

Hangok, hangulatok, gesztusok: magyar nyelvű dialógusok multimodális vizsgálata

Hunyadi László¹, Szekrényes István²

¹ Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék
hunyadi@unideb.hu

² Debreceni Egyetem, Filozófia Intézet
szekrenyes.istvan@arts.unideb.hu

1. Bevezetés

Egy, a Nyelvtudományi Intézettel, közelebbről annak számítógépes nyelvészeti osztályával és még közelebbről Váradi Tamással való évtizedes együttműködés eredménye az a HuComTech korpusz építésében és vizsgálatában megtestesülő, majd további irányokba vezető, 2009-ben indult kutatássorozat, amelynek elsődleges célja a multimodális kommunikáció vizsgálata elméleti alapjainak (Hunyadi, 2011; Németh, 2011) lefektetése volt. E munka sokszínűségét és az együttműködések sokféle szintjét jellemzi, hogy a Debreceni Egyetemen bölcsész-, informatikai és mérnöki karain és az MTA Nyelvtudományi Intézetén kívül részt vettek benne még a BME, az MTA TTK Pszichológiai Kutatóintézet, a Szegedi Tudományegyetem és a National Instruments Hungary szakemberei, sőt, egyes, mindezzel érintkező további kutatásokban-fejlesztésekben a Debreceni Egyetem orvosai és a Miskolci Egyetem mérnökei is.

A kutatás-fejlesztés célját természetesen jelentősen meghatározta, hogy a hazai (és nemzetközi) innovációs igényeknek az egyre hangsúlyozottabb előtérbe kerülésével nyilvánvalóvá vált, hogy bizonyos konkrét fejlesztésekhez egymástól látszólag távol eső szakterületek, tudományterületek, sőt szakmák együttes munkájára van szükség, határait egymáshoz kell közelíteni és átjárhatóvá tenni. Ehhez jó kiindulásnak bizonyultak a számítógépes nyelvészek, akik már addig is folyamatos dialógust folytattak szoftverfejlesztőkkel és mérnökökkel, de ugyancsak a mérnökök, akik beszédfeldolgozó algoritmusokon dolgoztak, vagy megsejtették, hogy az általuk épített robotoktól emberszerűbb viselkedést várnak a felhasználók. A pszichológusok is lelkesen csatlakoztak, hiszen az érzelmek, szándékok kutatása, beleértve azok felismerését és

szerepük tanulmányozását a kommunikáció sikerében vagy sikertelenségében, fontos elméleti és gyakorlati jelentőséggel bír számukra is. A nyelvész-pragmatikusokat mindenekelőtt a konverzáció folyamatának nyelvi vetülete érdekelte, kiegészítve mindezt a gesztusok és azok funkcióinak a tanulmányozásával, így multimodálissá szélesítve az addigi hagyományok verbálisközpontúságát. A fonetikusoknak lehetőségük nyílt arra, hogy informatikusokkal karöltve új algoritmusokat dolgozzanak ki a prozódia által közvetített tartalom felismerésére, egyebek között a gépi tanulás módszereivel feltárva a multimodális jelek közötti jellemző funkcionális összefüggéseket. A szintaxis kutatói is új lehetőségekhez juthattak azáltal, hogy a korpusz automatikus mondattani elemzése által első ízben kaptak lehetőséget a beszélt nyelv szintaxisának az eddigieknél átfogóbb igényű és terjedelmű megragadására. A bizonyos részleteiben egyre szélesedő kutatási palettán végül megjelentek fül-orr-gége szakorvosok is, akik a siketekkel való kommunikáció lehetőségeinek a bővítését tűzték ki célul a beszédakusztika elérhetővé tételével nem hallók számára, valamint annak az artikulációval való újszerű összekapcsolásával (Hunyadi és mtsai., 2015).

