

KLASIFICIRANJE BESEDILNIH ZVRSTI V JAPONŠČINI *

Andrej BEKEŠ

Univerza v Ljubljani in Univerza v Tsukubi

andrej.bekes@guest.arnes.si

Izveček

Članek v okvirju Hallidayeve sistemsko-funkcijske lingvistike obravnava nekatere pristope h klasificiranju besedilnih zvrsti oz. žanrov v japonščini v zadnjem času. Besedilne zvrsti se razlikujejo glede na značilnosti konteksta situacije, tj., področja, tona in načina. Klasifikacijski pristopi se večinoma osredotočajo na značilnosti področja, ki se odražajo v porazdelitvi polnopomenskih leksičnih enot v besedilu. Pristopi, predstavljeni v članku, se nasprotno opirajo na evidencialno-modalne pomene, izražane preko evidencialnih prislovov in pomožnih evidencialno-modalnih elementov na koncu povedi.

Ključne besede

klasificiranje, besedilna zvrst, evidencialni prislovi, evidencialno-modalni izrazi

Abstract

This paper deals with some recent approaches to Japanese text classification within the framework of Halliday's systemic functional linguistics. Text types differ according to the properties of their respective field, tenor and mode. Classification approaches usually center on field properties as reflected in content words and their distribution in texts. On the other hand, approaches introduced in the present paper are based on evidential-modal meanings, expressed by evidential adverbs and sentence-final auxiliary evidential-modal expressions.

Keywords

classification, text type, evidential adverbs, evidential-modal expressions

* Title in English: Text Type Classification in the Japanese Language

1. Uvod

Didaktika tujih jezikov, kjer se pogosto srečujemo z vprašanjem besedilne zvrstnosti, in kako to vedenje posredovati učečim se, je ena od pomembnih motivacij za proučevanje in klasificiranje besedilne zvrstnosti oz. besedilnih žanrov. To velja tudi za japonščino, kjer v zadnjem času dobiva velik poudarek poučevanje strokovne japonščine. Prav tako se je z uveljavljanjem obsežnih korpusov v raziskovanju japonskega jezika (npr. korpus BCCWJ, uravnoteženi korpus sodobnega japonskega pisnega jezika, Maekawa 2007) pojavila potreba po raznih zvrsteh besedil, ki naj bi bila vključena vanje. Pričujoči članek predstavlja nekaj novejših pristopov h klasificiranju besedilnih zvrsti v japonščini, predvsem iz kroga laboratorija Kikuko Nishina in njenih učencev, Irene Srdanović in Bora Hodoščka na Tokijskem tehnološkem inštitutu (Tokyo kogyo daigaku), kjer sem sodeloval kot gostujoči raziskovalec v letnem semestru 2010.

Ideja besedilne zvrsti je povezana z opažanjem, da se velik del človekovega delovanja v družbi ponavlja v bolj ali manj enakih situacijskih kontekstih, kar ima za posledico tudi ponavljajočo se rabo jezikovnih sredstev v enakih okoliščinah. V nadaljevanju se opiram na sistemsko-funkcijsko jezikoslovje. Le-to povezuje jezik kot vedenje (sistem) s kontekstom situacije, kjer se, glede na dejavnost, ki se v danem kontekstu odvija, jezik udejanja kot določena besedilna zvrst. Kontekst situacije sestavljajo trije vidiki, PODROČJE (dejavnost, ki se izvaja skozi jezik), TON (odnosi med soudeleženci, njihov status) in NAČIN (kako se dana jezikovna dejavnost udejanja kot besedilo). Ponavljajoči se jezikovni vidik konteksta situacije povzema pojem REGISTER. Register je "konfiguracija semantičnih sredstev, ki jih član neke kulture tipično povezuje z danim tipom situacije" (Halliday, 1978:111). Vidiki kontesta situacije so povezani tudi z naslednjimi pomenskimi kompleksi (meta-semantične funkcije): področje z ideacijsko funkcijo, ton z medosebno funkcijo ter način z besedilotvorno funkcijo (Halliday, 1978; Hasan, 2009).

