

Kutatási beszámoló

Nagyszótáras folyamatos magyar nyelvű beszéd felismerés nyelvmodellezési kérdései

Tarján Balázs
tanszéki mérnök
BME-TMIT

2017.

1 Bevezető

A gépi beszédfelismerő rendszerek egyik alapvető építőeleme a nyelvi modell, mely az egyes szószorozatok valószínűségét képes megbecsülni, így segít az akusztikai paraméterek alapján felmerülő hipotézisek (szószorozatok) közül kiválasztani a legnagyobb valószínűségűt. Doktori kutatásom keretében ezen nyelvi modellek továbbfejlesztését tűztem ki célul, elsődlegesen annak érdekében, hogy pontosabb beszédfelismerő rendszereket hozhassunk létre magyar nyelven.

Féléves kutatómunkám során televíziós feliratozó rendszerek nyelvi modelljének továbbfejlesztésén dolgoztam. Az első probléma, mellyel foglalkoztam elsősorban sportközvetítések gépi feliratozása során merül fel, nevezetesen, hogy a tanítás során hiába törekszünk rá, nem lehet az összes sportolónevet előre beépíteni a rendszerbe, hanem azokat szükségszerűen az egyes sportesemények ismeretében adaptálni kell. Vizsgálataim során a tavaly megrendezett Labdarúgó-Európa-bajnokság televíziós felvételeinek feliratozását javítottam egy speciális módszerrel, mellyel egy általános, labdarúgó-mérkőzések kommentárjainak felismeréshez tanított nyelvi modell szótárát a sporteseményhez hangoltam [1].

A második probléma, amivel foglalkoztam a jelenleg használt ún. n-gram modellek leváltására irányul. Azt vizsgáltam, hogy ún. rekurrens neurális hálózatok használatával milyen mértékű perplexitás javulás érhető el különböző televíziós feliratozási feladatok esetén. Azt találtam, hogy mind általános, mind feladatspecifikus feliratozási korpusz (híradó, időjárásjelentés) esetén igen jelentős javulás érhető le a visszacsatolt neurális hálózatokkal különösen, ha az n-gram modellekkel együtt, interpolálva használjuk őket.

2 Szótár-adaptáció labdarúgó-mérkőzések feliratozásához

A közéleti- és hírműsorokon túl a legnagyobb közérdeklődésre a sportműsorok, azon belül is a labdarúgó-mérkőzések számítanak. Kutatásunk célja az volt, hogy ennél a műsortípusnál is ki tudjuk váltani a költséges és gyakran hiányos feliratot eredményező gépelést, automatikus beszédfelismerő rendszerre épülő megoldással. Több lehetséges módszert is vizsgáltunk, melyek segítségével egy már betanított beszédfelismerő modellbe utólag is elhelyezhetők személynevek, abban az esetben is, ha az eredeti modell nem tartalmazott rájuk mintát. Elsőként bemutatjuk a legegyszerűbb, környezetfüggetlen megoldásokat, melyeknél a nyelvi modellbe önálló szavakként helyezzük el az új személyneveket. Ezután megmutatjuk, hogy hogyan lehetséges az új személyneveket az előfordulásuk lehetséges kontextusa alapján környezetfüggő módon modellezni. Mindkét modellezési megközelítést kipróbáltuk alanyesetben álló és esetragokkal ellátott személynevekkel is. Kiértékeléshez egy a 2016-os labdarúgó-Európa-bajnokság alatt rögzített kétórás tesztfelvételt használtunk, melyen a szokásos szóhiba-arányon túl a különböző modellezési módszerekkel elérhető személynév-visszakeresés hatékonyságát (felidőzés, pontosság, F1) is mértük.

2.1 Kezdeti nyelvi modell

Kísérleteink kiindulási alapja egy általános sport témakör alapján tanított nyelvi modell volt, melyhez a tanítószövegeket a 2016-os labdarúgó-Európa-bajnokság előtt gyűjtöttük, így biztosítva a tanítás és tesztelés közötti függetlenséget.

