

## Koreferenciaviszonyok vizsgálata enyhe kognitív zavarban szenvedők beszédátirataiban<sup>1</sup>

Kovács Viktória

SZTE BTK Nyelvtudományi Doktori Iskola  
viktoria.kovacs12@gmail.com

**Kivonat:** A jelen tanulmányban az enyhe kognitív zavarban szenvedők beszédátiratain végzett koreferenciaannotációt mutatom be, valamint az annotálás eredményéül kapott adatokon végzett statisztikai számításokat és egy gépi tanulási kísérletet. Az enyhe kognitív zavar lefolyása során megfigyelhető a memória működésének romlása, ez tetten érhető a koreferenciaviszonyok használatán. A láncok mennyisége és hossza, valamint a különböző szintaktikai és szemantikai szabályszerűségeken alapuló visszautalási stratégiák vizsgálata segíthet az enyhe kognitív zavar korai diagnosztizálásában, valamint a kapott adatokon végzett gépi tanulási kísérletek alapján az automatikus azonosításban is.

### 1 Bevezetés

A kutatás során arra a kérdésre kerestem a választ, hogy az enyhe kognitív zavar (a továbbiakban: EKZ) lefolyása során egyre hanyatló emlékezeti rendszerek hatására sérül-e a koreferencialáncok használatának képessége. Ehhez egy, a Szegedi Tudományegyetemen korábban zajló kutatáshoz készített korpuszt használtam fel (Vincze et al. 2015). A korpusz EKZ-ban szenvedő személyek és egy egészséges kontrollcsoport tagjai által elmesélt történetek átíratait tartalmazza, ezáltal alkalmassá válik arra, hogy a két csoport számítógépes nyelvészeti módszerekkel összehasonlítható legyen a koreferencialáncok használata alapján. Mivel az EKZ és a belőle kialakuló komolyabb mentális deficitek gyógyíthatatlanok, kiemelt fontosságú a betegség korai diagnosztizálása. A különböző nyelvészeti megközelítésekből történő vizsgálatok eredményei segíthetnek a szakembereknek a diagnózis felállításában és ezzel a betegség lefolyásának lassításában.

A továbbiakban ismertetem azokat a beszédben fellelhető és nyelvtudományi szempontból releváns tüneteket, amelyek segítenek az EKZ diagnosztizálásában. A tünetek bemutatása után a koreferencia definícióját fogom megadni és ezzel együtt a kutatásom motivációját, valamint az előzetes hipotéziseimet. Ezután bemutatom magát a kutatást, ezen belül is a kutatás alapjául szolgáló korpuszt, az annotálási módszereket, valamint a kapott adatokat és az azokon végzett statisztikai

---

<sup>1</sup> A kutatást az EFOP-3.6.1-16-2016-00008 azonosítójú, EU társfinanszírozású projekt támogatta.

számításokat. Végezetül pedig bemutatok egy gépi tanulási kísérletet, és levonom a kutatás konklúzióit.

## **1.1 Kognitív háttér**

A kutatás alapját képező korpuszt EKZ-ban szenvedők beszédátiratai alkotják. A következőkben a deficit tüneteit és a memóriával kapcsolatos fogalmakat szükséges tisztázni. Az EKZ-nak nem pusztán a beszédben megjelenő tünetei vannak, viszont mivel nyelvészeti szempontból vizsgáltam a szövegeket, csak azokat és a memóriával kapcsolatos tüneteket veszem sorra.

### **1.1.1 Az enyhe kognitív zavar tünetei**

Az EKZ az egészséges és a demens állapot közötti folyamatban jelenik meg köztes állomásként, tehát a diagnosztizálás során a demencia tüneteire kell odafigyelni (Negash et al. 2007). A demenciák közül az Alzheimer-kór az, amelynek a tünetei megjelenhetnek az EKZ-ban, egy sokkal enyhébb formában. Az Alzheimer-kór és az EKZ közötti erős összefüggést alátámasztja az a tény, hogy az idős, EKZ-ban szenvedők 50%-ánál három éven belül az Alzheimer-kórt is diagnosztizálják (Kempler 2005).

