséges, hiszen a tesztfelvételek minősége, zajossága, az artikuláció gondossága, stb. mind befolyásolja a felismerés eredményét, így az nem csak a nyelvi modultól függ.

A felvételek mindegyike – mind a betanításnál, mind a tesztelésnél – 16 kHz-en mintavételezett, 16 biten lineárisan kvantált jel, amely a megfelelő előfeldolgozás után kerül felismerésre.

Az **akusztikai modellek betanítását** az MRBA beszéd adatbázissal végeztük [9].

Végeredményben tehát a fonémaszintű felismerőnk 16 kHz mintavételezésű, 17 Bark frekvenciatérbeli derivált, + 17 időbeni derivált, + 17 időbeni második derivált, + energia bemeneti jelvektor mellett, 4-5 állapotú kvázi-folytonos, 24 lépcsős, rejtett Markov-modellekkel (QCHMM), fonéma alappal dolgozik. Az akusztikai, fonetikai szint optimalizálásáról már korábban beszámoltunk [8].

A **nyelvi betanításhoz** a budapesti SOTE II. sz. Belgyógyászati Klinikájától (2700 lelet) és a szegedi Orvostudományi Egyetemről (6365 lelet) gyűjtött korábbi leletanyag korpuszt használtuk. Ezen szöveg korpusz alapján készítettük el a teljes szóalak-szótárat, amely 14 331 szót tartalmaz, a kiejtés szótárat és ezek téma szerint osztott kisebb szótárait. A valamint a korpusz alapján morfémaszótárat is készítettünk, amelynek nagysága 6 824 morfémaelem.

Teszteléshez az orvosok által bemondott leletanyagot használtuk, ezek a SOTE II. sz Belgyógyászatián készültek, szakorvosok bemondásával, a rendszerhez illeszkedő mintavételi és kvantálási paraméterekkel. Az összes felvételtől a férfi orvosok bemondásaiból véletlenszerűen, öt beszélőtől egyenként négy-négy darab, azaz összesen 20 darab gasztroszkópiás felvételt válogattunk ki tesztelési célokra.

Lényegében a nyelvi modellhez bigram modelleket használtunk, de az egyik megoldásban a hagyományos szóalakok (lexémák) az alkotó elemek, a másik megoldásban viszont a morféma. A morfémaabontáshoz a Humor morféma elemzőt használtuk fel [5].

2 Bigram nyelvi modellek

2.1 Az endoszkópiai felismerő nyelvi modelljének leírása

A sokféle szómodell közül az angolszász területeken jól bevált n-gram szómodelleket használtuk a nyelvi szintű felismeréshez. Az n-gram modellek segítségével egy tetszőleges korpuszon minimálisra igyekszünk csökkenteni a perplexitás mértékét, aminek következménye a kevesebb hibázás.

Az n-gram modell szószekvenciáik valószínűségének halmazából áll:

$$\hat{P}(w_1, w_2, \dots, w_m) \quad (1)$$

A szekvencia valószínűsége ekkor:

$$P(w_1, w_2, \dots, w_m) = P(w_1) \prod_{i=2}^m P(w_i | w_{i-1} \dots w_1) \quad (2)$$

A kontextust limitálva:

$$P(w_1, w_2, \dots, w_m) \cong P(w_1) \prod_{i=2}^m P(w_i | w_{i-1} \dots w_{i-n+1}) \quad (3)$$

ahol $n > 0$ tetszőlegesen választott konstans egész. A nyelv olyan tulajdonságokkal rendelkezik, hogy a folyamat során egy későbbi állapot valószínűsége gyakorlatilag

független a kezdőfeltételektől, így n értékére nem kell nagy n értéket használni. (Tipikus értékek 2-től 6-ig). A fenti valószínűség ekkor a következőképpen számítható ki:

$$P(w_i | w_{i-1} \dots w_{i-n+1}) = \frac{N(w_i \dots w_{i-n+1})}{N(w_{i-1} \dots w_{i-n+1})} \quad (4)$$

ahol $N(\cdot)$ a megadott szekvencia előfordulásai száma a tanító szöveganyagban. Ehhez a számításához nem kell szegmentált hanganyagot használni, a célra legmegfelelőbbek a nagyméretű szöveges adatbázisok.