2. Az adatgyűjtés módszerei

Több hónap tervezés és előkészítő tevékenység után a kutatássorozat empirikus forrását jelentő HuComTech korpusz (Hunyadi és mtsai., 2012, 2016c) hang- és videóanyagát 2010 tavaszán, a Debreceni Egyetem Angol–Amerikai Intézetének stúdiójában, 111 (54 nő és 67 férfi, átlagéletkor: 22 év), főként egyetemista korú beszélő közreműködésével készítettük el. A mintegy 50 órát kitevő felvételanyagon 222 interjúbeszélgetést rögzítettünk, amelyek részben az adatközlőkkel készített szimulált állásinterjúkból, illetve az ezeket követő informális beszélgetésekből tevődtek össze. Az utóbbi esetében az interjúvezető egy előre kidolgozott kérdéssor segítségével, eltérő érzelmi töltetű reakciók kiprovokálásával (pl. „Kérlek, mesélj egy negatív élményről, amit mostanában átéltél!”) adott keretet a dialógusnak. A beszélgetések túlnyomó többségét ugyanazon személy vezette. Ennek előnye, hogy – egy esetleges további kutatás céljából – adott a lehetőség egyebek között a beszélgető partnerhez való sokféle alkalmazkodás vizsgálatához is.

A résztvevők ülő helyzetben történő beszélgetésének hanganyagát 2 darab Shure 16A típusú mérőmikrofon segítségével, 44 100 Hz-es mintavételezési frekvencia és 16 bites kvantálás mellett 2 csatornán rögzítettük, az annotáláshoz és az akusztikai elemzéshez később a felvételek egy

csatornára mixelt verzióját használtuk fel. Az interjúk képanyagát 3 pozícióból (2 kamerát irányítottunk az adatközlőre, egyet pedig az interjúvezetőre) nagy felbontásban vettük fel, 3 darab Sony HDRXR520VE típusú, statikus állványokra helyezett kamera használatával. A felvételeken a beszélők térdtől felfelé láthatóak.

A hangfelvételek elemzéséhez a *Praat* program (Boersma és Weenink, 2020) annotációs funkcióját használtuk, amely egy szöveges formátumú, más beszédtechnológiai platformok által is könnyen importálható és feldolgozható kimenetet produkál. A videófelvelelek annotálásához a DE ITK Képfeldolgozó Csoportja *QANNOT* néven fejlesztett egy saját alkalmazást (Pápay és mtsai., 2011), amely lehetővé tette a felvételek képkockáról képkockára történő, gördülékeny címkézését. A program az elemzéshez használt kategóriákat és a választható értékek hierarchikus szerkezetét egy külső XML-állományból dinamikusan olvasta be, amelynek elkészítése, illetve más annotációs feladatokra való átdolgozása, majd később a címkéket rendszerező relációs adatbázis struktúrájának kialakítása megkívánta a bölcsész kollégákkal való folyamatos konzultációt és egy közösen értelmezhető terminológia kialakítását.

A korpusz felvételeinek alapszintű annotálása mintegy két évet és egy tucatnyi annotátor együttes munkáját vette igénybe, ami magában foglalta a beszéd és a speciális beszédesemények (hezitáció, nevetés, levegővétel stb.) standard jelölésekkel történő leiratozását, az érzelmek, a fordulóváltások és a nonverbális gesztusok címkézését (Pápay és mtsai., 2011). Mindez később (további 6 év munka után) kiegészült a teljes szöveg fonetikai, morfológiai és szintaktikai leírásával, a dialógusok pragmatikai elemzésével és a prozódia automatikus annotálásával is.

Az automatikus morfológiai és szintaktikai elemzéshez a Szegedi Tudományegyetemen fejlesztett *magyarlanc* (Zsibrita és mtsai., 2013) alkalmazás kimeneteit használtuk fel. A korpusz teljes anyagát lefedő, speciális kódolási sémát alkalmazó manuális elemzés pedig Kiss Hermina munkájának köszönhető (Kiss, 2014). A multimodalitás mint alapvető szempont érvényesítése érdekében a CLARIN-D projekt WebMAUS (Kisler és mtsai., 2017) szolgáltatásával elkészítettük a korábban csak a megnyilatkozások és a tagmondatok szintjén szegmentált szöveg szószintű időillesztését, amivel lehetővé válik az egyes szavak, kifejezések vagy mondatok más, akár nem nyelvi attribútumokhoz (pl. gesztusokhoz, pragmatikai funkciókhoz), valamint a prozódiahoz való illesztése is.

