

## Szemantikus szerepek vizsgálata magyar nyelvű szövegek narratív pszichológiai elemzésében<sup>1</sup>

Ehmann Bea<sup>1</sup>, Lendvai Piroska<sup>2</sup>, Fritz Adorján<sup>3</sup>, Miháltz Márton<sup>2</sup>, Tihanyi László<sup>2</sup>

<sup>1</sup> MTA Pszichológiai Kutatóintézet  
1132 Budapest, Victor Hugó u. 18-22.  
{ehmannb}@mtapi.hu

<sup>2</sup> Nyelvtudományi Intézet,  
1068 Budapest, Benczúr u. 33.

{piroska, tihanyil}@nytud.hu; {mmihaltz}@gmail.com

<sup>3</sup> Pécsi Tudományegyetem, Pszichológiai Intézet  
7624 Pécs, Ifjúság útja 6.  
{kifino}@gmail.com

**Kivonat:** A narratív pszichológiai tartalomelemzés és a korpusznyelvészet több éve folytatott közös projektje a szemantikus szerepek és a narratív pszichológiai modulok összekapcsolása egyének és csoportok énelbeszéléseinek elemzéséhez. A két munkacsoport korábbi együttes fejlesztései a szemantikus szerepek felismerését úgy oldották meg, hogy a MetaMorpho nyelvi elemzés morfoszintaktikai és szemantikai kimenetét összekapcsolták a NooJ eszköz procedúráival. A jelen munka célja, hogy a korábbi törekvések továbbfejlesztésével magyar nyelvű szövegekben felismerhető váljon az ágencia, és ennek nyelvi kifejező elemeihez automatikusan hozzárendelhetőek legyenek az Ingroup/Outgroup pszichoszemantikai kategóriák. Ekképp a tudományos narratív pszichológia a *semantic role labeling* nyelvészeti terület új alkalmazójaként jelenik meg.

### 1 Miért fontos a tudományos narratív pszichológia számára a szemantikus szerepek vizsgálata?

A Tudományos Narratív Pszichológia (TNP) a szelf- és csoportnarratívákban azonosítható pszichológiai jelenségek longitudinális, kvantitatív vizsgálatára szolgáló, Magyarországon kifejlesztett elmélet, melynek számos empirikus alkalmazása létezik a szociálpszichológia, a személyiség- és a klinikai pszichológia területén [9]. Az elmélet módszere, a Narratív Pszichológiai Tartalomelemzés (NPTA) fejlődésének alapja a magyar korpusznyelvészekkel és nyelvtechnológusokkal történő együttműködése, melynek során a Narratív Pszichológiai Munkacsoport számos pszichoszemantikai taxonómiát és algoritmust fejlesztett ki [18,10,9] a NooJ nyelvészeti fejlesztési környezet keretében [16].

---

<sup>1</sup> A kutatást az OTKA 81633K pályázat támogatta.

Az eddig kifejlesztett NPTA-algoritmusok, TNP-modulok a következők: *Aktivitás-Passzivitás* [17], *Érzelem* [7], *Kognitív folyamatok* [21,20], *Értékelés* [1,2], *Intencionalitás* [6], *Tagadás; Én- és Mi Referencia* [8], *Perspektíva* [13], valamint a *Szubjektív Időélmény* [5].

E fejlesztésekről és a velük kapott empirikus eredményekről az elmúlt évek során a Munkacsoport a Számítógépes Nyelvészeti Konferenciákon és nemzetközi közleményekben is széleskörűen beszámolt<sup>2</sup>.

A Narratív Pszichológiai Tartalomelemző NooJ algoritmusok (modulok) a TNP két fő területén használatosak. A strukturális megközelítés azt vizsgálja, hogy a vizsgált kategóriák – elsősorban az elbeszélői perspektíva, az időélmény és az értékelés – miképpen változnak az énelbeszélések és a csoportelbeszélések egészének belső szerkezetében [4,14,13].

A másik vizsgálódási kör a mintázatelemzés, ami az egyes szógyakoriságok együttjárásából von le pszichológiai következtetéseket: ennek egyik példája, hogy kiscsoportok beszámolóiban a negatív érzelemmarkerek és a selfreferencia magas együttes aránya csoporton belüli konfliktust jelez; a negatív érzelemmarkerek és a mi-referencia magas aránya viszont együttesen a csoport fenyegetettségére utal [3].

