Practical Applications of Linguistic Research

edited by Anna Obrębska

Primum Verbum
Łódź 2012

RECENZJA NAUKOWA
prof. dr hab. Romuald Gozdawa-Gołębiowski

ISBN 978-83-62157-86-0

Łódź 2012

Wydawnictwo PRIMUM VERBUM
ul. Gdańska 112, 90-508 Łódź
e-mail: wydawnictwo@primumverbum.pl
www.primumverbum.pl

Table of Contents

Michał B. Paradowski*

LANGUAGE EMBODIMENT AND COROLLARIES FOR ROBOTICS

to turn print into exciting situations in their skulls
— Kurt Vonnegut *Slaughterhouse-Five* [1969: 205]

His eyes only see
His ears only hear
— Wisława Szymborska *No End of Fun* [1967]

1. Introduction

Until very recently, most language research has, in a Cartesian manner, traditionally regarded linguistic phenomena as internal, mental, isolationist and amodal (that is, separate and independent from perception, action and emotion systems, and the body); a view endorsed in psychology [e.g. Geschwind 1970; Newell & Simons 1972; Miller & Johnson-Laird 1976; Pylyshyn 1984; Levelt 1989; Kintsch 1998], philosophy [e.g. Katz & Fodor 1963; Katz 1972; Fodor 1975, 1983], and linguistics [e.g. early Chomsky – 1957, 1975; Nowak *et al.* 2000, 2002; Jackendoff 1983, 1997, 2002][1]. This could lead one into believing that in order to believably emulate linguistic behavior, it suffices to develop 'software' operating on (i.e. applying combinatorial rules such as Merge and Move to) abstract representations (terminal symbols, which can – subsequently or concurrently – be equipped with referential, meaning-bearing properties) that will work on any computational machine, and that its operations will be implementation-independent, functioning identically regardless of the physical hardware.

* Institute of Applied Linguistics, University of Warsaw.

[1] A *votum separatum* in this domain is the field of biolinguistics, which hypothesizes a strong genetic (or neurobiological) endowment for language (UG) and determination of its structure (e.g. postulating selectional – i.e. evolutionary fitness – advantages), treating the language faculty on a par with other biological systems [see e.g. Meader & Muyskens 1950; Lenneberg 1967; Piatelli-Palmarini 1989; Jenkins 2000; Hauser *et al.* 2002; Chomsky 2005; Di Sciullo & Boeckx 2011].

2. Embodied language in humans

If the system to be developed is to truly mimic human behavior, this picture is not very accurate for several reasons. One may be doubtful about modularity and the existence of a specifically dedicated innate language acquisition device, but must still acknowledge the reality of:

1) **lateralization** and **localization** of the language faculty in the brain, in the sense that linguistic capabilities have been shown to be limited to certain areas of the cerebrum, as evidenced primarily by various language disorders:

- receptive aphasia, commonly known as Wernicke's aphasia [Wernicke 1874]: damage to the medial temporal lobe destroying local language regions and cutting them off from most of the occipital, temporal and parietal regions [*cf.* e.g. Price 2000; Bookheimer 2002; Kolb & Whishaw 2003; Damasio *et al.* 2004];
- expressive aphasia [aka Broca's or agrammatic aphasia; Broca 1861][2];
- abnormal language developed in individuals with the left hemisphere removed [Dennis & Whitaker 1976][3];
- Specific Language Impairment (SLI), which is unrelated to other developmental disorders, mental retardation, brain injury, or deafness [e.g. Joanisse & Seidenberg 1998; Leonard 1998; Cohen 2002; Bishop & Snowling 2004; Archibald & Gathercole 2006];
- other cases of people with normal nonverbal abilities but impaired language, and 'normal' language but cognitive deficits (*cf.* the classic case studies of individuals with incommensurable linguistic and cognitive capacities: Genie [Curtiss 1981], Laura [Yamada 1990], Clive [Smith 1989], or Christopher [Smith *et al.* 1993]);

2) embodiment of language in neuronal circuitry – fMRI studies showing '**activation**' of certain brain areas involved in **language processing** [Osterhout 1997; Brown & Hagoort 1999; Hagoort *et al.* 1999; Embick *et al.* 2000; Hickock & Poeppel 2000, 2004, 2007; Hahne & Friederici 2002; Hagoort 2003; Horwitz *et al.* 2003; Pulvermüller & Shtyrov 2003; Ye *et al.* 2006; Pulvermüller & Assadollahi 2007; Rogalsky *et al.* 2008; for loci of syntax mainly in left-perisylvian language regions, especially Broca's and Wernicke's areas, but also adjacent neocortical areas, the insula, and subcortical structures including basal ganglia, *cf.* e.g. Ullman 2001; Friederici 2002; Hagoort 2005; Caplan 2006; Grodzinsky & Friederici 2006; for phonology *cf.* e.g. Diesch *et al.* 1996; Obleser *et al.* 2006; Uppenkamp *et al.* 2006];

[2] Interestingly, Broca's area is activated in both spoken and signed languages [Horwitz *et al.* 2003].

[3] Although one must be cautious about the conclusions since the cortical development in the subjects of the study was not normal in the first place [Chomsky 1980: 264].

3) genetic influence on language. While mutations of the Foxhead box protein 2 (**FOXP2** gene), deemed to cause a severe speech and language disorder [Lai *et al.* 2001; Fisher *et al.* 2003; MacDermot *et al.* 2005; Vargha-Khadem *et al.* 2005; Feuk *et al.* 2006; Vernes *et al.* 2006, 2008; Fisher & Scharff 2009],were initially taken as evidence for a 'language gene', it was later discovered that the protein impacts a wide range of phenotypic features all over the body (including facial motor control) and that the impairments of the family affected with the mutation went beyond language to other cognitive capacities. It is now more believed that it is networks of gene interactions rather than individual genes that have an influence on language [Knopka *et al.* 2009], but the neurobiological influence is there;

4) many Universal Grammar-based constraints now being reinterpreted as **learning** and **processing constraints**. That is, the difficulty in the acquisition of certain aspects of language are being accounted for by their complexity, the computational load under which the user/learner operates, his/her memory and attention limitations, or access to representations;

5) maturation and the **critical/sensitive period** [Lenneberg 1967; but consider e.g. Marinova-Todd *et al.* 2000 for a contradictory view];

6) the Chomskyan competence vs. **performance** distinction [Chomsky 1965][4], explaining mistakes in (originally native) language users' output (i.e., their actual deployment of the linguistic capacity) attributable to such psychosomatic states and factors affecting them as fatigue, tedium, intoxication, drugs, sudden changes of mind, haste, inattention, or external distractions;

7) interaction between (context-bound) language comprehension and production and sensorimotor activation, manifested in both directions by[5]:

- first by **motor resonance** observed in linguistic [Lakoff & Johnson 1980; Lakoff 1987], behavioral [primarily with priming – e.g. mention of tool and action concepts – modulating motor performance; e.g. Tanenhaus *et al.* 1995; Gentilucci & Gangitano 1998; Gentilucci *et al.* 2000; Spivey *et al.* 2001; Tucker & Ellis 2001; Glenberg & Kaschak 2002; Glover & Dixon 2002; Zwaan *et al.* 2002; Bergen *et al.* 2003; Richardson *et al.* 2003; Glover *et al.* 2004; Buccino *et al.* 2005; Kaup *et al.* 2005, 2007; Knoeferle *et al.* 2005; Boulenger *et al.* 2006, 2008; Zwaan & Taylor 2006; Scorolli & Borghi 2007; Lupyan & Spivey 2008; Nazir *et al.* 2008; Borghi & Scorolli 2009; Frak *et al.* 2010; Rueschemeyer *et al.* 2010b; for grammar *cf.* Madden & Zwaan 2003; Bergen &

[4] The distinction can be considered on the example of any organic system: "Studies of the digestive system, for example, distinguish between its structural properties and what it is doing after you ate a sandwich" [Noam Chomsky, p.c., Nov. 8 2011], and can actually be traced back to the classic Aristotelian dichotomy between δύναμις (potentiality) and ένέργεια (actuality).

[5] This seems to be a reflection of a more general phenomenon where "there is no animal in which there is known to be a complete segregation of sensory processing" [Stein *et al.* 1996].

Wheeler 2010], neuroimaging and TMS studies[6] [Zatorre *et al.* 1992; Fadiga *et al.* 1995, 2002; Paus *et al.* 1996; Chao & Martin 2000; Pulvermüller *et al.* 2001, 2003, 2006; Hauk & Pulvermüller 2004; Hauk *et al.* 2004; Oliveri *et al.* 2004; Shapiro 2004; Wilson *et al.* 2004; Buccino *et al.* 2005; Tettamanti *et al.* 2005, 2008; Aziz-Zadeh *et al.* 2006; Boulenger *et al.* 2006, 2009; Goldberg *et al.* 2006; Saccuman *et al.* 2006; Martin 2007; Rueschemeyer *et al.* 2007, 2010c; Fischer & Zwaan 2008; Kemmerer *et al.* 2008; Willems *et al.* 2010; Borghi & Cimatti 2010; Desai *et al.* 2010; for activation in visual areas *cf.* Martin *et al.* 1996; Chao *et al.* 1999; van Schie *et al.* 2003; Del Prado Martin *et al.* 2006; Pulvermüller & Hauk 2006; Simmons *et al.* 2007; in the olfactory cortex *cf*. González *et al.* 2006];

- **semantic resonance** [brain language areas getting activated during sensorimotor action; Bonda *et al.* 1994; Pulvermüller *et al.* 2005a, 2005b; Lindemann *et al.* 2006; Rueschemeyer *et al.* 2010a][7];
- verbalization of memory facilitated when assuming the original body position during recall [Dijkstra *et al.* 2007], linguistic tasks expedited when accompanied by action [Rieser *et al.* 1994], and sensorimotor experiences intertwined with cognition in episodic memory [Pfeiffer 2011];
- faster comprehension of depictions of spatial associations than of descriptions of spatial *dis*sociations [Glenberg *et al.* 1987]; speedier recognition of words with 'body-object interaction' than of ones without [Siakaluk *et al.* 2008];

[6] Somewhat importantly, motor resonance was not observed when the stimuli were used in idiomatic contexts [Rueschemeyer *et al*. 2010a] or metaphorical ones. Regarding the latter, Raposo et al. [2009] found activity in the pre- and motor cortex for literal-only usages of arm- and leg-related Vs, while Bergen *et al*. [2007] likewise demonstrated that visual imagery is triggered in sentence comprehension tasks (where general words of motion were employed) only where the utterances have literal spatial meaning. However, the picture is not completely clear-cut. Last year, Lacey *et al.* [2012] showed that textural metaphors do activate parietal operculum regions important to the sense of touch. To explain this discrepancy, one could posit a qualitative difference between 'directly' embodied sensory experiences (e.g. texture or temperature) and more 'indirect' ones such as those grounded in visual perception. The former are more 'primary':

i) sensed earliest – already in the womb, tactition being the first sense that begins to develop before 8 weeks gestational age together with the emergence of the nervous system [Montagu 1978], before taste and smell (14 weeks g.a.), hearing (16 weeks g.a.) [Shahidullah & Hepper 1992]) or vision (week 18 onwards),

ii) available in more 'primitive' organisms without vision or hearing,

iii) perceptible during half-sleep, and

iv) impacting our bodily functioning more strongly (the somatic reaction to extremely high or low temperatures, pressure or skin irritation is more likely to be stronger than e.g. to an unpleasant sight or sound).

This might account for the lack of activation in visual cortical areas.

[7] But see Bedny *et al.* [2008], Mahon & Caramazza [2008], Postle *et al.* [2008], Toni *et al.* [2008], or Kemmerer & Gonzalez-Castillo [2010] for opposing views.

- semantic interference and facilitation in the Stroop effect [longer RTs needed to name colour names written in incongruent ink hue; Jaensch 1929; Stroop 1935];
- clinical studies indicating that processing of action concepts degrades if action- or vision-related brain areas are lesioned in motor neuron diseases [Damasio *et al.* 1996; Tranel *et al.* 1997; Bak *et al.* 2001; Neininger & Pulvermüller 2003] and semantic dementia [Pulvermüller *et al.* 2010];
- comprehension of action words deteriorating after loss of procedural knowledge [*cf.* Boulenger *et al.* 2008 on Parkinson's disease patients; also Bak *et al.* 2006];

8) parallel emergence of speech and gesture in infancy (Iverson & Thelen 1999).

