

Fonetikus morfológiai elemző beszédfelismeréshez

Gyepesi György¹, Kertész Zsuzsa^{1,2}, Serény András¹,

¹ Alkalmazott Logikai Laboratórium,
1022 Budapest, Hankóczy J. u. 7.
{ggyepesi, kzsuzsa, sandris}@all.hu

² ELTE Angol Nyelvészeti Doktori Program,
1088 Budapest, Rákóczi út 5.

Kivonat: Ebben a tanulmányban azt mutatjuk be, hogy hogyan alakítottunk ki a magyar nyelvre **fonetikus morfológiai elemzőt**. Noha szegény morfológiájú nyelvek (mint amilyen az angol) esetében jó eredményt adnak a szóalak n-gramm nyelvmodellek, ragozó nyelvekhez olyan nyelvmodelleket érdemes kialakítani, amelyek a pusztán szórend helyett a szavak szerkezetét, alakját is figyelembe veszik. Ezek hatékony működéséhez azonban az szükséges, hogy bemenetük ne csupán a szóalak legyen, hanem az ahhoz tartozó morfológiai elemzés is. Ennek egyik módja az, ha fonematizáljuk a betűalapú morfológiai elemzőnket. Ennek alkalmazásával elkerülhető számos, a kiejtésszótár használatkor felmerülő buktató is.

1 Bevezetés

A beszédtechnológiai kutatásokban nem új keletű az a felismerés, hogy hatékony beszédfelismeréshez nem lehet elegendő csupán az akusztikus jelsorozatok „megfejtése”. A beszédfelismerő rendszer nem állhat pusztán az akusztikus komponensből, amely az akusztikus inputhoz egy fonémaszimbólum-sort (fonémajel-sorozat) rendel, mert így az adott nyelvben lehetetlen fonémakapcsolatok, szavak, szóorozatok ugyanolyan a priori valószínűséggel jelennének meg, mint a lehetségesek. Ahogyan az emberi beszédfelismerés során is támaszkodunk szókincsbeli és grammatikai ismereteinkre, a gépi felismerésben is olyan rendszert érdemes kialakítani, amely a lehető legpontosabban modellálja ezt a folyamatot. Emiatt az akusztikus szintet mindig valamilyen **kiejtésszótárral** és a mondattani mintákért felelős úgynevezett **nyelvmodellel** kell támogatni.

A klasszikus felépítése egy ilyen beszédfelismerő rendszernek a következő: az akusztikus jelsorozatot az **akusztikus komponens** dolgozza fel, az általa kimenetként adott fonémajel-sorozatra az úgynevezett **kiejtésszótár** elemeit illesztjük, végül az így kapott szóorozatok közül a **nyelvmodell** segítségével választunk. A kiejtésszótár olyan szólista, amely a szavak írásképe mellett felsorolja azok lehetséges kiejtéseit.

Az angolban például a *read* alaknál egyaránt szerepel a R IY D¹ (a jelen idejű alak kiejtése) és a R EH D (múlt idejű alak kiejtése) kiejtés.

A kiejtésszótár bemenetét tehát az akusztikus felismerő által adott **fonémajel-sorozatok** adják (például R IY D), a kimenete pedig az adott hangsor „szótári alakja” (*read*). Azokat a fonémajel-sorozatokat, amelyekre nem illeszthető értelmes szótári szavak sorozata, a felismerő eldobja.

A teljesnek talált illesztéseket (azaz a kiejtésszótár által adott szóorozatokat) kapja meg a nyelvmodell, amely valószínűségük szerint értékeli őket: amelyik az adott nyelvben valószínűtlen, ritkán előforduló sorrend, az kis értéket kap, a gyakori szó-együttállások értelemszerűen nagyot.

1.1 Mik a problémák a kiejtésszótárral?

A fent említett *read* ige példája máris felvet egy problémát a kiejtésszótár működési elvével kapcsolatban. Míg a kiejtett alaknál (az akusztikus komponens kimeneténél) még tudtuk, hogy R IY D vagy R EH D hangzott el, a kiejtésszótár a „neutralizált” *read* alakot fogja kiadni, mindenféle egyéb információ nélkül és ezt a szóalakot kapja meg a nyelvmodell. A nyelvmodell tehát nem használhat a szóalakokra vonatkozó egyéb információt.