Kakšne so konkretne značilnosti dane besedilne zvrsti, je odvisno od tega, kakšna jezikovna dejavnost se odvija v danem kontekstu situacije, to pa je odvisno od treh vidikov, tj. področja, tona in načina, njihove značilnosti pa se odražajo na pomenih aktiviranih iz prej omenjenih pomenskih kompleksov. Pričakovati je, da bodo merila za pripadnost neki zvrsti tesno povezana s področjem in tonom. Dano področje implicira določen nabor ideacijskih pomenov, ki se odražajo v rabi jezikovnih sredstev, ki pripadajo leksiki in v naklonskem sistemu predvsem evidencialnosti. Dani ton pa implicira določen nabor medosebnih pomenov, kar se v jezikovnih sredstvih odraža v naklonskem sistemu ter sistemu spoštljivosti.

V naslednjih razdelkih je predstavljen pristop h klasificiranju, temelječ na sistemu modalnosti oz. evidencialnih pomenov.

2. Dosedanji pristopi h klasificiranju besedil v japonščini

Večina dosedanjih pristopov h klasificiranju besedil v japonščini temelji na vidikih rabe polnopomenskega besedišča, skratka na analizi tega, kako se področje odraža v udejanjenem besedilu. Eden zgodnejših poskusov je Kabashima in Jugaku (1965), kjer avtorja predlagata kot merilo pogostnost posameznih besednih vrst, predvsem samostalnika, ter dodatno količnik med določili (*shûshokugo*; s tem terminom zajameta prislove, pridevnike in adnominalne) ter glagoli v besedilu. Besedila, kjer je pogostnost rabe samostalnikov visoka, naj bi grobo rečeno pripadala opisnim žanrom. Le-ti se glede na višino količnika med določili in glagoli nadalje delijo na dejavne (kadar je količnik nizek) in statične (kadar je količnik visok). Besedila, kjer je pogostnost rabe samostalnikov nizka, pa pripadajo povzermalnim žanrom. Hodošček (2011) je ta pristop preverjal na korpusu BCCWJ. Hodošček je potrdil ugotovitve Kabashima in Jugaku (1965), izjema je bil predvsem nov žanr spontanih internetnih besedil kot so blogi ipd., v katerih se distribucija besednih vrst razlikuje od drugih žanrov.

Drug pristop, Sano (2008), je novejše narave in temelji na pojmu informacijske gostote (*density of information*) (Halliday, 1990), oziroma besediščne gostote (*lexical density, goi mitsudo*). Besediščna gostota je kot parameter pridružena posameznemu besedilu. Sano (str. 85) jo definira kot razmerje med vsemi polnopomenskimi besedami v besedilu in vsemi stavki (*setsu, clause*), ki vsebujejo povedek. Besediščna gostota, kot je pokazal Sano, dobro korelira z značilnostmi raznih pisnih in govorjenih žanrov. Problem te metode pa je, da je z vidika obdelave podatkov in računanja dokaj zahtevna, štetje stavkov namreč zahteva ročno ali avtomatsko analizo besedila v sestavljene povedi in stavke ter zato ni dostopna za hitro klasificiranje posameznih besedil, ki bi ga npr. potreboval učitelj japonskega jezika pri sestavljanju jezikovnega gradiva za svoje učence.