9) co-speech gesture reducing cognitive load [Goldin-Meadow *et al.* 2001], and indications of a dual-task advantage for bimodal (signed-spoken) language production [i.e., production of code-blends, with elements of the signed and spoken languages appearing simultaneously; Kaufmann & Kaul 2012]; or

10) Conceptual Blending theory [Fauconnier & Turner 2002] explaining language creativity as a semantic process operating on the output of perception and interaction with the world to create new structures.

Thus, independently of theoretical persuasion, without taking into account both the architecture of the human brain, and embodiment – the interaction of the language faculty with the sensory apparatus and motor system – it is unrealistic to replicate accurately the processes which take place during language acquisition, comprehension, or production, or during non-linguistic actions. Cognitive mechanisms are synergistically intertwined with affective and somatic components, and largely inseparable [Ziemke 2011].

3. Corollaries for robotics

While evidently robots, even anthropomorphic ones, are far from isomorphic with humans in terms of both the 'brain' and the rest of the body, and robust artificial cognitive agents can bypass the human limitations inherent in most of the above points (just as they could overcome some contingencies resulting from the material properties of the human brain and bodily features such as synaptic speed and efficiency, the physical characteristics of the vocal tract, the auditory perception system, or muscular flexibility[8]), they could benefit from

[8] Just as robots can have an advantage when equipped with e.g. infrared or ultrasonic sensors.

strengthened associative connections owing to the motor and semantic resonance in both the optimization of their processes, and reactivity and sensitivity to environmental stimuli, across a range of tasks:

(i) in grounded language understanding [*cf.* e.g. Glenberg & Kaschak 2002; Feldman & Narayanan 2004; Bergen & Chang 2005; Gallese & Lakoff 2005; Kaschak *et al.* 2005; Knoeferle & Crocker 2006, 2007; Sato *et al.* 2008], where structuring the environment acts as scaffolding [Schöner 2011];

(ii) in automated articulation-based speech recognition (utilising motor information, i.e. combining spoken input with visual data – e.g. the shape of the speakers lips – and maybe even data such as strength of the incoming airstream);

(iii) while learning about context-dependent phenomena in the surrounding world [e.g. action sequences and argument structure in construction grammar; *cf*. Dominey 2007; since embodiment plays a constitutive role in the process of cognition; Vernon 2010], or in the process of language acquisition in general [because language – at least in the initial stages – is only acquired by situated embodied direct engagement with the world, and not just passive perception, e.g. watching television; *cf.* e.g. Steels 2009];

(iv) to help with storage and retrieval due to the benefits of episodic memory;

(v) to support action prediction, planning and anticipation [Koelewijn *et al.* 2008; Stapel *et al.* 2010; van Elk *et al.* 2010], including prediction of the next sensory feedback,

(vi) to support action execution (with linguistic input making the actor better aware of the affordances, i.e. physically feasible action possibilities), and

(vii) to reinforce feedback in 'soft robotics' and morphological computation, where there is no clear separation between the controller (or orchestrator) and the hardware (morphology), and the tasks are distributed between the brain, body, and environment [*cf.* e.g. Paul 2004; Pfeiffer 2011; which also has the aim of off-loading computation; Di Paolo 2009];

(viii) in cognitive developmental robotics, aiming at understanding human cognitive developmental processes by synthetic or constructive approaches [Asada *et al.* 2009, Asada 2011, Ishiguro *et al.* 2011];

(ix) in common grounding and alignment, which are crucial for fruitful situated human-machine interaction, and which are another area where sensory experience must be coordinated with linguistic interaction.

4. Towards a broader definition of multisensory integration

In order to form a meaningful experience and construct coherent, reliable and robust representations of the surrounding world, the human brain combines

prior knowledge with sensory input arriving from various modalities and integrates these at multiple levels of the neuraxis. This serves to maximize the efficiency of everyday performance and learning, enhancing the salience of the events, helping increase the detection and identification of the external stimuli, disambiguate them, compensate for incomplete information, and shorten reaction times. In view of the inseparability of language and the body, the concept of multisensory integration – whether in natural or artificial cognitive agents – should be extended and cover both the linguistic input and the complementary information that the brain combines from temporally coincident sensory impressions. This does not mean that we should 'dumb down' the statistical processes where they operate successfully; instead, where the input stream in one channel is too noisy, turning on auxiliary channels and interacting with the environment in an active manner may generate ancillary data and help e.g. disambiguate the signal and take the right decision [see also Pfeifer & Scheier 1997; Beer 2003]. An added benefit would then be significantly reduced programming costs.

Acknowledgments. The author wishes to thank Noam Chomsky, Anna Esposito, Taivo Lints, Richard Littauer, Gary Lupyan, Vincent C. Müller, Katerina Pastra, Michael Pleyer, Yulia Sandamirskaya, and Luc Steels for invaluable commentary, discussion and bibliographical references. Naturally, willingness to comment does not imply endorsement; all the usual disclaimers apply. Earlier version of this paper were presented at the 4th EUCogII Members' Conference "Embodiment – Fad or Future?", Anatolia College/American College of Thessaloniki, 11–12 Apr. 2011 and the symposium "Linguistic and Cognitive Approaches to Dialogue Agents", University of Birmingham, 7 Jul. 2012. The arguments are further elaborated and discussed in [Paradowski 2012].

References

Archibald, L. M. D. / Gathercole, S. E. [2006], *Prevalence of SLI in Language Resource Units*. J Res Spec Educ Needs 6, 1, 3–10.

Asada, M. / Hosoda, K. / Kuniyoshi, Y. / Ishiguro, H. / Inui, T. / Yoshikawa, Y. / Ogino, M. / Yoshida, C. [2009], *Cognitive developmental robotics: a survey*. IEEE Transactions on Autonomous Mental Development 1(1), 12–34.

Asada, M. [2011], *Can cognitive developmental robotics cause a paradigm shift?* In: Krichmar, J. L. / Wagatsuma, H. (eds.), Neuromorphic and Brain-Based Robots: Trends and Perspectives. Cambridge: Cambridge University Press, 251–73.

Aziz-Zadeh, L. / Wilson, S. M. / Rizzolatti, G. / Iacoboni, M. [2006], *Congruent embodied representations for visually presented actions and linguistic phrases describing actions*. Curr Biol 16, 1818–1823.

Bak, T. H. / O'Donovan, D. G. / Xuereb, J. H. / Boniface, S. / Hodges, J. R. [2001], *Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease – dementia – aphasia syndrome*. Brain 124, 103–120.

Bak, T. H. / Yancopoulou, D. / Nestor, P. J. / Xuereb, J. H. / Spillantini, M. G. / Pulvermuller, F. / Hodges, J. R. [2006], *Clinical, imaging and pathological correlates of a hereditary deficit in verb and action processing*. Brain 129 (2), 321–332.

Bedny, M. / Caramazza, A. / Grossman, E. / Pascual-Leone, A. / Saxe, R. [2008], *Concepts are more than percepts: the case of action verbs*. J Neurosci 28, 11347–11353.

Beer, R. [2003], *The dynamics of active categorical perception in an evolved model agent*. Adaptive Behav 11, 209–43.

Bergen, B. / Chang, N. [2005], *Embodied construction grammar in simulation-based language understanding*. In: Östman, J.-O. / Fried, M. (eds.), Construction grammars: Cognitive grounding and theoretical extensions. Amsterdam: Benjamins, 147–190.

Bergen, B. / Lindsay, S. / Matlock, T. / Narayanan, S. [2007], *Spatial and linguistic aspects of visual imagery in sentence comprehension*. Cogn Sci 31, 733–764.

Bergen, B. / Narayan, S. / Feldman, J. [2003], *Embodied verbal semantics: Evidence from an image-verb matching task*. In: Alterman, R. / Kirsh, D. (eds.), Proceedings of the 25th annual conference of the Cognitive Science Society. Boston, MA: Cognitive Science Society,139–144.

Bergen, B. / Wheeler, K. B. [2010], *Grammatical aspect and mental simulation*. Brain Lang 112, 150–158.

Bishop, D. V. M. / Snowling, M. J. [2004], *Developmental dyslexia and specific language impairment: same or different*? Psychol Bull 130, 858–888.

Bonda, E. / Petrides, M. / Frey, S. / Evans, A. [1994], *Frontal cortex involvement in organized sequences of hand movements: Evidence from a positron emission topography study*. Soc Neurosci Abstr 20, 353.

Bookheimer, S. [2002], *Functional MRI of language: New approaches to understanding the cortical organization of semantic processing*. Ann Rev Neurosci 25, 151–188.

Borghi, A. M. / Cimatti, F. [2010], *Embodied cognition and beyond: acting and sensing the body*. Neuropsychologia 48, 763–773.

Borghi, A. M. / Scorolli, C. [2009], *Language comprehension and hand motion simulation*. Hum Mov Sci, 28, 12–27.

Boulenger, V. / Hauk, O. / Pulvermüller, F. [2009], *Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension*. Cereb Cortex, 19(8), 1905–1914.

Boulenger, V. / Mechtouff, L. / Thobois, S. / Broussolle, E. / Jeannerod, M. / Nazir, T. A. [2008], *Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns*. Neuropsychologia 46 (2), 743–756.

Boulenger, V. / Roy, A. C. / Paulignan, Y. / Déprez, V. / Jeannerod, M. / Nazir, T. A. [2006], *Cross-talk between language processes and overt motor behavior in the first 200 ms of processing*. J Cogn Neurosci 18(10), 1607–1615.

Broca, P. [1861], *Remarques sur le siège de la faculté de la parole articulée, suivies d'une observation d'aphémie (perte de parole)*. Bull Soc Anatom (Paris) 36, 330–357.

Brown, C. M. / Hagoort, P. [1999], *The Neurocognition of Language*. Oxford: Oxford University Press.

Buccino, G. / Riggio, L. / Melli, G. / Binkofski, F. / Gallese, V. / Rizzolatti, G. [2005], *Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study*. Cogn Brain Res, 24(3), 355–363.

Caplan, D. [2006], *Aphasic deficits in syntactic processing*. Cortex, 42(6), 797–804.

Chao, L. L. / Haxby, J. V. / Martin, A. [1999], *Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects*. Nature Neurosci 2(10), 913–919.

Chao, L. L., / Martin, A. [2000], *Representation of manipulable man-made objects in the dorsal stream*. Neuroimage 12, 478–484.

Chomsky, N. [1957], *Syntactic structures*. The Hague: Mouton.

Chomsky, N. [1965], *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. [1975], *Reflections on language*. New York: Pantheon.

Chomsky, N. [1980], *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. [2005], *Three factors in language design*. Ling Inq 36(1), 1–22.

Cohen, N. J. [2002], *Developmental language disorders*. In: Howlin, P. / Udwin, O. (eds.), Outcomes in neurodevelopmental and genetic disorders, New York, NY: Cambridge University Press, 26–55.

Curtiss, S. [1981], *Genie: The Case of a Modern Wild Child*. New York: Academic Press.

Damasio, H. / Grabowski, T. J. / Tranel, D. / Hichwa, R. D. / Damasio, A. R. [1996], *A neural basis for lexical retrieval*. Nature 380, 499–505.

Damasio, H. / Tranel, D. / Grabowski, T. / Adolphs, R. / Damasio, A. [2004], *Neural systems behind word and concept retrieval*. Cognition 92: 179–229.

Del Prado Martin, F. / Hauk, O. / Pulvermüller, F. [2006], *Category specificity in the processing of color-related and form-related words: An ERP study*. Neuroimage 29, 29–37.

Dennis, M. / Whitaker, H. [1976], *Language acquisition following hemidecortication: Linguistic superiority of the left over the right hemisphere*. Brain Lang 3, 404–433.

Desai, R. H. / Binder, J. R. / Conant, L. L. / Seidenberg, M. S. [2010], *Activation of sensory-motor areas in sentence comprehension*. Cereb Cortex 20, 468–478.

Di Paolo, E. [2009, Oct 10], *From sensorimotor coordination to enaction: Agency, sense-making and sociality as horizons for embodied cognition*. 1st EU-CogII Members' Conference "Challenges for artificial cognitive systems", University Medical Center Hamburg-Eppendorf.