2.1.1 Szöveges tanító-adatbázisok

Kísérleti rendszerünk nyelvi modelljének betanításához három, különböző forrásból származó szövegkorpuszt használtunk fel (lásd **1. táblázat**):

- *Sportműsor leiratok*: Az MTVA által rendelkezésünkre bocsátott sportműsorok közül válogatott adatbázis kézi leiratai. Az adatbázis a 2016-os EB előtt rendezett labdarúgó-mérkőzéseket is tartalmaz.
- *Sporthír feliratok*: Az MTVA csatornáin sugárzott sporthírekhez tartozó feliratok
- *Sport webkorpusz*: A kisebb méretű, de feladatspecifikus tanítószövegek mellett kiegészítő adatbázisként egy tematikus weboldalokról gyűjtött korpuszt is felhasználtunk a kísérleti rendszerben

1. táblázat: A szöveges tanító-adatbázisok statisztikai adatai

	Sportműsor leiratok	Sporthír feliratok	Sport webkorpusz	Σ
Token [millió szó]	0,380	0,668	32,2	33,2
Type [ezer szó]	38	66	798	818

2.1.2 Kezdeti nyelvi modell tanítása

A szöveggörpuszok előkészítése során eltávolítottuk a nem lexikai elemeket (pl. számok, írásjelek, rövidítések), meghatároztuk a mondathatárokat, majd statisztikai módszer segítségével átalakítottuk a mondatkezdő szavakat, oly módon, hogy csak a tulajdonnevek őrizték meg a nagy kezdőbetűs írásmódot. Ezután bizonyos nem lexikai elemeket átírtunk kiejtett alakjukra (pl. '3', *három*), ezzel segítve a kiejtési modell generálását. A normalizált tanítószövegek alapján minden korpuszon független, 3-gram nyelvi modellt tanítottunk az SRI nyelvi modellező eszköz segítségével [2]. A kezdeti sport feliratozó nyelvi modellje ezután úgy készült, hogy az egyes modelleket lineáris interpoláció segítségével a beszédfelismerési feladathoz adaptáltunk a tesztanyag paraméterhangolási célokra elkülönített részén.

2.2 Szótárbővítési módszerek

Kísérleteink célja az volt, hogy az általános sport témakörben tanított és az előző fejezetben bemutatott kezdeti nyelvi modell szótárát kibővítsük a labdarúgó EB közvetítésében legnagyobb valószínűséggel elhangzó személynevekkel. A fejezet első felében bemutatjuk a szótárbővítéshez használt névlista összeállításánál alkalmazott módszereket, majd a második felében a szótárbővítési eljárásokat ismertetjük.

2.2.1 Személynév-listák összeállítása

Az első lista, amit összeállítottunk a szótárbővítéshez használt személynevek alany-esetben álló alakjait tartalmazta. Ez 595 néven alapult (EB résztvevő játékosok, bírók és szövetségi kapitányok), azonban a teljes neveken túl minden vezetéknev is még egyszer a listára került, mivel a kommentátorok gyakran így hivatkoznak az egyes személyekre (összesen 1190 személynév).

Nem csak alanyesetű személyneveket szerettünk volna azonban visszaadni, így a listán szereplő neveket a korpuszban előforduló leggyakoribb hat főnévi esetraggal is elláttuk, ezek gyakorisági sorrendben a következők: részes-, tárgy-, eszközhatározói-, ablativus-, delativus- és allativus esetek. Méréseink szerint ezzel a hat esettel a korpuszban előforduló ragozott főnevek 85%-át tudjuk modellezni. Mivel a magyar nyelvben a ragok egy része – mint a fentebbi hat eset is – váltakozó, tehát magánhangzó-harmónia szerint illeszkednie kell a szótóhoz, a ragok generálása illesztés alapján történt. A szótóvek magánhangzó-harmónia osztályba sorolása szabályalapú módszerrel zajlott: a nevek fonetikus átírataiban szereplő magánhangzók alapján. Három osztályt különböztettünk meg: hátulképzett (-hoz),

előlképzett kerekített (-höz) és előlképzett kerekítetlen (-hez). A tárgyeset rag változatának kiválasztásakor nem csak a szótó magánhangzó-harmónia osztálya, hanem mássalhangzóra végződés esetén a mássalhangzó fonetikus tulajdonságai is szerepet játszottak. A folyamat végén tehát az alanyesetben álló neveken túl még további hat darab az 1190 név különböző főnévi eseteit tartalmazó lista állt rendelkezésünkre.

