

Doktori (PhD) értekezés tézisei

**Prozódiai jellemzők gépi feldolgozása és
hasznosítása élőnyelvi korpuszok elemzésében**

Szekrényes István

Témavezető: Dr. Hunyadi László



DEBRECENI EGYETEM

Nyelvtudományok Doktori Iskola

Debrecen, 2019

1. Bevezetés

A prozódia a „beszéd zenei kifejezőeszközeiként” aposztrofált és a „kommunikáció nem nyelvi elemei” közé besorolt jelenségei - az idézett fordulatok pejoratív értelmezésében - könnyen kelthetik azt a benyomást, hogy az információközlés absztraktabb nyelvi szintjeihez képest a nyelvhasználat afféle addicionális, esztétika-retorikai kellékeivel állunk szemben a nyelvtudomány egy kimondottan perifériális területén. Ha tehát a nyelvi jel elsődleges, konstatív, logika tartalmára koncentrálunk, a beszéd zenei szerveződése tisztán formai adalék gyanánt (mint „színezet”) ignorálható. Néhány, a nyelvtudomány szempontjából belső és külső tényezőnek köszönhetően azonban a prozódia fönt vázolt státusza napjainkra jelentősen megváltozott.

Az egyik fontos, „belsőnek” nevezhető tényező a beszédaktus-elmélet (Austin, 1962), a pragmatika performatív fordulatként is emlegetett diszciplínájának megszületése, melynek keretében a kommunikáció alapegységeit képező megnyilatkozások nem pusztán mint proposícionális tartalmak hordozói, de mint ún. illokúciós erővel rendelkező performatív aktusok kerültek a vizsgálódás fókuszába. Az illokúciós erők azonosítása és tipologizálása során (az egyéb, főként kontextuális információk mellett) a prozódia szerepe és jelentősége is felértékelődik. Mindez nem jelenti azt, hogy a prozódiai ne játszana fontos szerepet a nyelvi szerkezetek formális szemantikában vizsgált konvencionális jelentésének meghatározásában is. Számos olyan példát találhatunk, ahol a szerkezeti többértelműségek (logikai operátorok hatóköre, fókusz és topik pozíció stb.) csak a szuprasegmentumok (itt elsősorban a hangsúlyra és az intonációra gondolunk) disztinktív szerepének játékba hozásával oldható fel. Habár a nyelvről alkotott generatív modellekben ez a modul a szintaxis-hoz képest háttérbe szorul, léteznek olyan megközelítések, melyek a prozódiai szerkezet függetlenségét és a logikai forma megfelelő értelmezésében betöltött szerepét kimondottan hangsúlyozzák (Jackendoff, 2002; Hunyadi, 2002). Ennek az a jelentősége, hogy a prozódia nem csupán a nyelvhasználat pragmatikai oldaláról, de a grammatika inherens részeként járulhat hozzá a felszíni szerkezet megformálásához. Nem kell meglepődnünk tehát, hogy a magyar nyelvre formális alapon kidolgozott fonológiai rendszerekben a hangsúly és az intonáció előkelő helyet kapott (Kornai & Kálmán, 1988; Varga, 1994).

„Külső” tényezőként könyvelhető el azoknak az új technológiáknak és módszereknek a megszületése és széles körű elterjedése, amelyek a beszéd akusztikai szerkezetét és szuprasegmentális alapelemeit (alapfrekvencia és intenzitás) számítógépes eszközökkel könnyen

mérhetővé és elemezhetővé teszik. A nyelvészek prozódiával kapcsolatos, többnyire introspekción alapuló megfigyelései és hipotézisei ezáltal természettudományos módszerekkel verifikálhatóvá, bővíthetővé, interdiszciplináris kutatások számára is könnyebben hozzáférhetővé, a nyelvtechnológiában pedig alkalmazhatóvá váltak. A modern kommunikációs technológiák fejlesztésében egyre markánsabb tendenciaként figyelhető meg a nyelvészeti és a kapcsolódó, interdiszciplináris alapkutatások eredményeinek felhasználása a kommunikációelmélet, az információtechnológia, a nyelvtudomány és a pszichológia képviselőinek hatékonyabb eljárások kialakítása érdekében folytatott közös munkájában.

A beszédtechnológiában már több esetben (Shriberg et al., 2000; Szaszák, 2010) nyert igazolást az a feltevés, hogy a prozódiai információk hasznos jellemzőként alkalmazhatók a beszéd szegmentálását és a szegmentumok feldolgozását célzó feladatokban (pl. beszéd- és érzelemfelismerés). A vonatkozó kutatások egyik fontos, leginkább távlati és alkalmazásorientált célja: az ember–gép kommunikáció hatékonyságának és természetességének növelése. Gördülékenyebben kezelhetővé tehető például a beszédfordulók kezelése, ha a felhasználó közléseinek interpretációjában, illetve válaszok generálásában a beszéd szuprasegmentális sajátosságai (beszéddallam, intenzitás, tempó stb.) is szerephez jutnak (Hirschberg, 2002).

