

# KOREFERENCIAVISZONYOK AZ ENYHE KOGNITÍV ZAVARBAN SZENVEDŐK BESZÉDÁTIRATAIBAN

Kovács Viktória

*SZTE Nyelvtudományi Doktori Iskola*



- I. Motiváció
- II. Enyhe Kognitív Zavar (EKZ)
  - 1, Általános jellemzés
  - 2, Az érintett emlékezeti rendszerek
  - 3, Beszédben megnyilvánuló tünetek
  - 4, A diagnosztizálás nehézségei
- III. Koreferencia
- IV. Előzetes hipotézisek
- V. Vizsgálat
  - 1, A korpusz adatai
  - 2, Módszertan
  - 3, Visszautalási kategóriák
- VI. Statisztikai adatok
  - 1, Szöveggel kapcsolatos statisztikai adatok
  - 2, A kategóriákkal kapcsolatos statisztikai adatok
- VII. Gépi tanulási kísérletek
- VIII. Összefoglalás

# I. Motiváció

- Számos kognitív betegség gyógyíthatatlan.
- A korai felismerés lassíthatja a betegség lefolyását.
- A mentális állapot legkönnyebben a beszéd segítségével mérhető fel.
- Hasznos információk nyerhetők ki a mentális állapotra nézve a szövegalkotás és szövegértés vizsgálatával.

## II. Az Enyhe Kognitív zavar (EKZ)

### *1) Általános jellemzése*

- Az egészséges és demens állapot közötti köztes állomás.
- Az idős, enyhe kognitív zavarban szenvedő betegek 50%-ánál három éven belül diagnosztizálják az Alzheimer-kórt (Kempler 2005).
- Hasonló tünetek enyhébb megjelenéssel.

## II. Az Enyhe Kognitív zavar (EKZ)

### 2) *Az érintett emlékezeti rendszerek*

- Epizodikus memória

Konkrét eseményekre való emlékezés (Baddeley–Eysenck–Anderson 2010) és a hozzájuk kapcsolódó személyes élményeink, érzéseink tárolása (Gósy 2005).

- Prospektív memória

Terveinkre való emlékezés

- Végrehajtó működések

Eltervezett célok végrehajtása

## II. Az Enyhe Kognitív zavar (EKZ)

### *3) Az eddig tapasztalt beszédbeli statisztikailag szignifikáns jellemzői*

- szavak száma
- hezitációk száma / aránya
- bizonytalan szavak száma / aránya
- ismeretlen szavak száma
- mondatbeli szavak száma
- néma szünetek száma
- nyújtások száma
- főnevek száma

(Vincze et al. 2015)

## II. Az Enyhe Kognitív zavar (EKZ)

### 4) *A diagnosztizálás nehézségei*

- Megkülönböztetési nehézségek a korosodással járó teljesítménycsökkenés és az EKZ tünetei között.
- Az egészséges idős és az EKZ-ban szenvedő személyek neuropszichológiai vizsgálatokon elért eredményei nem sokban térnek el egymástól.
- A betegség szűrésére hatékonyabb módszer Vincze, Hoffmann, Szatlóczki, Bíró, Gosztolya, Tóth, Pákáski, Kálmán a Szegedi Tudományegyetemen zajló projektje (Vincze et al. 2015), amely az EKZ automatikus azonosítására törekszik.

### III. Koreferencia

- 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 (Kiefer–Siptár 2003).
- Objektumok és események felismerése és azokra való folyamatos visszautalás.
- A visszautalás különböző szintaktikai és szemantikai törvényszerűségeken keresztül valósul meg.
- Láncokat alkotnak a szöveg mentén.



## IV. Hipotézisek

1. Az EKZ-ban szenvedők kevesebb koreferencialáncot fognak használni.
2. Az EKZ-ban szenvedők rövidebb koreferencialáncokat fognak használni.
3. A kontroll csoportban több lesz a szinonímia és egyéb, a két kifejezés közötti bonyolultabb szintaktikai/ szemantikai kapcsolatok felismerését feltételező visszautalási formák használata.

# V. Vizsgálat

## 1) Adatok

### Az adatközlők adatai:

- Az adatközlők magyar anyanyelvűek voltak.
- Rendelkezésemre állt a klinikai diagnózisuk.

### A korpusz adatai:

	Adatközlők	Átiratok száma	Szavak száma	Szövegek átlagos szóhossza
EKZ	51	102	10447	102,42
Kontroll	35	70	7078	101,11
Összesen	86	172	17525	101,89

# V. Vizsgálat

## 2) Módszertan

- Csak az átiratokat vettem figyelembe az annotáláshoz.
- Az átiratokat manuálisan annotáltam.
- Az MMAX2 annotációs eszközt használtam.
- Nem jelöltem az elliptikus szerkezeteket és a morfológia szintjén jelölt kapcsolatokat.