Angolra jól működnek a szóalak-alapú (tehát hozzáadott alaktani információt nélkülöző) n-gramm modellek (lásd többek között [4]), ezért kisebb a jelentősége annak, hogy az adott helyen a *read* ige jelen vagy múlt idejű alakja szerepelt-e.

Ugyanakkor egy erősen ragozó nyelvre, mint amilyen pl. a magyar, a finn vagy a török a szóalak n-gramm modelleknél jobb eredményt adnak azok a modellek, amelyek morfológiai információt is tudnak kezelni (erről részletesebben lásd a 2.1. pontot). Egy ilyen nyelv esetében nagyon nem mindegy, hogy a kiejtésben tükröződő morfológiai vagy jelentésbeli különbséget sikerül-e a nyelvmodell számára elérhetően reprezentálni².

A kiejtésszótárak nem képesek kódolni a szóhatárokon fellépő hasonulásokat³. Vegyük példának a brit angolban az *r* sajátos disztribúcióját és az emiatt fellépő bizonytalanságokat. Izoláltan az angolban szó végén nem fordulhat elő *r*; a szóvégi *r*-ek csak akkor jelennek meg, amikor egy magánhangzóval kezdődő szó vagy toldalék következik: *far away*.

A kiejtésszótár komoly akadályba ütközik, amikor a F AH R AX W EY fonémajel-sorozatot látja: ugyanis a két szó „között” megjelenő *r*-rel sehogy nem tud elszámol-

¹ Az angol szavak fonetikus átírását a DARPA fonetikus ábécé szerint végeztük, de itt az egyszerűség kedvéért a hangsúlyjelöléseket elhagytuk; a magyar szavak átírására a helyesírást nagyjából tükröző saját rendszert alakítottunk ki.

² A magyarban igen kevés olyan alak van, amelynél egy írott alakhoz több lehetséges kiejtés társul különböző jelentéssel (ezeket homográfoknak hívják a lexikológiában). Egy lehetséges példa az *egészség* szó két különböző ejtése: a gyakoribb, „betegség hiánya” jelentésben a *-ség* képző *s*-éhez hasonul az előző *-sz*; a ritkább (sokak szerint nem is létező), „valaminek az egész volta” jelentésű alakban ez nem történik meg.

³ Az egyszerűség kedvéért itt hasonulásnak nevezünk minden, morféma- vagy szóhatáron fellépő hangtani folyamatot (így a brit angolban megfigyelhető $r \sim \emptyset$ váltakozást is).

ni. Ha gondos kiejtésben hangzott el a szókapcsolat, akkor a F AH R AX W EY megoldásnak kell a legnagyobb akusztikus valószínűséggel bírnia; emellett persze megjelennek kisebb akusztikus valószínűséggel olyan (téves) megoldások is, mint például F AH AX W EY vagy F AH W EY stb. Ebből az a hiba származhat, hogy a kiejtésszótárban megjelenő (illetve nem megjelenő) alakok miatt épp a legnagyobb akusztikus valószínűségű – és valóban elhangzott – megoldást kell eldobni, és egy vagy több kisebb valószínűségűt megtartani és közvetíteni a nyelvmodell felé. Innen a nyelvmodellnek már lehetetlen korrigálnia a téves hipotézist.

Az ilyen problémákat a kiejtésszótáron belül csak úgy lehet orvosolni, ha minden szónak felvesszük az összes olyan alakját, amely egy szóhatáron fellépő folyamat miatt létrejöhet. Ez a megoldás viszont legalizálja azokat a fonémasorokat is, ahol egy szónak olyan kiejtése jelenik meg, amit a következő nem indokol (például a zöngéségi hasonulás miatt felvett alakokat elfogadja olyan esetben is amikor nincs hasonulás, olyanokat eredményezve, mint A B L A G A L A T⁴, *ablak alatt*). Emellett ez az út a kevésbé ragozó nyelvek esetében is nehézkes, és nagy bővítéssel járna együtt.