Di Sciullo, M. / Boeckx, C. [2011], *The Biolinguistic Enterprise. New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: Oxford University Press.

Diesch, E. / Eulitz, C. / Hampson, S., / Ross, B. [1996], *The neurotopography of vowels as mirrored by evoked magnetic field measurements*. Brain Lang 53(2), 143–168.

Dijkstra, K. / Kaschak, M. / Zwaan, R. [2007], *Body posture facilitates retrieval of autobiographical memories*. Cognition 102: 139–49.

Dominey, P. F. [2007], *Spoken language and vision for adaptive human-robot cooperation*. In: Hackel, M. (ed.), Humanoid robotics. Vienna: ARS International.

Embick, D. / Marantz, A. / Miyashita, Y. / O'Neil, W. / Sakai, K. L. [2000], *A syntactic specialization for Broca's area*. Proc Natl Acad Sci USA 97, 6150–6154.

Fadiga, L. / Craighero, L. / Buccino, G. / Rizzolatti, G. [2002], *Speech listening specifically modulates the excitability of tongue muscles: A TMS study*. Eur J Neurosci 15(2), 399–402.

Fadiga, L. / Fogassi, L. / Pavesi, G. / Rizzolatti, G. [1995], *Motor facilitation during action observation: A magnetic stimulation study*. J Neurophysiol 73(6), 2608–2611.

Fauconnier, G. / Turner, M. [2002], *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Feldman, J. / Narayanan, S. [2004], *Embodied meaning in a neural theory of language*. Brain Lang 89(2), 385–392.

Feuk, L. / Kalervo, A. / Lipsanen-Nyman, M. / Skaug, J. / Nakabayashi, K. / Finucane, B. / Hartung, D. / Innes, M. / Kerem, B. / Nowaczyk, M. J. / Rivlin, J. / Roberts, W. / Senman, L. / Summers, A. / Szatmari, P. / Wong, V. / Vincent, J. B. / Zeesman, S. / Osborne, L. R. / Cardy, J. O. / Kere, J. / Scherer, S. W. / Hannula-Jouppi, K. [2006], *Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia*. Am J Hum Genet 79(5), 965–972.

Fischer, M. H. / Zwaan, R. A. [2008], *Embodied language: A review of the role of motor system in language comprehension*. Q J Exp Psychol 61, 825–850.

Fisher, S. E. / Lai, C. S. / Monaco, A. P. [2003], *Deciphering the genetic basis of speech and language disorders*. Ann Rev Neurosci 26, 57–80.

Fisher, S. E. / Scharff, C. [2009], *FOXP2 as a molecular window into speech and language*. Trends Genet 25(4), 166–77.

Fodor, J. A. [1975], *The language of thought*. Cambridge: Harvard University Press.

Fodor, J. A. [1983], *The Modularity of Mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Frak, V. / Nazir, T. / Goyette, M. / Cohen, H. / Jeannerod, M. [2010], *Grip force is part of the semantic representation of manual action verbs*. PLoS ONE 5(3), e9728. doi:10.1371/journal.pone.0009728.

Friederici, A. D. [2002], *Towards a neural basis of auditory sentence processing*. Trends Cogn Sci 6(2), 78–84.

Gallese, V. / Lakoff, G. [2005], *The brain's concepts: The role of the sensory-motor system in reason and language*. Cogn Neuropsychol 22, 455–479.

Gentilucci, M. / Benuzzi, F. / Bertonali, L. / Daprati, E. / Gangitano, M. [2000], *Language and motor control*. Exp Brain Res 133(4), 468–490.

Gentilucci, M. / Gangitano, M. [1998], *Influence of automatic word reading on motor control*. Eur J Neurosci 10, 752–756.

Geschwind, N. [1970], *The organization of language and the brain*. Science 170 (961), 140–144.

Glenberg, A. / Meyer, M. / Lindem, K. [1987], *Mental models contribute to foregrounding during text comprehension*. J Mem Learning 26, 69–83.

Glenberg, A. M. / Kaschak, M. P. [2002], *Grounding language in action*. Psychon Bull Rev 9(3), 558–565.

Glover, S. / Dixon, P. [2002], *Semantics affect the planning but not control of grasping*. Exp Brain Res 146, 383–387.

Glover, S. / Rosenbaum, D. A. / Graham, J. / Dixon, P. [2004], *Grasping the meaning of words*. Exp Brain Res 154(1), 103–108.

Goldberg, R. F. / Perfetti, C. A. / Schneider, W. [2006], *Perceptual knowledge retrieval activates sensory brain regions*. J Neurosci 26(18), 4917–4921.

Goldin-Meadow, S. / Nusbaum, H. / Kelly, S. / Wagner, S. [2001], *Explaining math: Gesturing lightens the load*. Psychological Sci 12, 516–522.

González, J. / Barros-Loscertales, A. / Pulvermüller, F. / Meseguer, V. / Sanjuán, A. / Belloch, V. / Ávila, C. [2006], *Reading* cinnamon *activates olfactory brain regions*. Neuroimage 32(2), 906–912.

Grodzinsky, Y. / Friederici, A. D. [2006], *Neuroimaging of syntax and syntactic processing*. Curr Opin Neurobiol 16(2), 240–246.

Hagoort, P. / Ramsey, N. F. / Rutten, G. J. M. / van Rijen, P. C. [1999], *The role of the left anterior temporal cortex in language processing*. Brain Lang 69, 322–325.

Hagoort, P. [2003], *Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations*. J Cogn Neurosci 15, 883–899.

Hagoort, P. [2005], *On Broca, brain, and binding: A new framework*. Trends Cogn Sci 9(9), 416–423.

Hahne, A. / Friederici, A. D. [2002], *Differential task effects on semantic and syntactic processes as revealed by ERPs*. Cogn Brain Res 13, 339–356.

Hauk, O. / Johnsrude, I. / Pulvermüller, F. [2004], *Somatotopic representation of action words in human motor and premotor cortex*. Neuron 41, 301–7.

Hauk, O. / Pulvermüller, F. [2004], *Neurophysiological distinction of action words in the fronto-central cortex*. Hum Brain Mapp 21, 191–201.

Hauser, M. D. / Chomsky, N. / Fitch, W. T. [2002], *The language faculty: What is it, who has it, and how did it evolve?* Science 298(5598), 1569–1579.

Hickok, G. / Poeppel, D. [2000], *Towards a functional neuroanatomy of speech perception*. Trends Cogn Sci 4(4), 131–138.

Hickok, G. / Poeppel, D. [2004], *Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language*. Cognition, 92(1–2), 67–99.

Hickok, G. / Poeppel, D. [2007], *The cortical organization of speech processing*. Nat Rev Neurosci 8(5), 393–402.

Horwitz, B. / Amunts, K. / Bhattacharyya, R. / Patkin, D. / Jeffries, J. / Zilles, K. / Braun, A. R. [2003], *Activation of Broca's area during the production of spoken and signed language: A combined cytoarchitectonic mapping and PET analysis*. Neuropsychologia 41, 1868–1876.

Ishiguro, H. / Minato, T. / Yoshikawa, Y. / Asada, M. [2011], *Humanoid Platforms for Cognitive Developmental Robotics*. Intl J Humanoid Robotics 8(3), 391–418.

Iverson, J. / Thelen, E. [1999], *Hand, mouth, and brain: The dynamic emergence of speech and gesture*. J Consc Stud 6(11–12), 19–40.

Jackendoff, R. [1983], *Semantics and Cognition*. Cambridge: MIT Press.

Jackendoff, R. [1997], *The Architecture of the Language Faculty*. Cambridge: MIT Press.

Jackendoff, R. [2002], *Foundations of Language*. Oxford: Oxford University Press.

Jaensch, E. R. [1929], *Grundformen menschlichen Seins*. Berlin: Otto Elsner.

Jenkins, L. [2000], *Biolinguistics. Explaining the Biology of Language*. Cambridge: Cambridge University Press.

Joanisse, M. F. / Seidenberg, M. S. [1998], *Specific language impairment: A deficit in grammar or processing?* Trends Cogn Sci 2, 240–247.

Kaschak, M. / Madden, C. / Therriault, D. / Yaxley, R. / Aveyard, M. / Blanchard, A. / Zwaan, R. [2005], *Perception of Motion Affects Language Processing*. Cognition 94, B79–B89.

Katz, J. J. / Fodor, J. [1963], *The structure of a semantic theory*. Language 39, 170–210.

Katz, J. J. [1972], *Semantic Theory*. New York: Harper and Row.

Kaufmann, E. / Kaul, Th. [2012], *Language switch costs and dual-task costs in bimodal language production*. Paper presented at the conf. 'Formal and Experimental Advances in Sign Language Theory' (FEAST 2012), Univ. Warsaw.

Kaup, B. / Lüdtke, J. / Zwaan, R. A. [2005], *Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences*. In: Bara, B. G. / Barsalou, L. / Bucciarelli, M. (eds.), Proceedings of the 27th annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum, 1114–1119.

Kaup, B. / Yaxley, R. H. / Madden, C. J. / Zwaan, R. A. / Luedtke, J. [2007], *Experiential simulations of negated text information*. Q J Exp Psychol 60, 976–990.

Kemmerer, D. / Castillo, J. G. / Talavage, T. / Patterson, S. / Wiley, C. [2008], *Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI*. Brain Lang 107, 16–43.

Kemmerer, D. / Gonzalez-Castillo, J. [2010], *The two-level theory of verb meaning: an approach to integrating the semantics of action with the mirror neuron system*. Brain Lang 112, 54–76.

Kintsch, W. [1998], *Comprehension: A Paradigm for Cognition*. Cambridge, MA: Cambridge University Press.

Knoeferle, P. / Crocker, M. W. / Scheepers, C. / Pickering, M. J. [2005], *The influence of the immediate visual context on incremental thematic role assignment: Evidence from eye-movements in depicted events*. Cognition 95, 95–127.

Knoeferle, P. / Crocker, M. W. [2006], *The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking*. Cogn Sci 30, 481–529.

Knoeferle, P. / Crocker, M. W. [2007], *The influence of recent scene events on spoken comprehension: Evidence from eye movements*. J Mem Lang 57(4), 519–543.

Knopka, G. / Bomar, J. M. / Winden, K. / Coppola, G. / Jonsson, Z. O. / Gao, F. / Peng, S. / Preuss, T. M. / Wohlschlegel, J. A. / Geschwind, D. H. [2009], *Human-specific transcriptional regulation of CNS development genes by FOXP2*. Nature 462, 213–7.

Koelewijn, T. / van Schie, H. T. / Bekkering, H. / Oostenveld, R. / Jensen, O. [2008], *Motor-cortical beta oscillations are modulated by correctness of observed action*. Neuroimage 40, 767–775.

Kolb, B. / Whishaw, I. Q. [2003] *Fundamentals of Human Neuropsychology* (5th ed.). New York: Worth.

Lacey, S. / Stilla, R. / Sathian, K. [2012], *Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex*. Brain Lang 120(3), 416–421.

Lai, C. S. L. / Fisher, S. E. / Hurst, J. A. / Vargha-Khadem, F. / Monaco, A. P. [2001], *A forkhead-domain gene is mutated in a severe speech and language disorder*. Nature 413 (6855), 519–23.

Lakoff, G. / Johnson, M. [1980], *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. [1987], *Women, Fire, and Dangerous Things. What categories reveal about the mind*. Chicago: University of Chicago Press.

Lenneberg, E. H. [1967], *Biological foundations of language*. New York: Wiley.

Leonard, L. B. [1998], *Children with specific language impairment*. Cambridge, MA: MIT Press.

Levelt, W. J. M. [1989], *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lindemann, O. / Stenneken, P. / van Schie, H. T. / Bekkering, H. [2006], *Semantic activation in action planning*. J Exp Psychol Hum Percept Perform 32, 633–643.

Lupyan, G. / Spivey, M. J. [2008], *Ascribing meaning to unfamiliar items facilitates visual processing*. Curr Biol 18, R410–R412.

MacDermot, K. D. / Bonora, E. / Sykes, N. / Coupe, A. M. / Lai, C. S. / Vernes, S. C. / Vargha-Khadem, F. / McKenzie, F. / Smith, R. L. / Monaco, A. P. / Fisher, S. E. [2005], *Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits*. Am J Hum Genet 76(6), 1074–80.

Madden, C. J. / Zwaan, R. A. [2003], *How does verb aspect constrain event representations?* Mem Cognit 31, 663–672.