2.2...N-gram modellek simítása

A gyakorlatban a lehetetlen méretű adatbázisok készítése helyett az n-gram modellek statisztikai vizsgálata és különböző módszerekkel történő korrigálását alkalmazzuk [6].

A korrigálásra nemlineáris interpolációt használtunk. Mivel utóbbi esetben lényegesen jobb perplexitás-csökkenés érhető el, ezért a nemlineáris interpolációt használtuk [7], az *absolute discounting* funkciót. Tekintsük most példaként a bigram esetet, ahol a képlet a következő konkrét alakot ölti:

$$\hat{P}(w^{(j)} | w^{(i)}) = \max \left\{ \frac{N(w^{(j)}, w^{(i)}) - D_i}{N(w^{(i)})}, 0 \right\} + D_i \frac{|V| - n_0(w^{(i)})}{N(w^{(i)})} P(w^{(j)}) \quad (5)$$

($|V|$ a szótár számosságát jelöli, $n_0(w^{(i)})$ pedig azon szavak számát, amelyek egyszer

sem követték $w^{(i)}$ -t.) A nemlineáris interpoláció esetében a $q(k)$ eloszlás súlya arányosan megfelel $(K - n_0)$ -val, ami azon különböző események száma, amelyek legalább egyszer láthatóak voltak a szöveganyagban. Ez érdekes dologhoz vezet a feltételes valószínűségek modellezésekor: ha a megelőző szót (*predecessor word*) egy, vagy csak néhány szó követi, akkor a simítás 'kisebb' mértékű lesz, mintha sok szó követné azt. Erre utal a nemlineáris kitétel a módszer nevében. Ha $D=1$, akkor az egyszer látott eseményeket ugyanúgy kezeli az algoritmus, mint az egyszer sem látottak. Ha alkalmazzuk a *Leaving-one-out* elvet, akkor nem jelentkezik igazán lényegi különbség a perplexításban, ezért a D értékét a lényegesen egyszerűbben kivitelezhető abszolút modell alapján számítjuk, ahol:

$$D_i = \frac{|V| \cdot b}{n_0(w^{(i)})}, \text{ ahol } b = \frac{n_1}{n_1 + 2n_2} \quad (6)$$

Itt n_1 és n_2 azon bigramok száma, amelyek pontosan egyszer, illetve kétszer szerepeltek a betanító korpuszban. Kis betanító anyag esetén $\frac{|V|}{n_0(w^{(i)})}$ értéke közelítőleg

1, ezért ilyen korpuszok esetén lehet spórolni a számításokkal és $D_i=b$ helyettesítést végezni [6].

3... Nyelvi modell tesztelése, perplexitás vizsgálata

A bigram modellek elkészítéséhez – ún. betanításához – nagy méretű, szöveges, a felismerni kívánt szöveget jól közelítő összetételű és stílusú betanító anyagra van szükség. Esetünkben ez a korábban összegyűjtött és megfelelően feldolgozott (helyesírás ellenőrzés, egységesítés, rövidítések feloldása, fonetikai átírás, stb.) leletanyag volt. Ezt a leletanyagot négy csoportra osztottuk az alábbiak szerint:

1. *SOTE II. sz. Belgyógyászatról származó felső endoszkópiás leletanyag (budapesti gasztroszkópia)*
2. *SZTE Belgyógyászatáról származó felső endoszkópiás leletanyag (szegedi gasztroszkópia)*
3. *SOTE II. sz. Belgyógyászatról származó alsó endoszkópiás leletanyag (budapesti kolonoszkópia)*
4. *SZTE Belgyógyászatáról származó alsó endoszkópiás leletanyag (szegedi kolonoszkópia)*

A fenti négy csoportból természetesen lehetőség van kombinált anyagok összeállítására is, amely így nagy mennyiségű betanító anyagot szolgáltathat a bigram nyelvi modellezéshez.

MKBF akusztikai szint betanításait az: MRBA adatbázis férfi bemondásaival végeztük.

3.1...Tesztelési körülmények ismertetése:

A tesztelés megkezdése előtt felvetődött az a kérdés hogy a rendelkezésünkre álló betanító anyagok közül melyeket használjuk fel a felismerő nyelvi modelljének a betanítására. Az előzetes mérések alapján (1. táblázat) látható, hogy a rendelkezésre álló budapesti és szegedi leletek szókészlete kis mértékben korrelálnak egymáshoz.