3. Evidencialni pomeni kot kriterij besedilne zvrsti

3.1 Evidencialni prislovi in besedilne zvrsti

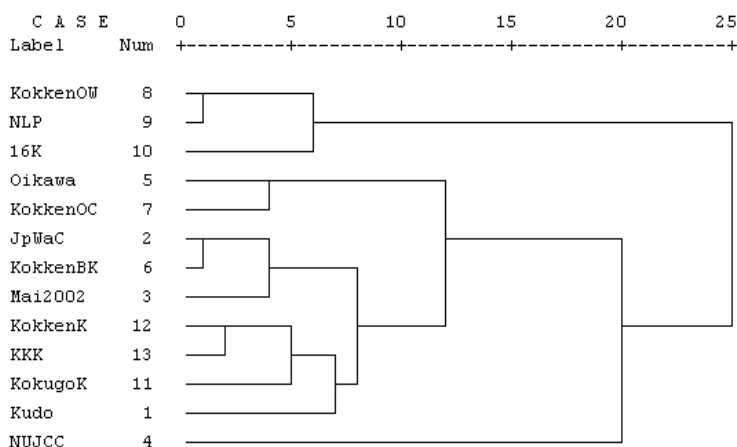
Kot ena od možnosti so se pokazali evidencialni pomeni, ki jih v japonščini posredujejo predvsem evidencialni povedni prislovi (*suiryôteki fukushi*) ter evidencialno-modalni izrazi na koncu povedi. Srdanović et al. (2008a) so, izhajajoč iz Kudô (2000) na obsežnem korpusnem gradivu pokazali, da Kudôjeva ugotovitev o sistematični korelaciji evidencialnih prislovov in modalnih izrazov na koncu povedi velja tudi v tem primeru. Prav tako, kot ugotavljajo Srdanović et al. (2008b), se porazdelitev gostote evidencialnih prislovov in evidencialnih modalnih izrazov razlikuje glede na različne vrste korpusov s tipsko homogenimi besedili (korpusi učbenikov, literarnih del, časopisnih člankov ipd.). Razlike v porazdelitvi so nazorno razvidne iz tabele 1 spodaj.

Tabela 1: Porazdelitev evidencialnih prislovov v različnih tipih besedil (Srdanović et al., 2008b)

Pridružena modalnost	Korpusi Evid. prislovi	NUJCC	Oikawa	KokkenOC	JpWaC	KokkenBK	Mai2002	KokkenOW	NLP	16K	KokugoK	KokkenK	KKK	Kudo
		N	kanarazu	7	14	4	8	15	25	5	23	42	12	28
N	zettai	52		14	9	6	11	2			3	2	4	5
N	zettaini			11	6	8	12	2		4	3	9	2	
N/M	kanarazushimo	1	5	2	6	6	8	84	66	39	0	10	6	
N/V	yohodo		1	2	2	3	2	0			3	1	2	4
N/V	yoppodo	2		2	1	1	1				1			
N/P	taigai	2	8	1	1	1	0							1
N/P	taitei	1		5	4	2	3	1	6	4	4	6	12	1
N/P	kitto	15	8	15	12	14	10		3		38	26	26	28
P	ookata		1	0	0	1	0				1	1	2	3
P	osoraku	1	8	1	13	12	9	1	3	7	2	5	10	19
P	sazo			0	0	1	1				4	1	2	5
P	tabun	3	39	26	16	11	6	2		3	3	4	8	10
V	doumo	6	7	6	8	7	5	0		1	15	2	4	5
V	douyara	2		3	5	5	3				3			5
V	angai	3	1	0	2	1	1	0		1	1	1		2
M	hyottoshitara	1	1	1	1	1	1				1	1	2	3
M	hyottosuruto				0									
M	kotonyoreba			0	0	0	0							
M	kotonyoruto				0	0	1	3		1				1
M	moshikashitara	5	8	5	3	3	1				2	2	2	5
M	moshikasureba			1	0									
M	moshikasuruto				1	1	0				1	1	2	

N: nujno; M: možno; V: verjetno; P: pričakovano; podatki so izraženi v odstotkih
 rdeče polje > 50%; oranžno polje: ~40%; rumeno polje: ~25%, lila polje: ~15%