Egy kevésbé feltérképezett, de a főntebb vázolt technológiai távlatok szempontjából sem érdektelen kutatási területhez sorolhatók azok a prozódia–pragmatika interfész vizsgálatára irányuló munkák (Nakajima & Allen, 1997; Zellers & Post, 2009), amelyek az élőbeszédbeli társas interakciók tematikus és prozódiai struktúrája között igyekeztek korrelációkat kimutatni. Nakajima & Allen (1997) kooperatív dialógusokban vizsgálták a témaszerkezet lehetséges prozódiai kísérőjegyeit, melynek során összefüggéseket állapítottak meg a megnyilatkozások kezdetén és végén mért alaphfrekvencia-értékek, illetve a környező megnyilatkozásokhoz fűződő propozicionális viszony (új téma bevezetése, új információ bevezetése a témán belül, előző megnyilatkozás kifejtése, téma folytatása) között. Az új társalgási témát bevezető megnyilatkozások elején az alaphfrekvencia magasabb és szélesebb terjedelemben változó értékeket mutatott, mint amit a téma folytatásaként később következő megnyilatkozásoknál mértek. Zellers & Post (2009) az alaphfrekvencián kívül egyéb prozódiai jellemzőket is vizsgálat alá vonva megállapították, hogy a tematikus szerkezettel való korrelációk feltárásához a jellemzők együttes vizsgálatára van szükség azok diszkrét elemzése helyett. A feltárt korrelációkból meglátásuk szerint így sem következik

biztosan, hogy bizonyos prozódiai minták egyértelmű proposícionális vagy pragmatikai vonatkozással bírnak. Baiat & Szekrényes (2012) egy esettanulmányban megállapította, hogy számukra ismeretlen (magyar, illetve perzsa) nyelven folyó dialógusok hallgatása közben a kísérleti alanyok meglepően jó arányban képesek érzékelni a társalgási téma megváltozását pusztán a dialógus dinamikai és prozódiai sajátosságaira hagyatkozva.

2. A kutatás célja

A jelen értekezés célja, hogy a fentebb bemutatott előzmények nyomvonalán haladva a beszélt nyelv kutatásához olyan újszerű módszertani eszközökkel és kísérletekkel járuljon hozzá, amelyek a beszéd prozódiai jelenségeit, mint a kommunikáció folyamatának önmagában is informatív akusztikai vetületét a percepció és a konverzációs analízis tükrében vizsgálják. A kutatás keretében megvalósított kísérletek háttérében az alábbi, intuitív hipotézis állt:

A prozódia az élőbeszédben nem csak a megnyilatkozások szintjén szerveződik, nem csak azok lehetséges szemantikai interpretációit modulálja, hanem az interakcióknak egyféle "rugalmas", kvázi zenei közeget biztosítva olyan önálló, bár a verbálishoz képest alulspecifikált, nem lexikális információtartalommal rendelkezik, ami a percepció számára az interakció értelmezésekor akkor is relevanciával bír, ha a konkrét verbális tartalom valamiért nem hozzáférhető.

Ez a prozodiában kifejeződő, alul-specifikált információtartalom akkor válik igazán érdekessé, ha a társas interakciók tényleges jellemzőinek és folyamatának rekonstruálásához adekvát kiindulópontot is jelent, ami alatt itt természetesen nem a konkrétumokat, a társalgás tárgyalási univerzumának pontos feltérképezését, hanem azokat a különböző intonációs mintákban, dinamikai változásokban, a beszélőváltás ritmusában tetten érhető, a *kommunikációs partnerek viszonyával, szerepével, a beszédhelyzettel* kapcsolatos globális tulajdonságokat, illetve a konverzáció folyamatának lokális eseményeit (pl. témaváltás és egyéb beszédaktusok) értem, amelyek bináris döntések formájában (formális vagy informális, megváltozott a téma vagy nem), a specifikus tartalom ismerete nélkül is kategorizálhatók.

A módszertan oldaláról a disszertáció legfontosabb célját egy egyszerűen algoritmizálható annotációs modell kidolgozása jelentette, ami a kutatás közvetlen kontextusát

adó HuComTech projekt (Hunyadi et al., 2012) egyéb elemzési sémái közé integrálva a beszélt nyelv prozódiai jelenségeinek lejegyzésével egészítik ki a korpuszban tárolt interakciók nem-verbális szerkezetét leíró adatokat. A kutatócsoport szempontjából ezek az információk a multimodális kommunikáció természetével kapcsolatos, az interakciókat az egyes modalitások együttműködése és formális alapszerkezete (Hunyadi, 2011) alapján karakterizáló vizsgálatok kiterjesztése miatt voltak érdekesek. Az automatikus prozódiai annotáció kialakításánál fontos szempontot jelentett, hogy kizárólag a hangfelvételek akusztikai elemzésére és a beszédfordulók a korpuszban manuálisan annotált szerkezetére támaszkodjon. Ezt nem csak az az esetleges tényező motiválta, hogy a fejlesztések idején a transzkripciók szótag- vagy tagmondatszintű szegmentálása nem volt teljes mértékben kivitelezett, de a megnyilatkozások prozódiai szerkezetének elemzését elméletsemlegességre törekedve eleve függetleníteni akartam minden olyan szintaktikai és fonológiai szemponttól, amelyek javarészt csak manuális elemzés révén reprodukálhatók, másrészt előfeltételezik a szupraszegmentumok egyéb nyelvi szinteknek való alárendeltségét. További fontos szempontot jelentett, hogy az annotáció kimenete a prozódia fizikailag mérhető változásait perceptuálisan is releváns, diszkrét kategóriákkal felcímkézhető kontúrok formájában reprezentálja. Mivel a fönti elvek alapján kidolgozott annotációs modell a prozódia deskriptív, nem nyelvspecifikus, kontúrelemző feldolgozását célozta meg, ez lehetőséget teremtett arra, hogy az algoritmus implementálása során a tanítóanyag elkészítését igénylő gépi tanulás helyett szabály-alapú megközelítéseket alkalmazzak.