1. táblázat: Szókészletek összehasonlítása

		Szegedi colonoscopia		Szegedi gastroscopia	
		szó megvan	szó nincs meg	szó megvan	szó nincs meg
Budapesti colonoscopia	szó megvan	1933	1174	1872	1235
	szó nincs meg	5089	6135	7067	4157
		Szegedi gastroscopia		Szegedi colonoscopia	
		szó megvan	szó nincs meg	szó megvan	szó nincs meg
Budapesti gastroscopia	szó megvan	2720	1594	2065	2249
	szó nincs meg	6219	3798	4957	5060

A későbbi vizsgálódások azt is megmutatták, hogy a felismerő tesztelésére kijelölt hanganyagok szótárkészlete újabb szavakat tartalmazott az írásos formában rendelkezésünkre álló budapesti- és szegedi endoszkópos leletekhez képest. A fenti táblázat egyértelműen mutatja, hogy a szegedi és budapesti leletek szóhasználata

között milyen nagy eltérés van és a tesztelésre kijelölt annotált felvételek budapesti kórházból származnak. Ennek ellenére a fent említett okok miatt az összesített leletanyaggal való betanítás ígérkezett megfelelőnek.

A tesztelésnél használatos mérőszámok:

Össz_ref: a felismerendő egységek száma, *Össz_rec*: a felismert egységek száma,

Helyes: a jól felismert egységek száma, *Ins*: a beszúrt egységek száma,

Del: a törölt,

Subs: a helyettesített egységek száma

$$CORR = \frac{Helyes}{Össz_rec}, Acc = \frac{Helyes - Ins}{Össz_rec}, Wer = 1 - CORR$$

3.2...Perplexitás alapú WER becslés

A gépi beszéd felismerés felismerési pontosságát a szakirodalomban – a fentiekben leírt a Word Error Rate (WER) indikátorral szokásos jellemezni. A Word Error Rate egy költséges művelet eredménye, ezért szükségessé vált egy olyan indikátor bevezetése, amely a beszéd felismerés akusztikai szintjétől függetlenül becslést tudna adni a felismerés pontosságára. Így a nagy felismerési idő és a költséges WER számítás kikerülhetne a beszéd felismerés nyelvi modelljének vizsgálata esetén. Egy ilyen becslési módszer a – szakirodalomban is jól ismert – perplexitás, melynek segítségével vizsgálni tudjuk a nyelvi modellt. Bár különböző kutatások rávilágítanak hogy készíthető paraméterfüggő (betanítás, nyelvi modell, akusztikai modell) becslési eljárás [3,4], mégis a konkrét paraméterek ismeretének hiányában a perplexitást találtuk olyan becslési eljárásnak, amely a szakirodalomban elfogadott és számítási módja ismert. A perplexitás számítási módját az alábbi képletben ismertetem:

$$PP = \frac{1}{\left(\prod_{i=1}^N P(W_i | W_{i-1}) \right)^{\frac{1}{N}}} \quad (7)$$

ahol W_i a i . szava, W_{i-1} a $i-1$. szava, N a szavak számát alkotó szavak száma.

A perplexitás képletét bigram alapú nyelvi modell formájában használtuk. A szavak jelentése lexéma szintű felismerés esetén lexéma, míg morféma alapú felismerés esetén morféma. A perplexitás értékészlete egy 1-nél nagyobb valós szám. A tesztanyag nyelvi modul általi felismerése annál tökéletesebb, minél jobban közelít a perplexitás értéke 1-hez. Minél nagyobb a perplexitás értéke, annál kevésbé fedti a nyelvi modell a tesztelő szöveget.

4...A tesztelési eredmények ismertetése

A következő táblázatokban a felismerő tesztelési eredményei láthatóak:

2. táblázat: Gasztroszkópiás felvételek tesztelési eredményei lexéma alapú összegzett betanítású nyelvi modell esetén, orvosok bemondásában

Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER
1173	1580	750	451	22	401	25,4	36,1

3. táblázat: Colonoszkópiás felvételek tesztelési eredményei lexéma alapú összegzett betanítású nyelvi modell esetén, orvosok bemondásában

Össz z ref	Össz z rec	Helyes	Ins	De l	Subs	Acc	WE R
890	1326	504	822	8	370	35,7	43,4