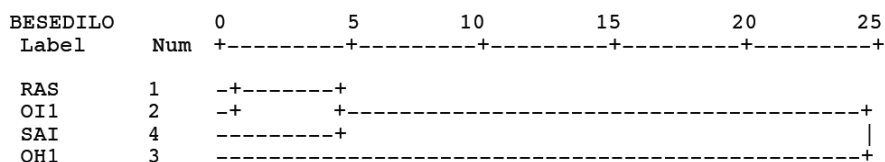
Še izrazitejše postanejo, če rezultate v tabeli analiziramo s pomočjo algoritmov razvrščanja v skupine, kot je prikazano na sliki 1 spodaj. Intuitivno je sprejemljivo, da imamo podobno porazdelitev evidencialnih prislovov v zbirki uradnih besedil iz uradnih “belih knjig” (KokkenOW) ter v znanstvenem diskurzu v člankih o računalniški obdelavi naravnih jezikov (NLP), ki se jim pridružuje mini korpus naravoslovnih učbenikov (16K). Kot je razvidno iz tabele 1, vse tri odlikuje zelo močna prisotnost evidencialnega prislova *kanarazushimo* ([ne] nujno), ki se tipično nanaša na negativno možnost. Pri vseh treh gre za objektivna, formalna besedila. Na nasprotnem koncu je prav tako sprejemljivo, da korpus neformalnih pogovorov NUJCC izstopa iz ostale mešane skupine.



Slika 1: Razvrstitev korpusov iz tabele 1 v skupine (Srdanovič et al., 2008b)

Skratka, razlike v porazdelitvi gostote izrazov glede na tip besedila bi načeloma lahko služile kot osnova za klasificiranje besedilnih zvrsti, ali posameznih besedil ali pa korpusov homogenih besedilnih zvrsti.

Bekeš je v pilotski študiji (2008) preverjal zamisel na omejenem gradivu manjših govornih in pisnih korpusov. Izkazalo se je, da je razlikovanje med besedilnimi zvrstmi na osnovi porazdelitve evidencialnih prislovov dokaj zadovoljivo glede na omejen obseg podatkov iz posameznih besedil.



Slika 2: Razvrščanje korpusov skupine glede na besedilne zvrsti

Kot je razvidno iz slike 2 zgoraj, porazdelitev evidencialnih prislovov, sicer aglomerirana glede na tip evidencialnih pomenov, ostro razlikuje med korpusom neformalnih spontanih konverzacij (OH1), ter med ostalimi tremi: novelo Rashomon (RAS), korpusom formalnih intervjujev (OI1) ter učbenikom celične biologije (SAI). Nadalje, znotraj grupe treh je učbenik (SAI) jasno profiliran nasproti noveli (RAS) in formalnim intervjujem (OI1).

Pri izražanju evidencialnih pomenov se pojavljajo evidencialni prislovi 20- do 100-krat manj pogosto kot evidencialno-modalni izrazi na koncu povedi, s katerimi

pomensko korelirajo. Tako ti prislovi predstavljajo sicer grob a hiter način, kako preveriti, v katero besedilno zvrst uvrstiti dano besedilo ali korpus homogenih besedil.

3.2 Evidencialno-modalni izrazi na koncu povedi in besedilne zvrsti

Hodošček et al. (2009) so se lotili zahtevnejše naloge analizirati besedilne zvrsti na osnovi evidencialno-modalnih izrazov na koncu povedi. Težave tod nastopijo, kot je bilo rečeno, tako zaradi bistveno večje pogostnosti takšnih izrazov, še bolj pa zaradi njihove raznolikosti in strukturne nedoločenosti.

Naloga zahteva intenzivno obdelavo jezikovnih podatkov, zaradi strukturne nedoločenosti evidencialno-modalnih izrazov na koncu povedi je velik del obelave, zlasti presojanje, kaj vse spada v evidencialno-modalni izraz, še vedno ročen. Obdelanih je bilo pet korpusov, od tega štirje podkorpusi korpusa BCCWJ (časopisni članki /ČAS/, zapisi govorov v državnem zboru /DZB/, šolski učbeniki /UČB/ ter »bele knjige« /BEL/), poleg njih pa še neodvisen korpus znanstvenih člankov /ČLA/.