Mahon, B. Z., / Caramazza, A. [2008], *A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content*. J Physiol-Paris 102(1–3), 59–70.

Marinova-Todd, S. / Marshall, D. / Snow, C. [2000], *Three misconceptions about age and L2 learning*. TESOL Quarterly 34, 9–34.

Martin, A. / Wiggs, C. L. / Ungerleider, L. G. / Haxby, J. V. [1996], *Neural correlates of category-specific knowledge*. Nature 379, 649–652.

Martin, A. [2007], *The representation of object concepts in the brain*. Ann Rev Neurosci 58, 25–45.

Meader, C. L. / Muyskens, J. H. [1950], *Handbook of Biolinguistics*. New York: Wiley.

Miller, G. A. / Johnson-Laird, P. N. [1976], *Language and Perception*. Cambridge, MA: Harvard University Press

Montagu, A. [1978], *Touching: The Human Significance of the Skin*. New York: Harper & Row.

Nazir, T. A. / Boulenger, V. / Roy, A. C. / Silber, B. Y. / Jeannerod, M. / Paulignan, Y. [2008], *Language-induced motor perturbations during the execution of a reaching movement*. Q J Exp Psychol 61(6), 933–943.

Neininger, B. / Pulvermüller, F. [2003], *Word-category specific deficits after lesions in the right hemisphere*. Neuropsychologia 41(1), 53–70.

Newell, A. / Simons, H. A. [1972], *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nowak, M. A. / Komorova, N. L. / Niyogi, P. [2002], *Computational and evolutionary aspects of language*. Nature 417, 611–617.

Nowak, M. A. / Plotkin, J. B. / Jansen, V. A. A. [2000], *The evolution of syntactic communication*. Nature 404, 495–498.

Obleser, J. / Boecker, H. / Drzezga, A. / Haslinger, B. / Hennenlotter, A. / Roettinger, M. / Eulitz, C. / Rauschecker, J. P. [2006], *Vowel sound extraction in anterior superior temporal cortex*. Hum Brain Mapp 27(7), 562–571.

Oliveri, M. / Finocchiaro, C. / Shapiro, K. / Gangitano, M. / Caramazza, A. / Pascual--Leone, A. [2004], *All talk and no action: a transcranial magnetic stimulation*

study of motor cortex activation during action word production. J Cogn Neurosci 16(3), 374–381.

Osterhout, L. [1997], *On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences*. Brain Lang 59, 494–522.

Paradowski, M. B. [2012], *Developing embodied multisensory dialogue agents*. In: Rzepka, R. / Ptaszyński, M. / Dybala, P. (eds.), *Linguistic and Cognitive Approaches To Dialogue Agents*. Birmingham: The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 6–14.

Paul, C. [2004], *Morphology and computation*. *Proc of Int Conf on the Simulation of Adaptive Behavior: From animals to animats*. Cambridge, MA: MIT Press, 33–38.

Paus, T. / Perry, D. W. / Zatorre, R. J. / Worsley, K. J. / Evans, A. C. [1996], *Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges*. Eur J Neurosci 8(11), 2236–2246.

Pfeifer, R. / Scheier, C. [1997], *Sensory-motor coordination: The metaphor and beyond*. Robotics and Autonomous Systems 20, 157–78.

Pfeiffer, R. [2011, Apr 11], *The emergence of cognition from the interaction of brain, body, and environment*. 4th EUCogII Members' Conference "Embodiment – Fad or Future?". Thessaloniki: Anatolia College.

Piatelli-Palmarini, M. [1989], *Evolution, selection and cognition: from "learning" to parameter setting in biology and in the study of language*. Cognition 31, 1–44.

Postle, N. / McMahon, K. L. / Ashton, R. / Meredith, M. / de Zubicaray, G. I. [2008], *Action word meaning representations in cytoarchitectonically defined primary and premotor cortices*. Neuroimage 43, 634–644.

Price, C. J. [2000], *The anatomy of language: contributions from functional neuroimaging*. J Anat 197, 335–359.

Pulvermüller, F. / Assadollahi, R. [2007], *Grammar or serial order? Discrete combinatorial brain mechanisms reflected by the syntactic mismatch negativity*. J Cogn Neurosci 19(6), 971–980.

Pulvermüller, F. / Härle, M. / Hummel, F. [2001], *Walking or talking? Behavioural and neurophysiological correlates of action verb processing*. Brain Lang 78, 143–168.

Pulvermüller, F. / Hauk, O. / Nikulin, V. V. / Ilmoniemi, R. J. [2005a], *Functional links between motor and language systems*. Eur J Neurosci 21(3), 793–797.

Pulvermüller, F. / Hauk, O. [2006], *Category-specific processing of color and form words in left fronto-temporal cortex*. Cereb Cortex 16(8), 1193–1201.

Pulvermüller, F. / Huss, M. / Kherif, F. / Moscoso del Prado Martin, F. / Hauk, O. / Shtyrov, Y. [2006], *Motor cortex maps articulatory features of speech sounds*. Proc Natl Acad Sci USA 103(20), 7865–7870.

Pulvermüller, F. / Pye, E. / Dine, C. / Hauk, O. / Nestor, P. / Patterson, K. [2010], *The word processing deficit in semantic dementia: All categories are equal but some categories are more equal than others*. J Cogn Neurosci 22(9), 2027–2041.

Pulvermüller, F. / Shtyrov, Y. / Ilmoniemi, R. [2005b], *Brain signatures of meaning access in action word recognition*. J Cogn Neurosci 17, 884–892.

Pulvermüller, F. / Shtyrov, Y. / Ilmoniemi, R. J. [2003], *Spatio-temporal patterns of neural language processing: An MEG study using minimum-norm current estimates*. Neuroimage 20, 1020–1025.

Pulvermüller, F. / Shtyrov, Y. [2003], *Automatic processing of grammar in the human brain as revealed by the mismatch negativity*. Neuroimage 20, 159–172.

Pylyshyn, Z. W. [1984], *Computation and cognition*. Cambridge, MA: MIT Press.

Raposo, A. / Moss, H. E. / Stamatakis, E. A. / Tyler, L. K. [2009], *Modulation of motor and premotor cortices by actions, action words and action sentences*. Neuropsychologia 47, 388–396.

Richardson, D. C. / Spivey, M. J. / Barsalou, L. W. / McRae, K. [2003], *Spatial representations activated during real-time comprehension of verbs*. Cogn Sci 27, 767–780.

Rieser, J. / Garing, A. / Young, M. [1994], *Imagery, action and young children's spatial orientation: It's not being there that counts, it's what one has in mind*. Child Dev 45, 1043–56.

Rogalsky, C. / Matchin, W. / Hickok, G. [2008], *Investigating the relationship between Broca's area, sentence processing and working memory: An fMRI study of sentence comprehension during articulatory suppression*. Paper presented at the Society for Neuroscience, Washington, DC .

Rueschemeyer, S.-A. / Brass, M. / Friederici, A. D. [2007], *Comprehending prehending: Neural correlates of processing verbs with motor stems*. J Cogn Neurosci 19, 855–865.

Rueschemeyer, S.-A. / Lindemann, O. / van Rooij, D. / van Dam, W. / Bekkering, H. [2010a], *Effects of intentional motor actions on embodied language processing*. Exp Psychol 57(4), 260–266.

Rueschemeyer, S.-A. / Pfeiffer, C. / Bekkering, H. [2010b], *Body Schemantics: On the role of the body schema in embodied lexical representations*. Neuropsychologia 48(3), 774–781.

Rueschemeyer, S.-A. / Rooij, D. / Lindemann, O. / Willems, R. / Bekkering, H. [2010c], *The function of words: Distinct neural correlates for words denoting differently manipulable objects*. J Cogn Neurosci 22, 1844–1851.

Saccuman, M. C. / Cappa, S. F. / Bates, E. A. / Arevalo, A. / Della Rosa, P. / Danna, M. / Perani, D. [2006], *The impact of semantic reference on word class: An fMRI study of action and object naming*. Neuroimage 32, 1865–1878.

Sato, M. / Mengarelli, M. / Riggio, L. / Gallese, V. / Buccino, G. [2008], *Task related modulation of the motor system during language processing*. Brain Lang 105(2), 83–90.

Schöner, G. [2011, Apr. 12], *Putting representation back into embodied cognition: Dynamical Field Theory*. 4th EUCogII Members' Conference "Embodiment – Fad or Future?". Thessaloniki: Anatolia College.

Scorolli, C. / Borghi, A. M. [2007], *Sentence comprehension and action: Effector specific modulation of the motor system*. Brain Res 1130, 119–124.

Shahidullah, S. / Hepper, P. G. [1992], *Hearing in the Fetus: Prenatal Detection of Deafness*. Intl J Prenatal and Perinatal Studies 4(3/4), 235–40.

Shapiro, L. A. [2004], *The Mind Incarnate*. Cambridge, MA: MIT Press.

Siakaluk, P. / Pexman, P. / Aguilera, L. / Owen, W. / Sears, C. [2008], *Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect*. Cognition 106: 433–43.

Simmons, W. K. / Ramjee, V. / Beauchamp, M. S. / McRae, K. / Martin, A. / Barsalou, L. W. [2007], *A common neural substrate for perceiving and knowing about color*. Neuropsychologia 45(12), 2802–2810.

Smith, N. V. / Tsimpli, I.-M. / Ouhalla, J. [1993], *Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant*. Lingua 91, 279–347.

Smith, N. V. [1989], *The Twitter Machine: Reflections on Language*. Oxford: Blackwell.

Spivey, M. J. / Tyler, M. J. / Eberhard, K. M. / Tanenhaus, M. K. [2001], *Linguistically mediated visual search*. Psychol Sci 12, 282–286.

Stapel, J. C. / Hunnius, S. / van Elk, M. / Bekkering, H. [2010], *Motor activation during observation of unusual vs. ordinary actions in infancy*. Soc Neurosci 5, 451–460.

Steels, L. [2009, Nov 26], *The origins and evolution of languages: Darwin's unsolved mystery*. International workshop "150 Years after Darwin: From Molecular Evolution to Language", Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca.

Stein, B. E. / London, N. / Wilkinson, L. K. / Price, D. D. [1996], *Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis*. J Cogn Neurosci 8(6), 497–506.

Stroop, J. R. [1935], *Studies of interference in serial verbal reactions*. J Exp Psychol 18 (6), 643–662.

Tanenhaus, M. K. / Spivey-Knowlton, M. J. / Eberhard, K. / Sedivy, J. C. [1995], *Integration of visual and linguistic information in spoken language comprehension*. Science 268, 632–634.

Tettamanti, M. / Buccino, G. / Saccuman, M. C. / Gallese, V. / Danna, M. / Scifo, P. / Fazio, F. / Rizzolatti, G. / Cappa, SF. / Perani, D. [2005], *Listening to action-related sentences activates fronto-parietal motor circuits*. J Cogn Neurosci 17, 273–281.

Tettamanti, M. / Manenti, R. / Della Rosa, P. A. / Falini, A. / Perani, D. / Cappa, S. F. / Moro A. [2008], *Negation in the brain: Modulating action representations*. Neuroimage 43, 358–367.

Toni, I. / de Lange, F. P. / Noordzij, M. L. / Hagoort, P. [2008], *Language beyond action*. J Physiol-Paris 102, 71–79.

Tranel, D. / Damasio, H. / Damasio, A. R. [1997], *A neural substrate for the retrieval of conceptual knowledge*. Neuropsychologia 35, 1319–1327.

Tucker, M. / Ellis, R. [2001], *The potentiation of grasp types during visual object categorization*. Visual Cognition 8, 769–800.

Ullman, M. T. [2001], *A neurocognitive perspective on language: The declarative/procedural model*. Nature Reviews Neuroscience 2(10), 717–726.

Uppenkamp, S. / Johnsrude, I. S. / Norris, D. / Marslen-Wilson, W. / Patterson, R. D. [2006], *Locating the initial stages of speech-sound processing in human temporal cortex*. Neuroimage 31(3), 1284–1296.

van Elk, M. / van Schie, H. T. / Zwaan, R. A. / Bekkering, H. [2010], *The functional role of motor activation in language processing: motor cortical oscillations support lexical-semantic retrieval*. Neuroimage 50, 665–677.

van Schie, H. T. / Wijers, A. A. / Kellenbach, M. L. / Stowe, L. A. [2003], *An event-related potential investigation of the relationship between semantic and perceptual levels of representation*. Brain Lang 86, 300–325.

