

Označevanje vrste diskurznih označevalcev

Darinka Verdonik

Fakulteta za elektrotehniko, računalništvo in informatiko
Smetanova 17, SI-2000 Maribor
darinka.verdonik@uni-mb.si

Povzetek

Razvoj zahtevnejših jezikovnotehnoloških aplikacij in uporaba korpusov v pragmatičnih in diskurznih raziskavah spodbujata označevanje različnih diskurznih in pragmatičnih pojavov v jezikovnih virih. V tem prispevku obravnavam označevanje diskurznih označevalcev in predlagam ločevanje štirih vrst: ideacijskih označevalcev, interakcijskih označevalcev, označevalcev procesov tvorjenja in interpretacijskih označevalcev. Predstavljena shema je osnova za nadaljnjo korpusno podprto analizo diskurznih označevalcev in za evalvacijo tako označenih virov v zahtevnejših jezikovnotehnoloških aplikacijah.

Annotating discourse marker type

With the demand for more powerful NLP applications and for the use of corpora in pragmatic and discourse studies comes a need for discourse and pragmatic attributes in language resources. In this paper, I focus on the annotation of discourse markers. I propose a classification of discourse markers which consists of four categories, ideational markers, interactional markers, markers of production processes and interpretation markers. The classification is a foundation for further corpus based analysis of discourse markers and for the evaluation of language resources in NLP applications.

1. Uvod

Diskurzna in pragmatična raven postajata vse bolj pogosto predmet označevanja v jezikovnih virih, zlasti ko skušamo razvijati zahtevnejše jezikovnotehnološke aplikacije, kot so razpoznavanje spontanega govora, prevajanje govora, dialog ipd. Najdemo lahko vedno več poskusov, kjer skušajo razviti sheme za označevanje teh ravni v jezikovnih virih (Carlson et al., 2003; Mitkov et al., 2000; Muller et al., 2002; Byron et al., 1997; Heeman, Allen, 1999; Miltsakaki et al., 2002). Čeprav se diskurz in pragmatika uvrščata tudi v pregledne publikacije s področja jezikovnih tehnologij (npr. Mitkov, 2003), pa je na področju diskurznega in pragmatičnega označevanja jezikovnih virov še veliko prostora za nove raziskave in razprave, saj še ni jasno niti, ali bosta diskurz in pragmatika dve različni ravni označevanja ali ena sama oziroma ali in kako naj se standardno označujejo različni diskurzni in pragmatični pojavi, kot so navezovanje, kohezija, diskurzni označevalci, struktura pogovora, govorna dejanja, struktura informacije, retorična razmerja, namernost idr.

V tem prispevku se bom osredotočila na označevanje samo enega od navedenih pojavov, diskurznih označevalcev. V preteklosti je že bilo predstavljenih nekaj shem in poskusov označevanja diskurznih označevalcev (Heeman, Allen, 1999; Miltsakaki et al., 2002; Verdonik et al., 2007), vendar nadaljnje raziskave tovrstnih izrazov, ki so v analizi diskurza zelo aktivne, prinašajo vedno nova spoznanja in vedno širši in raznovrstnejši nabor diskurznih označevalcev (npr. Schiffrin, 1987; Redeker, 1990; Fraser, 1996; Blakemore, 2002; Overstreet, 2005; Swerts, 1998; Clark, Fox Tree, 2002; od slovenskih raziskav npr. Gorjanc, 1998; Smolej, 2004; Pisanski Peterlin, 2005; Schlamberger Brezar, 2007). Kot bomo videli v nadaljevanju, funkcije, ki jih opravljajo posamezni diskurzni označevalci, pokrivajo zelo širok spekter in segajo na različne ravni diskurza, prav tako so izrazi v vlogi diskurznih označevalcev kar se da raznovrstni. Zato se že dalj časa kaže potreba po njihovi nadaljnji razvrstitvi.

V tem prispevku predlagam podkategorijo vrst pri označevanju diskurznih označevalcev v jezikovnih virih in predstavim eno od možnih razvrstitev. V shemah za označevanje jezikovnih virov se skušata zagotavljati enostavnost in nedvoumnost, kar omogoča homogeno, hitro in v kasnejših fazah zadovoljivo uspešno avtomatsko označevanje. Zato predstavljena razdelitev vrst diskurznih označevalcev teži k preprostosti in poenostavitvam, ki temeljijo na najbolj izrazitih skupnih lastnostih. Shema ni mišljena kot natančna jezikoslovna tipologija diskurznih označevalcev, ampak kot osnova za nadaljnjo korpusno podprto analizo diskurznih označevalcev v jezikoslovju, na področju tehnologij pa pomeni osnovo za evalvacijo uporabnosti atributa diskurznih označevalcev.