2.2.2 Környezetfüggetlen szótárbővítés

A környezetfüggetlen szótárbővítés lényege, hogy az új szavakat (jelen esetben személyneveket), környezetükből kiragadva, izoláltan helyezzük el a nyelvi modellben.

Elhelyezés a tanítószövegben: Az első szótárbővítési módszer, amit vizsgálunk, az a szélsőségesen egyszerű megoldás, amikor az új szavak listáját egyszerűen felsoroljuk a nyelvi modell tanítószövegében. Bár ezzel minden új szó nullánál nagyobb valószínűséget kap, valódi statisztikai tanítás nem történik, így az új szavak valószínűsége távol fog esni a valós gyakoriságtól. Ezt a módszert leginkább kíváncsiságból és csupán az alanyesetben álló névlistával próbáltuk ki.

Interpoláció unigram modellel: A második környezetfüggetlen szótárbővítési módszer lényege az, hogy az új szavak listájából építünk egy unigram nyelvi modellt uniform valószínűségekkel, és ezt a modellt interpoláljuk a kezdeti rendszerben bemutatott nyelvi modellel. Elsőre nagyon hasonlónak tűnik az előző megoldáshoz, de nagy előnye, hogy az új szavak nyelvi modelljéhez tartozó interpolációs súly segítségével hangolni tudjuk az új szavak valószínűségét.

Kétféle szótárbővített modellt készítettünk ezzel a módszerrel. Az első esetben az unigram modell az alanyesetben lévő 1190 szavas névlista alapján lett összeállítva. A második esetben nem csak az alanyesetű személyneveket, hanem az előző pontban bemutatott módszerrel generált egyéb ragozott alakokat is felvettük a modellbe, mely így összesen 8330 szótárelemmel bővült.

2.2.3 Környezettfüggő szótárbővítés

A környezetfüggő szótárbővítés lényegi gondolata, hogy azonosítjuk a tanítószövegben azokat a szavakat, melyek az új szótári elemekkel azonos szerepet töltenek be. Ezután az új szavak listáját elhelyezzük a nyelvi modellben minden olyan kontextusba, melyet relevánsnak azonosítottuk. Így a környezetfüggetlen modellel ellentétben az új szavak a valóságos használatukat jól közelítő kontextussal kerülhetnek be a felismerő hálózatba.

A feliratozó rendszer mögött dolgozó gépi beszédfelismerő súlyozott, véges állapotú átalakítókat (WFST) [3] használ a különböző szintű nyelvi információk integrációjához. Ez a technológia kiválóan alkalmas arra is, hogy a tanítószövegben elhelyezett címkéket a névlistákkal helyettesítsük.

Először a címkékkel ellátott kézi leiratokból osztály n-gram nyelvi modellt tanítottunk, melyben a személynevek helyén az esetrag nélküli vagy esetraggal ellátott címkék álltak. Ezután a címkéket tartalmazó nyelvi modellt interpoláltuk az összes korábban bemutatott nyelvi modellel. Az immáron személynév címkéket is tartalmazó interpolált nyelvi modellt WFST-vé alakítottuk és az ún. kompozíció művelettel [3] egyesítettük a névlistákból alkotott WFST-vel. Az így létrejött WFST már tartalmazta a megfelelő esetben álló személynevek listáját a címkék helyén. Ez a WFST képezi az alapját az új felismerő hálózatnak, mely a névlistákon található személyneveket a főnévi esetüknek megfelelő kontextusban képes modellezni.

2.3 Eredmények

Ez a fejezet kutatásunk eredményeit foglalja össze. Ismertetjük a kezdeti és szótárbővített modellek feliratozási és személynév-felismerési hatékonyságát, valamint ezek meghatározáshoz használt módszereket.

2.3.1 Kiértékelés

A feliratozó rendszer tesztelésére az MTVA által sugárzott 2016-os labdarúgó-Európa-bajnokság magyar nyelvű kommentárjait használtuk fel. Egy összesen két-órás adatbázist állítottunk össze különböző nemzetek által vívott mérkőzések véletlenszerűen válogatott részleteinek felhasználásával. Miután elkészítettük a pontos szöveges átiratot, az adatbázist két részre osztottuk: 29 percet félreraktunk a fejlesztés során felmerülő szabad paraméterek hangolásához, 94 percet pedig a végső kiértékeléshez használtunk.