A tudományos narratív pszichológia annyiban lép túl a hagyományos pszichológiai tartalomelemzési koncepción, hogy nem elégszik meg a pszichológiai tartalmak puszta számlálásával és strukturális vagy mintázatelemzésével, hanem azt is vizsgálja, hogy az adott érzelem, kogníció vagy cselekvés milyen cselekvőhöz, illetve milyen elszenvedőhöz tartozik. Minthogy a Narratív Pszichológiai Munkacsoport kiemelt kutatási területe a nemzeti és európai identitás vizsgálata, sarkalatos kérdés, hogy valamely történelmi esemény vagy korszak beszámolóiban a TNP által vizsgált kategóriák a saját csoporthoz vagy a külső csoporthoz tartoznak.

A cselekvő és az elszenvedő kérdésköre a pszichológiában hagyományosan az ágenciakutatás területéhez kapcsolódik. A személyiség- és a klinikai pszichológiában ez főként az énhatékonyság megítélésében fontos, a szociálpszichológiában pedig a humán ágens és a humán elszenvedő egyén vagy csoport nyelvi megjelenítése vagy ennek hiánya a társas-társadalmi-hatalmi felelősség felvállalását, hátrítását vagy elkenődését teszi vizsgálhatóvá.

Ezért fontos a TNP számára a szemantikus szerepek (Semantic Role Labeling) vizsgálatára szolgáló elemzőeszköz kifejlesztése.

## 2 A magyar és európai történelem narratív pszichológiai korpuszai

Az MTA Pszichológiai Intézetének Oral History Archívumában a következő elektronikus korpuszok állnak rendelkezésre:

- *Történelemkönyv korpusz*: a magyar történelemről szóló könyvek részletei a 10 legfontosabb eseményről, 1900-tól 2000-ig, 10 éves bontásban (kb. 200000 szó);

<sup>2</sup> Cf. <http://narrativpszichologia.pte.hu>

- *Történelmi regény korpusz*: nemzetek közötti konfliktusokról szóló 6 történelmi regény teljes szövege (kb 700000 szó);
- *Történelem tankönyv korpusz*: általános és középiskolai tankönyvek részletei a 10 legfontosabb eseményről (kb. 210000 szó);
- *Néphistóriai korpusz*: Félig strukturált interjúk a legpozitívabb/legnegatívabb magyar és európai történelmi eseményekről 500 fős rétegzett mintán (kb. 120000 szó).

A két utolsó korpuszt az MTA Nyelvtudományi Intézetének korábban már átadtuk; ezek annotálása számos vonatkozásban már megtörtént. Ezek szolgálnak alapul a szemantikus szerepek vizsgálatára szolgáló fejlesztésekhez.

### 3 A pszicho-szemantikai szerepek vizsgálatának problematikája

Adott tehát egy elméleti paradigma (a tudományos narratív pszichológia), egy kutatási módszer (a narratív pszichológiai tartalomelemzés), egy magyar nyelvű szövegkorpusz (a történelemszövegek). Amit első körben keresünk, az az, hogy terjedelmes szövegkorpuszokból automatikusan olyan konkordanciákat hozzunk létre, melyek kilistázzák, hogy ki cselekszik, ki érez, ki gondol és ki értékeli.

A történelem szövegek vizsgálatokor a 'ki' nem csupán személy lehet – például 'Mátyás király' – hanem csoport is – például 'a tatár hadak'. A narratív szociálpszichológia számára nem az fő kritérium, hogy személy vagy csoport cselekszik-e/érez-e, stb., hanem az, hogy saját csoport (ingroup) vagy külső csoport (outgroup) teszi-e ezt. Ezért az egyes személyek is ingroup, illetve outgroup címkét kapnak.

Kutatásunk egyik konceptuális nehézsége, ami majd a finomabb vizsgálatoknál fog szerepet játszani, hogy az 'ingroup' és az 'outgroup' kategória egyaránt több alhalmazból tevődik össze, például attól függően, mennyire aprólékos, illetve milyen kognitív doménra fókuszál a szövegíró megközelítése. Többnyire egységes ingroupként jelennek meg 'a magyarok' vagy 'az Árpád-házi királyok' olyan szövegrészletekben, amikor „külsőkről”, vagyis a szervezeten kívüli azonos nagyságrendű outgroup szereplőkkel való interakcióról olvasunk (pl. 'besenyők', 'jászok', 'keleti lovas népek', 'kun törzsek', 'Európa hatalmasságai', 'német császár', 'orosz fejedelemség' 'újak', stb). Természetesen a „belügyekről” szóló tudósításokor a magyar színtér szereplői finomodnak ('a főurak', 'Béla király', 'a trónörökös István herceg' stb.), és ezek kétpólusú csoportba sorolásának automatizálása komoly szakmai kihívást jelent úgy a pszichológus, mint a nyelvtechnológus számára.