2. Definicija diskurznih označevalcev in metoda

Obstajajo različne interpretacije diskurznih označevalcev, tj. izrazov, kot so npr. *ja*, *mhm*, *aha*, *no*, *dobro*, *v redu*, *glejte*, *eee*, *in*, *torej* in še številni drugi. Med temeljne pristope lahko štejemo interakcijsko-socialingvističnega (Schiffrin, 1987; Redeker, 1990), slovnično-pragmatičnega (Fraser, 1996) ter relevantnostnega (Blakemore, 2002). Če jih primerjamo, ugotovimo, da avtorji praviloma ločujejo med dvema temeljnima ravnema diskurza: predstavno, ideacijsko oz. propozicijsko ter proceduralno, pragmatično oz. komunikacijsko. Pri tem so diskurzni označevalci vedno predstavljeni kot izrazi, ki ne prispevajo veliko k vsebini in ne vplivajo bistveno na pomen sporočila, ampak opravljajo predvsem komunikacijske, pragmatične oz. proceduralne funkcije. To sicer zelo ohlapno definicijo lahko štejemo kot skupno izhodišče preučevanja diskurznih označevalcev.

V nadaljevanju predlagam, da tako definirane diskurzne označevalce delimo na štiri vrste: ideacijske, interakcijske, označevalce procesov tvorjenja in interpretacijske označevalce. Delitev skušam utemeljiti na treh ravneh: z označevanjem in analizo avtentičnega gradiva, s sintezo spoznanj tujih raziskav diskurznih označevalcev ter s teoretičnim izhodiščem, po katerem je diskurz zgrajen iz več temeljnih ravni.

V nadaljevanju najprej utemeljum delitev s teoretičnimi izhodišči in sintezo tistih spoznanj drugih raziskav, ki podpirajo predstavljeno razdelitev, v drugem delu članka pa s predstavitev rezultatov označevanja vrst diskurzni označevalcev v korpusnem gradivu.

3. Teoretična izhodišča

Številne raziskave diskurza in jezikovne rabe se osredotočajo na različne ravni diskurza oz. jezikovne rabe, pri čemer lahko ločimo vsaj naslednje: predstavno oz. ideacijsko raven, interakcijsko oz. sociološko raven ter kognitivno oz. mentalno raven.

Kot kažejo lastne predhodne raziskave (npr. Verdonik, 2006; Verdonik et al., 2007), na vseh teh ravneh učinkujejo tudi diskurzni označevalci; običajno učinkujejo na več ravneh hkrati, vendar je, kot ugotavlja že Schiffrin (1987), praviloma ena od ravni bolj poudarjena in zato primarna. Glede na raven diskurza, na kateri primarno učinkujejo, ločim ideacijske, interakcijske in kognitivne diskurzne označevalce, slednje pa naprej na označevalce procesov tvorjenja in interpretacijske označevalce.

3.1. Ideacijski in interakcijski označevalci

Schiffrin (1987) ločuje pet ravni diskurza, na katerih delujejo diskurzni označevalci: menjavanje vlog, struktura dejanj, okvir sodelovanja, predstavna struktura, informativnost. Vendar na koncu ugotavlja, da označevalci s semantičnim pomenom, kot so vezniki in časovni deiksi, delujejo predvsem na ravni predstavnostne strukture, in tisti brez semantičnega pomena na ostalih ravneh. Takšne ugotovitve vodijo k zaključku, da obstaja večja razlika med označevalci predstavnostne strukture (npr. *in, pa, torej, in sicer ...*) in vsemi ostalimi označevalci (npr. *mhm, ja, no, dobro, glejte ...*). Tako sklepa tudi Redekerjeva (1990), ki z nadgradnjo modela Schiffrinove (1987) loči označevalce ideacijske strukture in označevalce pragmatične strukture. Tukaj za označevalce pragmatične strukture uporabljam izraz interakcijski označevalci, saj lastne predhodne raziskave (Verdonik et al., 2007; Verdonik et al., v tisku) kažejo, da je njihova osrednja lastnost vzpostavljanje in razvijanje odnosa med sogovorniki, izraz pragmatičen pa je lahko zelo široko in različno razumljen. Označevalci, ki imajo primarno povezovalno vlogo in so usmerjeni predvsem v besedilo diskurza, manj pa v medosebne odnose, so večinoma vezniškega in prislovnega izvora in jih imenujem ideacijski. Poudarjena povezovalna vloga različnih vezniških in prislovnih, pa tudi drugih besed v slovenskem jeziku je bila raziskana npr. že v Gorjanc (1998), vendar nas tukaj v nasprotju z navedeno raziskavo zanima povezovalna vloga teh izrazov samo na ravni diskurza in v skladu z definicijo diskurzni označevalcev.