A értekezés keretében megvalósított percepció, illetve neuronhálókval végzett kísérletek célja, hogy a prozódia társalgás-szintű, a verbális tartalomtól független informatívásával kapcsolatos alaphipotézist konkrét klasszifikációs feladatok megoldásával igazolja. A prozódiai annotáció kimenetén tanított neurális hálózatok esetén tehát nem az összes elérhető jellemzőt felhasználó és a maximális hatékonyságot garantáló eljárás kivitelezése, hanem annak a tesztelése jelentette az elsődleges szempontot, hogy a verbális szemantikai tartalom és a vizualitás kizárásával, pusztán prozódia információkra támaszkodva milyen eredmények érhetők el.

3. Felhasznált anyag és módszerek

A kutatás elsődleges adatforrását és tesztanyagát a 111 adatközlő részvételével készült, 50 órányi spontán beszédet tartalmazó, a *The Language Archive* és a *Meta-Share* projekteken keresztül kutatási célokra közzétett *HuComTech* multimodális előnyelvi korpusz¹ formális és informális dialógusokat rögzítő hangfelvételei, illetve a hozzájuk készített annotációk alkották.

Az elsőként kivitelezett percepciók kísérletekben Baiat & Szekrényes (2012) munkáját követve a *HuComTech* korpusz felvételeihez hasonló társalgási sémát követő, de perzsa nyelven folyó dialógusok hanganyagát, illetve ezek újrászintetizált, a beszéd szegmentális tartalmától megfosztott, csak a prozódiai szerkezetet megőrző verzióit használtam fel. A kísérletben részt vevő alanyok számára ismeretlen perzsa nyelv használatát a verbális tartalom kizárása indokolta. Az első kísérletben arra voltam kíváncsi, hogy a nyelv ismerete nélkül, pusztán a prozódiaira hagyatkozva milyen benyomások alakulnak ki bennük az interakciók előző szakaszban taglalt jellemzőiről. A második kísérlet közvetlenül a prozódiai szerkezet mentális reprezentációinak feltérképezését tűzte ki célul: mennyire tudják az alanyok élőbeszédben követni a hangmagasság, a beszédintenzitás és a beszédtempó változásait.

Egy későbbi kísérlet kidolgozásához Varga László a magyar hanglejtés fonológiai igényű leírására kidolgozott rendszerét (Varga, 1994) használtam fel. A kísérlet célja a rendszerben szereplő karakterdallamok disztinktív erejének mérésére volt: lehetséges-e a hanglejtésre hagyatkozva beazonosítani az eredeti kontextusból izolált megnyilatkozások attitűdtartalmát. A kísérletet három különböző kondícióban végeztem el: a megnyilatkozásokat rögzítő eredeti hangfelvételek, az alaphangfrekvencia-értékek alapján újrászintetizált szinuszhangok, illetve az előbbieket stilizált verziójának használatával.

A prozódia automatikus annotációjára kidolgozott, Praat program (Boersma & Weenink, 2016) beépített szkript nyelvén implementált alkalmazás fejlesztését leginkább Christophe d'Alessandro és Piet Mertens a tonális kontúrok pszichoakusztikai alapokon (Hart, 1976) történő stilizálására kidolgozott (d'Alessandro & Mertens, 1994), később *Prosogram* néven (Mertens, 2004) implementált módszere inspirálta, de az alaphang modulációinak kategorizálására használt módszereket tekintve sokat merített a *Tilt*² modell (Taylor,

¹<https://tla.nytud.hu>

²http://www.cstr.ed.ac.uk/projects/speech_tools/manual-1.2.0/c16909.htm

2000) paramétereiből is. Az előbbi eljárás lényege, hogy a dallam ténylegesen érzékelhető modulációit a beszéd intenzitása és bizonyos – a zenetudományból is kölcsönzött – küszöbértékek segítségével emeli ki a frekvenciaértékek csak géppel mérhető változásai közül. Az eredményül kapott reprezentáció az alaphangszint szótagszinten stilizált mozgásait (az eredeti adatsorok értékpárokra redukált vektorait) tartalmazza. Saját algoritmusom kialakítása során a szótagok szintjén túllépve a beszéddallam holisztikusabb, adott esetben akár több szótagot is egyetlen „dallameseményben” integráló leképezésére törekedtem, ahol az eredményül kapott annotációs címkék a hanglejtés formájáról (emelkedő, eső, ereszkedő stb.) és a beszélő egyedi hangterjedelmében elhelyezett relatív (kezdő és záró) hangmagasságáról adnak számot. A dallamformák kategorizálása során figyelembe vett küszöbértékek számításához a beszélők egyedi hangterjedelmét, átlagos hangmagasságát és hangmagasság-ingadozását vettem alapul, de fontos támpontot jelentettek a percepció kísérletek eredményei is.