2 Fonematizált morfológia

2.1 Miért kell morfológiai elemzés?

Morfológiailag gazdag nyelvek esetében is meg lehet valósítani – bár feltételezhetően kevésbé hatékonyan – a felismerést anélkül, hogy morfológiai elemzést vennénk igénybe. Nézzük, hogyan nézne ki a fenti rendszer a magyarra.

Először is ki kell alakítani a kiejtésszótárat, amely ez esetben egy szóalaktár. Ez azt jelenti, hogy elvileg az összes lehetséges szóalakot és azok kiejtéseit tartalmazza. Ez első ránézésre reménytelennek tűnik: egy magyar névszó lehetséges alakjainak száma száznál is több lehet, az igék esetében ez a szám pedig még több.

A probléma azonban csak látszólagos: a korpuszok vizsgálatával ugyanis azt lehet kideríteni, hogy az előforduló szóalakok száma mindössze kb. 2-3-szorosa a szótári alakoknak [6]. Ez azt jelenti, hogy minden szónak 2-3 (de mondjuk maximum 5-6) alakját kellene felsorolnunk, és ez már nem tűnik olyan lehetetlennek. Ehhez persze a korpuszhoz kellene igazítani a szótárt, ez pedig nem kis munka, de kivitelezhető.

Mivel a magyarra viszonylag egyértelműek és világosak a kiejtési szabályok, a szóalaktár fonematizálása – azaz a „kiejtésszóalaktár” – létrehozása sem okoz nagyobb gondot. Tehát a kiejtésszótár létrehozása megoldható: okozhat ugyan némi bizonytalanságot a felismerésnél – ha például olyan alakkal találkozunk, amely véletlenül épp nincs benne a korpuszhoz igazított szótárban – de ez nem okoz jelentős romlást az eredményekben.

Egy másik alternatíva a Mathias Creutz által [2] bemutatott *Morfessor* nevű alkalmazás elvén működő elemzés: a szótárban nem csak szavak, hanem toldalékok is szerepelnek. Ha ezeket fonematizáljuk, kész a kiejtésszótárunk a ragozó nyelvre is.

⁴ Technikai okok miatt a szegmentumok hosszúságát nem kezeljük.

Ezzel ugyan elfogadhatóvá válnak értelmetlen szavak is, de mivel az akusztikus be-
menetben ritkán fordulnak elő nem létező alakok, elvileg itt is bizhatunk abban, hogy
a felismerést ez nem rontja számottevően.⁵

Azonban hiába működnének ezek a megoldások, mindegyiknél megmarad a szóha-
táron előforduló hasonulások problémája. Ennek megoldására csak az a lehetőség
marad, hogy a különböző alakokat felsoroljuk a szótárban (lásd például F AH mellett
F AH R). Ennek hiányában torzulhat az akusztikus felismerés, ahogyan már az angol
példán bemutattuk.

A legfőbb motivációnk azonban arra, hogy a morfológiai elemzést ne kerüljük
meg, az a **nyelvmodell** jellege. Mint azt már fentebb említettük, a magyarra és a
ragozó nyelvekre az n-gramm modellek rosszabb eredményt nyújtanak, mint egy
morfológiailag egyszerűbb nyelvre. Ennek több, a nyelvi tipológiából fakadó oka
van. Egyrészt a gazdag morfológia általában – bár nem feltétlenül – együtt jár a
szórendi kötöttségek elmaradásával. Egy angol mondat szórendje szigorúbb, mint egy
magyar mondaté, és még ha a magyar szórendi variációk többnyire fontos grammati-
kai információt hordoznak is (lásd például topik, fókusz, stb.), a nyelvmodell szem-
pontjából ez kevésbé lényeges. Csak az látható számára, hogy ugyanarra a szóhal-
mazra a szavak számától függően akár 4-6 legális szórend is előfordulhat (lásd példá-
ul *Peti tavaly abbahagyta a hegedülést; Peti hagyta abba a tavaly a hegedülést; Peti
tavaly hagyta abba a hegedülést; Peti tavaly a hegedülést hagyta abba*).