Az EKZ során jelentkező deficit az emlékezeti rendszerek tekintetében az epizodikus memóriát, a prospektív memóriát és a végrehajtó működéseket érinti (Tariska 2002; Turi–Németh–Hoffmann 2014). Az epizodikus memória az explicit (deklaratív) emlékezeti rendszer része, amely a konkrét eseményekre való emlékezést biztosítja (Baddeley–Eysenck–Anderson 2010), és ezekhez az eseményekhez kapcsolódó érzéseink tárolására alkalmas (Gósy 2005). A prospektív memória a terveinkre való emlékezést biztosítja (Baddeley–Eysenck–Anderson 2010), ezeknek az emlékeknek alacsonyabb az információtartalma, spontán beszédben nem is igazán fellelhető a romlása. A végrehajtó működések a terveink végrehajtásáért felelnek, legyen az egy megnyilatkozás kimondása vagy egy fizikai cselekvés végrehajtása.

A deficittel kapcsolatban meg kell említeni a munkamemóriát, amely ugyan az EKZ-ban még nem érintett, de az enyhe Alzheimer-kórban már igen. A munkamemória a rövid távú memória része, az új ismeretek feldolgozásáért, az adatrögzítésért és a szabályalkotásért felel, kapacitása egyénenként eltérő. A munkamemória-modell egyik fő komponense a központi végrehajtó, amely a feladatmegoldásért is felel, pl. a következtetésért, a nyelvi megértésért, a problémamegoldásért, valamint azért, hogy kapcsolatba tudjuk hozni az egymástól független információkat. A munkamemóriának az egyik perifériás alrendszere a fonológiai hurok, amely a beszédszerű információk feldolgozásáért felelős (Gósy 2005).

### **1.1.2 A beszédben megjelenő tünetek**

Az EKZ diagnosztizálási nehézségei abból fakadnak, hogy az egészséges idős egyének neuropszichológiai vizsgálatokon elért eredményei nem sokban térnek el az enyhe kognitív zavarban szenvedők teljesítményétől. Ezzel párhuzamosan a spontán

beszédben is csak nehezen figyelhetők meg a deficitre utaló jellegzetességek, hiszen az egészséges adatközlők is alkalmaznak különböző stratégiákat a szókeresésre.

Az EKZ eredményesebb szűréséhez hozzájárul egy, a Szegedi Tudományegyetemen zajló projekt (Vincze et al. 2015; Vincze et al. 2016), amely a deficit automatikus azonosítására törekszik beszédben fellelhető jellemzők statisztikai vizsgálata alapján.

Kutatásuk során az alábbi jellemzők statisztikailag szignifikánsnak bizonyultak az EKZ szűrésében: a szöveg szavainak száma, a hezitációs jelenségek száma és aránya, a bizonytalan (körülíró, megnevezési nehézségre utaló) szavak száma és aránya, az ismeretlen szavak (az adatközlők által alkotott szavak) száma, mondatbeli szavak száma, a néma szünetek száma, a nyújtások száma, a főnevek száma, az írásjelek száma (Vincze et al. 2015).

## 1.2 Nyelvtudományi háttér

Az enyhe kognitív zavar tehát egy olyan kognitív deficit, amelynek az egyik legjellemzőbb tünete a rövidtávú emlékezet zavara. Olyan beszédből kinyerhető további információkat szükséges jelölni az annotáció során, amelyek az emlékezet zavarára utalhatnak, és nyelvészeti szempontból vizsgálhatók.

### 1.2.1 A koreferencia

A memória romlását jól mutathatja az is, hogy hogyan utalunk vissza egy korábban megemlített objektumra. A koreferencia két szövegbeli szó vagy szóelem, szókapcsolat között jön létre a szövegvilágon belüli ugyanazon dologra utalással, melynek alapja a referencia. A második szóelem az elsőre utal vissza, és rajta keresztül referál a már említett dologra (Tolcsvai 2003). Tehát a koreferencia objektumok és események említése és az azokra való folyamatos visszautalás. Ezek a visszautalási formák különböző szintaktikai és szemantikai szabályszerűségeken keresztül valósulnak meg a kifejezések között. Az azonos objektumokra és eseményekre vonatkozó kifejezések ezáltal láncokat alkotnak a szöveg mentén. A koreferencia bonyolult pszicholingvisztikai, valamint logikai jelenség, hiszen a mondandónk részben ez által lesz logikusan felépített, tartalmas egység.