Vargha-Khadem, F. / Gadian, D. G. / Copp, A. / Mishkin, M. [2005], *FOXP2 and the neuroanatomy of speech and language*. Nature Reviews Neuroscience 6(2), 131–137.

Vernes, S. C. / Newbury, D. F. / Abrahams, B. / Winchester, L. / Nicod, J. / Groszer, M. / Alarcon, M. / Oliver, P. L. / Davies, K. E. / Geschwind, D. H. / Monaco, A. P. / Fisher, S. E. [2008], *A functional genetic link between distinct developmental language disorders*. N Engl J Med 359, 2337–2345.

Vernes, S. C. / Nicod, J. / Elahi, F. M. / Coventry, J. A. / Kenny, N. / Coupe, A. M. / Bird, L. E. / Davies, K. E. / Fisher, S. E. [2006], *Functional genetic analysis of mutations implicated in a human speech and language disorder*. Hum Mol Genet 15(21), 3154–67.

Vernon, D. [2010, Jan 29], *Cognitive development and the iCub humanoid robot*. 2nd EUCogII Members' Conference "Development of Cognition in Artificial Agents", University of Zürich.

Wernicke, C. [1874], *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau: Kohn & Weigert.

Willems, R. M. / Hagoort, P. / Casasanto, D. [2010], *Body-specific representations of action verbs: Neural evidence from right- and left-handers*. Psychological Science 21, 67–74.

Wilson, S. M. / Saygin, A. P. / Sereno, M. I. / Iacoboni, M. [2004], *Listening to speech activates motor areas involved in speech production*. Nature Neuroscience 7(7), 701–702.

Yamada, J. [1990], *Laura: A Case for the Modularity of Language*. Cambridge, MA: MIT Press.

Ye, Z. / Luo, Y. / Friederici, A. D. / Zhou, X. [2006], *Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials*. Brain Research 1071, 186–196.

Zatorre, R. J. / Evans, A. C. / Meyer, E. / Gjedde, A. [1992], *Lateralization of phonetic and pitch discremination in speech processing*. Science 256, 846–849.

Ziemke, T. [2011, Apr. 12], *Human embodied cognition. Scientific evidence and technological implications*. 4th EUCogII Members' Conference "Embodiment – Fad or Future?". Thessaloniki: Anatolia College.

Zwaan, R. A. / Stanfield, R. / Yaxley, R. [2002], *Language comprehenders mentally represent the shapes of objects*. Psychological Science 13, 168–171.

Zwaan, R. A. / Taylor, L. J. / de Boer, M. [2010], *Motor resonance as a function of narrative time: further tests of the linguistic-focus hypothesis*. Brain Lang 112, 143–149.

Michał B. Paradowski*

THE ROLE OF DATA-DRIVEN SOCIOLINGUISTIC RESEARCH IN DEVELOPING HUMAN-MACHINE AND MACHINE-MACHINE INTERACTION

Why is it that when robots are stored in an empty space,
they will group together, rather than stand alone?
— Dr. Alfred Lanning in *I, Robot* (2004)

1. The Original Turing Test

The year 2012 marked the centenary of the birth of Alan Mathison Turing. Yet it is an unjust legacy that to this day few, even in the AI field, seem aware of what Turing originally had in mind when, in his 1950 paper *Computing machinery and intelligence* [14], he introduced his concept of a test examining a machine's ability to exhibit intelligent (or, to be more precise, humanlike) behavior. For it was not merely a matter of whether a computer could interact with a human in such a way that the interrogator would be deceived into thinking s/he were 'conversing' with another human being. The design of the imitation game was much more subtle: to see whether a computer pretending to be a woman could be more convincing than a man also pretending to be a woman[1].

2. Language Simulations

Regrettably, this necessary proviso of 'other things being equal', so forcefully emphasized in Turing's original scenario, all too often seems to be overlooked in much of current research literature which either has the ambition to serve as input for developing AI, or which could potentially be applied by the field.

* Institute of Applied Linguistics, University of Warsaw.

[1] For an interesting distant but convincing analogy in popular culture, think for instance of the title protagonist of Mankiewicz's 1950 film *All About Eve*, whose faked femininity is more compelling than that of the remaining, genuinely heterosexual heroines.

For instance, over the past decade much space has been devoted to language simulations, from workshops devoted exclusively to that topic[2] to articles posted on arXiv and published across scientific journals. A major share of the papers, devoted to phenomena such as language evolution, language competition, language spread, and semiotic dynamics, have repeatedly been based on regular-lattice *in silico* experiments that are glaringly inadequate, especially to the scenery of the 21st century:

— the models take into account only Euclidean relationships (whereas the current telecommunication technology and the global accessibility of mass media mean that more and more of our linguistic input reaches us from afar, and – especially with services such as VoIP calls and social networking sites – spatial proximity can no longer be equated with social proximity);

— are 'static' (while mobility has been a distinctive feature of humankind – but also the animal world—as evidenced by warriors, refugees, missionaries, civil servants, and tradespeople long before the time of the Hanseatic League);

— assume a limited, identical number of 'neighbors' for every agent ($4 \leq 8$; first of all, an underestimate, secondly, again unrealistic given that persons vary in terms of the number of their close friends, acquaintances, or relatives – suffice it to think of the growing number of nuclear and patchwork families, the multi-generation families of the not-so-distant past, or the divide in China between urban couples who have had to abide by the one-child policy and the rural countryside where the restriction was not stringently enforced, but where in turn male offspring have often been valued more than female);

— presuppose identical perception of the prestige of a given individual by each of its neighbors (while, again, take a single person known to a group, be it a celebrity or an insider, and their perceived prestige and respect is again going to fluctuate from individual to individual), as well as

— invariant intensity of interaction between the different agents;

— absence of multilingual agents (with a few notable exceptions, e.g. the work of Castelló *et al.* [2]);

— and sometimes more technical issues such as lack of memory effect or zero noise (while noise may be the indicator and initiator of pattern change).

3. Alternatives

This is why there still remains much work in front of the AI circles to move from coarser-grained game-theoretic (e.g. [7]) and agent-based models (e.g. [6]) which not infrequently only manage to capture the initial and final states and the general trend of the phenomena they are purported to describe, towards increasingly accurate and sophisticated work based on the results of

[2] E.g. the GIACS Workshop on Language Simulations, which took place at the University of Warsaw in the year 2006.

rigorous data-driven research and empirical studies that recreate the necessary conditions and parameters as faithfully as possible. One solution is experimental designs involving actual cognitive agents. A new quandary that arises with many designs involving interactions between and learning by (embodied) artificial intelligent agents (for a good overview *cf.* e.g. [12], [13]) is the fact that they are often restricted to dyadic scenarios. This can naturally be justified when the process in which the robots engage is akin to the initial stages of language acquisition in humans, where a baby can conceivably find him-/herself in situations where s/he only interacts with a single caretaker. However, sooner or later the child's interaction becomes more social, with an increased number of input sources and persons against whom linguistic hypotheses can be tested. This calls for research paradigms involving more agents engaging in interactions with the subject under investigation, and luckily more and more robotics teams are moving in this direction.

Another, often more time-, cost- and resource-effective alternative is rigorous research fuelled by data from genuine human interactions. In the case of linguistic phenomena such as language learning and the uptake of new linguistic expressions, such data can be gleaned by either interviewing each member of a community and additionally verifying their responses against a more objective benchmark such as e.g. standardized test scores (in the case of foreign language acquisition; *cf.* e.g. [8] – admittedly still a time-consuming process, and one laden with the limitations posed by self-assessment – or, an easier way, utilize readily available repositories of user-generated content such as Web 2.0 sites.

The recent availability of user-generated text provides evident and unprecedented new research opportunities. The data stored on the Internet is virtually unregulated, essentially uncensored, spontaneous, immediately registered, interconnected, and amenable to relatively easy search and exploration with the use of statistical and concordancing tools. Web 2.0 services, with content (co)generated by the users, especially the ones which allow enriching their analyses with information concerning the structure of the connections and interactions between the participating users, are particularly useful for multi-angle explorations of language and social phenomena, such as humans' communicative behavior. By tapping into the repositories of language data nearly perfectly suited to fine-grained large-scale dynamic linguistic analyses and applying novel, transdisciplinary research methodology, most of the formerly-mentioned limitations can be addressed and bypassed.

4. A Case in Point

In a recent project [9], we investigated the creation and adoption of tags (metalabels) on a Polish microblogging site (roughly analogous to Twitter),

with special emphasis on neological expressions. Among other things, we introduced two measures to quantify how Internet users are influenced to adopt tags from one another. The data were analyzed in Python.

The preliminary analyses conducted involved calculating thresholds for all tag adoptions (i.e., the first usages of tags). We describe the user-tag network with a bipartite graph $G = G(U,T,E)$, where U is the set of users, T is the set of tags, and E represents the edges between users and tags. The user-user network we define using a directed graph $D = D(U,H)$, where H is the set of edges. To every $e_{u\to t} \in E$ edge connecting user u to tag t added in time $\tau_{u\to t}$ we assign a variable $a(e_{u\to t})$, such that

$$a(e_{u\to t}) \begin{cases} 1 \text{ if in time } \tau_{\mathrm{u\to t}} \text{ there is a neighbour of } u \text{ who is already connected to tag } t, \\ 0 \text{ else} \end{cases}$$

We capture the adaptive ('copycat') behavior of user u with the statistical variable $\alpha_u \in \langle 0,1 \rangle$

$$\alpha_u = \frac{\sum_{e_{u\to t} \in E(u)} a(e_{u\to t})}{|E(u)|}, \qquad (1)$$

where $E(u) \in E$ is the set of connections of user u. A high value of α_u means that the user tends to introduce new information into the system. The distribution of the alpha measures across the Blip population is represented in Fig. 1:

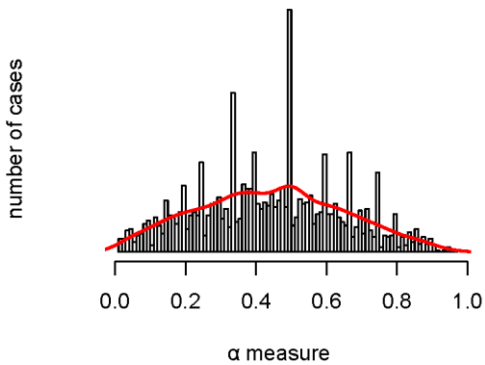

Fig. 1. Microblog users' copying behavior

We can see that the general innovativeness of the users scales not like a power law, but a hump-shaped unimodal [10].

Using the above notation, we calculate the (mean) measure of the number of alters (neighbours) who had adopted new linguistic content before user u (in line with Granovetter's [3] threshold model of collective behavior and Valente's [15] theory of network exposure). We only consider first usages:

$$\beta_u = \frac{\sum_{e_{u\to t}\in E(u)} \frac{A(e_{u\to t})}{H_{(t)}(u)}}{|E(u)|}, \qquad (2)$$

where:

- $A(e_{u\to t})$ is the number of neighbors of u who are already connected to t at time $\tau_{u\to t}$;
- $H_{(t)}(u)$ is the number of neighbors of u at time t;
- $E(u)$ is the total number of (unique) tags used by u.