A viszonylag rossz eredmények oka (tipikusan a kötőszavak tévesztése nagy), hogy bár a szótárkészlet ezen betanítóanyag választása mellett biztosítja a legnagyobb fedést, ennek ellenére bigram szókapcsolatok nem fedik a tesztelési bigram szókapcsolatokat. Ha megfigyeljük az 1. táblázatbeli eredményeket, és összevetjük a tapasztaltakkal, akkor megállapíthatjuk következtetésképpen, hogy a teljes anyaggal történő betanításkor (szegedi, budapesti, gasztroszkópiái, colonoszkópiái) olyan nagy mértékű hamis szókapcsolat-statisztikát vittünk be a rendszerbe, hogy az a bigram valószínűségi mezőben zaj keletkezett, így hiába lettek betanítva ezen szókapcsolatok, mégis rossz lett a felismerés. (lásd 1. táblázat budapesti és szegedi szókészlet eltérését.)

Szűkítve a betanítási anyagot, a betanításra csak a budapesti gasztroszkópiás betanítóanyagot az eredmények javulnak, amint azt a 4. táblázatban mutatjuk.

4. táblázat: Gasztroszkópiás felvételek tesztelési eredményei lexéma alapú budapesti gasztroszkópiás betanítású nyelvi modell esetén, orvosok bemondásában

Össz z ref	Össz z rec	Helyes	Ins	De l	Subs	Acc	WE R	PP
1150	1417	799	283	8	343	44,8	30,5	73,59

A 2 táblázat eredményei orvosok bemondásai alapján elkészített tesztelési eredmények. A lelet-felvételek meghallgatásakor azt tapasztaltuk, hogy a felvételek igen zajosak és kiejtés szempontjából is igen rossz minőségűek. Így felvetődött azon lehetőség - a nagyobb felismerési pontosság elérése érdekében - hogy limitált szintű zajkörnyezetben felvett felvételekkel teszteljünk és a felvétel során ügyeljünk a helyes artikulációra. Ennek érdekében a budapesti gasztroszkópiás leleteket –amelyeket az orvosok is bemondtak (20 lelet)– bemondtuk a laboratóriumban és ezen felvételeket használtuk fel az MKBF tesztelésére. A tesztelési eredményeit az 5. táblázatban közöljük.

5. táblázat: Kiszajú, laboratóriumi gasztroszkópiás felvételek, tesztelési eredményei lexéma alapú budapesti gasztroszkópiás betanítású nyelvi modell esetén

Össz z ref	Össz z rec	Helyes	Ins	De l	Subs	Acc	WER	PP
1173	1451	922	280	1	250	54,7	21,3	73,59

Az eredményeket összevetve a 4. táblázatával, látható hogy a szótévesztési arány (WER) javult, mivel az akusztikai szintű felismerés javult, ezáltal az egész felismerés is pontosabbá vált.

Továbbá olyan leletbemondással tesztelünk, ami szerepelt a nyelvi modell betanításában. Ennek megvizsgálása érdekében – a fentiekben említett akusztikai feltételek biztosítása mellett – budapesti gasztroszkópiás leleteket rögzítettünk, amelyeket a budapesti gasztroszkópiás írott leletanyagból olvastunk fel, és a nyelvi modellt a 5. táblázathoz hasonlóan budapesti gasztroszkópiás írott leletanyaggal tanítottuk be. Az eredményeket a 6. táblázatban adjuk meg.

6. táblázat Laboratóriumi, budapesti, gasztroszkópiás felvételek tesztelési eredményei lexéma alapú, budapesti, gasztroszkópiás, betanítású nyelvi modellekkel

Össz z ref	Össz z rec	Helyes	In s	De l	Sub s	Acc	WE R	PP
416	444	380	28	0	36	84,6	8,6	9,36

A 6. táblázat egyértelműen mutatja, hogy a szótévesztési arány (WER) javul, hiszen olyan leletanyagot teszteltünk, amely a betanításban szerepelt. Ha a perplexitást vizsgáljuk, akkor 4. és 5. táblázatokban található perplexitás-értékekhez képest nagy arányú perplexitás csökkenést tapasztalhatunk.

4.1 A lexéma alapú tesztelési eredmények kiértékelése

Az eredmények alapján az alábbi következtetéseket vonhatjuk le.