### 1.2.2 Visszautalási kategóriák

A visszautalási kategóriák a koreferencialánc egymást követő tagjai közötti szintaktikai, illetve szemantikai kapcsolatokon és a visszautaló szó szófaján alapulnak. Számos jellemző felsorolható, amelynek a segítségével antecedens választható a visszautaló szóhoz, pl. a két kifejezés közötti szemantikai kapcsolat, a kötési elvek vagy a megnyilatkozás kontextusából származó információk. Az annotálás szempontjából olyan kategorizálást kell alkalmazni, amely a lehető legáttekinthetőbb és leggazdaságosabb. A kategóriák megválasztása során az egyik szempont az volt, hogy elkülönítsem egymástól a deiktikus kifejezéseket, ezen belül is a térre és időre, valamint a személyekre és objektumokra vonatkozó kifejezéseket.

Egy másik szempont az események és objektumok elkülönítése. A harmadik pedig az objektumok közötti szemantikai kapcsolatok figyelembevétele.

Az annotálás során ezek alapján a következő kategóriákat különböztettem meg: Adverbial (határozó), Apposition (értelmező), Epithet (jelző), Meronym (meronímia), Hypernym (hiperonímia), None (egyik sem), Other (egyéb), Pronominal (névmás), Repetition (ismétlés), Synonym (szinonímia), Verbal (igei).

A következő példákban a visszautaló szót félkövérrel emeltem ki, a (...) jelölés a példa megszakítottóságát jelöli.<sup>ii</sup> Mivel a kutatás középpontjában a memória működése és az abban bekövetkező változások felismerése áll, ezért a visszautalási kategóriát mindig a visszautaló szó és a láncban őt megelőző elem közötti kapcsolat határozta meg.

Az Adverbial (határozó) kategóriába tartozott a határozói visszautalásnak az összes típusa. Ezek többnyire helyhatározók voltak, de előfordult a szövegekben időhatározó is, mivel a térre és időre vonatkozó információk az elmesélt történetek kontextusát határozták meg, egyaránt ebbe a kategóriába kerültek.

[1] Hát egy bizonyója egy vendéglő volt, **ahol** leült ugye a vendég.

Az Apposition (értelmező) kategória az értelmezést és a kiegészítő magyarázat típusú elemeket tartalmazza. Azok a kifejezések kerültek ebbe a kategóriába, amelyek valamilyen információ hozzáadásával azonosították az objektumot vagy az eseményt, amire visszautaltak, tehát többlettartalmat adtak hozzá, értelmezték azt. Ezek a visszautalások azért nem egyszerűen az ismétlés kategóriába kerültek, mert a többlettartalom hozzáadása fontos volt az objektum beazonosításának szempontjából. A többlettartalom hiányában nem lett volna egyértelmű, hogy melyik láncba tartozik az elem.

[2] Hát valahol ... egy asztal két szék ... egy férfi már ül ... a másik meg jön befelé. (...) **Az aki ült férfi** elővette a botját.

Epithet (jelző) kategóriába kerültek azok a visszautalások, ahol egy objektumot egy tulajdonsága alapján azonosított be az adatközlő, és ennek segítségével utalt vissza rá. Ez a kategória is az objektumok elkülönítésében játszott szerepet, viszont ebben a kategóriában nincs ismétlés.

[3] Férfi ilyen nyeles fával volt (...). De **a hálónál** van egy virág...

A Hypernym (hiperonímia) kategóriába a két kifejezés közötti fogalmi alá-fölé rendeltség alapján kerültek a visszautalások.

[4] ... de valami madzagon volt ez a bagoly (...) hogy talán el akarta fogni **a madarat**.

Azokban az esetekben, amelyekben az egyik szó által jelölt dolog része volt a másik szó által jelölt dolognak, a visszautaló szó a Meronym (meronímia) kategóriába került. Ezekben a szövegekben ez leggyakrabban testrészeket és bútordarabokat

<sup>ii</sup> A hanganyagok átiratai számítógépes feldolgozásra készültek, ezért a további központosítási jelek a számítógépes feldolgozás szempontjából jelentősek, mivel a koreferenciaannotációt nem befolyásolják, változtatás nélkül közlöm a példákat. Bővebben az átiratok elkészítésénél figyelembevett módszertani elvekről: (Vincze et al. 2015, Vincze et al. 2016).

jelentett. Ezek a kifejezések nem alkottak külön koreferencialáncot, mivel csak egy-egy alkalommal említették őket az adatközlők.