# V. Vizsgálat

## 3) *Visszaautalási kategóriák*

- Adverbial (határozó)  
*(1) Hát egy bizonyája **egy vendéglő** volt, **ahol** leült ugye a vendég.*
- Apposition (értelmező)  
*(2) Egy kihalt fa elülső része ilyen **ággal**, (...)ilyen **száraz ággal** volt valami .*
- Epithet (jelző)  
*(3) **Férfi** ilyen nyeles fával volt (...). De **a hálósnál** van egy virág...*
- Hypernym (hiperonímia)  
*(4)...de valami madzagon volt ez **a bagoly**(...) hogy talán el akarta fogna **a madarat***

- Meronym (meronímia)  
*(5) Akkor .. utána ugye jött **egy** .. **hölgy** , gondolom én , hogy .. helyet foglalt és akkó kínálta volna , **fejére** öntötte az italt ...*
- Pronominal (névmás)  
*(6) És megjelent **egy pincér** , **aki** eléggé ilyen mókás figura.*
- Repetition (ismétlés)  
*(7) Behívja **az inast** aki (...) jön utána a ... az **az inas** ...*
- Synonym (szinonímia)  
*(8) Aztán **a hölgy** le akar ülni (...) Ésö lekuporodik **a nő** és fölötte vívnek.*
- Verbal (igei)  
*(9) Hát az , hogy **kiment** ez a részeg ember , ki **ki** . .**osont** vagy **ki** . .**suttyogott** **sunnyogott**.*
- Other (egyéb)  
*(10) ...**valószínű** hogy **gyűrűt** vett a férfi elő . Kis dobozba ... volt **valami**.*