Az alkalmazott módszereket kiterjesztettem az intenzitás és a beszédtempó hasonló vizsgálatára. Az algoritmus fejlesztése és tesztelése a HuComTech korpusz hangfelvételeinek és szöveges átiratainak felhasználásával történt, de bizonyos javítások és átdolgozások, melyeknek a program jelenlegi flexibilitását köszönheti, a *SegCor* projekt³ közreműködésével, a FOLK korpusz (Schmidt, 2016) német nyelvű, változatos kondíciók között (2–14 adatközlővel) készített hangfelvételeinek elemzése során valósultak meg. Az alkalmazás kimenete a Praat program által támogatott TextGrid, illetve XML formátumban érhető el, ami később az XML fájlok XSLT transzformációjával a dallamkontúrok SVG grafikaként történő vizualizációjával is kiegészült. Ezek a vizualizációk elsősorban az elemzés kimenetének szubjektív értékelésekor hasznosultak, mivel a hallás és a látás modalitásainak a percepcióban való összekapcsolására vonatkozó korábbi kutatásaink (Hunyadi et al., 2014, 2015) alapján várhattuk, hogy a két modalitás egy időben való értelmezése (kongruens vagy inkongruens), az automatikus annotáció vizuális kimenete könnyebben összevethető a hallható intonációval. A kísérletben néhány, a Debreceni Egyetem Néprajztudományi Tanszékének archívumából kölcsönzött hangfelvételt használtam fel.

A prozódiai annotáció kimenetét két mély neuronhálók végzett kísérletekben használtuk fel (Szekrényes & Kovács, 2017; Kovács & Szekrényes, 2019), amelyek célja a percepció vizsgálatára kidolgozott tesztekhez hasonló osztályzási feladatok automatizálása

³<http://www1.ids-mannheim.de/prag/muendlichekorpora/segcor.html>

volt. Az első kísérlet a HuComTech korpusz formális és informális dialógusainak megkülönböztetésére irányult, amelyet három jellemzőcsoport alapján végeztünk el: a beszélőváltás szekvenciái, nyers F0-adatok és prozódiai annotáció kimenetében felcímkézett hanglejtésformák.⁴ A második, szintén mély neuronhálókkal megvalósított kísérlet célja a HuComTech korpusz dialógusaiban elhangzó interjúkérdések motiválatlan témaváltásként történő lokalizálása volt. A neuronhálók tanításához itt is a beszélőváltás szekvenciáit, illetve a ProsoTool kimenetének beszélőnként normalizált, a beszédtempót és a beszédintenzitást is felhasználó numerikus változatát használtuk fel.

4. A kutatás eredményei

Az automatikus prozódiai annotációval kapcsolatos fejlesztések kezdeti szakaszának rész-eredményei a HuComTech korpusz feldolgozásának új lehetőségeként a VIII. Magyar Számítógépes Nyelvészeti Konferencián kaptak először nyilvánosságot (Szekrényes et al., 2011). A későbbi publikációk elsősorban a beszédhallam automatikus lejegyzésére szolgáló eljárás fejlesztésének teoretikus és módszertani hátterét (Szekrényes, 2014), illetve a konkrét technikai kivitelezés részleteit (Szekrényes, 2015) tárgyalják. Az alkalmazás aktuális verziója a beszélődiarizálás kimenetére épülő modulként integrálva az e-magyar projekt⁵ (Várad et al., 2017) keretében vált elérhetővé (Kornai & Szekrényes, 2017) a terület kutatói számára. A beszéd hangerőváltozásait és tempóját elemző algoritmusok jelenleg is fejlesztés alatt állnak. A prozódiai annotáció módszereire támaszkodó tudományos eredmények három könyvfejezet (Hunyadi et al., 2016a,b; Kovács & Szekrényes, 2019), két konferenciaközlemény (Kovács & Várad et al., 2017; Szekrényes & Kovács, 2017) és egy sikeresen megvédett doktori disszertáció (Abuczki, 2014) keretében kerültek publikálásra.

Az alaphékvencia feldolgozását végző algoritmus kimenetének szubjektív ellenőrzésére kidolgozott interaktív felületen eddig 9 személy nyújtott be értékeléseket, akik az F0 értékek stilizálása alapján generált vizualizációkat vetették össze a megnyilatkozások általuk érzékelt hanglejtésével. Az eddigi értékelések alapján a 25 megnyilatkozásra kapott automatikus elemzés átlagosan 82,9 százalékban volt összeegyeztethető a hallható hanggal. Az értékeléshez használt módszerek és eredmények további részletei a disszer-

⁴Azonos tesztanyag és kondíciók használatával a vizsgálat egy percepciók kísérlettel is kiegészült.

⁵<http://e-magyar.hu/hu>

táció **4.2.6.** fejezetében találhatók meg. Maga az interaktív felület, mint a stilizált alaphfrekvencia-kontúrok webes környezetben történő vizualizálására kidolgozott technológia szintén önálló, módszertani eredménynek tekinthető. Távlati terveim között szerepel, hogy *github* repozitóriumban jelenleg is elérhető Praat scriptek és XSL sablonok letöltése helyett a prozódiai címkék generálása és vizualizálása közvetlenül az e-magyar projekt webes felületén, egy tetszőleges hangfájl és a hozzá tartozó szöveges átirat feltöltése után kipróbálható legyen.