A korábbi fejezetekben ismertetett modellekből minden esetben súlyozott, véges állapotú átalakítókat építettünk, melyekből a SpeechTex VOXerver [4] segítségével készítettük el a tesztanyag feliratait. A feliratozási pontosság megállapításához a beszéd felismerő rendszerek kiértékelésénél leggyakrabban alkalmazott metrikát a szóhiba-arányt használtuk. A személynevek felismerési hatékonyságának megállapításához először elkészítettük a vizsgált felirat és a referencia leirat szószintű összerendelését, majd ez alapján kiszámoltuk a nevek felismerésének pontosságát és felidézési arányát. Csak teljes egyezést fogadtunk el találatnak, azaz például egy hibás raggal felismert név egy fals pozitív és egy fals negatív hibát is eredményezett.

2. táblázat: A tesztadatbázis statisztikai adatai

Teszthalmaz	Hangfelvétel hossza [perc]	Szavak száma	Személynevek száma
Fejlesztő	29	2189	268
Kiértékelő	94	6975	989

2.3.2 Feliratozási eredmények

A kiértékelés első szakaszában a hagyományos beszéd felismerési metrikákat használtuk a különböző szótár-bővítési módszerek vizsgálatához. A perplexitást és a szótáron kívüli szavak arányát a kiértékelő teszhalmazon mértük, és mindkettő a nyelvi modellek szöveges illeszkedéséről szolgáltat információt. A szóhiba-arány a gépi felirat szövegének a referenciához mért hibaarányát mutatja meg.

3. táblázat: A nyelvi modellek és a feliratozás kiértékelése

Szótár-bővítő névlista	Nyelvi modell	Perplexitás [-]	Szótáron kívüli szavak aránya [%]	Szóhiba-arány [%]
-	Kezdeti modell (EB nevek kiejtése nélkül)	630	2.3	33.8
-	Kezdeti modell	630	2.3	32.3
EB névlista	Elhelyezés a tanítósz.-ben	580	2.3	31.3
EB névlista	Unigram interpoláció	534	2.3	29.9
EB névlista	Környezetfüggő m.	490	2.3	29.6
EB névlista +esetragok	Unigram interpoláció	606	1.7	30.5
EB névlista +esetragok	Környezetfüggő m.	513	1.7	29.3

A **3. táblázat** eredményeit vizsgálva azt láthatjuk, hogy az EB névlistákkal történő szótárbővítés önmagában nem csökkentette a szótáron kívüli szavak arányát, mivel az alanyesetű személynevek többségét már eleve tartalmazta a kezdeti modell. Lényeges csökkenést csak akkor látunk, amikor a ragozott személynevekkel is bővítjük a szótárat, tehát a ragozott alakokból már keveset tartalmazott az eredeti tanítószöveg. Ezzel szemben a **perplexitás határozottan csökken** ahogy egyre komplexebb szótárbővítési technikákat vezetünk be, ami arra utal, hogy ezek a technikák valóban nagyobb pontossággal becsülik a személynevek valószínűségét.

A feliratozás szóhiba-aránya három esetben csökkent számottevően. Az első csökkenés ahhoz köthető, amikor megadjuk az EB névlista szereplőinek kiejtését a modellben. Másodsorra akkor csökken, amikor az EB személyneveket elhelyezzük a tanítószövegben, majd ezután akkor, amikor unigram modell interpolációt is alkalmazunk. A környezetfüggő modellezés jelentős javulást ezen felül már nem hoz. Egészen addig igaz ez, míg nem követeljük meg a személynevek esetragjainak visszaadását. A ragozott személynevek visszaadásakor a **környezetfüggő modellezés egyértelmű előnyre tesz szert**. Feltehetően az az oka ennek, hogy a ragozott személynevekkel feltöltött listák akusztikailag könnyen összetéveszthető szavakat tartalmaznak, így ezek elhelyezésénél a kontextus ismerete felértékelődik.