A nem verbális kommunikációs szintek közül annotáltuk az arc, a tekintet, a felsőtest, a fej és a kéz mozgásait, ezekhez fizikai jellemzőket

(pl. mozgás vagy változás iránya) illetve, de ugyancsak hozzáadva az érzelmi és pragmatikai attribútumokat is. A megfigyelő által értelmezett érzelmeket annotáltuk multimodálisan a hang és a videó együttes érzékelésével és unimodálisan is, egyedül a hang alapján. A sokrétű pragmatikai annotálásból, amely magában foglalt minden lényeges és hagyományos, szövegalapú jellemzőt (beszédváltás, különböző beszédaktusok, új és régi információ) újdonságként kiemeljük a beszélés elkezdésének (videóban és/vagy hangban érzékelhető) szándékát, ami nem feltétlenül esik egybe a beszélés valóságos kezdetével.

A beszéddallam automatikus elemzését egy saját fejlesztésű, a *Praat* program szkriptnyelvén implementált algoritmus (Szekrényes, 2014, 2015) segítségével végeztük el. A fejlesztés során arra törekedtünk, hogy az intonáció perceptuálisan releváns változásait az alaphangfrekvencia-görbe nagyobb dallamtrendekre történő stilizálásával, szegmentálásával és a beszélő egyéni sajátosságaihoz adaptált kategorikus címkézésével ragadjuk meg. A később XML-formátummal és vizuális megjelenítésre alkalmas XSL-stíluslapokkal is kiegészített, eredetileg *Praat TextGrid* formátumú kimenet a mért értékek mellett számot ad a dallamszegmentumok különböző karakteréről (pl. „emelkedő”, „eső”, „szinttartó”), illetve a beszélő 5 tartományra felosztott hangterjedelmében elfoglalt relatív pozíciójáról. A módszert később kiterjesztettük az intenzitás és a beszédtempó hasonló céllal történő vizsgálatára is. A beszéddallam elemzését végző eljárás később az *e-magyar* projekt (Váradi és mtsai., 2017) keretében, *emPros*¹ néven vált részévé egy nyílt forráskódú megoldásokat adoptáló beszédelemző lánc moduljainak (Kornai és Szekrényes, 2017). Itt az *e-magyar* projektet vezető Váradi Tamás és a beszédfeldolgozó alprojektet irányító Kornai András érdeméért kell kiemelnünk, hogy a korpuszban tárolt adatok mellett egy, addig csak belső használatra szánt automatikus eljárás is publikusan elérhető vált az érdeklődő szakmai közönség számára. Az algoritmus flexibilitásának javításában előzetesen nagy segítséget jelentettek a *SegCor* projekt² munkatársai is, akik lehetővé tették a FOLK korpusz (Schmidt, 2016) hangfelvételein való tesztelést.

¹ <http://e-magyar.hu/hu/speechmodules/empros>

² <https://segcor.cnrs.fr/>

3. A korpusz közzététele

A több millió címke lejegyzése önmagában korlátozott jelentőséggel bír, ha – mivel a kommunikáció alapvető tulajdonsága, hogy időben zajlik – a címkékben és kapcsolataikban hordozott információ nem kereshető vissza és nem elemezhető más címkék jelenlétének/hiányának időbeli függvényében. Az első adatbázis, amelyet az adatok elemzéséhez építettünk, SQL-alapú volt, amely így lehetővé tette standard SQL-lekérdezések alkalmazását. Az adatelemzésnek ez a módszere azonban még a lekérdezéshez készített grafikus interfésszel is megkövetelte a felhasználotól az adatbázis struktúrájának, a mögöttes technikai megoldásoknak a pontos ismeretét, ezért csak a szűkebb kutatócsoporton belül tudtuk hasznosítani. A tágabb kutatóközönség kiszolgálásához más módszerek alkalmazására, a metaadatok standard formában történő rögzítésére volt szükség.