A perzsa nyelvű hangfelvételeken alapuló kísérletei esetében a stimulusok kis mennyisége, a függő változók kontrollálásával kapcsolatos problémák, illetve a feladat nehézsége miatt nehéz határozott, számszerűsíthető eredményeket felmutatni. A prozódiához köthető, alul-specifikált információtartalom relevanciájával kapcsolatos hipotézist sem megerősíteni, sem pedig cáfolni nem tudtam, ugyanakkor az adatok alapján megalapozottnak látom a tesztek gondosabban kontrollált kondíciók mellett történő folytatását. Továbbá a prozódiai jellemzők élőbeszédben való követésénél érdekes eredményként könyvelhető el, hogy habár a jellemzők valós időben történő követése – azok gyakori változásai miatt – az összes kondíció esetén rendkívül nehéznek bizonyult, az alanyoktól kapott jelzések (szintlépések) átlagos mennyisége az eredeti beszédet rögzítő felvételek hallgatása közben volt a legalacsonyabb. Ez a megfigyelt jelenség utalhat arra, hogy a kísérleti alanyokban az idegen nyelv használata ellenére is beindulhatott egy magasabb szintű feldolgozási folyamat, amely megnehezítette a prozódiai jellemzők fizikai változásaira való koncentrációt. Ha tehát a felvételeken izolálható beszédhangokat és szavakat hallottak, akkor a percepcióban ösztönös módon a prozódiai jellemzők is a beszélők megnyilatkozásainak és attitűdjeinek értelmezését segítő nyelvi, illetve kommunikatív funkciója felől nyerhettek megvilágítást, amely a fizikai viszonyokról kialakult képzetek átrendeződését, perceptuális prioritásának csökkenését is maga után vonta.

A Varga (1994) által tipologizált karakterdallamok disztinktív erejével kapcsolatos kísérlet eredményei azt igazolták, hogy az alaphfrekvencia-kontúr feldolgozására használt eljárások (simítás és stilizálás) a megnyilatkozások attitűdtartalmának értékelése szempontjából nem jelentenek szignifikáns információvesztést. A kísérleti alanyok ugyanolyan hatékonysággal tudták azonosítani a megnyilatkozásnak leginkább megfelelő kontextust, mint a nyers F0-adatok alapján generált stimulusok esetén. Az „eső” és „féleső” karakter elkülönítéséhez már kis számú minta esetén is fontos tényezőt jelentett a relatív

hangmagasság.

A mély neuronhálókkal végzett kísérletek egy nem várt, de fontos eredményének tekinthető, hogy a HuComTech korpusz formális és informális dialógusainak megkülönböztetése céljából végzett klasszifikáció pusztán a beszélőváltás szekvenciáinak ritmusa alapján is meglepően hatékony. A 108 dialógusrészletet tartalmazó teszhalmaz osztályozása a beszélőváltás alapján 81,5 százalékos pontossággal zárult, ami további 3,7 százalékot javult (ami 20% relatív hibaarány csökkenést jelent), ha a jellemzőkészletet a prozódiai annotáció címkéivel (dallamkontúrok és relatív hangmagasságszintek) is kiegészítettük. Ezzel szemben a nyers F0-adatok bevonása az eredmény 1.1 százalékkal történő romlásához vezetett. A motiválatlan témaváltások automatikus detektálására tett kísérletek a prozódiai jellemzők hasznosíthatósága szempontjából pozitív, a hasonló céllal és kondíciók között megvalósított nemzetközi kutatásokkal is összehasonlítható eredménnyel zárultak (lásd **6.2.** fejezet).

5. Az értekezés tézisei

A disszertáció keretében folytatott kutatás eddigi eredményei alapján levont következtetések az alábbi tézisek formájában foglalhatók össze:

1. Az előbeszéddel kapcsolatos percepció és információtartalom tekintetében releváns prozódiai változások szabály alapú algoritmusokkal való felcímkézése (stilizálás, relatív alaphangfrekvencia-viszonyok vizsgálata) nem csak lényegesen gyorsabb, mint a humán annotációt és gépi tanulást alkalmazó eljárások, de az eredmény minőségében is összemérhető velük.
2. Kimutattuk, hogy egy deskriptív, a kontúrok elemzésében fonológiai szempontokat nélkülöző, szabály alapú annotációs modell magyar nyelvű anyagon alkalmas az érzékelhető prozódiai változások számítógépes feldolgozására. Mivel a módszer nyelv-specifikus elemeket nem tartalmaz, lehetségesnek tartom, hogy más nyelvekre is alkalmazható.
3. Az alaphangfrekvencia-kontúr megfelelő stilizálása a percepció vonatkozásában nem eredményez szignifikáns információvesztést a különböző hanglejtés-minták által kifejezésre jutó attitűdtartalmak sikeres azonosításában.