- [5] Akkor .. utána ugye jött egy .. hölgy , gondolom én , hogy .. helyet foglalt és akkó kínálta volna , **fejére** öntötte az italt ...

A None kategóriába minden esetben a láncok első elemei kerültek, mivel ezek még semmire sem utaltak vissza.

A Pronominal (névmás) kategóriába azok a visszautalások kerültek, amelyekben a visszautaló szó személyes, mutató vagy vonatkozó névmás volt. Ezek a visszautalások a jellemzőkészlet gazdaságossága érdekében kerültek egy kategóriába, de elválasztva a határozói visszautalásokhoz tartozó névmásoktól és a határozatlan névmásoktól.

- [6] Hát szerintem... az illető az süketnéma volt ... mer.. azé mutogatott **neki** a másik

Az egyszerű ismétléssel történő visszautalások a Repetition (ismétlés) kategóriába kerültek. Ebbe beletartoznak azok a visszautalások is, amelyek egyébként más kategóriába tartoztak volna.

- [7] Behívja az inast aki (...) jön utána a ... az **az inas** ...

- [8] És akkor ő odavitte neki. Nem, először vitt **neki** virágot.

A Synonym (szinonímia) kategóriába a két kifejezés közötti szinonímián alapuló visszautalások kerültek.

- [9] Aztán a hölgy le akar ülni (...) És lekuporodik **a nő** és fölötte vívna.

Az eseményekre való visszautalás a Verbal (igei) kategóriába tartozik, mivel ilyen esetekben a lánc feje legtöbbször ige. Elhanyagolható mennyiségű példa volt arra, hogy pontosan ugyanarra az eseményre utaljon vissza az adatközlő, viszont az áttekinthetőség érdekében érdemes elkülöníteni az objektumokra történő visszautalásoktól.

- [10]Hát az , hogy kiment ez a részeg ember , ki **ki** . **osont** vagy **ki** . **suttyogott** sunnyogott.

Other (egyéb) kategóriába azok a visszautalások kerültek, ahol valamely kifejezésnél megnevezési nehézségei akadtak az adatközlőnek, ezt legtöbbször határozatlan névmással fejezte ki, de ettől függetlenül egyértelmű volt az, hogy mire utal vissza.

- [11]...valószínű hogy gyűrűt vett a férfi elő . Kis dobozba ... volt **valami**.

### 1.3 Hipotézisek

Mivel a koreferenciaviszonyok használatához szükséges a memória épsége, ezért feltételeztem, hogy az enyhe kognitív zavarban szenvedők az egészségesektől eltérő stratégiát alkalmaznak a visszautalásra ugyanazon történet elmesélése során. Elsődleges hipotézisem az volt, hogy az enyhe kognitív zavarban szenvedők kevesebb és rövidebb koreferencialáncot fognak használni ugyanazon történet elmeséléséhez.

Másodsorban pedig, hogy az enyhe kognitív zavarban szenvedők csoportjában az ismétlés a legtöbb, a kontrollcsoportban pedig a szinonímia és egyéb, a két kifejezés közötti bonyolultabb kapcsolatok felismerését feltételező visszautalási formák használata a leggyakoribb.

## 2 Kísérleti személyek, anyag, módszer

A kutatás a már említett Szegedi Tudományegyetemen zajló projektet veszi alapul, amely az enyhe kognitív zavar automatikus azonosítására törekszik (Vincze et al. 2015; Vincze et al. 2016).

### 2.1 Korpusz

A kutatáshoz készített korpuszt a kísérleti személyek által elmesélt történetek beszédátiratai alkotják. A hanganyagok felvétele minden esetben a Szegedi Memóriaambulancián történt, majd ezeket később jegyezték le nyelvészek. A hanganyagok lejegyzése során figyelembe vették a spontán beszéd sajátosságait, mint a néma szünetek: *előveszi a ... .. görbebotját*, vagy kitöltött szünetek: *Őő és akkor*, a kétszer indítások: *ért ... . értettem belőle*, fonológiai nyújtások: *képrű*. Az átiratok nem tartalmaznak azonban az intonációs jelenségekre vonatkozó információkat.