3.2. Označevalci procesov tvorjenja in interpretacijski označevalci

Pri nekaterih diskurzni označevalcih je bolj kot interakcijska ali predstavna poudarjena kognitivna raven. Sem spadajo večinoma izrazi, ki tradicionalno (npr. v Schiffrin, 1987; Fraser, 1999; Blakemore, 2002) niso obravnavani kot diskurzni označevalci.

Prva podskupina kognitivnih označevalcev so označevalci procesov tvorjenja. Sem spadajo predvsem t. i. zapolnjevalci vrzeli oz. mašila (*eee, mmm* idr.), s katerimi govorec opozarja, da išče ustrezno besedo, se

odloča, kaj bo rekel v nadaljevanju, pa tudi skuša ohraniti vlogo oz. prevzeti vlogo (Clark, Fox Tree, 2002; Swerts, 1998; Verdonik, 2007). Prav tako je kazanje na raven tvorjenja primarna funkcija različnih diskurzni izrazov z glagoli rekanja in vedenja (*bom rekel, ne vem ...*). Čeprav tudi označevalci procesov tvorjenja do neke mere delujejo na interakcijski ravni, pa je bolj poudarjena funkcija razkrivanja procesov tvorjenja, ki za interakcijske označevalce ni značilna.

Druga podskupina kognitivnih označevalcev, interpretacijski označevalci, ima primarno vlogo na ravni interpretacije. Gre predvsem za izraze, ki so bili v angleščini raziskovani pod termini *general extenders, co-ordination tags, set markers, discourse extenders* idr. in se rabijo praviloma na koncu izjave, začenjajo pa se ali z besedico *in* oz. *pa* ali z *ali* (npr. *in tako naprej, pa tako, ali pa kaj*). Overstreetova (2005) definira naslednje osnovne funkcije teh izrazov: signalizirajo predpostavko, da naslovnik ve, kaj ima tvorec v mislih, in da zato nadaljnje tvorjenje v nakazani smeri ni potrebno; spodbujajo naslovnika k solidarnosti, naj se vživi situacijo, ki jo tvorec opisuje; nakazujejo, da bi lahko rekli še veliko več o predmetu pogovora oz. da bi lahko rekli še več, ampak je tisto nepomembno; opozarjajo, da to, kar je bilo rečeno, ni povsem natančno; ublažijo izjave, ki bi lahko prizadele naslovnika; poudarjajo povedano in spodbujajo odgovor. Sklenemo lahko, da je skupna primarna funkcija teh izrazov, da usmerjajo naslovnika pri interpretaciji, zato jih štejem za interpretacijske označevalce.

4. Gradivo

Gradivo vključuje dva pogovorna žanra, ki se razlikujeta v stopnji spontanosti in formalnosti ter prenosniku. Prvi žanr predstavljajo telefonski pogovori med stranko in informatorem v turistični agenciji, turistični pisarni in hotelski recepciji. Gradivo je izbrano iz korpusa *Turdis* (Verdonik, Rojc, 2006) in poimenovano *Turdis-2*. Natančnejši podatki so v tabeli 1.

	Št. pog.	Povprečna dolžina		Skupna dolžina	
		minute	besede	minute	besede
Agencija	38	3,40	525	129,23	19936
TIC	12	3,63	529	43,58	6350
Hotel	15	2,78	417	41,68	6261
Skupaj	65	3,30	501	214,49	32547

Tabela 1: Število in dolžina pogovorov v *Turdis-2*.

Drugi žanr predstavljajo televizijski intervjuji o aktualnih dogodkih v dnevnoinformativni oddaji, v katerih sodelujejo novinar ter en ali dva intervjuvanca, iz obdobja 1999 do 2005. Gradivo je izbrano iz baze *BNSI Broadcast News* (Žgank et al., 2004) in poimenovano *BNSIint*. Natančnejši podatki so v tabeli 2.