4. A nyers akusztikai jellemzők szabály-alapú feldolgozása és kategorikus változók (hanglejtés-minták) formájában történő reprezentációja a neurális hálózatok tanulási idejét hatékonyan csökkenti és nem eredményez olyan információvesztést, ami az előszóban zajló társas interakciónak a beszédhelyzet (formális vagy informális) szempontjából történő osztályozását rontaná.
5. A beszélőváltás szünetekkel és átfedőbeszédekkel, illetve prozódiai információkkal is kiegészített szekvenciáinak ritmusa a verbális tartalom hiányában is hatékony jellemzőként használható fel a kommunikációs partnerek viszonyával, szerepével és a beszédhelyzettel kapcsolatos globális tulajdonságok meghatározásához.
6. A témaváltás verbális tartalomtól független, pusztán a prozódiai szerkezetre támaszkodó detektálása motiválatlan témaváltások esetén hatékonyan implementálható, ha mély neuronhálók tanításánál és az osztályzás során nem csupán az új témát bevezető megnyilatkozást vizsgáljuk (kezdeti F0-érték, hangterjedelem), hanem az azok környezetében előforduló hanglejtés szekvenciák tulajdonságait és mintázatait is.

Hivatkozott irodalom

- Abuczki, A. (2014). *A core/periphery approach to the functional spectrum of discourse markers in multimodal context* (PhD dissertation). Debreceni Egyetem, Nyelvtudományok Doktori Iskola.
- Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.
- Baiat, G. E., & Szekrényes, I. (2012, Dec). Topic change detection based on prosodic cues in unimodal setting. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (p. 527-530). doi: 10.1109/CogInfoCom.2012.6422037
- Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer [computer program]. version 6.0.22*. <http://www.praat.org/>. (retrieved 15 November 2016)
- d'Alessandro, C., & Mertens, P. (1994). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language*, 9(3), 257–288.
- Hart, J. t. (1976). Psychoacoustic backgrounds of pitch contour stylisation. *IPO-APR*, 11, 11-19.

- Hirschberg, J. (2002). Communication and prosody: functional aspects of prosody. *Speech Communication*, 36, 31–43.
- Hunyadi, L. (2002). *Hungarian sentence prosody and universal grammar*. Frankfurt: Peter Lang Verlag.
- Hunyadi, L. (2011). Multimodal human-computer interaction technologies. theoretical modeling and application in speech processing. *Argumentum*, 7, 240–260.
- Hunyadi, L., Földesi, A., Szekrényes, I., Staudt, A., Kiss, H., Abuczki, A., & Bódog, A. (2012). Az ember–gép kommunikáció elméleti–technológiai modellje és nyelvtechnológiai vonatkozásai. In *Általános nyelvészeti tanulmányok XXIV: Nyelvtechnológiai kutatások* (pp. 265–309). Budapest: Akadémiai Kiadó.
- Hunyadi, L., Kiss, H., & Szekrényes, I. (2016a, 01). Incompleteness and fragmentation: Possible formal cues to cognitive processes behind spoken utterances. In (p. 231–257). doi: 10.1007/978-3-319-21209-8_14
- Hunyadi, L., Kiss, H., & Szekrényes, I. (2016b). Prosody enhances cognitive infocommunication: Materials from the hucomtech corpus. In A. Esposito & C. L. Jain (Eds.), *Toward robotic socially believable behaving systems - volume I: Modeling emotions* (pp. 183–204). Cham: Springer International Publishing.
- Hunyadi, L., Szekrényes, I., Czap, L., & Sziklai, I. (2014). Seeing the sounds? *Argumentum*, 10, 325–338.
- Hunyadi, L., Szekrényes, I., & Sziklai, I. (2015). Vizuális percepció és nyelvi feldolgozás. *Beszédkutatás*, 23, 186–208.
- Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford University Press.
- Kornai, A., & Kálmán, L. (1988). Hungarian sentence intonation. In H. Hulst van Der & N. Smith (Eds.), *Autosegmental studies on pitch accent* (pp. 183–195). Dordrecht: Foris Publications.
- Kornai, A., & Szekrényes, I. (2017). e-magyar beszédarchívum. In A. Tanács & V. Vincze (Eds.), *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY2017)* (pp. 103–109). Szeged: JATEPress.

- Kovács, G., & Szekrényes, I. (2019). Applying neural network techniques for topic change detection in the hucomtech corpus. In L. Hunyadi & I. Szekrényes (Eds.), *The temporal structure of multimodal communication: Theory, methods and applications* (pp. 147–162). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-22895-8_8 doi: 10.1007/978-3-030-22895-8_8
- Kovács, G., & Váradi, T. (2017). A különböző modalitások hozzájárulásának vizsgálata a témairányítás eseteinek osztályozásához a hucomtech korpuszon. In A. Tanács & V. Vincze (Eds.), *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY2017)* (pp. 103–109). Szeged: JATEPress.
- Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Proceedings of speech prosody*.
- Nakajima, S., & Allen, J. F. (1997). Prosodic features of utterances in task-oriented dialogues. In *Computing PROSODY* (pp. 377–386). New York: Springer-Verlag.
- Schmidt, T. (2016). Good practices in the compilation of folk, the research and teaching corpus of spoken german. In J. M. Kirk & G. Andersen (Eds.), *Compilation, transcription, markup and annotation of spoken corpora, special issue of the international journal of corpus linguistics [ijcl 21:3]* (pp. 396–418).
- Shriberg, E., Stolcke, A., Hakkani-Tür, D., & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32, 127–154.
- Szaszák, G. (2010). A prozódia szerepe a beszédfelismerésben. In *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek* (pp. 390–392). Budapest: Akadémiai Kiadó.
- Szekrényes, I., & Kovács, G. (2017). Classification of formal and informal dialogues based on turn-taking and intonation using deep neural networks. In A. Karpov, R. Potapova, & I. Mporas (Eds.), *Speech and computer* (pp. 233–243). Cham: Springer International Publishing.
- Szekrényes, I. (2014). Annotation and interpretation of prosodic data in the hucomtech corpus for multimodal user interfaces. *Journal on Multimodal User Interfaces*, 8:(2), 143–150.