### 2.2 Kísérleti személyek

Összesen 86 adatközlő vett részt a felmérésben, ez 172 átíratot jelent, amelyből 35 adatközlő, azaz 70 átírat alkotta a kontrollcsoportot. A 35 kontrollszemély közül 23 volt nő és 12 férfi, az 51 enyhe kognitív zavarban szenvedő beteg közül 33 volt nő és 18 férfi. Az adatközlők mind magyar anyanyelvűek voltak, és rendelkezésemre állt a klinikai diagnózisuk. A feladatuk egységesen az volt, hogy két fekete-fehér rövidfilm megnézése után elmeséljék az első filmet, majd beszéljenek a legutóbbi napjukról, ezután meséljék el a második filmet.

### 2.3 Módszertan

Az annotálás az MMAX2 annotációs eszközzel és kizárólag manuálisan történt a hibázási lehetőségek kiküszöbölése miatt. A kutatáshoz a hanganyagok semmilyen formában sem lettek felhasználva. Az elliptikus szerkezetek és a morfológia szintjén jelölt kapcsolatok – mivel az annotációs eszköz nem volt rá alkalmas – nem lettek jelölve. Azokat az eseteket, amelyek nem voltak egyértelműek, vagy hezitációnak minősíthetőek voltak, nem kerültek bele a koreferencialáncokba az áttekinthetőség és egyértelműség végett. A két csoport átíratait véletlenszerű sorrendben, anonimizálva annotáltam, hogy a klinikai diagnózis ismerete ne befolyásolja az annotációt.

Az annotálás után az átíratokat két csoportba soroltam a klinikai diagnózis alapján, majd először az átíratoknak olyan általános tulajdonságait gyűjtöttem ki, mint a szövegek hossza, a bennük található láncmennyiség és lánc hossz. Ezek után

gyűjtöttem ki a két csoportban külön a használt visszautalási kategóriák számát. A kapott adatokon ezután statisztikai számításokat végeztem, hogy megvizsgáljam, melyek azok az átiratokból kinyerhető jellemzők, amelyek segíthetnek az enyhe kognitív zavar azonosításában és az egészséges állapottól való elkülönítésében.

A kategóriákra vonatkozó adatok alapján ezután gépi tanulási kísérleteket végeztem a Weka szoftver (Hall et al. 2009) segítségével. Ennek alapja a már említett Szegedi Tudományegyetemen zajló projekt volt (Vincze et al. 2016). A gépi tanítás során ők a következő jellemzőkészletet használták fel: A spontán beszéden alapuló jellemzők voltak a kitöltött és kitöltetlen szünetek száma, a hezitációk száma és aránya a tokenek számához képest, a névelő és a tartalmas szó közötti szünetek száma és a nyújtások száma. A morfológiai jellemzők a tokenek és szavak száma, a lemmák száma és aránya, a központozási jelek száma, a főnév, ige, melléknév, névmás, kötőszó száma és aránya, az egyes szám első személyű igék száma, az ismeretlen szavak száma. Szemantikai tulajdonságaik a bizonytalan szavak száma és aránya az összes tokenhez képest, az emlékezési tevékenységgel kapcsolatos szavak száma és aránya, a tagadó szavak száma, a tartalmas és funkciószavak száma és aránya, a film tartalma szempontjából kulcsszavak száma. A demográfiai jellemzők a nem, életkor és képzettség (Vincze et al. 2016). Az általuk használt jellemzőkészletet egészítettem ki az annotált visszautalási kategóriákkal. Minden beszélő esetében hozzáadtam jellemzőként az általa használt egyes visszautalási típusok számát az egyes szövegekre lebontva.

### 3 Eredmények

A következőkben a szövegekkel és a kategóriákkal kapcsolatos adatokat, valamint az azokon végzett statisztikai számításokat ismertetem, majd a gépi tanulási kísérletek közül is bemutatom azt, amelyik a legeredményesebb lett.