A szórendi kötöttségekhez az angolban az is hozzátartozik, hogy a jelentésanilag
és/vagy szintaktikailag összetartozó szavak – vagy szószerkezetek, frázisok – nem,
vagy csak bizonyos mértékben szakíthatók el egymás mellől: az igét többnyire köz-
vetlenül követik a legfontosabb bővítményei (tárgy vagy egyéb kötelező vonzatok), a
különböző határozók (idő-, mód-, hely-, stb.) nem férkőzhetnek be ezek közé. A
magyarban ez utóbbi „beférkőzésnek” gyakorlatilag nincs akadálya: emiatt az össze-
tartozó szószerkezetek olykor igen messze, több szó távolságra is kerülhetnek egy-
mástól, amit egy szóalak n-gramm modell nem tud kezelni. Noha a magyarban is
akadnak ilyen „szétszakíthatatlan” szerkezetek, ezek jóval ritkábbak, mint az angol-
ban.

Másrészt a ragozás során előálló viszonylag sokféle szóalak miatt adatritkasággal
találkozunk, ami ismét ront a nyelvmodell tanulási hatékonyságán. Azt gondoljuk,
hogy ezen okok miatt olyan nyelvmodellel kell dolgoznunk, amely nem hagyja fi-
gyelmen kívül a szavak belső szerkezetét. Egy ilyen nyelvmodell (legtípikusabb
morfológiai nyelvmodellek az úgynevezett faktorizált nyelvmodellek lásd pl. [1])
például képes morfológiai szabályszerűségeket tanulni: például a *fut+nak* szóalak
statisztikáiból következtetni tud az esetleg ritkábban előforduló *szalad+nak* alakra is.
Az viszont világos, hogy ez egy szóalak alapú nyelvmodellnél nem történhet meg,
ugyanis amíg a szavak belső szerkezete láthatatlan a nyelvmodellnek, addig az ilyen
analógiákat nem lehet vele felfedeztetni⁶.

5 Természetesen ha nem izolált szavakról, hanem szószorozatokról van szó, akkor igenis okoz-
hat problémát az, hogy nem létező szavakat is elfogad az elemző, ugyanis előfordulhat, hogy
tévesen azonosítja a morfémákat és ezáltal a szavakat is.

6 Meg kell azonban jegyezni, hogy az említett „távoli” grammatikai összefüggéseket a
faktorizált nyelvmodellek sem képesek jól kezelni. Ez a probléma még mindig nem megol-
dott a nyelvmodellek kialakításában.

2.2 Miért legyen fonematizált a morfológiai elemző?

A fentiek alapján láthatjuk, hogy olyan eszköz megvalósítása lenne gyümölcsöző, amely egyrészt morfológiai annotációval tud bemenetet nyújtani a nyelvmodellnek, másrészt pedig a szó- és morfémahatáron történő hasonulásokat is jól tudja kezelni. Egy olyan morfológiai elemzőt képzelünk tehát el, amelynek a bemenete betűsor helyett fonémajelsor, a kimenete pedig a szokásos módon az adott sztring morfológiai elemzése.

Tehát, adva van egy fonémajelekből álló input (pl. B AA NY A); a morfológiai elemzőnk (ispell-alapú, lásd később) erre kiadja a következő elemzéseket

1. bány/NOUN
2. bán/VERB<DEF>
3. bán/VERB<SUBJUNC-IMP><<DEF>
4. bán/NOUN<POSS>

Ezután már a nyelvmodellen múlik, hogy melyiknek tud nagyobb valószínűséget adni.

Mik az előnyei egy ilyen fonetikus (fonematizált) morfológiai elemzőnek? Egyrészt mivel a magyar fonematizálási szabályok és a morféma- és szóhatáron lejátszódó hangtani folyamatok egyaránt kódolva vannak benne, nincs szükség arra, hogy felvegyük azokat a szóalakokat, amelyeket más esetben csak a szóhatáron történő hasonulások miatt kellene felvenni a szótárba (lásd az angol *far away* példát). Másrészt mivel a morfológiai elemző (ha jól működik), elő tudja állítani az összes lehetséges szóalakot – ráadásul ideális esetben csakis a létezőket állítja elő –, fel sem merül a szóalakok felsorolásának igénye. Harmadszor pedig fontos megemlíteni, hogy ezáltal meg lehet spórolni az egész kiejtésszótárt, eggyel kevesebb modulunk lesz az építményben, ami miatt az egész rendszerünk egyszerűbb, elegánsabb, és vélhetően gyorsabb is lehet.