- Szekrényes, I. (2015). Prosotool, a method for automatic annotation of fundamental frequency. In *6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 291–296). New York: IEEE.
- Szekrényes, I., Csipkés, L., & Oravecz, C. (2011). A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei, automatikus prozódiai annotáció. In V. Vincze (Ed.), *VIII. magyar számítógépes nyelvészeti konferencia* (pp. 190–198). Szeged: JATEPress.
- Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America*, *107*(3), 1697–1714.
- Varga, L. (1994). A hanglejtés. In *Struktúrális magyar nyelvtan 2. Fonológia* (pp. 468–549). Budapest: Akadémia Kiadó.
- Várad, T., Simon, E., Sass, B., Gerőcs, M., Mittelholcz, I., Novák, A., ... Vincze, V. (2017). Az e-magyar digitális nyelvfeldolgozó rendszer. In V. Vincze (Ed.), *Xiii. magyar számítógépes nyelvészeti konferencia (mszny2017)* (pp. 49–60). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.
- Zellers, M., & Post, B. (2009). Fundamental frequency and other prosodic cues to topic structure. In *Proceedings of IDP* (pp. 377–386). Paris.



Nyilvántartási szám: DEENK/367/2019.PL
Tárgy: PhD Publikációs Lista

Jelölt: Szekrényes István
Neptun kód: SQGGTJ
Doktori Iskola: Nyelvtudományok Doktori Iskola
MTMT azonosító: 10038180

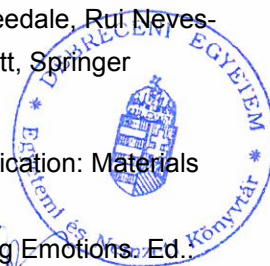
A PhD értekezés alapjául szolgáló közlemények

Magyar nyelvű könyvrészletek (2)

1. Hunyadi, L., Földesi, A., **Szekrényes, I.**, Staudt, A., Kiss, H., Abuczki, Á., Bódog, A.: Az ember-gép kommunikáció elméleti-technológiai modellje és nyelvtechnológiai vonatkozásai.
In: Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai kutatások. Szerk.: Kenesei I., Prószycki G., Várady T, Akadémiai Kiadó, Budapest, 265-309, 2012. ISBN: 9789630593083
2. **Szekrényes, I.**: Automatizált F0-mérések beágyazott mondat szerkezetek rekurziós sajátosságainak vizsgálatához.
In: Rekúzió a nyelvben I. Prozódiák megközelítés. Szerk.: Hunyadi László, Tinta Könyvkiadó, Budapest, 156-169, 2011, (Segédkönyvek a nyelvészet tanulmányozásához, ISSN 1419-6603 ; 123.) ISBN: 9789639902824

Idegen nyelvű, külföldi könyvrészletek (3)

3. Kovács, G., **Szekrényes, I.**: Applying Neural Network Techniques for Topic Change Detection in the HuComTech Corpus.
In: The Temporal Structure of Multimodal Communication: Theory, Methods and Applications. Ed.: Hunyadi László, Szekrényes István, Springer International Publishing Ag, Cham, 147-162, 2019. ISBN: 9783030228941
4. Hunyadi, L., Kiss, H., **Szekrényes, I.**: Incompleteness and Fragmentation: Possible Formal Cues to Cognitive Processes Behind Spoken Utterances.
In: Intelligent Decision Technology Support in Practice. Ed.: Jeffrey W. Tweedale, Rui Neves-Silva, Lakhmi C. Jain, Gloria Phillips-Wren, Junzo Watada, Robert J Howlett, Springer International Publishing Ag, Cham, 231-257, 2016. ISBN: 9783319212081
5. Hunyadi, L., **Szekrényes, I.**, Kiss, H.: Prosody Enhances Cognitive Infocommunication: Materials from the HuComTech Corpus.
In: Toward Robotic Socially Believable Behaving Systems. Vol. I : Modeling Emotions. Ed.: Anna Esposito, Lakhmi C. Jain, Springer International Publishing Ag, Cham, 183-204, 2016. ISBN: 9783319310565





Magyar nyelvű tudományos közlemények hazai folyóiratban (1)

6. Hunyadi, L., **Szekrényes, I.**, Sziklai, I.: Vizuális percepció és nyelvi feldolgozás.

Beszéd kutatás. 23, 186-208, 2015. ISSN: 1218-8727.

Idegen nyelvű tudományos közlemények hazai folyóiratban (2)

7. Hunyadi, L., **Szekrényes, I.**, Czap, L., Sziklai, I.: Seeing the sounds?

Argumentum (Debr.). 10, 325-338, 2014. EISSN: 1787-3606.