Részben ez az oka annak, hogy egy egyetemes vagy kutatási igény szerint különböző történelmi korokra lebontott, robusztus 'Ingroup-Outgroup' szólista elkészítése korántsem triviális, hiába tűnik úgy, hogy pl. a 'Tatár Outgroup' szótár nagyjából véges számú elemből és azok variálódásából áll össze ('Dzsingisz kán', 'Batu kán', 'nagykán', 'mongolok', 'mongol törzs', 'nomád sereg', 'tatár hadak', 'tatár hordák', 'tatárok', stb.). A kézzel összeállított szólisták a variabilitás miatt nagy időbefektetés árán tudják csak a releváns entitásokat lefedni a korpuszokban (pláne az

egyelőre feldolgozatlanokban) előfordulókból. Továbbá, leggyakrabban csakis szöveggörnyezet vagy egyéb, nem objektív kritérium/megegyezés alapján lehet eldönteni, hogy egy entitás melyik csoportba tartozik. Ezért fejlesztésünk során a digitális bölcsészeti kutatásokra jellemző félautomatikus módszerrel dolgozunk, ami a kutatási és implementációs folyamat fontos részeként az automatikus feldolgozás részeredményei után, meghatározott fázisokban, lokális vizsgálattal elvégzett kézi egyértelműsítést és javítást foglal magába.

A továbbiakban a probléma nyelvtechnológiai megközelítéséről és modellezéséről számolunk be. Módszerünk a MetaMorpho nyelvi elemzés morfoszintaktikai és szemantikai kimenetét kapcsolja össze a NooJ eszköz procedúráival. Nyelvészeti szempontból az egyik legnehezebb feladat a szövegekben a koreferencia (illetve az anafora) feloldása, mivel az egyes szereplőket több kifejezés is jelezheti (például: 'IV. Béla leánya' = 'a király leánya' = 'Árpádházi Szent Margit'.) További aspektusa az entitások felismerésének a metonimikus használat, vagyis, hogy önmagukban élettelen dolgok is ágensként, aktív szereplőként említődhetnek – például 'az egri vár hősiessen ellenállt', stb. Ezekkel a jelenségekkel jelenleg csak marginálisan foglalkozunk, mivel többszörösen összetett technológiai megoldást igényelnek.

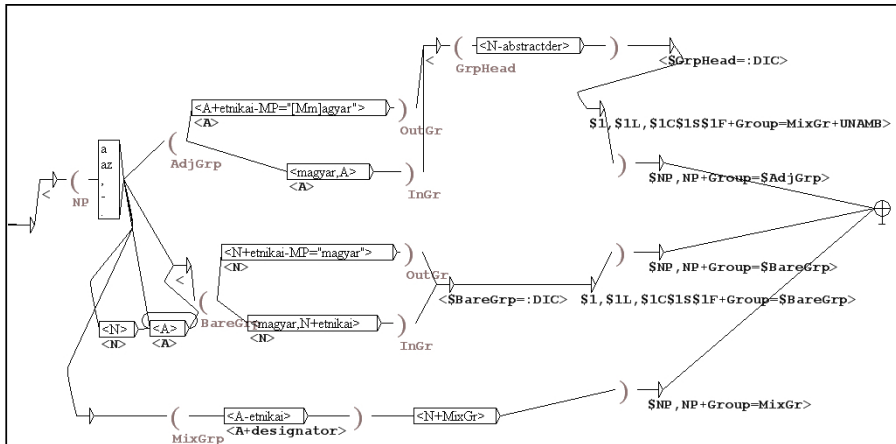
#### 4 Pszichoszemantikai szerepek annotációja