The resultant smoothed distribution of adoption thresholds is considerably skewed, with a median of 0.11 and a long tail of higher values[3], suggesting more susceptibility to social influence – at least in low-stakes scenarios – than may erstwhile have been expected:

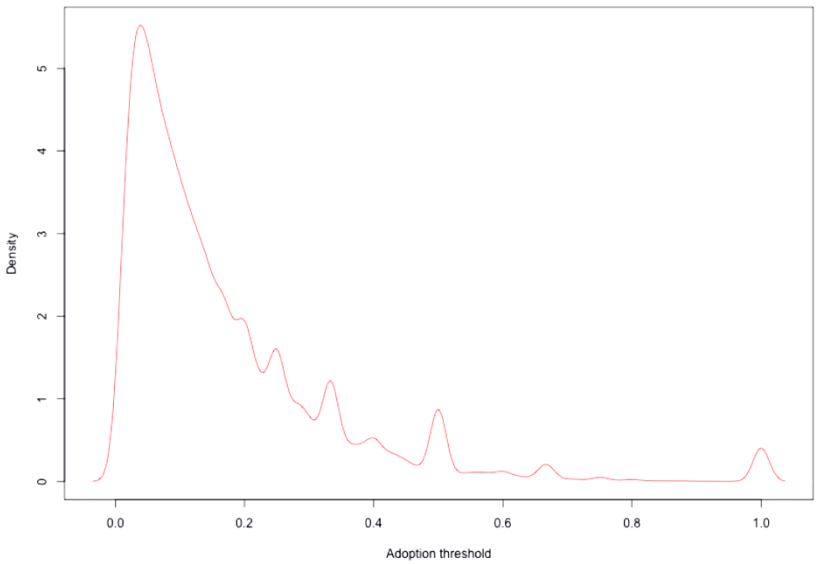

Fig. 2. Distribution of tag adoption thresholds in Blip [11]

5. Insight for AI?

Such rigorous data-driven research offers the chance of not only approximating to descriptive adequacy, but also moving beyond explanatory adequacy to approaching principled explanation. Results like the above are conceivably useful to answer questions how a) creative, and b) susceptible to the influence

[3] The "humped" feature of the distribution tail stems from the skewed distribution of the variables used to calculate the threshold values.

of alters an artificial cognitive agent should optimally be. The established parameters can also be helpful in the development of interactive dialogue systems, HCI, and intelligent machines that acquire knowledge via interaction with other (human and non-human) agents (rather than all their knowledge being put in by their creators). Naturally, the question needs to be posed to what degree observed online behavior – which may naturally be affected by the medium – can be treated as a realistic proxy for offline behavior. If we grant the assumption that any difference that may exist is insubstantial, such and similar data-driven research can have practical import for the discipline of artificial (social) intelligence, providing a reference point for at least three aspects of cognitive systems' behavior:

(i) interaction[4],
(ii) learning, and
(iii) collective intelligence.

Where the agents are expected to pass off as humans, exhibiting performance indistinguishable from that of mankind, e.g. in affective contexts (where, for instance, their task is that of companions), the established data could then be used to emulate human behavior as closely as possible (bearing in mind the desirability of optimal distinctiveness (*cf.* [1], [4]) and the uncanny valley problem; [5]). In other scenarios, it may be more desirable for the agents to outperform humans[5] (think, for instance, of Deep Blue defeating Kasparov in 1997, or IBM's another wunderkind, Watson, the computer capable of answering natural-language queries, which in February 2011 won the *Jeopardy* quiz show against two of its all-time human champions). In that case, it is still useful to have a reference point or benchmark. Only subsequently, given the growing sophistication of tools for ABMs, can fine-grained simulations be employed to try to emulate and explain the behavior observed. This is what Alan Turing would have appreciated[6].

Acknowledgments. The author's research is supported by the Polish Society of Social Psychology, grant-in-aid № 32-01-00-501/86-DSM-100809 from the Ministry of Science and Higher Education, and the 3rd European Network for the Advancement of Artificial Cognitive Systems, Interaction and Robotics. The computations in the project cited were performed by Łukasz Jonak. Further results and discussion can be found in [Paradowski & Jonak 2012a].

[4] And each utilizable cognitive system must be interactive.

[5] To use the words of Gigolo Joe from *A.I.* [2001], "Man made us better at what we do than was ever humanly possible."

[6] Even if computational cognitive systems may be *non*-Turing, with non-terminating computations, interactivity, and non-uniform evolution [16].

References

Brewer, M. B. [1991], *The social self: On being the same and different at the same time*. Pers. Soc. Psychol. Bull. 17(5), 475–482.

Castelló, X. / Loureiro-Porto, L. / Eguíluz, V. M. / San Miguel, M. [2007], *The fate of bilingualism in a model of language competition*. In: Takahashi, S. / Sallach, D. / Rouchier, J. (eds.), Advancing Social Simulation: The First World Congress. Berlin: Springer, 83–94.

Granovetter, M. [1978], *Threshold Models of Collective Behavior*. Am J Sociol 83(6), 1420–1443.

Leonardelli, G. J. / Pickett, C. L. / Brewer, M. B. [2010], *Optimal distinctiveness theory: A framework for social identity, social cognition and intergroup relations*. In: Zanna, M. / Olson, J. (eds.), Advances in experimental social psychology, vol. 43. New York: Elsevier, 65–115.

Mori, M. [1970] 不気味の谷 [*Bukimi NoTani*]. Energy 7(4), 33–35.

Nolfi, S. / Miroli, M. (eds.) [2010], *Evolution of communication and language in embodied agents*. Berlin: Springer.

Nowak, M. A. / Komarova, N. L. / Niyogi, P. [2001], *Evolution of universal grammar*. Science 291, 114–118.

Paradowski, M. B. / Jonak, Ł. [2012a], *Understanding the social cascading of geekspeak and the upshots for social cognitive systems*. In: Galton, A. / Wood Z. (eds.), Understanding and Modelling Collective Phenomena. Birmingham: University of Birmingham, 27–32.

Paradowski, M. B. / Jonak, Ł. [2012b], *Diffusion of linguistic innovation as social coordination*. Psychology of Language and Communication 16(2), 53–64.

Paradowski, M. B. / Jonak, Ł. / Kuscsik, Z. [2009, Nov 23], *The evolution of speech – towards network models of language development and spread*. International Workshop "150 Years after Darwin: From Molecular Evolution to Language". Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca.

Paradowski, M. B. / Jonak, Ł. / Kuscsik, Z. [2010, Jan. 27], *Follow the white rabbit, or how complexity science and linguistics can inform AI*. 4th International Conference on Cognitive Systems, Institute of Robotics and Intelligent Systems, ETH Zürich.

Paradowski, M. B. / Jonak, Ł. / Kuscsik, Z. [2011, Jun 10], *Milgram's legacy and lipstick on a pig – how to track linguistic epidemics in online social networks*. International School and Conference on Network Science, Central European University/Hungarian Academy of Sciences, Budapest.

Steels, L. [2011], *Modeling the cultural evolution of language*. Phys. Life Rev. 8(4), 339–356.

Steels, L. (ed.) [2012], *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Turing, A. M. [1950], *Computing machinery and intelligence*. Mind 59, 433–460.

Valente, T. W. [1995], *Network Models of the Diffusion of Innovations*. Hampton Press: Cresskill, NJ.

Wiedermann, J. [2011, Apr 7], *The singularity might indeed be near, but the next interesting level of intelligence is too far*. AISB'11 Convention, University of York.

Anna Obrębska*

THE AUTOMATIC PROCESSING OF THE NATURAL LANGUAGE VIA PRINCIPLES OF *COMMUNICATIVE GRAMMAR*. THEORY – PRACTICE – PROSPECTS

Introduction

As rightly pointed out by Zygmunt Vetulani [2004: 19], "the production of scientific theories (usually) serves as an explanation of reality. This in turn usually acts as a response to certain demand of an intellectual or practical nature". It seems that computer science contributes its dynamic development to its response to intellectual, as well as practical, demands.

Approximately 50 years ago, cybernetics turned its direction to linguistic theories due to unsatisfactory results in the field of programming machines for natural language processing. As noted by Vetulani, "practical usage of computational technologies for the automatic analysis and generation (more generally, 'processing') of text in a natural language depends on subjecting the grammar description of these languages to a certain regime, i.e. subjecting this description to formalization" [Vetulani 2004: 29]. An attempt to conduct such language formalization was made by Noam Chomsky, whose linguistic assumptions have been used in computer linguistics for 50 years. According to Chomsky's theory, a sentence has a tree form phrase structure, in which branches are described with the types of phrases and the leaves refer to the words. An algorithm which turns a sentence into a syntactic tree is referred to as a parsing algorithm, and the program which realizes this algorithm – a parser. An important role in Chomsky's concept is attributed to the so-called transcription principles, which in formal grammar is used to determine operations performed on certain alphabets which consist of a finite set of terminal symbols and an infinite set of non-terminal (auxiliary) symbols. Terminal symbols include specific words in the natural language, whereas non-terminal symbols

* Department of Journalism and Social Communication, University of Łódź.

represent formal categories, such as NP (nominal phrase). It appears that programs which use principles of transformational-generative grammar and concentrate on problems related to the grammar aspect of the language are frequently used to indicate the formal structure of utterance, or to control grammar correctness. Although later modifications of generativism, mainly due to Lakoff's and Fillmore's efforts, took the semantic component more into consideration than the early varieties of generativism, it seems that the problem of proper formalization of the language semantic layer was not adequately solved within the framework of generativism and thus it's not capable to complete the tasks that computational linguistics impose on modern linguistics. However, it needs to be mentioned that Chomsky's concept of language as a generative system was a turning point for linguistics and enabled the development of computational linguistics.

These days, the production of various linguistic computer programmes is based on generativism principles. Other attempts to formalize language in order to facilitate text processing are made as well. In Poland, problems related to natural language processing are being intensively investigated by Leonard Bolec's team from the Institute of Computer Science PAN, which is working on a formal description of the Polish language using HPSG (Head-driven Phrase Structure Grammar) [*cf.* Przepiórkowski i inni 2002, Przepiórkowski 2008][1].

One of the new linguistic concepts used for the automatic processing of the natural language is Aleksy Awdiejew and Grażyna Habrajska's *Communicative Grammar*, to which I have devoted the rest of the article.

Theory

The primary purpose of communicative grammar is "a description of the organization and generation (creation) of communicative utterance and the process of its receipt" [Awidiejew, Habrajska 2004: 14] and "this description refers not only to a determination of the grammatical correctness of utterances, which can be achieved via a formal analysis, but also to their acceptability from the perspective of communicative value" [Awdiejew, Habrajska 2004: 14].

For grammar to be useful in computational linguistics it has to be properly formalized, a particular problem is posed by the formalization of ideational content which is related to the representational level of the language. According to Aleksy Awdiejew:

[1] The current advances in computational linguistics are described in detail in Z. Vetulani's book, *Komunikacja człowieka z maszyną. Komputerowe modelowanie kompetencji językowej*, as well as in numerous articles, e.g. in M. Piasecki's publication, *Cele i zadania lingwistyki komputerowej* [Piasecki 2009].

> A problem arises when capturing and expressing a direct, non-observable semantic representation, conceptual structure or, using the increasingly popular term, cognitive structure, in a given transcript [Awdiejew 1992: 113].

This task is not easy, but the mentioned capture and expression of the cognitive structure is of crucial importance when it comes to constructing computer systems whose task is to generate (grammatically and semantically) acceptable utterances.

It seems that the primary difficulty in constructing the AI system is not the fact that the mental representations created in our minds have a distorted character, but rather the genesis of ideational images created in human minds. Sensory stimuli are the primary source of ideational images and to transfer the images, which arise naturally in a human's mind, to the computer it is necessary to formalize them properly. Aleksy Awdiejew's semantic standards provide invaluable practical advantages in terms of ideational image recording. In communicative grammar terms, semantic standards are generalized representations of sense with a predicate-argument structure, which input lexemes in contexts that are typical for them and bring information about their lexical and grammar connectedness [Awdiejew, Habrajska 2004]. As noted by Aleksy Awdiejew and Grażyna Habrajska:

> From an informative perspective, a text is organized in such a way that it falls into individual wholes of sense – messages, which do not have to correspond to textual units [Awdiejew, Habrajska 2004].

As noted previously, semantic standards form an organized transcription of fixed cognitive images, which are provided to our mind by sensory stimuli and are preserved by interaction with society and its culture. Unfortunately, these obvious, mind-rooted images are not provided to the computer programme in such a natural way. Therefore, placement of semantic standards in a computer programme is the basis for detrivialization processes that a computer program performs according to predefined principles in order to generate informative (non-trivial) messages. Similarly, the predicate-argument structure, in which semantic standards are captured, brings practical advantages when it comes to the creation of linguistic software. According to Vetulani:

> Compatibility between the notional apparatus of language description and the notional apparatus of extralinguistic reality representation provides favourable circumstances for the automatic analysis of text [Vetulani 2004: 54].

The predicate-argument approach is also known thanks to works by such individuals as Fillmore, Karolak, Polański, Topolińska. As rightly noted by Karolak, "notions perform defined functions in systems, i.e. complex notional

structures" [Karolak 2002: 26]. The predicate's role is to constitute a complex structure where argument fulfils the predicate's implications and thus becomes an internal element of a complex structure [por. Karolak 2002: 27]. Karolak's definition assumes that the predicate-argument structure is a system consisting of a predicate (constitutive notion) and one or more arguments.