8. Pápay, K., Szeghalmy, S., **Szekrényes, I.**: HuComTech Multimodal Corpus Annotation.

Argumentum (Debr.). 7, 330-347, 2011. EISSN: 1787-3606.

Idegen nyelvű tudományos közlemények külföldi folyóiratban (2)

9. **Szekrényes, I.**: Post-processing T-patterns Using External Tools From a Mixed Method Perspective.

Front. Psychol. 10, 1-12, 2019. EISSN: 1664-1078.

DOI: <http://dx.doi.org/10.3389/fpsyg.2019.01680>

IF: 2.129 (2018)

10. **Szekrényes, I.**: Annotation and interpretation of prosodic data in the HuComTech corpus for multimodal user interfaces.

J. Multimodal User Interfaces. 8 (2), 143-150, 2014. ISSN: 1783-7677.

DOI: <http://dx.doi.org/10.1007/s12193-013-0140-1>

IF: 0.797

Magyar nyelvű konferencia közlemények (2)

11. Kornai, A., **Szekrényes, I.**: e-Magyar beszédarchívum.

In: XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017). Szerk.: Vincze Veronika, Szegedi Tudományegyetem, Szeged, 103-109, 2017. ISBN: 9789633065181

12. **Szekrényes, I.**, Csipkés, L., Oravecz, C.: A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei. Automatikus prozódiai annotáció.

In: VIII. Magyar Számítógépes Nyelvészeti Konferencia : MSZNY 2011. Szerk.: Tanács Attila, Vincze Veronika, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 190-198, 2011. ISBN: 9789633061213

Idegen nyelvű konferencia közlemények (6)

13. Hunyadi, L., Váradi, T., Kovács, G., **Szekrényes, I.**, Kiss, H., Takács, K.: Human-human, human-machine communication: on the HuComTech multimodal corpus.

In: Selected papers from the CLARIN Annual Conference 2018. Ed.: Inguna Skadina, Maria Eskevich, Linköping University Electronic Press, Linköping, 56-65, 2018, (Linköping Electronic Conference Proceedings, ISSN 1650-3686) ISBN: 9789176850343





14. **Szekrényes, I.**, Kovács, G.: Classification of Formal and Informal Dialogues Based on Turn-Taking and Intonation Using Deep Neural Networks.
In: Speech and Computer : 19th International Conference, SPECOM 2017: Proceedings. Ed.: Alexey Karpov, Rodmonga Potapova, Iosif Mporas, Springer International Publishing Ag, Cham, 233-243, 2017. ISBN: 9783319664286
15. Hunyadi, L., Váradi, T., **Szekrényes, I.**: Language technology tools and resources for the analysis of multimodal communication.
In: Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH). Ed.: Erhard Hinrichs, Marie Hinrichs, Thorsten Trippel, University of Tübingen, Tübingen, 117-124, 2016. ISBN: 9784879747082
16. **Szekrényes, I.**: ProsoTool, a method for automatic annotation of fundamental frequency.
In: 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). Ed.: Baranyi Péter, Csapó Ádám, IEEE, New York, 291-296, 2015.
17. Hunyadi, L., **Szekrényes, I.**, Borbély, A., Kiss, H.: Annotation of spoken syntax in relation to prosody and multimodal pragmatics.
In: 3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2012 : Proceedings, December 2-5, 2012 Košice, Slovakia. Ed.: Péter Baranyi, IEEE, Danvers, 537-541, 2012. ISBN: 9781467351881
18. Esfandiari Baiat, G., **Szekrényes, I.**: Topic change detection based on prosodic cues in unimodal setting.
In: 3rd IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2012 : Proceedings, December 2-5, 2012 Košice, Slovakia. Ed.: Péter Baranyi, IEEE, Danvers, 527-530, 2012. ISBN: 9781467351874

További közlemények

Idegen nyelvű, hazai könyvek (1)

19. Csűry, I., Hunyadi, L., Abuczki, Á., Esfandiari Baiat, G., Földesi, A., **Szekrényes, I.**: Elements of Electronic Information and Document Processing: An introduction to informatics (not only) for the humanities. Debrecen Univ. Press, Debrecen, 139 p., 2016. ISBN: 9789633185643

Magyar nyelvű tudományos közlemények hazai folyóiratban (2)

20. **Szekrényes, I.**: Egy apolitikusan felfogott élet esélyei.
Nagyerdei Almanach. 1, 1-8, 2010. EISSN: 2062-3305.
21. **Szekrényes, I.**: A király meztelen: avagy Egy boldogtalan kritikus "apológiája".
Forrás. 39 (7-8), 161-167, 2007. ISSN: 0133-056X.





Idegen nyelvű konferencia közlemények (1)

22. Hunyadi, L., Bertók, K., Németh, T. E., **Szekrényes, I.**, Abuczki, Á., Nagy, G., Nagy, N., Németi, P., Bódog, A.: The outlines of a theory and technology of human-computer interaction as represented in the model of the HuComTech project.
In: Proceedings of 2nd Cognitive Infocommunications International Conference. Ed.: by Baranyi Péter, IEEE Computer Society, New York, 249-255, 2011. ISBN: 9781457718069

A közlő folyóiratok összesített impakt faktora: 2,926

**A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre):
2,926**

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudománymetriai ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2019.11.06.