Az automatizált annotációs folyamat kiindulópontja a 'Mixgroup' kategória, vagyis az olyan lexikális elemek, amik az Ingroup vagy Outgroup kategóriához egyaránt tartozhatnak, ilyen például a 'sereg', 'lakosság', 'ország', 'lovasság', stb. A szövegekben ezek megjelenhetnek önmagukban, illetve pl. névelővel és/vagy jelzőkkel együtt, egy NP fejeként. Korábbi munka során elkészült az etnikai főnevek/mellékevek szótára, amit a NooJ fejlesztői környezetben a Mixgroup entitások modellezésében felhasználunk. Létrehoztunk egy NooJ egyértelműsítő prototípus-gráfot (l. 1. ábra), ami

- eldönti, hogy mikor áll főnévi és mikor jelzői szerepben egy etnikai entitás (pl. 'a törökök', ill. 'a török szultán'), és ezeket az NP-eket InGr, illetve OutGr szemantikai címkével látja el;
- begyűjti az összes olyan fejet, ami etnikai jelzővel áll és N+MixGr-ként címkézi őket;
- a MixGr címkéjű főneveket módosító, de eddig az etnikai szólistában nem szereplő jelzőket egy speciális osztályba sorolja, ami azt fogja jelezni, hogy utána egy potenciálisan InGr vagy OutGr elem következhet.

Így egy kb. 600 szóból álló NP halmazhoz jutunk csak a "tatár korpuszon", amit a NooJ-ban félautomatikus szótárrá alakítunk: az NP-k lemmájukkal és morfológiai jegyeikkel együtt egy külön szótárban eltároljuk. A szótárt a következő elemzési fázisokban, illetve új korpuszok elemzésekor használjuk fel. A fent leírt eljárás azért fontos, mert a szintaktikai szövegelemzők gyakran csak az NP-fejet írják ki; mi olyan gyakorlatias megközelítést választottunk, ami feltételezi, hogy a magasszintű (szemantikai) elemzés során egy már meglévő előfeldolgozó eszköz kimenetére

támaszkodunk, amibe nincs lehetőségünk belenyúlni (vagyis ‘black box’-ként érhető el).



1. ábra: NooJ szintaktikai-szemantikai egyértelműsítő prototípusgráf.

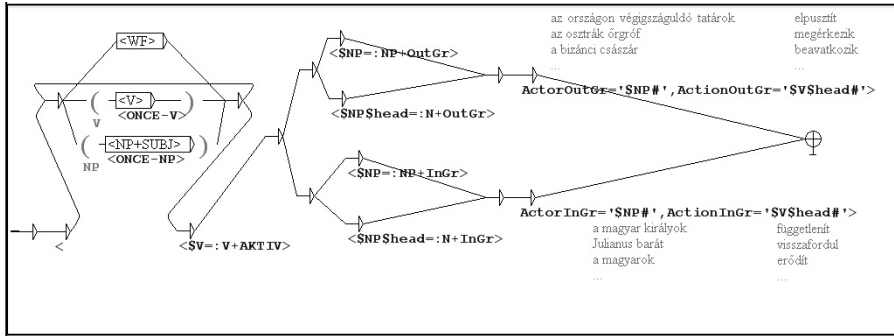
Az entitásdetektáló kör további lépesei:

- az egyértelműsítő/címkéző gráfnak a nagyobb lefedés érdekében történő kiterjesztése;
- egyéb mechanizmusok beépítése, pl. hogy megtaláljunk olyan képzett szavakat, amelyeket a NooJban morfológiai lekérdezés alapján az aktivitás stb. szótárba felvehetünk: ‘a mongolok előretöréséről’ > ‘előretör’;
- a szemantikailag is anaforikus MixGr kifejezések kézi egyértelműsítése NooJ konkordancia alapján, ahol szövegösszefüggésből lehet eldönteni, hogy az adott csoport saját csoportot vagy külső csoportot jelöl: pl. ‘Batu kán visszavonta a *katonáit*’. Sokszor mondathatáron túl átnyúló anaforáról van szó (pl. ‘Az *uralkodó* nehéz helyzetbe került.’), ekkor a NooJ kontextusablakának a méretét nagyobb szószámra lehet állítani. A NooJ-ban az illetett kategóriákhoz tartozó lexikai elemekhez egyben az annotációk is hozzárendelhetők.