1. A budapesti és a szegedi leletanyag annak ellenére hogy mind a két korpusz azonos témájú, azonban használatban, stílusban jelentősen különböznek, hogy a vegyes anyag alapján készített bigram nyelvi modellel, noha elviekben robosztusabb, a gyakorlatban mégis gyengébb felismerés érhető el. A *WER* eredmény 5.54%-kal rosszabb a csak budapesti anyag alapján tanított bigramhoz képest.
2. A beszéd felismerés során különösen fontos a szöveg gondos, folyamatos bemondása, így a *WER* értéke akár 10%-kal is lejjebb szorítható.
3. A jelenlegi, budapesti gasztroszkópiás leletanyag alapján készült bigram nyelvi modell nem fedi kellőképpen a kívánt alkalmazási területet. Ezt igazolja a szakorvosok és a saját bemondásban készült leletek hibaarányainak nagyfokú korrelációja, azaz a hibaarány jelentős része a bigram nem megfelelő fedéséből, és nem az akusztikai jel minőségéből adódik.
4. Megjegyezzük, hogy a szakorvosok által végzett bemondásokban a PI kódjelű bemondó halk, az általánosan elvárhatónál gyengébb beszédproduktumot adott a felvételek során. Amennyiben az ezekkel a felvételekkel kapott hibaarányt figyelmen kívül hagyjuk, az összesített *WER* értéke 30,52%-ról 24,27%-ra javul, amely utóbbi véleményünk szerint a hitelesebb adat.
5. Az *Acc* paraméter esetenkénti alacsony értéke arra enged következtetni, hogy az adott tesztfelvétel a nyelvi modell számára ismeretlen, vagy a bigram betanító anyagában nem kellő mértékben előfordult szót tartalmaz. Ilyenkor a beszúrások megszorodnak, amely jellemzően több rövid, felolvasva az elhangzó, de fel nem ismert szóéhoz hasonló hangélményt ad.

4.2 A perplexitás vizsgálata és ennek összevetése a lexéma – valamint morféma alapú beszédfelismerő tesztelési eredményeivel

A 1.2 fejezetben ismertettük és az 1.3 fejezet utolsó eredményeivel szemléltettük, hogy a perplexitással jósolható a felismerés pontossága. Ebben a részben azt vizsgáltuk, hogy a felismerés hatékonysága mennyire jósolható a nyelvi modell szimulálására szolgáló perplexitás-számítással. Ennek vizsgálata érdekében megmértük, hogy mennyire konvergálnak a perplexitás alapú becslési értékek a Word Error Rate értékekkel abban az esetben, amikor olyan anyaggal tesztelünk, ami részét képezi a nyelvi modell betanításának. A budapesti gasztroszkópiás leletekkel kapott mérés eredményeit 1. ábrán a szemléltetjük.

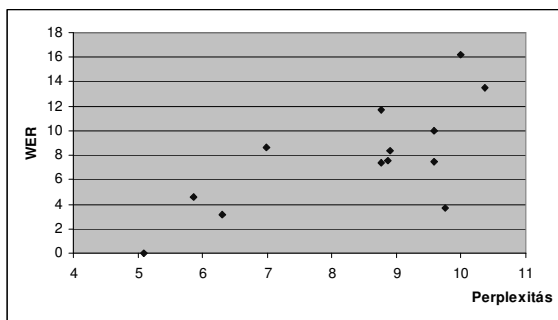


Fig1. Perplexitás és WER értékek korrelációja kiszajú, laboratóriumi, budapesti gasztroszkópiás felvételek esetén, budapesti gasztroszkópia nyelvi modell tanítva

Amint az 2. ábra eredményei szemléltetik, a perplexitás – WER értékpárosok által meghatározott pontok szórnak, de az összefüggés a perplexitás és a WER értékek között jól látható [1]. Az, hogy az értékek szórnak, érthető, hiszen nem a teljes magyar nyelvet vizsgáltuk, hanem csak egy igen szűk tématerületet, ami – a tapasztalatok alapján – a szakterületi mellett hétköznapi, valamint irodalmi nyelvet is tartalmaz. A másik lehetséges oka abban rejlik, hogy a perplexitásszámítás folyamán nem számolunk a fonémátévesztéssel és a fonémátévesztés hatására eltolódik a WER-Perplexitás kapcsolat [4].