A strong correspondence to the predicate-argument approach is displayed by *Słownik syntaktyczno-generatywny* edited by Kazimierz Polański. In its introduction, Polański points out that a "language grammar description must correspond to its lexical description, in other words, grammar and dictionaries should complement each other" [Polański 1980: 6]. According to Polański, when we construct an utterance we are restricted by strictly semantic factors, like rationality of lexical connections and compliance with a pragmatic intention, as well as connectedness principles, which often have a strictly conventionalized character. The mentioned *Słownik syntaktyczno-generatywny* takes account of conventional connectedness principles; however, as pointed out by its editor, the proposed set of semantic features is neither exhausting nor final. The development in research to which Polański contributed is continued by Aleksy Awdiejew and Grażyna Habrajska. *Communicative Grammar* also corresponds to the thematic relations theory, which refers to "both dictionary (lexicon) and syntactical levels; its characteristic feature is that it deals with elements of lexical semantics in a syntactic structures context" [Stalmaszczyk 2001: 79]. The theory of thematic relations, predicate-argument structure, and the formalization of representational content are further developed by arguments; performed within the framework of communicativeness, separation of primary and secondary predicates and primary and secondary meta-predicates as well as the division of language units into therms and auxiliaries [zob. Habrajska 2004: 44–58].

Practice

Using linguistic tools developed on the grounds of *Communicative Grammar*, we have created a basic computer programme which generates sentences and trivial, acceptable and unacceptable communicative utterances, as well as a computer programme which automatically produces word-formation paraphrases. I provide a brief description of the selected computer programmes which were created using the principles of *Communicative Grammar*[2].

[2] A more detailed description of these programmes can be found in Anna Obrębska's *Algorytmy słowotwórczego parafrazowania*, Łódź 2010, as well as in the article *Reprezentacja kognitywna a sztuczna inteligencja*, [w:] *Język trzeciego tysiąclecia III. Tendencje rozwojowe współczesnej polszczyzny*, t. 1, red. Grzegorz Szpila, Kraków 2005. Some of the problems considered in the article were discussed in detail in previous works by the author of this article.

1. A computer programme for generating messages and recognizing their acceptability

The lexical basis of the programme consists of 5 verbs (GRAĆ, CAŁOWAĆ, ROBIĆ, BRAĆ, KUPOWAĆ) and 103 nouns. Verbs occupy the predicate's position whereas nouns are variables like a1 and a2 which stand for arguments.

For each of the five verbs, we have established a rule that the first argument takes a value from 1 to 103 for each second argument. This assumption means that the programme is capable of generating 53 045 different sentences, but not all of them are communicatively acceptable. In order for the programme to divide the generated utterances into trivial, acceptable and unacceptable[3], we had to equip it with principles of arguments for standard selection.

The computer programme works thanks to a set of formalizing rules, like in the case of the predicate GRAĆ; each lexeme which fulfils the condition from the P1[4] level, i.e. the one which belongs to the N_{HUM} (name of a human being) substitutive class and occupies the position of the first argument, will create a configuration which is communicatively acceptable. The N_{HUM} is a symbol of a name which can be adopted by a human being; this group also includes all lexemes which come from subclasses forming the main category, i.e. job titles, etc.

In the case of the predicate GRAĆ, for each first argument that is, e.g., a lexeme from the $N_{HUM\text{-}ZAW\text{-}ART\text{-}OBR}$ substitutive class and whose position in the second argument is occupied by a term from the $N_{ART\text{-}OBR}$ class, a sentence with the following structure: a1 [$N_{HUM\text{-}ZAW\text{-}ART\text{-}OBR}$] + PRED + a2 [$N_{ART\text{-}OBR}$] will be a trivial sentence (*Aktor gra w filmie*).

However, trivial utterances are usually communicatively useless so acceptable message schemes and means of detrivializing semantic standards appear to be more important. Through the interruption of trivial cognitive images, the speaker expresses informative content and compels the recipient to recognize a communicative purpose and sense of crossing the boundaries established by the semantic standard.

Methods of derivation can be varied. However, if we assume that the sentences generated by the program have the following structure: a1 + PRED + a2 (argument + predicate + argument), we have to reject the possibility of

[3] According to the assumptions accepted during the software design process, trivial sentences are those which express ideas that are obvious for an average language user, constitute an element of language system and generally do not appear in natural communication. (For example, *Trawa jest zielona*, *Ryby żyją w wodzie*). Acceptable sentences are utterances which have an informative power, are not meaningless, and are grammatically correct (e.g., *Ksiądz gra w szachy*). Unacceptable sentences are utterances which are ontologically contradictory (e.g., *Szafa całuje piekarza*), they are logically unacceptable (e.g., *Piekarz całuje szafę*) or are grammatically incorrect.

[4] The division into levels is well-illustrated with tables attached to *Reprezentacja kognitywna a sztuczna inteligencja...* article.

detrivialization by introducing update operators, which even in the case of trivial argument correlation can create an informative utterance. For example, if the trivial and meaningless utterance, *On ma nogi,* is completed by an operator which updates the position, *w rzece*, then the sentence will become detrivialized and will have the following form: *On ma nogi w rzece*. Hence, even at the level of arguments there are possibilities of detrivialization. If in a given class, a term which makes the utterance trivial is replaced with a class of lexemes at the same level of detail, e.g., P3, and, at the same time the two classes have a common category which links them at the P2 level, the replacement will result in the creation of acceptable sentences. For example, if in the following sentence structure, a1 [$N_{HUM\text{-}ZAW\text{-}ART\text{-}MUZ}$] + PRED + (na) a2 [$N_{RES\text{-}INSTR\text{-}MUZ}$], which can be realized in the sentence *Muzyk gra na skrzypcach*, the a1 class argument, i.e. $N_{HUM\text{-}ZAW\text{-}ART\text{-}MUZ}$, is replaced with a related category $N_{HUM\text{-}ZAW\text{-}ART\text{-}OBR}$ and the sentence adopts the following structure, a1 [$N_{HUM\text{-}ZAW\text{-}ART\text{-}OBR}$] + PRED + (na) a2 [$N_{RES\text{-}INSTR\text{-}MUZ}$], its triviality is broken and the sentence adopts a non-trivial form: *Aktor gra na skrzypcach*.

If a given category of lexemes is replaced with words from classes at the P3 or P2 levels, this will lead to the creation of structures of sentences that are communicatively acceptable, but if the replacement takes place at the highest level of categorization, i.e. P1, then the substitution boundaries are crossed and we obtain an unacceptable sentence scheme. For example, if in a general scheme like a1 [N_{HUM}] + PRED + (na) a2 [$N_{RES\text{-}INSTR\text{-}MUZ}$] the N_{HUM} category, which occupies the first argument position, is replaced with another class of lexemes from the same P1 level, N_{RES}, then the structure a1 [N_{RES}] + PRED + (na) a2 [$N_{RES\text{-}INSTR\text{-}MUZ}$] will become a source of unacceptable sentences like *Lodówka gra na skrzypcach*.

2. Programme for automatic utterance paraphrasing with word-formation morphemes

Based on the methodology of *Communicative Grammar* and my separation of mechanisms for word-formation paraphrasing [*cf*. Obrębska 2010], a computer programme for automatic paraphrasing was created. The programme was designed using the Turbo Delphi environment. It has four bookmarks forming the programme panel. The first panel includes a view of the whole database and allows users to complete it. The first panel's column provides the type of paraphrasing, which is related to a given derivative and a derivative's number in a given category. This information is also displayed at the top of the panel. The next column, *wersja A*, includes a word form on which the program should react when it encounters it in the text. The *wersja B* column displays information on the type of reaction an element encountered in the text and included in the column. The reaction involves adding information about the category

and grammar case to the word. For example, if the programme's search engine encounters a *skrypt* form, which is entered in its database's *kolumna A*, it will replace it with a form from *kolumna B*, which is "NRESB skrypt", where "N" stands for noun, "RES" refers to the object, and "B" states that a given form is in accusative. In some cases, the *wersja A* and *wersja B* columns remain empty and it occurs when the next column includes a language form whose word-formation paraphrasing doesn't influence the grammar form of words with which it has syntactic relations. Behind the *wersja A* and *wersja B* columns there is the *konstrukcja A* column which includes a form which may undergo word-formation paraphrasing along with information about grammar form that the paraphrased expression imposes on other words. For example, in the mentioned column the following structure can be seen: "kartkował NRESB", which means that the predicate *kartkował* attaches an argument which stands for an object and imposes an accusative form on it. In this case, the NRESB symbol may stand for *skrypt*, *zeszyt*, *podręcznik*, *książka*, etc. It would be difficult to enter each possible connection of predicate *kartkować* with each possible argument; a much better solution involves transforming the argument part of the predicate-argument structure into a symbolic form, which can represent a whole class of arguments. A more detailed symbolic marking of arguments is not required as the predicate *kartkować* will be followed by an argument denoting an object which can be browsed. The *konstrukcja B* column includes a form in which a word from the *konstrukcja A* column can be changed, and thus in the discussed case "kartkował NRESB" can be replaced with "przekładał kartki NRESD" where NRESD means that the argument which is attached by the word is in the genitive case. After this operation, the paraphrased form adopts the "przekładał kartki NRESD skrypt" form. Therefore, we need to match the case with the argument *skrypt* and remove the markings of the auxiliary symbol. At this stage, the computer programme scans the text for the argument's compliance with its preceding marking and when it finds the "NRESD skrypt" form, it adjusts the argument's case to its symbolic marking and removes the marking which is no longer needed. Eventually, the initial structure, "kartkował skrypt", is replaced with "przekładał kartki skryptu". Here the paraphrasing mechanism consists of three independent processes of text scanning and three independent form replacements. The first scanning included the *kolumna A* column and here the program found the *skrypt* form, on whose presence in the text it reacted and replaced with a related form from the *kolumna B* column, which attaches a symbolic marking to the argument found in the *kolumna A* column.

The algorithm almost works according to the principle of dominos, which can match in various ways. This solution limits the amount of data to be entered in the database. In individual phrases the programme only considers the

predicate and symbolic marking of argument which it matches, or argument's name with its symbolic marking. In situations when paraphrasing influences grammar form of the matched arguments, the program doesn't consider the whole phrase as this would involve entering very large numbers of possible matches in predicate-argument structures. Of course, when paraphrasing doesn't involve substantial changes in predicate-argument structures, automatic paraphrasing is not restricted to the automatic replacement of one form with another. For example, the predicate *owocuje* is paraphrased via a simple replacement of this form with the expression *wydają owoce.*

Prospects

The undertaken attempts to implement *Communicative Grammar* produced results which allow us to say that Aleksy Awdiejew and Grażyna Habrajska's *Communicative Grammar* is a useful methodology when it comes to the computational processing of text. The computer programmes discussed in this article are being constantly developed. In addition, completely new projects, which are supposed to contribute to more advanced language processing, have been launched. The designed (and described) computer programme for the generation of messages and determination of their acceptability created only simple sentences which consisted of a predicate and two arguments and its lexical basis was fairly restricted. Current works on the programme development aim to determine algorithms, which will generate and categorize simple sentences consisting of ideational, interactional and organizational auxiliaries as well as complex sentences according to their communicative acceptability. In the more distant future, the program is supposed to create all kinds of text combinations from its lexical basis and assess their acceptability. The second programme for automatic paraphrasing has also been developed, however, the very process of establishing linguistic rules for paraphrasing is very arduous and time-consuming. Particular difficulties are experienced with automatic syntactic paraphrases, where syntactic changes – which usually profoundly modify the utterance's structure – also provoke changes on other language levels (especially on the word-formation level). Another program which has been created from the ground up is a system which uses the PHP script programming language and MySQL technology. Such a solution enables various people engaged in the system development to complete the database via the Internet. The project is in its initial stage. However, basic functions like adding, editing and removing records in two tables are available on its website. The first table includes arguments and predicates as well as their symbols, which are useful during the process of automatic language processing, e.g., matka – NHUM-N. Such a general level of categorization will not be sufficient in most

applications so several levels of categorization will be introduced in the future. The second table includes generalized schemes of predicate-argument structures and consists of the following columns: *predicate*, *argument 1*, *argument 2*, *argument 3*, *argument 4*, *argument 5*, *scheme*, *edit/remove*. In the *predicate* column, concrete predicates are placed as they are supposed to create gaps in predicate-argument structures for arguments. In *argument 1, argument 2, argument 3…* columns, there are generalized symbols of arguments which occupy positions in the syntactic structure constituted by a defined predicate.