## 5 Thematikus szerep hozzárendelése

Nyelvtechnológiai megközelítésben a thematikus szerep-felismerés (‘semantic role labeling’) kínálkozik alkalmas megoldásként az ágencia detektálására, viszont magyar nyelvre egyelőre nincs létrehozva thematikus szerep-felismerésben felhasználható strukturált szemantikai erőforrás vagy annotált korpusz [lásd pl. 11]. A MetaMorpho [15] képes bizonyos igékhez tartozó thematikus szerepeket felismerni, amelyet szabályalapú koreferenciafeloldással is támogat [12], habár csak viszonylag kevés számú ige mellett. A Vincze Orsolya és Gábor Kata által megalkotott NooJ protézisgráf a MetaMorpho által thematikus szereppel felcímkézett igei bővítményeket találja meg

[19]. Ezen túlmutatva, jelen munkánk célja, hogy a már meglévő erőforrásokhoz igazodva úgy ismerjük fel az ágenciát, hogy ehhez felhasználjuk a történelmi szövegek lexikális elemeihez gráfokkal automatikusan hozzárendelt pszichoszemantikai kategóriákat. A MetaMorpho által felismert főnévi csoportokat és azok mondatban betöltött szerepét kódoló XML-fájlt importáljuk a NooJ-ba, mely után a szótárakkal, illetve szintaktikai mintaillesztésével az In/Outgroup entitások Ágens szerepét igyekszünk meghatározni (l. 3. ábra).



2. ábra: Az ágenciát pszichoszemantikus csoportok alapján szűrő NooJ gráf.

1. Az Aktivitás NPTA gráfot a NooJ elemzőfolyamatban kiemelt szintaktikai elemzőként beállítva a "+AKTIV" címkét kapott igékre szorítjuk a keresést. Így automatikusan kiszűrjük a találatok közül az olyan tartalmú mondatokat, ahol az alany nem cselekvő, pl. 'László király a kunok között érezte jól magát.', 'A főurak közül sokan örültek a király bajának'. Bár megjegyezzük, hogy a történelmi szövegekben előforduló entitások főképp cselekvőként vagy szenvedőként szerepelnek, és a nem aktív igék viszonylag ritkán, illetve nem az általunk vizsgált "etnikai" entitásokkal fordulnak elő, pl. 'A páncélos katonaság mellett nőtt a könnyűlovasság száma is.', 'A mongol sereget nem egészen helyesen, általánosítva nevezük - előcsapataikról - tatár seregek'.
2. A fent elkészített Group szótár alapján lehetővé tesszük a keresést In/Outgroup entitásokra lebontva, anélkül, hogy ezek lexikai alakjait a gráfba kódolnánk, illetve a szótár új korpuszokon történő iteratív bővítése alapján egy növekvő főnévi lemma- és NP lista, és a szövegek kézzel egyértelműsített annotálása alapján.

## 6 A fejlesztés további lépései és alkalmazási lehetőségei

Soron következő lépésünk egyrészt a szereplői szótárak és a narratív pszichológiai tartalomelemzési modulok (jelen esetben főként az érzelem, a kogníció és az értékelés) összeillesztése és továbbfejlesztése lesz, másrészt az igei argumentumok tematikus szerepének meghatározása a NooJ-ban írt lokális grammatikák alapján, például:

*ha V+AKTIV és Group(SUBJ) == OutGr,  
akkor Th\_role(OBJ) = Undergoer és Group(OBJ) = InGr.*

Ezzel a terjedelmes szövegkorpuszokból automatikusan olyan konkordanciák hozhatók létre, melyek nem csupán azt listázzák, hogy ki cselekszik, ki érez, ki gondol és ki értékkel, hanem azt is, hogy ki mindezeknek a tárgya vagy elszenvetője/kedvezményezettje (agent vs. patient/undergoer). Ezek statisztikai feldolgozása révén vonhatók le az egyéni és a csoportidentitással kapcsolatos narratív pszichológiai következtetések.