Várad Tamás a CESAR (Várad, 2012) és a CLARIN projekt hazai koordinátoraként, később a HunCLARIN megalapítójaként szerzett szakmai tapasztalatai és javaslatai a kutatómunka ezen fázisában hatalmas segítséget jelentettek. A Nyelvtudományi Intézet a CESAR projekt keretében folyó újabb együttműködés keretében a korpusz metaadatait és XML-formátumba konvertált állományait először a META-SHARE online felületén tettük közzé. Később a *The Language Archive* projekt³ (a továbbiakban: TLA) által preferált IMDI metaadatsémára (Broeder és Wittenburg, 2006) áttérve az ARBIL program (Withers, 2012) és a LAMUS (Broeder és mtsai., 2006) rendszer használatával minden adatot elérhetővé tettünk a TLA nyílt gyűjteményében, ahol bárki hozzáférhet a korpuszhoz a TROVA kereső⁴ és az ANNEX (Berck és Russel, 2006) lekérdezőfelületen keresztül (a médiaanyagokhoz előzetes engedély alapján). Fontos továbblépés volt, amikor a korpuszban használt annotációs sémákról készített részletes útmutató⁵ elkészítése után a TLA HuComTech teljes anyagát tükröztük a Nyelvtudományi Intézet szerverén,⁶ a hazai adatbázisok gyűjteményében is, ezzel hozzájárulva a magyarországi nyelvészeti kutatások-fejlesztések széleskörű bemutatásához. Ezen adatokat már eddig is számos munka, köztük eddig két megvédett PhD-értekezés (Abuczki, 2014; Szekrényes, 2020) használta fel, valamint beszédtechnológiai fejlesztéshez is alkalmazták.

³ <https://archive.mpi.nl/tla/>

⁴ <http://tla.mpi.nl/tools/tla>

⁵ <https://tla.nytud.hu/info/hucomtech/guide.html>

⁶ <https://tla.nytud.hu>

A hozzáférés továbbra is biztosított szerteágazó kutatások jövőbeli specifikus céljaira. Az adatok elemzésére az ANNEX kereten kívül alkalmas a közismert, szabad hozzáférésű ELAN (Wittenburg és mtsai., 2006) szoftver is, amely az adatfájlokat saját gépre letöltve ugyancsak kényelmes elemzőeszköznek bizonyul.

4. Az adatok elemzése

A korpusz adatait a deskriptív, a különböző modalitások alá tartozó címkék gyakoriságát és együttállásait vizsgáló statisztikák mellett a Nyelvtudományi Intézet és az MTA-SZTE Mesterséges Intelligencia Kutatócsoport bevonásával gépi tanulással végzett kísérletekhez is felhasználtuk. Ezek egy része az interjúkban jelölt témaváltások automatikus, szövegfüggetlen detektálására irányult (Kovács és Váradi, 2017; Kovács és Szekrényes 2019), amit többféle, az annotálás során használt elemzési szintek (a videón megfigyelt nonverbális gesztusok, a megnyilatkozások prozódiai és szintaktikai szerkezete) címkéit összefogó jellemzőcsoport alapján is kipróbáltunk. Egy másik kísérlet az interjúk formális és informális felvételekre történő osztályozását célozta, amelyhez kizárólag a prozódia és a beszélőváltások ritmusát reprezentáló jellemzőkre hagyatkoztunk (Szekrényes és Kovács, 2017). Ezeknek a kutatásoknak és fejlesztéseknek a célja elsősorban egyik esetben sem az adott feladatra maximális hatékonyságot garantáló eljárás kivitelezése volt, hanem az egyes modalitások együttműködésére, informativitására vonatkozó hipotéziseinknek az ellenőrzése.

Az adatok elemzésében jelentős előrelépésnek számít, hogy csatlakoztunk a MASI nemzetközi hálózathoz (Multimodal Analysis of Social Interactions) és használóivá váltunk a *Theme* szoftvernek (Magnusson, 2000). E kifejezetten a viselkedés időbeli multimodális mintázatainak a feltáráására létrehozott szoftver válasz más (így idősoros) elemzési módszerek azon korlátozottságára, hogy azokkal szemben képes azonosítani olyan viselkedési mintázatokat is, amelyeknek az egyes elemei csupán opcionálisak, és időbeli jellemzőik (kezdet, vég, időtartam) sem állandóak. A *Theme* szoftver mint kutatási keretrendszer segítségével így számos olyan viselkedési mintázatot sikerült feltárnunk, amelyek – a mintázatok egyes összetevő elemeinek opcionálitása és a figyelembe vett események közötti idő variabilitása miatt – jobbra észrevétlenek maradnának. Így a korpusz egy részkorpuszán leírtuk az élőbeszéd töredezettségének szintaktikai jellemzőit (Hunyadi és mtsai., 2016a), a prozódia és a beszélt szintaxis összefüggéseit (Hunyadi és mtsai., 2016b), az *egyértés/egyét*

nem értés (Hunyadi, 2019), valamint az *öröm* mint kommunikációs esemény multimodális mintázatait (Hunyadi, 2020). A *Theme* alapján kapott mintázatoknak önállóan, valamint az ELAN annotációs és feldolgozó rendszerben való további vizsgálatára egy SQL-alapú, de könnyen használható webes felületet⁷ is létrehoztunk és nyílt felhasználásúvá tettünk (Szekrényes, 2019). A korpusz, köszönhetően komplexitásának és méreteinek, valamint elérhetőségének, remélhetően még további sokrétű és multidiszciplináris vizsgálatok gazdag lehetőségét fogja nyújtani.