3 Megvalósítás

3.1 A transzducer felépítése

Ahhoz, hogy a fonetikus morfológiai elemzőt meg tudjuk valósítani, szükség van (i) egy fonematizáló algoritmusra, és (ii) egy speciális szöveges morfológiai elemző transzducerre. A fonematizáló feladata megadni egy tetszőleges szöveg fonetikus átíratát. A morfológiai elemző transzducer abban az értelemben speciális, hogy az élein nem karakterek, hanem szótövek és toldalékok (tehát morféma) szerepelnek.

Első lépésben azt az algoritmust ismertetjük, amely a fonematizáló és a speciális transzducer használatával kialakítja a fonetikus morfológiai transzducert. Végül

pedig azt mutatjuk be, hogyan építettünk fonematizálót és speciális morfológiai transzducert a magyar nyelvre.

Inicializáljuk a morfológiai elemző transzducert úgy, hogy vesszük a speciális morfológiai transzducert, és minden élen kicseréljük az inputot (a szótövet vagy toldalékot) annak fonetikus átírával. Ezzel együtt megőrizzük az eredeti, karakteres (tehát „helyesírás szerinti”) inputokat is. A fonetikus átíratokat a fonematizáló algoritmus adja.

Ezek után a transzducer *minden* egymást követő élpárját sorra vesszük. Legyen E és F két egymásutáni él, továbbá *ei* az E él inputja, *eo* az E él outputja; *fi* az F él inputja, *fo* pedig az F él outputja. Az output jelen esetben morfológiai annotációt, elemzést jelent. Ezenkívül legyen *epi* az E él inputjának, *fpi* pedig az F él inputjának fonematizált alakja.

Az *ei* és *fi* konkatenációjaként létrehozott *ei^fi* karaktersorozatot fonematizáljuk, az eredményt nevezzük *p*-nek.

Amennyiben a *p* fonémajelsorozat megegyezik *epi^fpi*-vel (tehát az *ei* és *fi* fonetikus alakjainak egymás után fűzésével), akkor nem történik semmi, mert ez azt jelenti, hogy az E és F éleken megjelenő inputok olyanok, hogy konkatenációjukkor a fonematizálásban nem történik hasonulás (pl. *ház+ban* = H AA Z B A N).

Ha azonban *p* eltér *epi^fpi*-től, akkor a fonematizált transzducerhez hozzáveszünk **egy új élet** E kezdőpontjából F végpontjába. Az új él inputja *p*, outputja pedig *eo^fo*: *ablak+ban* = A B L A G B A N, lásd az 1. ábrát.

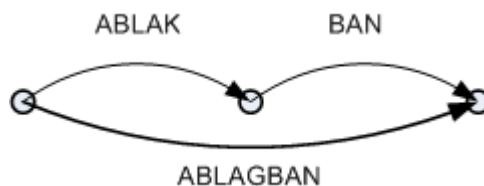


Fig. 1. A morféma- illetve szóhatáron történő hasonulásokat úgy reprezentáljuk, hogy az első él kezdőpontjából a második él végpontjába behúzzunk egy harmadik élet, amelyen a két morféma konkatenációjának fonematizált alakja szerepel.

Mit érünk el ezzel a módszerrel? Azt, hogy az új transzducer a hasonulást tükröző *p* fonémajelsorozatot elfogadja, és azt az outputot generálja hozzá, mint az eredeti, hasonulást nem reprezentáló változathoz.

Az élpárok mentén való hasonítás időnként kihagy olyan hasonulásokat, amelyek három vagy több él mentén jelennek meg. A magyarban igen ritkák az olyan hasonulások, amelyekhez élhármasokat kell figyelembe venni: ilyen például a *hajt+s+d*, ahol egyrészt a *t+s* összeolvadását (*cs*), valamint a *d* által kiváltott visszafelé történő zöngésségi hasonulást is kezelni kell valahogy (*hajdzsd*). Az algoritmusnak paraméte-

re az, hogy élhányasokkal dolgozzon, tehát nagyobb számra is beállítható; a magyarra azonban az esetek legnagyobb részére elégségesnek találtuk az élpárokat.