#### 3.1 Az átiratokkal kapcsolatos adatok

Először az átiratok adatait vizsgáltam meg az egyes adatközlők alapján. Mivel a két történet elmesélése között eltelt idő is befolyásolhatja az eredményeket, így a két történet átiratait is elkülönítettem.

	EKZ		Kontroll	
	1-es szöveg	2-es szöveg	1-es szöveg	2-es szöveg
<b>Láncok átlagos száma</b>	4,53	5,02	5,66	5,40
<b>Láncok átlagos hossza</b>	2,31	2,32	2,32	2,42
<b>Leghosszabb szöveg (szó)</b>	369	222	206	202
<b>Legrövidebb szöveg (szó)</b>	24	29	37	52
<b>Átlagos szöveg (szó)</b>	109,14	95,71	104,37	97,86

1. táblázat. A szövegek adatai. EKZ: enyhe kognitív zavar.

A szöveggel kapcsolatos adatok alapján elmondható, hogy a kontrollcsoport valóban több koreferencialáncot használ átlagosan az egyes szövegekben, a láncok hossza között azonban nem mutatkozik különbség. Az enyhe kognitív zavarban

szenvedők csoportjának átiratai – a deficitből adódóan – valószínűleg nagyobb mértékben tartalmaznak hezitációs jelenségeket, mint a kétszer indítás és a gyakori szóismétlés, ez pedig befolyásolja a leghosszabb és legrövidebb elmesélt történet jellemző eredményeit az egyes csoportokban. A leghosszabb szöveg jellemző nagy eltérést mutat a két csoport között, az EKZ-csoport kiugró értékét a hezitációs jelenségek okozhatják. Tehát az, hogy az adatközlő sokkal több szót használ, nem jelenti azt, hogy tartalmasabban fejezte ki magát, ezért ez a jellemző nem alkalmas a két csoport összehasonlítására. A legrövidebb szöveg jellemző alapján a kontrollcsoport tartalmasabban fejezte ki magát. Az átlagosan használt szavak száma szövegenként a leghosszabb szöveg jellemző miatt szintén nem alkalmas az összehasonlításra.

### 3.2 A kategóriákkal kapcsolatos adatok

Az átiratokon végzett annotáció után a jelölt kategóriákat kinyertem az egyes átiratokból és csoportonként összesítettem.

	EKZ		Kontroll	
<b>Adverbial</b>	13	1,15%	15	1,64%
<b>Apposition</b>	48	4,26%	33	3,61%
<b>Epithet</b>	6	0,53%	5	0,55%
<b>Hypernym</b>	16	1,42%	10	1,09%
<b>Meronym</b>	43	3,64%	40	4,36%
<b>Other</b>	76	6,75%	42	4,60%
<b>Pronominal</b>	404	35,88%	284	31,04%
<b>Repetition</b>	407	36,15%	410	44,81%
<b>Synonym</b>	111	9,86%	73	7,98%
<b>Verbal</b>	4	0,36%	5	0,55%

2. táblázat. A kategóriák adatai. EKZ: enyhe kognitív zavar

A kategóriákkal kapcsolatos adatok alapján kiemelkedő mind a két csoportban az ismétlések (repetition) száma. Megfigyelhető, hogy amíg az EKZ-csoport ismétlései az összes visszaütésnek a 36,15%-át teszik ki, addig a kontrollcsoport ismétlései a 44,81%-át. Ennek fényében az a hipotézisem, hogy az enyhe kognitív zavarban szenvedő adatközlők több ismétlést fognak használni a kontrollcsoportnál, nem helytálló.

Érdemes azonban a többi kategóriát is megvizsgálni. A második leggyakrabban használt kategória szintén egyöntetűen mind a két csoportnál a névmási (pronominal) kategória volt. Ez az EKZ-csoportban 35,88%-ot jelentett, tehát majdnem megegyezett az ismétlések számával. A kontrollcsoportban azonban csak 31,04%-ot ért el, ami elég nagy eltérés az ismétlések arányától. A harmadik leggyakoribb kategória a szinonímia volt, ez az EKZ-csoportban 9,86%, a kontrollcsoportban pedig 7,98%.