Najprej je bila analizirana gostota porazdelitve evidencialnih prislovov po posameznih korpusih. Rezultati so tukaj podobni kot tisti v tabeli 1.

Zaradi velikega obsega evidencialno-modalnih izrazov na koncu povedi je bila analiza teh izrazov omejena na izraze, ki se sopoljavljajo skupaj z najbolj pogostimi evidencialnimi prislovi, in sicer *kanarazushimo* ([ne] nujno), *kanarazu* (zagotovo), ter *osoraku* (verjetno), ki sami nosijo levji delež izražanja evidencialnosti v obdelanem gradivu. Tukaj je prikazan samo rezultat analize sopoljavljanja z *osoraku*.

Tabela 2: *Osoraku* (verjetno) in gostota evidencialno-modalnih izrazov na koncu povedi po posameznih korpusih (Hodošček et al., 2009)

<i>osoraku</i>							
DZB	%	ČAS	%	ČLA	%	UČB	%
darou	9.52	darou	29.27	Ø	40.00	Ø	33.33
Ø	7.94	Ø	19.51	darou	13.33	darou	33.33
toomou	6.35	nodarou	19.51	toomowareru	13.33	nodewanaidarouka	33.33
nodewanaika	5.56	nitigainai	4.88	monodarou	6.67		
daroutoiufuuniomou	2.38	hazu	4.88	monodatosuisokusareru	6.67		
nodewanaikatoomou	2.38	nai	4.88	daroune	6.67		
toomounodane	2.38	nodewanaidarouka	4.88	tokangaerareru	6.67		
nodewanaikatoiuufuunikanjiteiru	2.38	toomowareta	2.44	nai	6.67		
nodehanaidarouka	2.38	dewanaidarouka	2.44				
nodewanaikato	2.38	dewanaika	2.44				
nodarou	2.38	dewanaikatoomou	2.44				
darou to omou	2.38	mai	2.44				
nodane	2.38						
rezultati pod 2% izpuščeni; v BEL se osoraku ne pojavi							

Iz tabele 2 je razvidna velika disperznost izrazov v korpusu besedil iz državnega zbora (DZB) – odraz ndr. načina, t.j. ustnega komuniciranja. Vsi drugi korpusi so korpusi pisnih besedil. Disperznost porazdelitve se zmanjšuje od časopisnih člankov (ČAS), preko znanstvenih člankov (ČLA) do učbenikov (UČB). V teh korpusih pa so tudi očitne razlike v porazdelitvi gostote evidencialno-modalnih izrazov. V časopisnih člankih (ČAS) *osoraku* najpogosteje sovпада s pričakovano obliko *darou* (verjetno). V znanstvenih člankih (ČLA) sovпада z ničelno modalno obliko, medtem ko se v učbenikih (UČB) sicer pojavlja redko, a v treh različnih oblikah.

Iz tabele 2 je razvidno, da variabilnost sopojavljanja evidencialno-modalnih izrazov z evidencialnimi prislovi sovпада s posameznimi besedilnimi zvrstmi. Na podrobnejšo analizo, v kakšni povezavi je ta variabilnost z značilnostmi konteksta situacije, to je s področjem, tonom in načinom, bo treba še počakati. Je pa disperznost porazdelitve evidencialno-modalnih izrazov na koncu povedi v sopojavljanju s posameznimi evidencialnimi prislovi bistveno večja, kot bi bilo pričakovati na osnovi rezultatov v Kudô (2000), in ponuja, ob ustrezni operacionalizaciji in teoretski utemeljitvi, obetavno možnost za klasificiranje besedilnih zvrsti v japonščini.