Št. pog.	Povprečna dolžina		Skupna dolžina	
	minute	besede	minute	besede
30	6,61	1041	198,35	31236

Tabela 2: Število in dolžina pogovorov v *BNSIint*.

5. Rezultati korpusne analize

V korpusnem gradivu so bile predstavljene vrste diskurznihi označevalci ročno označene in s tem tudi razdvoumljene v primerih, ko je lahko isti izraz v vlogi diskurznega označevalca ali ne. V nadaljevanju podajam pregled izrazov v vlogi diskurznihi označevalci po vrstah in ločeno za vsak žanr ter podatke o pogostosti rabe v številu konkordanc (št. rab) in v odstotkih glede na število vseh besed v korpusih (% besed).

5.1. Ideacijski označevalci

Ideacijski označevalci so predvsem nekateri priredni vezniki in prislovi. Najpogostejši ideacijski označevalci so bili *in*, *pa*, *torej* in *tako da*. Več podatkov je v tabeli 3.

Ideacijski	Turdis-2		BNSIint	
	št. rab	% besed	št. rab	% besed
<i>in</i>	73	0,224	151	0,483
<i>pa</i>	121	0,372	1	0,003
<i>torej</i>	4	0,012	44	0,141
<i>tako da</i>	63	0,387	1	0,006
<i>ampak</i>	9	0,028	3	0,010
<i>in sicer</i>	16	0,098	1	0,006
<i>namreč</i>	2	0,006	16	0,051
<i>potem</i>	60	0,184	0	0,000
<i>pol</i>	13	0,040	0	0,000
<i>sicer</i>	6	0,018	0	0,000
<i>vendar (pa)</i>	0	0,000	4	0,016
SKUPAJ	367	1,370	221	0,717

Tabela 3: Ideacijski označevalci v Turdis-2 in BNSIint.

Ideacijski označevalci so v naših korpusih v primerjavi z drugimi vrstami diskurznihi označevalci dokaj redki, saj predstavljajo le okoli 1 % vseh besed. V telefonskih pogovorih v turizmu so nekoliko pogostejši kot v televizijskih intervjujih, vendar je razlika veliko manjša kot pri interakcijskih označevalcih.

5.2. Interakcijski označevalci

Med interakcijske označevalce sodijo večinoma izrazi, ki so tudi tradicionalno obravnavani kot diskurzni označevalci. Z njimi sogovorniki signalizirajo, da se poslušajo in razumejo, da se strinjajo oz. ne strinjajo in se dogovarjajo o poteku diskurza. Najpogostejši interakcijski označevalci v našem gradivu so bili oporni signali, *ja*, *(a/ali) ne?*, *no*, *dobro/v redu/okej/prav* ter *aha*. Podrobnejši podatki so v tabeli 4.

Interpretacijski	Turdis-2		BNSIint	
	št. rab	% besed	št. rab	% besed
<i>ja</i>	603	1,853	68	0,218
<i>aha</i>	234	0,719	0	0,000
<i>aja</i>	14	0,043	0	0,000
<i>mhm</i>	97	0,298	7	0,022
<i>(a/ali) ne?</i>	604	1,902	50	0,179
<i>dobro/v redu/okej/prav</i>	227	0,977	22	0,074
<i>no</i>	92	0,283	109	0,349
<i>(po)(g)lejte/(g)lej</i>	79	0,243	48	0,154
<i>(a) veste</i>	27	0,095	6	0,022
<i>zdaj</i>	208	0,639	3	0,010
<i>tako</i>	54	0,166	6	0,019

oporni signali*	1048	3,220	34	0,109
SKUPAJ	3287	10,437	353	1,156

* Z opornimi signali udeleženci v pogovoru signalizirajo, da poslušajo, da razumejo ali da se strinjajo z govorcem, ne uvajajo pa (daljše) menjave vlog niti udeleženci z njimi ne nakažejo namena, da prevzamejo vlogo. V gradivu so oporni signali posebej označeni.

Tabela 4: Interakcijski označevalci v Turdis-2 in BNSIint.

Interakcijski označevalci so v našem gradivu najpogostejši diskurzni označevalci. Opazimo pa veliko razliko v pogostosti rabe: v telefonskih pogovorih predstavljajo več kot 10 % vseh besed, v televizijskih intervjujih jih je skoraj 10-krat manj in le nekaj več kot ideacijskih označevalci. Podrobnejša analiza razlogov za te razlike je bila predstavljena v Verdonik et al. (v tisku).