### 3.3 Statisztikai vizsgálatok

A statisztikai adatok számítása során a University of Washington által kiadott elveket követtem (Writing an APA Empirical (lab) Report, 2010). Mivel a kontrollcsoporton belül a kategóriák használati arányai között ekkora különbségek voltak, de ez az enyhe kognitív zavarban szenvedők csoportjára nem volt jellemző, a kategóriákon alkalmaztam a  $\chi^2$  statisztikai eljárást. Az eredmény azt mutatja, hogy szignifikáns az eltérés a kategóriák eloszlásának tekintetében a két csoport között ( $p = 0,010$ ).

A kapott adatokon egyéb statisztikai vizsgálatokat is végeztem, hogy meg tudjam vizsgálni az egyes kategóriák szerepét a két csoport tagjainak elkülönítésében. Ezért minden egyes kategóriára, szövegenként kétmintás  $t$ -próbát végeztem. A kutatás során a két minta szabadságfoka kontrollcsoport: 34, EKZ-csoport 50. A kategóriákhoz tartozó szignifikanciaszinteket ( $p$  értékeket) a következő táblázat mutatja. Ennek alapján elmondható, hogy egyedül a repetition, azaz az ismétlés mutat statisztikailag szignifikáns különbséget a két csoport között.

	<i>t</i> -próba az 1-es szövegen	<i>t</i> -próba a 2-es szövegen
<b>Adverbial</b>	0,4837	0,0688
<b>Apposition</b>	0,2985	0,2102
<b>Epithet</b>	0,1565	0,2116
<b>Hypernym</b>	0,3560	0,3407
<b>Meronym</b>	0,2714	0,0925
<b>Other</b>	0,3432	0,1771
<b>Pronominal</b>	0,3847	0,4949
<b>Repetition</b>	0,0135	0,0234
<b>Synonym</b>	0,4549	0,3581
<b>Verbal</b>	0,1565	0,3945

3. táblázat. A kategóriák szignifikanciaszintjei

### 3.4 Egy gépi tanulási kísérlet

A már említett, a Szegedi Tudományegyetemen zajló projektet kiegészítve gépi tanulási kísérleteket is végeztem. Az általuk alkotott teljes jellemzőkészletet felhasználva a legeredményesebb módszer számukra az SVM algoritmus (Cortes–Vapnik 1995) volt, amellyel 69,1%-os pontosságot értek el. Ez azt jelenti, hogy 84 esetből 58-szor állapított meg a rendszer pontos diagnózist (Vincze et al. 2016).

Az általuk használt teljes jellemzőkészletet kiegészítettem az annotált visszautalási kategóriákkal. Minden beszélő esetében hozzáadtam jellemzőként az általa használt egyes visszautalási típusok számát, az egyes szövegekre lebontva. Így egy nagyobb mintát felhasználva az SVM algoritmus 75,58%-os pontosságot ért el. A mérésben a 86 adatközlő eredményei szerepeltek. Az összes esetből 65-ször állapított meg helyes diagnózist a rendszer. A 21 hibás diagnózisból 11-szer fals pozitív, 10-szer fals negatív állapított meg, ami azt jelenti, hogy a rendszer 11 személyt tévesen enyhe kognitív zavarban szenvedőnek, 10 személyt pedig tévesen egészségesnek minősített.

## 4 Következtetések

Az általam végzett vizsgálat alapján elmondható, hogy az enyhe kognitív zavarban szenvedők az egészségesektől részben eltérő stratégiát alkalmaznak a visszautalásra ugyanazon történet elmesélése során. Az elsődleges hipotézisem az volt, hogy az EKZ-sok kevesebb és rövidebb koreferencialáncot fognak használni: ez csak részben igazolódott. Az EKZ-s adatközlők valóban kevesebb koreferencialáncot használtak átlagosan, azonban a hosszai között nem mutatkozott különbség.

