

Koreferenciaviszonyok az enyhe kognitív zavarban szenvedők beszédátirataiban

Bevezetés

Az enyhe kognitív zavar az egészséges és a demens állapot közötti folyamatban jelenik meg köztes állomásként, ezért fontos szerepe van az Alzheimer-kór és egyéb demenciák minél hamarabbi felismerésében (Negash et al. 2007). Az egészséges idős és az enyhe kognitív zavarban szenvedő személyek különböző neuropszichológiai vizsgálatokon elért eredményei nem sokban térnek el egymástól, ez megnehezíti a betegség felismerését.

Mivel a betegség egyik vezető tünete a rövidtávú memória működésének romlása, az emlékezés spontán beszédprodukcióban való megnyilvánulására pedig az egyik legjobb példa a koreferencialáncok használata, célszerű megvizsgálni ezt a nyelvi jelenséget.

Hipotéziseim a következők: (1) Az enyhe kognitív zavarban szenvedők koreferenciahasználata jelentős eltéréseket fog mutatni az egészséges emberekétől, és ez segítségül szolgálhat a betegek diagnosztizálásában. Ezek a különbségek a következők lehetnek: (2) a koreferencialáncok hossza, (3) a láncok mennyisége, (4) és a visszautalási típusok egymáshoz viszonyított aránya. További hipotézisem az, hogy ezek a különbségek segíteni fognak az enyhe kognitív zavarban szenvedők automatikus szűrésében.

Módszertan

A vizsgálatot egy már a Szegedi Tudományegyetemen folyó kutatásba (Vincze et al. 2015) bekapcsolódva végeztem el. Az adatfelvételre a szegedi memóriaambulancián került sor, melynek keretében két történetet meséltek el az adatközlők. A hanganyagok lejegyzett változatai követik a spontán beszéd sajátosságait. Az összes adatközlő magyar anyanyelvű volt, és rendelkezésemre állt a betegségük klinikai diagnózisa is. 86 adatközlő 172 elmesélt története közül 35 adatközlő alkotta a kontroll csoportot, ezek az arányok követik a nemzetközi tendenciát. Ezt a már kész korpuszt használtam fel és annotáltam az MMAX2 nevű annotációs eszközzel. Az annotálás során a következő koreferenciakategóriákkal dolgoztam: adverbial, apposition, epithet, meronym, hypernym, pronominal, repetition, synonym, verbal. Az átiratokon gépi tanulási kísérleteket is végeztem a Szegedi Tudományegyetemen zajló projektet (Vincze et al. 2016) kiegészítve. A Weka szoftver (Hall et al. 2009) segítségével több tanuló algoritmust is kipróbáltam.

Eredmények

A következő hasonlóságokat tapasztaltam az enyhe kognitív zavarban szenvedő és a kontroll csoport között: a koreferencialáncok hossza, valamint a kategóriák használatának gyakorisági sorrendje. Különbségek adódtak a koreferencialáncok mennyisége, valamint egyes használt kategóriák csoporton belüli eloszlása tekintetében a csoportok között. Minden egyes kategóriára szövegenként kétmintás t-próbát végeztem, hogy statisztikailag is bizonyítható legyen a kategóriák szerepe a két csoport elkülönítésében, ennek eredményeként elmondható, hogy az ismétlés tekintetében szignifikáns eltérést mutat a két csoport. A tanuló algoritmusok közül a legeredményesebb az SVM algoritmus (Cortes–Vapnik 1995) 75,58%-os pontosságot ért el.

Következtetések

Míg a kontroll csoport tagjai közel azonosan teljesítették a feladatot, az enyhe kognitív zavarban szenvedők csoportjának tagjai között kiugró eltérések voltak. Ebből azt a következtetést vonhatjuk le, hogy a rövidtávú memória sérülésével egyértelműen sérül a koreferencia használatának stratégiája is, a nagy eltéréseket a különböző stratégiahasználat eredményezheti. Jelen kutatásom eredményei javítják az enyhe kognitív zavar felismerésének lehetőségét.

Szakirodalom

- Negash, S., Petersen, L.E., Geda, Y.E., Knopman, D.S., Boeve, B.F., Smith, G.E., Ivnik, R.J., Howard, D.V., Howard, Jr J.H., Petersen, R.C. (2007) Effects of ApoE genotype and Mild Cognitive Impairment on implicit learning. *Neurobiology of Aging* 28(6), 885–893.
- Vincze V., Hoffmann I., Szatlóczki G., Bíró E., Gosztolya G., Tóth L., Pákáski M., Kálmán J. (2015) Az enyhe kognitív zavar automatikus azonosítása beszédátiratok alapján. In: *XI. Magyar Számítógépes Nyelvészeti Konferencia*, Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 249–256.
- Vincze V., Gosztolya G., Tóth L., Hoffmann I., Szatlóczki G., Bánréti Z., Pákáski M., Kálmán J. (2016) Detecting Mild Cognitive Impairment by Exploiting Linguistic Information from Transcripts. In: Katrin Erk, Noah A. Smith (szerk.) *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Berlin, 181–187.
- Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. (2009) The WEKA data mining software: an update. *SIGKDD Explorations* 11(1), 10–18.
- Cortes, C., Vapnik, V. (1995) Supportvector networks. *Machine Learning* 20(3), 273–97.