Az algoritmus eredménye tehát olyan fonetikus morfológiai transzducer, amelynek élein fonémajelsorozatok vannak.

2.2 A transzducer a felismerésben

Van tehát egy transzducerünk, amelynek élein egy véges ábécéből (a fonémajelekből) alkotott sorozatok szerepelnek inputként. Egy transzducer esetében azonban definíció szerint egy véges ábécé jelei az inputok, minden élen egy jellel: a bemutatott transzducer látszólag nem felel meg ennek a definíciónak. Ugyanakkor véges sok éle van, a véges sok élen pedig véges sok input, tekintsük tehát ábécének azt a halmazt, amelynek elemei a transzducer élein szereplő inputok.

Hogyan működik ez a transzducer mint felismerő? Legyen adott egy véges fonémajelsorozat P . A transzducerünk input ábécéjében fonémajelsorozatok vannak, P viszont *fonémajelek* és nem *fonémajelsorozatok* sorozata. Nem tudunk mást tenni, mint hogy P -t feldaraboljuk minden lehetséges módon, és minden darabolást felismertetünk a transzducerünkkel, összegyűjtjük a kapott kimeneteket, és ezek halmazát adjuk vissza.

Így is megvalósítható a felismerés, de ez a módszer exponenciálisan lassabb, mint ha olyan transzducerünk lenne, aminek minden élen egy jel (azaz egy fonémajel) az input.

A transzducerünket tehát **betűsíteni** kell: ha adott egy transzducer, amelynek élein egy *véges ábécé jeleiből álló sorozatok* az inputjelek, akkor a betűsítő algoritmussal készítjük el azt az eredetivel ekvivalens transzducert, amelynek élein az *ábécé jelei* az inputok. A betűsítő algoritmus szétvágja az egynél hosszabb éleket annyi élre, amennyi jeltől az eredeti él inputja áll, az utolsó élre írja az eredeti él kimenetét, és természetesen elkészíti a szétvágáshoz szükséges új csúcsokat. Az első él indulóállapota az eredeti él indulóállapota, az utolsó él végállapota az eredeti él végállapota, a közbenső csúcsok az újak, és az újak egyike sem elfogadó.

A speciális morfológiai elemző transzducert a [3]-ban ismertetett módon készítjük el ispell, illetve magyarra a **morphdb** [8] erőforrásból. Megjegyezzük, hogy a speciális transzducerből fent ismertetett betűsítő algoritmussal készítünk karakter-inputú morfológiai transzducert.

Fonematizáló transzducert a magyar nyelv fonetikus szabályaiból az [5] szerint építettük.

Bibliográfia

1. Bilmes, J. A., Kirchhoff, K.: Factored language models and generalized parallel backoff. In Proceedings of HLT/NAACL, (2003) 4–6.
2. Creutz, M.: Induction of the Morphology of Natural Language: Unsupervised Morpheme Segmentation with Application to Automatic Speech Recognition. Doctoral Dissertation (2006)

3. Halácsy P. et al. Végesállapotú transzducerek mindenkinek. V. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, 2007. december 6–7.
4. Jurafsky, M. Martin, J. H.: *Speech and Language Processing: an Introduction to Natural Language Processing, Speech Recognition, and Computational Phonology.* (2000) Prantice-Hall.
5. Kaplan, R., Kay, M.: Regular models of phonological rule systems. *Computational Linguistics*, 20:3 (1994) 331-379.
6. Kornai, A.: Frequency in morphology. In Kenesei I. (ed): *Approaches to Hungarian IV.* (1992) 246-268.
7. Koskenniemi, K.: Two-level morphology: A general computational model of word-form recognition and production. Tech. rep. Publication No. 11, Department of General Linguistics, University of Helsinki. (1983).
8. Trón V. et al.: morphdb.hu: magyar morfológiai nyelvtan és szótári adatbázis. In Alexin Zoltán, Csentes Dóra (szerk.). *III. Magyar Számítógépes Nyelvészeti Konferencia.*, SZTE, Szeged, 2005, p. 169-179.