A második hipotézisem az volt, hogy az EKZ-s csoportban az ismétlés használata a legjellemzőbb, a kontrollcsoportban pedig a szinonímia és egyéb visszautalási formák. A vizsgálat alapján ez a hipotézis sem helytálló, mind a két csoportban az ismétlés volt a leggyakoribb visszautalási forma. Ennek egyszerű magyarázata lehet, hogy az elménkben emlényomokként tárolt információ könnyebben hívható elő, minél többször ismétlődik egy szó, annál erősebb és frissebb az emlényom. Ezért lehetséges az, hogy a leggyakoribb kategória a kontroll- és az EKZ-csoportban is az ismétlés volt. Emellett a kontrollcsoportban közel 8%-kal több lett az ismétlés használata, mint az EKZ-s csoportban. Ennek oka lehet az, hogy mivel a kimondott szó a fonológiai hurokban tárolódik (Gósy 2005), így egy ép emlékezeti rendszerrel rendelkező egyénnek könnyebb onnan előhívni a már egyszer kimondott szavakat, míg egy sérült emlékezeti rendszerrel rendelkező egyénnek ugyanakkora erőfeszítés az ismétlés használata, mint más visszautalási formáé.

A hipotézisek ugyan nem igazolódtak, de a gépi tanítás alapján érdemes a koreferenciára vonatkozóan további kutatásokat végezni a mentális betegségek kapcsán. Mivel az annotált kategóriák közül nem mind mutatott szignifikáns különbséget a két csoport között, ezért gazdaságosabbá lehetne tenni a jellemzőkészletet egyes kategóriák összevonásával. További kategóriák felvételével szintén árnyaltabbá lehetne tenni a különbséget. Mivel a névmási kategóriát elég nagy arányban használta mind a két csoport, érdemes lenne megvizsgálni azt, hogy ezen a kategórián belül mekkora az előfordulása az egyes névmási visszautalásoknak.

Mivel a gépi tanítás során az olyan kategóriák javítottak jelentősen az elért eredményen, amelyek egyébként nem mutattak szignifikáns különbséget használati arányaik szempontjából, a minta növelése is indokolt lenne.

## 5 Összegzés

Kutatásom során az enyhe kognitív zavarban szenvedők és egészséges adatközlők által elmesélt rövid történetekben fellelhető koreferencialáncokat hasonlítottam össze. A két csoport között eltérést mutatott a koreferencialáncok száma, az ismétlés mint visszautalás használatának aránya és az egyes visszautalási kategóriák csoporton belüli eloszlása. Nem mutat különbséget a használt koreferencialáncok hossza és a többi visszautalási stratégia használati aránya a két csoport között. A kapott eredményeim javítják a különböző automatikus felismerésre törekvő eljárásokat.

## Irodalom

- Baddeley, A., Eysenck M. W., Anderson M. C. 2010. *Emlékezet*. Budapest: Akadémiai Kiadó.
- Cortes, C., Vapnik, V. 1995. Supportvector networks. *Machine Learning*, 20(3): 273–297.
- Gósy, M. 2005. *Pszicholingvisztika*. Budapest: Osiris Kiadó.
- 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.
- Kempler, D. 2005. *Neurocognitive disorders in aging*. Thousand Oaks: Sage Publications.
- Tolcsvai Nagy G. 2003. A szöveg. In: Kiefer, F. (szerk.) *A magyar nyelv kézikönyve*. Budapest: Akadémiai Kiadó. 245–256.
- 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.
- Tariska, P. (szerk.) 2002. *Kortünet vagy kórtünet? Mentális zavarok az időskorban*. Budapest: Medicina Könyvkiadó.
- Turi, Zs., Németh D., Hoffmann I. 2014. Nyelv és emlékezet. In: Pléh, Cs., Lukács, Á. (szerk.) *Pszicholingvisztika: Magyar pszicholingvisztikai kézikönyv*. 2. kötet. Budapest: Akadémiai Kiadó. 743–776.
- Vincze, V., Hoffmann I., Szatlóczki G., Biró E., Gosztolya G., Tóth L., Pákáski M., Kálmán J. 2015. Az enyhe kognitív zavar automatikus azonosítása beszédátiratok alapján. In: Tanács, A., Varga V., Vincze V. (szerk.) *XI. Magyar Számítógépes Nyelvészeti Konferencia. MSZNY2015*. Szeged: JATEPress. 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 (Berlin, 2016. augusztus 7–12.)*. Berlin: Association for Computational Linguistics. 181–187.
- University of Washington, Psychology Writing Center. 2010. *Reporting Results of Common Statistical Tests in APA Format*. Elérhető: [https://depts.washington.edu/psych/files/writing\\_center/stats.pdf](https://depts.washington.edu/psych/files/writing_center/stats.pdf). Letöltve: 2017. október 6.