## Hivatkozások

1. Bigazzi S., Csertő I., Nencini, A.: A személy- és csoportközi értékelés pszicholingvisztikája. In: Alexin Z., Csendes D. (szerk.): IV. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem Informatikai Tanszékcsoport (2006) 267–276
2. Csertő I.: A személy- és csoportközi értékelés pszichológiai szempontú elemzése elbeszélő szövegekben. In: Alexin Z., Csendes D. (szerk.): VI. Magyar Számítógépes Nyelvészeti Konferencia, Szegedi Tudományegyetem Informatikai Tanszékcsoport (2009) 272–284
3. Ehmann, B., Balázs, L., Fülöp, É., Hargitai, R., Kabai, P., Péley, B., Pólya, T., Vargha, A., Vincze, O., László, J.: Narrative Psychological Content Analysis as a Tool for Psychological Status Monitoring of Crews in Isolated, Confined and Extreme Settings. *Acta Astronautica*, Vol. 68, No. 9-1 (2011) 1560–1566
4. Ehmann, B., Garami, V.: Narrative Psychological Content Analysis with NooJ: Linguistic markers of time experience in Self reports. In: *Proceedings of the 2008 International NooJ Conference*. Cambridge Scholar Publishing (2010) 180–190
5. Ehmann, B., Garami, V., Naszódi, M., Kis, B., László, J.: Subjective Time Experience: Identifying Psychological Correlates by Narrative Psychological Content Analysis. *Empirical Text and Cultural Research* Vol. 3 (2007) 14–25
6. Ferenczhalmy R., László J.: Az intencionalitás modul kidolgozása NooJ tartalomelemző programmal. In: Alexin Z., Csendes D. (szerk.): IV. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem Informatikai Tanszékcsoport (2006) 285–295
7. Fülöp É., László J.: Az elbeszélések érzelmi aspektusának vizsgálata tartalomelemző program segítségével. In: Alexin Z., Csendes D. (szerk.): IV. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem Informatikai Tanszékcsoport (2006) 296–304
8. Hargitai, R., Naszódi, M., Kis, B., Nagy, L., Bóna, A., László, J.: Linguistic Markers of Depressive Dynamics in Self Narratives: Negation and self reference. *Empirical Text and Cultural Research* Vol. 3 (2007) 26–38
9. László, J.: *The Science of Stories: An introduction to Narrative Psychology*. Routledge, London, New York (2008)
10. László, J., Ehmann, B., Péley, B., Pólya, T.: Narrative psychology and narrative psychological content analysis. In: László, J., Stainton Rogers, W. (eds.): *Narrative Approaches in Social Psychology*. New Mandate, Budapest (2002) 9–25
11. Márquez, L., Carreras, X., Litkowsky, K. C., Stevenson, S.: Semantic Role Labeling: An Introduction to the Special Issue. *Computational Linguistics* Vol. 34, No. 2 (2008) 145-159
12. Miháltz, M.: Knowledge-based Coreference Resolution for Hungarian. In: *Proceedings of The Sixth International Conference on Language Resources and Evaluation*. Marrakesh, Morocco (2008)

13. Pólya, T., Kis, B., Naszódi, M., László, L.: Narrative perspective and the emotion regulation of a narrating person. *Empirical Text and Cultural Research* Vol. 3 (2007) 50–61
14. Pólya, T., László, J. and Forgas, J. P.: Making sense of life stories: The role of narrative perspective in communicating hidden information about social identity. *European Journal of Social Psychology* Vol. 35, No. 6 (2005) 785–796
15. Prószték, G., Tihanyi, L.: MetaMorpho: A Pattern-Based Machine Translation System. In: *Proceedings of the 24th 'Translating and the Computer' Conference*. ASLIB, London, United Kingdom (2002) 19–24
16. Silberstein, M.: NooJ Manual (2003) Elérhetőség: [www.nooj4nlp.net](http://www.nooj4nlp.net)
17. Szalai, K., László, J.: Activity as a Linguistic Marker of Agency: Measuring in-Group versus Out-group Activity in Hungarian Historical Narratives. *Empirical Text and Culture Research* RAM-Verlag: 4 (2010) 50–58
18. Váradi, T.: The Hungarian National Corpus. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas de Gran Canaria (2002) 385–389
19. Vincze O., Gábor K., Ehmann B., László J.: Technológiai fejlesztések a NooJ pszichológiai alkalmazásában. In: *VI. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Egyetemi Kiadó (2009) 285–294
20. Vincze, O., László, J.: Narrative Means of Intergroup Relations: Cognitive States and their role in reducing or increasing intergroup conflict. In: *General Meeting of the European Association of Social Psychology (EASP)*. Stockholm (2011) 30
21. Vincze, O., Tóth, J., László, J.: Representations of the Austro-Hungarian Monarchy in the history books of the two nations. *Empirical Text and Cultural Research* Vol. 3 (2007) 62–71