5. Összegzés

A Nyelvtudományi Intézet másutt aligha tapasztalt értékes módon járul hozzá e kutatások kiszélesítéséhez, a HuComTech korpusz adatainak kivételes léptékű feldolgozásához és a módszer szélesebb körökben való elterjesztéséhez: amellett, hogy kezdeményez és helyt ad meghívásoknak, konzultációknak, szakemberek cseréjének, az Intézet a gazdája annak a virtuális számítógéprendszernek is, amely komoly kapacitásával a felhőben végzi adataink feldolgozását.

Bibliográfia

- Boersma, Paul, Weenink, David: Praat: doing phonetics by computer [Computer program]. Version 6.1.36, retrieved 6 December 2020 from <http://www.praat.org/>
- Broeder, D., Claus, A., Offenga, F., Skiba, R., Trilsbeek, P., Wittenburg, P.: LAMUS: The language archive management and upload system. In: Proceedings of LREC 2006. pp. 2291–2294 (2006)
- Broeder, D., Wittenburg, P.: The IMDI metadata framework, its current application and future direction. *International Journal of Metadata, Semantics and Ontologies* 1/2. pp. 119–132 (2006)
- Hunyadi, L.: Multimodal human-computer interaction technologies. theoretical modeling and application in speech processing. *Argumentum* 7. pp. 240–260 (2011)
- Hunyadi, L., Földesi, A., Szekrényes, I., Staudt, A., Kiss, H., Abuczki, A., Bódog, A.: Az ember–gép kommunikáció elméleti-technológiai modellje és nyelvtechnológiai vonatkozásai. In: Általános nyelvészeti tanulmányok XXIV: Nyelvtechnológiai kutatások. pp. 265–309. Akadémiai Kiadó, Budapest (2012)
- Hunyadi, L., Kiss, H., Szekrényes, I.: Incompleteness and fragmentation: Possible formal cues to cognitive processes behind spoken utterances. In: Jeffrey W. Tweedale, Rui, Neves-Silva, Lakhmi C. Jain, Gloria, Phillips-Wren, Junzo Watada, Robert J. Howlett (szerk.) *Intelligent Decision Technology Support in Practice*. pp. 231–257. Springer International Publishing, Cham (2016a)