4. Zaključek

V članku so prikazani nekateri pomembnejši rezultati analize besedilnih zvrsti, temelječe na sistemu modalnosti oz. evidencialnih pomenov. Pristopi, temelječi na porazdelitvi polnopomenskih besed, so glede na področje lahko uspešnejši, a z vidika obdelave jezikovnih podatkov zahtevni. Pristop na osnovi evidencialnih prislovov je v tem pogledu nezahteven in se izkaže za hitro a grobo metodo določanja besedilnih zvrsti. Pristopi na osnovi evidencialno-modalnih izrazov na koncu povedi obetajo več, zaradi pogostnosti pojavljanja teh oblik so načeloma uporabni tudi za manjše korpusne in posamezna besedila, a so zopet zahtevni z vidika obdelave podatkov.

Zaključimo lahko, da bi bilo za natančno določanje besedilnih zvrsti verjetno treba kombinirati posamezne metode. Vsaka od metod se opira samo na omejen nabor značilnosti konteksta situacije analiziranih besedil, širši zajem značilnosti s kombinacijo posameznih metod bi tako nedvomno prispeval k natančnejšemu klasificiranju besedilnih zvrsti.

Reference

- Bekeš, A. (2008). Suppositional adverbs as indicators of discourse genre. *Asian and African Studies*, 12, 5-16.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, M. A. K. (1990). Some grammatical problems in scientific English. *Annual Review of Applied Linguistics*, 6, 13-37.

- Hasan, R. (2009). The place of context in a systemic functional model. In M.A.K. Halliday & J. J. Webster (Eds.) *Continuum companion to systemic functional linguistics* (pp. 166-189), New York: Continuum.
- Hodošček, B., Bekeš, A., & Nishina, K. (2009). Suiiryôteki fukushi no kyoôkijôhō ni motozuita *genre* betsu no bunmatsu hyôgen no bunseki. *Gengo shori gakkai dai 15 kai nenji taikai happyô ronbunshû* (Proceedings of The Fifteenth Annual Meeting of The Association for Natural Language Processing), 598-601.
- Hodošček, B. (2011). Word class ratios and genres in written Japanese: Revisiting the modifier verb ratio. *Acta Linguistica Asiatica*, 1(1), 53-62.
- Kabashima, T., & Jugaku, A. (1965). *Buntai no kagaku* [Stilistika]. Sogetsu.
- Kudô, H. (2000). Fukushi to bun no chinjutsu no taipu (Adverbs and the type of sentence-final modality). In Y. Nitta & T. Masuoka (Eds.) *Nihongo no bunpô 3 – modariti* (Japanese grammar 3: modality) (pp. 161-234). Tokyo: Iwanami shoten.
- Maekawa, K. (2007). Kotonoha and BCCWJ: Development of a balanced corpus of contemporary written Japanese. *Proceedings of the First International Conference on Korean Language, Literature, and Culture*, 2, 158 – 177.
- Sano, M. (2008). Text-classification system for large scale balanced corpus: A systemic functional approach. *Proceedings of the 2008 General Meeting of the MEXT grant-in-aid for Scientific Research Priority Area Program "Japanese Corpus"* [in Japanese].
- Srdanović Erjavec, I., Bekeš, A., & Nishina, K. (2007). Cluster analysis of suppositional adverbs and clause-final modality. *Asian and African Studies*, 21-31.
- Srdanović Erjavec, I., Bekeš, A., Nishina, K. (2008a) Distant collocations between suppositional adverbs and clause-final modality forms in Japanese language corpora. In *Large-scale knowledge resources: Construction and application; Lecture Notes in Computer Science*, 4938 (pp. 252-266). Berlin: Springer.
- Srdanović Erjavec, I., Bekeš, A., & Nishina, K. (2008b) Adverbs and clause-final modality collocations in various corpora. In *Tokutei ryôiki kenkyû 'Nihongo koopasu': Heisei 19 nendo kôkai waakushoppu (kenkyûseika hôkokukai) yokôshû: 15.3.-16. 3.* (pp. 223-230) Tokyo: Monbukagakushô kagakukenyûhi tokuteiryôiki kenkyû 'Nihongo koopasu' Sôkatsu ban.