2.3.3 Személynév-felismerési eredmények

A személynév-felismerési eredmények úgy készültek, hogy a tesztanyag összes személynevét kigyűjtöttük és információ-visszakeresési problémaként értelmezve a feladatot megvizsgáltuk, hogy milyen hatékonysággal adja vissza őket a gépi felirat. Fontos megjegyezni, hogy a tesztanyag névlistája természetesen eltér a modellbővítéskor használt névlistától, hiszen a tesztanyag nem tartalmazza az összes lehetséges EB személynevet, ellenben sok olyan nevet is tartalmaz, melyeket a bővítőlista nem.

4. táblázat: Személynév felismerési eredmények

Szótárbővítő névlista	Nyelvi modell	Pontosság [%]	Felidézési arány [%]	F1 [%]
-	Kezdeti modell (EB nevek kiejtése nélkül)	90.0	32.0	47.2
-	Kezdeti modell	91.4	39.8	55.5
EB névlista	Elhelyezés a tanítósz.-ben	91.1	44.6	59.9
EB névlista	Unigram interpoláció	90.8	51.9	66.0
EB névlista	Környezetfüggő m.	90.3	51.1	65.3
EB névlista +esetragok	Unigram interpoláció	89.3	48.1	62.6
EB névlista +esetragok	Környezetfüggő m.	91.6	52.7	66.9

A személynév felismerési eredményeket vizsgálva (**4. táblázat**) azt vehetjük észre, hogy a személynevek felismerési **pontossága minden módszer esetén 90% körül mozog**, attól csak kis mértékben tér el. Azaz a feliratozó rendszer csak kb. minden tizedik felismert személynevet helyettesít másik névvel, és még ezek többsége is a főnévi eset meghatározásához köthető hiba.

A felidézési arányok változatosabb képet mutatnak. Látható, hogy a kezdeti rendszer csak minden harmadik nevet ismer fel a tesztanyagban, míg a szótár bővítés után már minden másodikat. A felidézés esetén is elmondható, amit már szóhiba-aránynál is kifejtettünk: az EB személynevek kiejtésének megadása, a nevek elhelyezése a tanítószövegben, majd az unigram interpoláció mind érdemben növeli a detektált nevek arányát. A környezetfüggő szótár bővítés hatására az alanyesetű nevek felidézési aránya kicsit visszaesik az unigram interpolációhoz képest. Ahogy azonban a szóhiba-arány esetén is megfigyelhettük, ha az esetragokat is modellezzük, újra a környezetfüggő megoldás kerül előnybe. Összességében tehát az **esetragokkal ellátott, környezetfüggő modell a legjobban teljesítő megoldás**.

3 Nyelvmodellezési kísérletek rekurrens neurális hálózatokkal

A statisztikai nyelvi modellek a szószorozatok valószínűségét modellezik, melynek minél pontosabb ismerete elengedhetetlen a korszerű beszédfelismerő rendszerekben. A jelenlegi rendszerek döntő többsége ún. n-gram nyelvi modellt használ, mely n-1 korábbi szó alapján becsüli a következő szótári elem valószínűségét. Gyakorlatban a 3-gram és 4-gram modellek használata az elterjedt. Könnyű belátni azonban, hogy ilyen rövid előtörténet (tipikusan két korábbi szó) nem minden esetben elegendő. A magyar nyelvben nagyon gyakran fordul elő, hogy egy mondat egymással összefüggő elemei jóval távolabb kerülnek egymástól. Egy érdekes új kutatási irány képviselői az ún. visszacsatolt (rekurzív) neurális hálózatokat (Recurrent Neural Network - RNN) használják nyelvi modellezési célokra [5], [6]. Az RNN alkalmazásának óriási előnye, hogy szakít az n-gram-okkal és felépítésének köszönhetően nagyon hosszú függőségeket tud modellezni a szövegben.

3.1 Adatbázisok

Kísérleteinkhez az MTVA által rendelkezésünkre bocsátott feliratokat használtuk fel. Az egyes tanító és tesztelőkorpuszok méretét az 5. táblázatban mutatjuk be. Az általános és híradós korpuszt ugyanazon az általános teszt szövegen értékeltük ki.