⁷ <https://altnyelv.unideb.hu/ThemeToMySQL/login.php>

- Hunyadi, L., Kiss, H., Szekrényes, I.: Prosody enhances cognitive infocommunication: Materials from the hucomtech corpus. In Esposito, A., Jain, C. L. (eds.) *Toward robotic socially believable behaving systems – volume I: Modeling emotions*. pp. 183–204. Springer International Publishing, Cham (2016b)
- Hunyadi, L., Váradi, T., Szekrényes, I.: Language technology tools and resources for the analysis of multimodal communication, In: Erhard Hinrichs, Marie Hinrichs, Thorsten Trippel (eds.) *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH 2016)*. 117–124. University of Tübingen, Tübingen (2016c)
- Hunyadi, L., Szekrényes, I., Sziklai, I.: Vizuális percepció és nyelvi feldolgozás. *Beszédkiutató* 23, 186–208 (2015)
- Hunyadi, L.: Agreeing/Disagreeing in a Dialogue: Multimodal Patterns of Its Expression. *Frontiers in Psychology* 10, 1–9 (2019)
- Hunyadi, L.: Happy hour: the multimodal analysis of ‘being happy’ in a conversation (2020, kézirat)
- Kisler, T., Reichel U. D., Schiel F.: Multilingual processing of speech via web services, *Computer Speech & Language* 45, pp. 326–347 (2017)
- Kiss, H.: A HuComTech audio adatbázis szintaktikai szintjének multimodális vizsgálata. In: Tanács, A., Varga, V., Vincze, V. (szerk.) *X. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2014)* pp. 27–38. Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged. (2014)
- Kornai, A., Szekrényes, I.: e-Magyar beszédarchívum. In: Tanács, A. Vincze, V. (szerk.) *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY 2017)*. pp. 103–109. JATEPress, Szeged (2017)
- Kovács, G., Váradi, T.: A különböző modalitások hozzájárulásának vizsgálata a témairányítás eseteinek osztályozásához a hucomtech korpuszon. In: Tanács, A., Vincze, V. (szerk.) *XIII. magyar számítógépes nyelvészeti konferencia (MSZNY 2017)* pp. 103–109. JATEPress, Szeged (2017)
- Kovács, Gy.: Classification of Formal and Informal Dialogues Based on Emotion Recognition Features. In: Sojka, P.; Horák, A.; Kopeček, I., Pala, K. (eds.) *Text, Speech, and Dialogue: 21st International Conference, TSD 2018, Brno, Czech Republic, September 11–14, 2018, Proceedings*. pp. 518–526. Springer Nature, Cham (2018)
- Kovács, G., Szekrényes, I.: Applying neural network techniques for topic change detection in the hucomtech corpus. In: Hunyadi, L., Szekrényes, I. (eds.) *The temporal structure of multimodal communication: Theory, methods and applications*. pp. 147–162. Springer International Publishing, Cham (2019)
- Magnusson, M. S.: Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers* 32/1, 93–110 (2000)
- Németh, T. E. (szerk): *Ember-gép kapcsolat. A multimodális ember-gép kommunikáció modellezésének alapjai*. Budapest: Tinta Könykiadó (2011)
- Pápay, K., Szeghalmy, S., Szekrényes, I.: Hucomtech Multimodal Corpus annotation. *Argumentum* 7, 330–347 (2011)

- Schmidt, T.: Good practices in the compilation of folk, the research and teaching corpus of spoken German. In: Kirk, J. M., Andersen, G. (eds.) *Compilation, transcription, markup and annotation of spoken corpora*, special issue of the international journal of corpus linguistics [IJCL 21:3] pp. 396–418 (2016)
- Székrenyess, I.: Annotation and interpretation of prosodic data in the hucomtech corpus for multimodal user interfaces. *Journal on Multimodal User Interfaces* 8/2, 143–150 (2014)
- Székrenyess, I.: Prosotool, a method for automatic annotation of fundamental frequency. In: 6th IEEE International conference on cognitive Infocommunications (CogInfo-Com). pp. 291–296. IEEE, New York (2015)
- Székrenyess, I., Kovács, G.: Classification of formal and informal dialogues based on turn-taking and intonation using deep neural networks. In: Karpov, A., Potapova, R., Mporas, I. (eds.), *Speech and computer*. pp. 233–243. Springer International Publishing, Cham (2017)
- Székrenyess, I.: Post-processing T-patterns Using External Tools From a Mixed Method Perspective. *Frontiers in Psychology* 10, 1–12 (2019)
- Székrenyess, I.: *Prozódiai jellemzők gépi feldolgozása és hasznosítása élőnyelvi korpuszok elemzésében*. PhD-értekezés. Debreceni Egyetem, Nyelvtudományok Doktori Iskola (2020)
- Váradí, T.: Central and South-East European Resources in META-SHARE. In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*. pp. 431–438 (2012)
- Váradí, T., Simon, E., Sass, B., Gerócs, M., Mittelholcz, I., Novák, A., Indig, B., Prószéky, G., Vincze, V.: *Az e-magyar digitális nyelvfeldolgozó rendszer*. In: Tanács, A., Vincze, V. (szerk.) XIII. magyar számítógépes nyelvészeti konferencia (MSZNY 2017). pp. 103–109. JATEPress, Szeged (2017)
- Withers, P.: Metadata management with Arbil. In: V. Arranz, D. Broeder, B. Gaiffe, M. Gavrilidou, M. Monachini (eds.) *Proceedings of the workshop describing LRs with metadata: Towards flexibility and interoperability in the documentation of LR at LREC 2012*. pp. 72–75. ELRA (2012)
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.: ELAN: a professional framework for multimodality research. In: Calzolari, N. et al. (eds.) *Proceedings of LREC 2006*. pp. 213–269. ELRA (2006)