4.3 Lexéma vagy morféma felismerés

A bigram statisztikákat egyszer lexéma egyszer morféma alapokon számítottuk. A tesztelések hasonló eredményekre vezettek mindkét esetben.

Ami a morféma alapú felismerés mellett szól:

Ha a morféma szintű felismerést választjuk a szótárméret jelentősen csökken, így kisebb bigram valószínűségi mezőt kell kezelni. Lexéma alapú betanítás esetén a leletkorpusz alapján 14331 (lexéma) egység jön létre, míg morféma alapú betanítás esetén 6706 egység (morféma).

Mivel a bigram valószínűség mező egy négyzetes diszkrét valószínűségi mező, így tárolási szempontból körülbelül 4,5-szeres tárcsökkenést eredményez, és még a leg-

rosszabb esetben (simítás esetén) is már átlagosan 2,13-szeres valószínűségérték növekedés érhető el.

8. táblázat: tesztelési eredmények, betanítás a budapesti gasztroszkópiás anyaggal, tesztelés kiszajú, laboratóriumi, budapesten felvett gasztroszkópiás felvételekkel

Morféma alap								
Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER	PP
1631	2045	1241	778	9	355	28,3	23,9	27,31
Lexéma alap								
Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER	PP
1173	1451	922	280	1	250	54,7	21,3	73,59

Ami a lexéma szintű felismerés mellett szól:

Morféma szintű felismerés esetén komoly problémát jelent a toldalékok határain fellépő hasonulások, összeolvadások, hangrendilleszkedések hangzókiesések kezelése. Ennek leírása egyelőre úgy tűnik, csak manuálisan oldható meg.

5 Felismerési pontosság növelése perplexitás alapú szimulálás segítségével

Amint azt az 1. fejezetben ismertettük, a perplexitással becsülni lehet a felismerés pontosságát. A 5. táblázat egy olyan tesztelési eredményeket tartalmazó táblázat, ahol a tesztelésnek kinevezett állomány nem szerepelt a nyelvi modell betanító leletei között. A 6. táblázat viszont olyan tesztelési eredményeket tartalmaz ahol a tesztanyag részét képezte a nyelvi modell betanító anyagának, tehát biztosítottak voltak azon szókapcsolatok betanításai amelyek a tesztelésnek kinevezett anyagokban szerepeltek. Ha a 5. táblázat eredményeit összevetjük a 6. táblázat eredményeivel, azt tapasztalhatjuk hogy a hibaszázalékok kisebbek a 6. táblázatban. Így felvetődött az a kérdés, hogy ha betanítanánk a 5. táblázathoz tartozó tesztelési mondatokból azon szókapcsolatokat amelyek nem szerepeltek a betanításnál, akkor a felismerési pontosság várhatóan növekedni fog e. Ehhez csupán ezen hiányzó szókapcsolatokat kell megkeresni és a betanítóanyagban elhelyezni.

A hiányzó szókapcsolatok betanításánál azt a technikát választottuk, hogy a tesztelő anyagból azon **szóláncokat** kerestük meg, amely szóláncok bármely bigram szókapcsolatát tekintve, egyik sem szerepelt a betanításban. Tehát a szóláncok kiválasztása a következő:

<utolsó betanításban szereplő szó> <betanításban nem szereplő szó>⁺ <első olyan szó ami a betanításban szerepelt>

+ jel jelenti, hogy 1-nél többször is szerepelhet egymás után betanításban nem szereplő szó, a reguláris kifejezéseknél használt jelölésekhez hasonlóan

Ezen módszer választása mellett a szándékunk az, hogy a hiányzó bigram valószínűségeket betanítsuk anélkül hogy a már meglévő bigram valószínűségeket jelentősen torzítanánk.

Azt tudjuk, hogy mely részekkel kell a betanítást bővíteni, azonban azt nem, hogy ezen hiányzó részek betanítását hányszor kell megismételni. Az ismétlések számának

meghatározása az általunk használt perplexitás alapú nyelvi modell hatékonyságának becslése alapján történt (6. ábra). Előállítottunk különböző betanító anyagokat, amelyek felépítésüket tekintve a következőképpen alakultak:

Betanítás=<budapesti gasztroszkópiás leletek> + <hiányzó szóláncok >*

* jel jelenti, hogy a hiányzó szóláncok 0..n- szer szerepelhetnek a budapesti gasztroszkópiás leletek után, reguláris kifejezéseknél használt jelölésekhez hasonlóan.