5.3. Označevalci procesov tvorjenja

Označevalci procesov tvorjenja so predvsem različni izrazi, ki so pogosto imenovani mašila ali zapolnjevalci vrzeli. Podaljšani polglasnik, podaljšani fonem *m* ali *n* ipd. so najpogostejši, poleg teh pa še *mislim*, različni večbesedni izrazi z glagoli rekanja ter z glagolom *vedeti*. Več podatkov je v tabeli 5.

Označevalci tvor.	Turdis-2		BNSIint	
	št. rab	% besed	št. rab	% besed
<i>eee/mmm/eeem ...</i>	1264	4,047	1293	4,139
<i>Mislim</i>	24	0,074	2	0,006
<i>(kako) bi rekel/-la</i>	0	0,000	15	0,099
<i>bom (jaz) rekel/-la</i>	7	0,046	13	0,083
<i>moram reči/rečt</i>	0	0,000	9	0,064
<i>da/če (tako) rečem</i>	1	0,006	4	0,038
<i>kaj (jaz) vem</i>	6	0,052	1	0,010
<i>ne vem</i>	23	0,141	3	0,019
SKUPAJ	1325	4,203	1340	4,460

Tabela 5: Označevalci tvorjenja v Turdis-2 in BNSIint.

Označevalci procesov tvorjenja so edina skupina diskurznihi označevalci, ki so pogosteje rabljeni v televizijskih intervjujih kot v telefonskih pogovorih, vendar je razlika minimalna. Po pogostosti rabe sodijo med bolj pogoste, saj dosegajo skoraj 5 % vseh besed.

5.4. Interpretacijski označevalci

Interpretacijski označevalci so večinoma izrazi, ki se začenjajo ali z besedico *in*, z besedico *pa* (predvsem v manj formalnih konverzacijah), redkeje tudi z besedico *ali*. Podrobneje so predstavljeni v tabeli 6.

Interpretacijski označevalci	Turdis-2		BNSIint	
	št. rab	besed v %	št. rab	besed v %
<i>in tako/tko naprej</i>	3	0,028	24	0,231
<i>in tako dalje</i>	3	0,028	0	0,000
<i>in podobno/-ega</i>	4	0,025	3	0,019
<i>in te/takšne stvari</i>	1	0,009	2	0,013
<i>in to/tega</i>	8	0,049	0	0,000
<i>in vse/vsega</i>	2	0,012	0	0,000

<i>in vse to</i>	0	0,000	1	0,010
<i>pa tako/ko naprej</i>	1	0,009	0	0,000
<i>pa te/take stvari</i>	3	0,028	0	0,000
<i>pa to/tega/temi/teh</i>	5	0,031	0	0,000
<i>pa tako/ko/tak</i>	15	0,092	0	0,000
<i>pa teh zadev</i>	1	0,009	0	0,000
<i>pa vse skupaj</i>	2	0,018	0	0,000
<i>ali (pa) kaj</i>	6	0,046	0	0,000
<i>ali (pa) kaj/kej</i>	6	0,068	0	0,000
<i>takega/takšnega/tazga</i>				
<i>ali (pa) (ne)kaj podobno/-ega</i>	5	0,052	1	0,010
<i>ali (pa) kakorkoli (že)</i>	1	0,012	1	0,006
SKUPAJ	66	0,516	32	0,288

Tabela 6: Interpretacijski označevalci v Turdis-2 in BNSInt.

Interpretacijski označevalci so v našem gradivu najmanj pogosti: v telefonskih pogovorih obsegajo komaj pol odstotka vseh besed, v televizijskih intervjujih pa še enkrat manj. Pri tem moramo upoštevati tudi, da so večinoma večbesedni in že zaradi tega dosežejo še nekoliko večjo pogostost. Tudi tukaj velja, da so v (manj formalnih) telefonskih pogovorih bolj pogosti kot v televizijskih intervjujih.

6. Zaključek

V članku sem predstavila shemo za označevanje vrste diskurznihih označevalcev v jezikovnih virih, in sicer ideacijskih označevalcev, interakcijskih označevalcev, označevalcev procesov tvorjenja in interpretacijskih označevalcev. Shema izhaja iz empiričnih analiz in je podprta s teoretičnim ločevanjem temeljnih ravni diskurza. Zaradi potreb označevanja jezikovnih virov je kar se da robustna, zato upošteva le najbolj pogoste in splošno prepoznane značilnosti diskurznihih označevalcev in se izogiba razvrščanju istega izraza v več vrst.