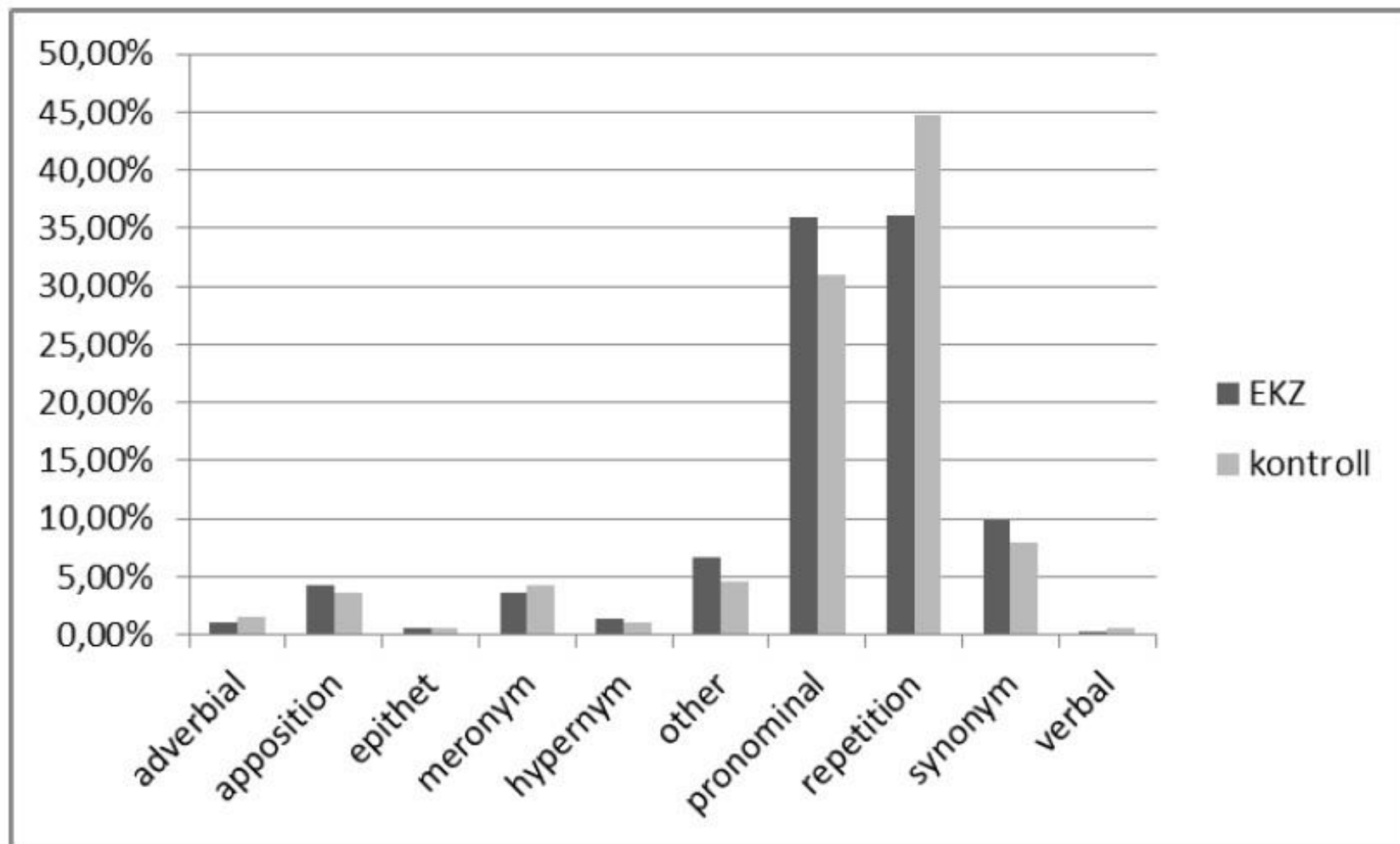
# VI. Statisztikai adatok

## 1) Szöveggel kapcsolatos statisztikai adatok

	EKZ		Kontroll	
	1-es szöveg	2-es szöveg	1-es szöveg	2-es szöveg
Láncok száma	231	256	198	189
Láncok átlagos hossza	2,30736	2,32422	2,32323	2,41800
Leghosszabb szöveg (szó)	369	222	206	202
Legrövidebb szöveg (szó)	24	29	37	52
Átlagos szöveghossz (szó)	109,14	95,71	104,37	97,86

# VI. Statisztikai adatok

## 2) Kategóriákkal kapcsolatos adatok



Kategóriák eloszlása a két csoportban

Khi<sup>2</sup> p= 0,0100

Kategóriák	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%



Kategóriák	T-próba az 1-es szövegen	T-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

## VII. Gépi tanulási kísérletek

- Vincze, Hoffmann, Szatlóczki, Bíró, Gosztolya, Tóth, Pákáski, Kálmán Szegedi Tudományegyetemen zajló projektjét (Vincze et al. 2016) kiegészítve gépi tanulási kísérleteket végeztem.
- Az általuk használt jellemzőkészletből statisztikailag szignifikánsnak bizonyult a szavak száma, hezitációk száma, bizonytalan szavak száma, ismeretlen szavak száma, hezitációk aránya, bizonytalan szavak aránya, mondatbeli szavak száma, néma szünetek száma, nyújtások száma, főnevek száma, írásjelek száma.
- A teljes jellemzőkészletük használatával 69,1%-os pontosságot értek el az SVM algoritmussal (Cortes–Vapnik 1995).

- A Weka szoftvert használtam fel (Hall et al. 2009).
- Az SZTE-n futó projekt jellemzőkészletét egészítettem ki.
- Minden beszélő esetében hozzáadtam jellemzőként az általa használt egyes koreferenciatípusok számát, az egyes szövegekre lebontva.
- Nagyobb mintát felhasználva az SVM algoritmus 75,58%-os pontosságot ért el.

## VIII. Konklúzió

1. Az EKZ-ban szenvedők kevesebb koreferencialáncot használnak.
2. A koreferencialáncok hosszai között nem mutatkozott nagyobb eltérés.
3. A két csoport között szignifikáns eltérést mutat az ismétlés használata, valamint a kategóriák eloszlása csoporton belül.
4. A kontroll csoport közel azonos teljesítménnyel végezte el a feladatot, az EKZ-ban szenvedők csoportjának tagjai között nagyobb teljesítménybeli eltérések vannak.
5. A vizsgálat javítja a korai diagnosztizálás lehetőségét.

# Felhasznált irodalom

- Kempler, D. (2005) *Neurocognitive disorders in aging*. CA Sage Publications, Thousand Oaks.
- Gósy Mária (2005) *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Baddeley, A., Eysenck, M. W., Anderson, M. C. (2010) *Emlékezet*. Akadémiai Kiadó, Budapest.
- 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. *Accept: Association for Computational Linguistics*, Berlin.
- 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.
- Kiefer Ferenc, Siptár Péter (2003) *A Magyar nyelv kézikönyve*. Akadémiai Kiadó, Budapest.
- Cortes, C., Vapnik, V. (1995) Supportvector networks. *Machine Learning*, 20(3) Kluwer Academic Publishers, Boston, 273–97.
- 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) ACM, New York 10–18.