A perplexitás értékeket különböző betanító anyagoknál nem lehet összehasonlítani, esetünkben viszont az összehasonlítás elvégezhető, mivel a betanítóanyagot csak kismértékben módosítottuk, szókészletek megegyeznek, a bigram valószínűségi mező csupán eloszlási értékeiben csak kismértékben különbözik egymástól.

5.1 A betanításszám meghatározása nyelvi modell szimulálás segítségével

Kiválasztottunk 4 leletet tesztelésre, a betanítóanyaggal összehasonlítva meghatároztuk a hiányzó szóláncokat. A betanításnál ezen hiányzó szóláncokat szerepeltettük 0..n –szer. Ezen leletek mellett figyeltük a többi leletet is, hiszen a cél a bigram valószínűség mező felismerésének erősítése, nem pedig a torzítása. Amint a 2.1 ábrából megfigyelhető a hiányzó szólánc ismétlési számának növelésével a perplexitás értékek javulnak azon tesztanyag esetén, amely alapján a hiányzó szólánc elő lett állítva. A 2.2 ábrán szemléltettem azon gasztroszkópiás leletek perplexitás értékeinek alakulását, amelyekből nem lett hiányzó szólánc véve. Látható, hogy a 19-20 –szoros betanításig folyamatos perplexitás javulás van, e feletti betanításnál viszont csak romlás tapasztalható.

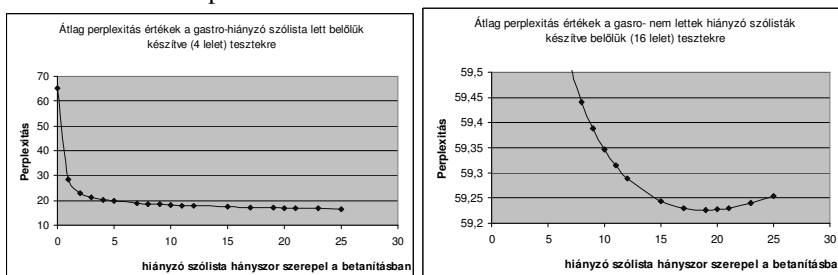


Fig2.1-2.2: Perplexitás átlagok alakulása

A hússzoros betanítással kiegészített budapesti gasztroszkópiás anyaggal újra elvégeztük a tesztelést.

Betanítás:

1: budapesti gasztroszkópiás írott leletanyagok+<különbözeti szólánc>

2: budapesti gasztroszkópiás írott leletanyagok + 20*hiányzó szólánc+<különbözeti szólánc>

A 9.1 táblázat az első betanítás tesztelési eredményeit szemlélteti, míg a 9.2 táblázat a második betanítás eredményeit szemlélteti.

9.1 táblázat: Tesztelési eredmények budapesti gasztroszkópiás írott leletek esetén.

betanítás : hiányzó szólánc nem szerepel a betanításban									
Lelet	Bemondó	Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER
3	BG	95	116	67	21	0	28	48,4	29,4
33	BG	63	80	48	17	0	15	49,2	23,8
53	SG	55	68	50	13	0	5	67,2	9,1
92	SG	55	62	46	7	0	9	70,9	16,3
átlagos_WER:		19,6%	átlagos Acc:		58,9%				

9.2 táblázat: Tesztelési eredmények budapesti gasztroszkópiás írott leletek + 20*hiányzó szólánc esetén. Tesztanyag: 3, 33, 53, 92

betanítás : hiányzó szólánc 20 szor szerepel a betanításban									
Lelet	Bemondó	Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER
3	BG	95	110	78	17	1	16	64,2	17,8
33	BG	63	69	61	6	0	2	87,3	3,1
53	SG	55	58	53	3	0	2	90,9	3,6
92	SG	55	61	49	6	0	6	78,1	10,9
Átlagos_WER:		8,9%	átlagos Acc:		80,1%				

A táblázatok azt mutatják, hogy a szóláncok 20-szoros megismétlésével a szótévesztés erősen lecsökkent (9.2 táblázat), a **19,7%-os szótévesztés 8,9 %-ra javult.**

Azonban kérdéses, hogy a tesztelésnél megjelölt többi lelet esetében mi lett az ilyen betanítás mellett az eredmény.