5. táblázat: A szöveges tanító-adatbázisok mérete

Korpusz típusa	Híradó	Időjárás	Általános
Tanító [millió szó]	1	0,476	15
Teszt [ezer szó]	24	1,2	24

3.2 Tanítási módszerek

A nyelvi modellek tanításához három különböző modellező eszközt használtunk. A kiinduló, hagyományos n-gram modelleket az SRILM toolkit-tel [2] készítettük. A rekurrens neurális modellek tanításához az **rnnlm** [5] és az **rwthlm** [6] eszközöket használtuk. Az rekurrens nyelvi modellek esetén nem használható a tanítószöveg teljes szótára a tanítás során, hanem azt valamilyen módszerrel csökkentenünk kell. Az **rnnlm**-mel tanított nyelvi modellek esetén a beépített, gyakoriságalapú osztályozást használtuk, míg az **rwthlm** esetén Brown osztályozást [7] készítettünk a korpuszokon. Fontos különbség a két eszköz között, hogy míg az **rnnlm** hagyományos aktivációs függvényeket alkalmaz, addig az **rwthlm** ún. long short-term memory (LSTM) cellákat [8] használ, mely a korábbi tapasztalatok alapján pontosabb valószínűségbecslést tesz lehetővé.

3.3 Eredmények

A különböző nyelvi modellezési technikákkal kapható perplexitásjavulást 3 különböző tanítószóveggel mértük ki, mely kísérletek eredményeit a következőkben ismertetjük.

3.3.1 Perplexitás eredmények a híradós tanítószóvegen

A híradó felvételekhez készült feliratkorpusz alapján öt nyelvi modellt tanítottunk. Az eredményeket a 6. táblázatban foglaltuk össze. Mint látható a kezdeti, n-gram modelleket interpoláció nélkül csak az rwthlm-mel tanított LSTM RNN-ekkel sikerült túlszárnyalni. Az n-gram és rekurrens nyelvi modellek interpolációja minden esetben pontosabb modellt eredményezett, mint a kiinduló n-gram.

6. táblázat: Perplexitás eredmények a híradó tanítószóveg alapján tanított nyelvi modellekkel

Nyelvi modell típusa	Osztályok száma	Neuronok száma	Teszt perplexitás	
			Önmagában	Interpoláció 5-gram modellel
3-gram	–	–	760	–
4-gram	–	–	749	–
5-gram	–	–	748	–
rnnlm	300	100	793	599
rwthlm	2000	200	643	533

3.3.2 Perplexitás eredmények az időjárás-jelentés tanítószóvegen

Az időjárás tanítókorpuszon a korpusz kis méretéből kifolyólag csak az rnnlm toolkit-tel tanítottunk rekurrens hálózatot. A visszacsatolt neurális hálózat itt is lényegesen jobban működött, mint a kezdeti n-gram-alapú modellek. Eredményeket lásd az 7. táblázatban.

7. táblázat: Perplexitás eredmények az időjárás tanítószóveg alapján tanított nyelvi modellekkel

Nyelvi modell típusa	Osztályok száma	Neuronok száma	Teszt perplexitás	
			Önmagában	Interpoláció 5-gram modellel
3-gram	–	–	49	–
4-gram	–	–	47	–
5-gram	–	–	47	–
rnnlm	1000	200	41	36

3.3.3 Perplexitás eredmények az általános tanítószövegen

A különböző felirattípusok vegyítésével kaptuk az általános tanítószöveget, melyet ugyanazon a tesztszövegen értékeltünk ki, mint a híradós korpusz modelljeit. A három tanítószöveg közül ez a legnagyobb méretű, így az ezzel kapott eredmények adják vissza legpontosabban az egyes modellezési módszerek tulajdonságait. A **8. táblázat** alapján azt mondhatjuk, hogy mindkét rekurrens tanítóeszközzel sikerült szignifikánsan javítani a perplexitást. Bár meg kell jegyeznünk, hogy ehhez az rnnlm esetén magasabb neuronszámra volt szükség, melyet feltehetően az egyszerűbb aktivációs függvények használata okozott. A hagyományos n-gram és neurális nyelvi modellek interpolációjával itt is további jelentős perplexitás-csökkenés volt mérhető.