V korpusnem gradivu se je pokazalo, da so najpogostejši interakcijski označevalci, zatem označevalci procesov tvorjenja, manj pogosti so bili ideacijski označevalci, najmanj pogosti pa interpretacijski. Vendar so bile velike razlike med obema uporabljenima korpusoma: v splošnem so bili diskurznihih označevalci v televizijskih intervjujih veliko manj pogosti (skupaj 4,748 % vseh besed) kot v telefonskih pogovorih v turizmu (skupaj 16,526 % vseh besed).

Predstavljena shema je podlaga za nadaljnjo korpusno podprto analizo diskurznihih označevalcev v jezikoslovju, na področju tehnologij pa pomeni osnovo za evalvacijo uporabnosti atributa diskurznihih označevalcev.

7. Literatura

Blakemore, D. (2002) *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.

Byron, D. K., P. A. Heeman (1997). *Discourse marker use in task-oriented spoken dialog*. 5th European Conference on Speech Communication and Technology (Eurospeech), Rhodes, Greece.

Carlson, Lynn, Daniel Marcu, and Mary Ellen Okurowski, 2003. *Current Directions in Discourse and Dialogue*, chapter Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. Kluwer Academic Publishers.

Clark, H.H., Fox Tree, J.E.: Using uh and um in spontaneous speaking. *Cognition* 84, 73–111 (2002)

Fraser, B. (1996) 'Pragmatic markers', *Pragmatics*, 6/2, 167–190.

Gorjanc, Vojko, 1998: Konektorji v slovnicih in opisu znanstvenega besedila. *Slavistična revija* 46/4. 367–388.

Heeman, Peter, James Allen (1999). *Speech repairs, intonational phrases and discourse markers: modeling speakers' utterances in spoken dialog*. *Computational Linguistics*, 25(4).

H. van den Heuvel, D. Iskra, E. Sanders in F. de Vriend. 2008. Validation of spoken language resources: an overview of basic aspects. *Language Resources and Evaluation*, 42: 41–73.

Miltsakaki, E., R. Prasad, A. Joshi, B. Webber (2002). *The Penn Discourse Treebank. Proceedings of 4th LREC*, Lisbon, Portugal.

Mitkov, Ruslan, 2000. Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC 2000)*, Lancaster, U.K..

Mitkov, 2003...

Muller, Christoph, Stefan Rapp, and Michael Strube, 2002. Applying co-training to reference resolution. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Philadelphia PA.

Overstreet, M.: And stuff und so: Investigating pragmatic expressions in English and German. *Journal of Pragmatics* 37, 1845–1864 (2005)

Pisanski Peterlin, A. (2005) 'Text-organising metatext in research articles: An English-Slovene contrastive analysis', *English for Specific Purposes* 25: 307–319.

Redeker, G. (1990) 'Ideational and pragmatic markers of discourse structure', *Journal of Pragmatics* 14: 367–381.

Schiffirin, D. (1987) *Discourse Markers*. Cambridge: Cambridge University Press.

Schlamberger Brezar, M. (2007) 'Vloga povezovalcev v govornem diskurzu', *Jezik in slovnost* 52(3-4): 21–32.

Smolej, M. (2004) 'Členki kot besedilni povezovalci', *Jezik in slovnost* 49(5): 45–57.

Swerts, M., 1998: Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485–496.

Verdonik, D., Rojc, M. (2006) 'Are you ready for a call? – Spontaneous conversations in tourism for speech-to-speech translation systems', in *Proceedings of 5th LREC*, Genoa, Italy.

D. Verdonik, M. Rojc in M. Stabej. 2007a. Annotating discourse markers in spontaneous speech corpora on an example for the Slovenian. *Language Resources and Evaluation*, 41, 147–180.

D. Verdonik, A. Žgank in A. Pisanski Peterlin. 2007b. Diskurznihih označevalci v dveh pogovornih žanrih. *Jezik in slovnost*, 52/6, 19–32.

D. Verdonik, A. Žgank in A. Pisanski Peterlin. V tisku. The impact of context on discourse marker use in two conversational genres. *Discourse Studies*.

Žgank, A., Rotovnik, T., Sepesy Maučec, M., Verdonik, D., Kitak, J., Vlaj, D., Hozjan, V., Kačič, Z., Horvat, B. (2004) 'Acquisition and annotation of Slovenian Broadcast News database', in *Proceedings of 4th LREC*, Lisbon, Portugal, pp. 2103–2106.