10.1 táblázat: Tesztelési eredmények budapesti gasztroszkópiás írott leletanyagok esetén. Tesztanyag: 94, 38, 174

betanítás : hiányzó szólánc nem szerepel a betanításban									
Lelet	Bemondó	Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER
94	SG	51	64	41	13	0	10	54,9	19,6
38	ZT	34	41	11	11	2	21	0	67,6
174	SG	118	167	70	49	0	48	17,7	40,6
átlagos_WER:		42,6%	átlagos Acc:		24,2%				

10.2 táblázat: Tesztelési eredmények budapesti gasztroszkópiás írott leletanyagok + 20*hiányzó szólánc esetén. Tesztanyag: 94, 38, 174

betanítás : hiányzó szólánc 20 szor szerepel a betanításban									
Lelet	Bemondó	Össz ref	Össz rec	Helyes	Ins	Del	Subs	Acc	WER
94	SG	51	64	41	13	0	10	54,9	19,6
38	ZT	34	41	11	11	2	21	0	67,6
174	SG	118	166	68	48	0	50	16,9	42,3
átlagos_WER:		43,2%	átlagos Acc:		23,9%				

.A 10.1 és 10.2 táblázatokat összehasonlítva láthatjuk, hogy nem változott a felismerés pontossága, tehát itt is beigazolódott az előzetes becslés.

Tehát a perplexitás elemzéssel, és a betanító anyag egyszerű módosításával egy meghatározott szótárkészletű, nyelvi szöveg felismerését jelentős mértékben javítani tudjuk.

6 Kiértékelés

Tehát a példaként bemutatott perplexitás átlag vizsgálata alapján sikerült a szólán-cok ismétlési számát optimálisra beállítani, úgy, hogy a felismerés lényegesen jobb lett, 90% fölötti. A vázolt eljárás sok esetben javíthatja a felismerést.

Figyelembe kell azonban venni, hogy a módszerünk nem ad valóságos megoldást, hiszen a gyakorlatban, az erősen agglutináló nyelveknél tisztán statisztikai n-gram modellel dolgozva, mindig lehet új elem az új bemondások között, ami a betanító anyagban nem szerepelt, és ez hibázáshoz vezet. Azonban, egy közepes szótár méretű, kötött témában kialakítandó felismerő létrehozásában jelentős segítség lehet.

Bibliográfia

1. Máté Szarvas, Sadaoki Furui : Evaluation of the Stochastic Morphosyntactic Language Model on a One Million Word Hungarian Dictation Task. EUROSPEECH (2003) GENOVA, 2297-2300
- 2 Chen S., Beferman D., Rosenfeld R. : Evaluation Metrics For Language Models, In: DARPA98 , National Institute of Standards and Technology (NIST),
Elérhető: www.nist.gov/speech/publications/darpa98/html/lm30/lm30.htm
3. Clarkson P., Robinson T. : Towards improved language model evaluation measures. Elérhető: <http://Citeseer.ist.psu.edu/clarkson99toward.html>
4. Deng Y., Mahajan M., Acero A. : Estimating Speech Recognition Error Rate without Acoustic Test Data. Elérhető: <http://research.microsoft.com/srg/papers/2003-milindm-eurospeech.pdf>
- 5 HUMOR Morfológiai elemző. Elérhető: http://www.morphologic.hu/h_humor.htm
- 6 Becchetti C., Ricotti L. P.: Speech Recognition, Theory and C++ implementation. Fondazione Ugo Bordoni, Rome, (1999). ISBN 0-471-97730-6
- 7:Ney, H., Essen, U., Kneser, R.: On Structuring Probabilistic Dependencies in Stochastic Language Modeling. Computer Speech and Language, (1994). 8:1-38.
- 8 Velkei Szabolcs, Vicsi Klára: Beszédfelismerő modellépítési kísérletek akusztikai, fonetikai szinten, kórházi leletező beszédfelismerő kifejlesztése céljából, MSZNY (2004). 307-315.
- [9] Vicsi Klára, Kocsor András, Teleki Csaba, Tóth László: Beszédatadátbázis irodai számítógépfelhasználói környezetben, II. Magyar Számítógépes Nyelvészet Konferencia, 2004. 315-319 oldal