8. táblázat: Perplexitás eredmények az általános tanítószöveg alapján tanított nyelvi modellekkel

Nyelvi modell típusa	Osztályok száma	Neuronok száma	Teszt perplexitás	
			Önmagában	Interpoláció 5-gram modellel
3-gram	–	–	406	–
4-gram	–	–	386	–
5-gram	–	–	380	–
rnnlm	1000	100	510	328
rnnlm	500	300	365	284
rwthlm	2000	100	326	259

4 Összefoglalás

Féléves kutatómunkám során két kísérletsorozatot is végrehajtottam, melyek célja magyar nyelvű feliratozórendszerek nyelvi modelljének pontosságnövelése volt. Az első kísérletsorozatban egy labdarúgó-mérkőzések gépi úton történő, élő feliratozásához fejlesztett rendszert mutattam be. A rendszer kiértékeléséhez a 2016-os labdarúgó-Európa-bajnokság mérkőzéseinek magyar nyelvű kommentárjaiból állítottunk össze egy tesztadatbázist, melyen a felirat és a személynevek felismerésének pontosságát is mértük. Általános következtetésként azt vonhatjuk le, hogy a környezetfüggetlen szótárbővítés egy egyszerű és hatékony alternatíva abban az esetben, ha csak alanyesetben álló személyneveket akarunk visszaadni. Amennyiben azonban ragozott személyneveket szeretnénk látni a feliratban, a környezetfüggő modellek alkalmazása tűnik jobb választásnak.

A második kísérletsorozatban kézi feliratok korpuszából tanítottunk nyelvi modelleket hagyományos, n-gram-alapú módszerrel és rekurrens neurális hálózatok segítségével is. A kipróbált két neurális nyelvmodell-tanító eszköz közül az rwthlm bizonyult pontosabbnak elsősorban az aktivációs függvény helyén alkalmazott LSTM celláknak köszönhetően. Az n-gram és neurális nyelvmodellek interpolációjával további jelentős perplexitás-javulást értünk el. További kísérleteink során azt kívánjuk majd feltárni, hogy ez az itt bemutatott módszerekkel tanított neurális nyelvi modellek egy beszédfelismerő rendszerben alkalmazva milyen mértékű beszédfelismerési hibacsökkenést hoznak létre.

Végezetül fontosnak tartjuk megjegyezni, hogy minden a magyar nyelv minél szélesebb körű és jobb minőségű beszéd- és nyelvtechnológiai támogatottsága felé tett lépés kiemelt jelentőséggel bír, mivel hozzásegít minket kultúránk megőrzéséhez a digitális korban [9].

Hivatkozások

- [1] B. Tarján, L. Szabó, A. Balog, D. Halmos, T. Fegyó, and P. Mihajlik, “Élő labdarúgó-közvetítések gépi feliratozása,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia: MSZNY 2017*, 2017.
- [2] A. Stolcke, “SRILM – an extensible language modeling toolkit,” in *Proceedings International Conference on Spoken Language Processing*, 2002, pp. 901–904.
- [3] M. Mohri, F. Pereira, and M. Riley, “Weighted finite-state transducers in speech recognition,” *Comput. Speech Lang.*, vol. 16, no. 1, pp. 69–88, 2002.
- [4] B. Tarján, P. Mihajlik, A. Balog, and T. Fegyó, “Evaluation of lexical models for Hungarian Broadcast speech transcription and spoken term detection,” in *2nd International Conference on Cognitive Infocommunications (CogInfoCom)*, 2011, pp. 1–5.
- [5] T. Mikolov and S. Kombrink, “RNNLM-Recurrent neural network language modeling toolkit,” *Proc. 2011 ...*, 2011.
- [6] M. Sundermeyer, R. Schl, and H. Ney, “rwthlm – The RWTH Aachen University Neural Network Language Modeling Toolkit,” no. September, pp. 2093–2097, 2014.
- [7] P. F. Brown, P. V deSouza, R. L. Mercer, V. J. Della Pietra, and J. C. Lai, “Class-based N-gram Models of Natural Language,” *Comput. Linguist.*, vol. 18, no. 4, pp. 467–479, Dec. 1992.
- [8] S. Hochreiter and J. J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1–32, 1997.
- [9] E. Simon, P. Lendvai, G. Németh, G. Olaszy, and K. Vicsi, *A magyar nyelv a digitális korban - The Hungarian Language in the Digital Age*. Springer, 2012.