For example, in the case of the predicate DAWAĆ in the *argument 1* column, there is a NHUM-N symbol, which means that this position in the structure is reserved for a man's name in a nominative form; the verse in the *argument 2* column is occupied by the NRES-ACC symbol, which stands for any object of activity in the accusative form, whereas in the *argument 3* column there is a symbol of the activity's beneficiary, NHUM-D. The *schemat* column includes a generalized transcription of the whole predicate-argument scheme: DAWAĆ [NHUM-N, NRES-ACC, NHUM-D].

We hope that the further development of this system, as well as other previously discussed programmes, will allow us to produce solutions which are useful in the processes of automatic processing of natural language.

References

Awdiejew, Aleksy [1992], *Składnik wyjściowy w gramatyce komunikacyjnej*. In: I. Nowakowska-Kempna (ed.), *Język a kultura*, t. 8, Wrocław, 113–124.

Awdijew, Aleksy [1998], *Standardy semantyczne a znaczenie leksykalne.* In: J. Anusiewicz, J. Bartmiński (ed.), *Język a kultura*, t. 12, Wrocław, 53–62.

Awdiejew, Aleksy [1999], *Standardy semantyczne w gramatyce komunikacyjnej (teoria i zastosowanie)*. In: A. Awdiejew (ed.), *Gramatyka komunikacyjna*, Kraków, 33–68.

Awdiejew, Awdiejew [1999], *Leksykon w gramatyce komunikacyjnej*. In: A. Awdiejew (ed.), *Gramatyka komunikacyjna*, Kraków, 133–160.

Awdiejew, Awdiejew [2004], *Gramatyka interakcji werbalnej*, Kraków.

Awdiejew, Awdiejew; Habrajska, Grażyna [2004], *Wprowadzenie do gramatyki komunikacyjnej*, t. 1, Łask.

Awdiejew, Awdiejew; Habrajska, Grażyna [2006], *Wprowadzenie do gramatyki komunikacyjnej*, t. 2, Łask.

Chomsky, Noam [1982], *Zagadnienia teorii składni,* Wrocław.

Fillmore, Charles [1968], *The Case for Case*, New York.

Habrajska, Grażyna [2004], *Komunikacyjna analiza i interpretacja tekstu*, Łódź.

Karolak, Stanisław [2002], *Podstawowe struktury składniowe języka polskiego*, Warszawa.

Obrębska, Anna [2005], *Reprezentacja kognitywna a sztuczna inteligencja*. In: G. Szpila, *Język trzeciego tysiąclecia III. Tendencje rozwojowe współczesnej polszczyzny*, t. 1, Kraków.

Obrębska, Anna [2010], *Algorytmy słowotwórczego parafrazowania*, Łódź.

Piasecki, Maciej, *Cele i zadania lingwistyki informatycznej*. In: P. Stalmaszczyk (ed.), *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje*, Kraków.

Przepiórkowski, Adam / Kupść, Anna / Marciniak, Małgorzata / Mykowiecka, Agnieszka [2002], *Formalny opis języka polskiego, teoria i implementacja*, Warszawa.

Przepiórkowski, Adam [2008], *Powierzchniowe przetwarzanie języka polskiego*, Warszawa.

Słownik syntaktyczno-generatywny czasowników polskich [1980], (ed.) K. Polański, Kraków.

Stalmaszczyk, Piotr [2001], *Język i gramatyka w programie minimalistycznym i semantyce pojęciowej.* In: G. Habrajska, *Język w komunikacji*, t. 1, Łódź.

Vetulani, Zygmunt [2004], *Komunikacja człowieka z maszyną. Komputerowe modelowanie kompetencji językowej*, Warszawa.

Mirosław Bańko, Ignacy Doliński,
Jacek Duda, Milena Hebal-Jezierska*

COLLOCATION IMAGES OF HUNGARIANS IN SLAVONIC LANGUAGES – RESEARCH REPORT

1. Introductory remarks

The paper presents a research report of the collocation image of Hungarians. This topic was taken up in the framework of the work in the newly established Team of Corpus Linguistics of Slavonic Languages at the Institute of Western and Southern Slavonic Studies at University of Warsaw. The idea came from common interest in the problems of stereotypes and the past attempts to use corpora for this type of research. The task was taken on by the members of the Slavists' team: Ignacy Doliński (Bulgarist), Jacek Duda (Serbist) and Milena Hebal-Jezierska (Bohemist). They were joined by Mirosław Bańko (Polonist), which extended the scope of research.

At the outset, it should be noted that the image of Hungarians presented in the article is not a complete image, but a partial one. It was created with only one of the methods belonging to the corpus-driven methodology, having its origins in the English method of collocation profiles. The collocates are analyzed and divided into semantic groups. Collocations are generated from the software available in the corpus (including the use of various statistical tests, most MI-score or Chi2). This entails a broad term of the collocate. We then analyze not only the word which is connected (and makes sense) with the main lexeme (e.g. *Hungarian goulash*), but also the words which create a part of meaningful connection that contains the main lexeme (e.g. the word *conservative* in the following construction: *Hungarian Conservative Party*).

Research on Hungarians' stereotype was carried out in five Slavonic languages: in Bulgarian (Ignacy Doliński), Czech and Slovak (Milena Hebal-Jezierska),

* Mirosław Bańko – Instytute of Polish Language, University of Warsaw; Ignacy Doliński – Institute of Western and Southern Slavic Studies, University of Warsaw; Jacek Duda – Institute of Western and Southern Slavic Studies, University of Warsaw; Milena Hebal-Jezierska – Institute of Western and Southern Slavic Studies, University of Warsaw.

Polish (Mirosław Bańko), Slovene (Jacek Duda). This choice was determined not only by knowledge of these languages, but most of all the possibilities/functionality of the national corpora of the various languages. Specific images will be presented separately. The exception are the Czech and Slovak languages which are presented together because their collocations are arranged in similar categories, but in dissimilar contexts. The order of presentation is dictated by the clarity of the images.

2. The collocation profile of Hungarians in the Czech and Slovak languages

Images of Hungarian were reconstructed on the basis of collocations retrieved from the electronic synchronic and balanced (one-typed, written) journalistic texts corpora of the Czech (SYN2009PUB) and Slovak languages (prim-5.0.-public-inf), comprising respectively 800 million and 514 million tokens. The research was carried out on Czech and Slovak equivalents of the noun and adjective *Hungarian*. These words were searched using the queries which let us find all the inflectional forms of a lexeme, for Czech language [lemma = "Maďar"], [lemma = "maďarský"]; for the Slovak language, in the attribute "lemma", look for words *maďar* and *maďarský*. The first conclusion reached after receiving the number of occurrences is the difference in frequency of the lexemes studied. In the Slovak language corpus there are four more occurrences of the noun *Hungarian* (22 450 occurrences = 0.4%) than in the Czech language corpus (8481 = 0.1%). A similar situation concerns the adjective: in the journalistic corpus prim-5.0.-public.-inf, although the Slovak corpus is smaller than the Czech one, there are 96 067 occurrences (1.8%), while in the Czech counterpart there are – 24 108 (0.3%). Difference is evident also in frequency of the collocations (obtained from the MI-test score). In the Slovak corpus, co-occurring auto-semantic words (with the proper value of mutual information) are twice more frequent than in Czech. Semantic groups of collocation are similar, but they differ in the weight of particular categories.

In both corpora, a significant part of collocations are adjectives denoting geographical area, country or city inhabited by the Hungarian minority. In the Czech language corpus there are statistically representative words referring to Serbia, Romania and Slovakia (*vojvodinský* 'Vojvodinian', *sedmihradský* 'Transylvanian', *slovenský* 'Slovak', *rumunský* 'Romanian'). Similar lexemes were found in the Slovak language corpus: *vojvodinský* 'Vojvodinian', *hornozemský* 'highlander', *sedmohradský* 'Transylvanian', *neslovanský* 'non-Slavic', *podkarpatský* 'Carpathian', *rumunský* 'Romanian'.

These lexemes are embedded in the context of national minority, often as an organization unit or political party which affects (or undertakes such effort) the politics of the country inhabited by them, and the politics of the homeland,

such as *Demokratický svaz vojvodinských Maďarů* 'Democratic Union of Vojvodinian Hungarians', *Demokratický svaz rumunských Maďarů* 'Democratic Union of Romanian Hungarians', *Zväz vojvodinských Maďarov* 'Union of Vojvodinian Hungarians', *Demokratické společenstvo vojvodinských Maďarov* 'Democratic Community of Vojvodinian Hungarians'.

A detailed analysis of the occurrences highlights the diversity of topics taken up in the Czech and Slovak journalism, depending on the situation of a national minority. The most common theme is connected with autonomy, double citizenship, current problems and the issue of displacement and compensation. Thus, the Hungarian minority appears in the press as a powerful organization, fighting for their rights (often for autonomy) and as an object of concern for their homeland. This is to a large extent perceived as a threat, especially in the Slovak language corpus. Some contexts point to the Hungarian minority as a victim of persecution, most point to displacement.

> SME 2010/06 2010-06-10 Stalo sa tak z iniciatívy europoslanca strany Csanáda Szegediho, ktorý sa podľa vlastných slov mieni angažovať za realizáciu územnej autonómie <sedmohradských Maďarov> a rozšírenie používania maďarského jazyka na celom území Sedmohradska.[1]

> Mladá fronta DNES, 19. 6. 1999 PUB Orbán po setkání s předsedou Demokratického svazu <vojvodinských Maďarů> Andrásem Ágostonem prohlásil, že autonomie je nejlepší způsob, jak zabezpečit budoucí vývoj této oblasti.[2]

> Lidové noviny, č. 109/1998, PUB Osud české menšiny žijící na Slovensku je v mnohém podobný osudu <slovenských Maďarů>. Snad jediný rozdíl spočívá v tom, že české spoluobčany žádná z vládních koaličních stran nepodezírá z úmyslu odtrhnout některá území od Slovenska a připojit je ke své někdejší domovině.[3]

In both corpora, one significant collocation that appears in the contexts of the Hungarian minority is *zahraniční/zahraničný* Maďar 'foreign Hungarian', denoting a Hungarian living outside Hungary. This connection is related to the so-called Hungary Card helping preserve the ethnic group living in exile. In the Czech journalistic corpus the mentioned collocation occurs primarily in the context of the situation in Slovakia. Co-occurring words are: *zákon* 'act', *průkaz* 'identity' and *status* 'status'. Occurrences include not only the description of privileges for a person of the described status, but also the negative reactions of representatives of the countries hosting the Hungarian minority. Therefore,

[1] This was done by initiative of MEP Csanád Szegedi who claims to intend to contribute to the implementation of Transylvanian Hungarians' territorial autonomy and to extend the use of Hungarian language to the entire territory of Transylvania.

[2] Orban, having met with the chairman of the Democratic Union of Vojvodinian Hungarians András Agoston, said that autonomy is the best way to secure the future development of this area.

[3] The fate of the Czech minority living in the Slovak Republic is in many ways similar to Slovak Hungarians' fate. Perhaps the only difference lies in the fact that Czech citizens are not suspected by any government coalition parties to intend to break away from some of the territory of Slovakia and connect to their former homeland.

the card is presented here as an outbreak of a dispute between Hungary and particular countries.

> Respekt, č. 17/2002PUB Výsledek se hodnotí výlučně optikou Benešových dekretů a tzv. krajanského zákona o <zahraničních Maďarech>, což jsou horká témata, v nichž je Bratislava s Budapeští ve sporu.[4]

The Slovak press devotes a lot of space to these issues. Concordances, including the collocation *zahraničný Maďar*, are eight times more frequent than in the Czech language corpus, due to the fact that the problem of the Hungarian minority in Slovakia is growing. Occurrences often point to dispute and the threatening nature of the Hungarian Card. Czech occurrences often refer to the Slovak situation or represent the Slovak point of view.

> Dvojaké občianstvo je veľká hrozba pre Európu Plán maďarskej vlády udeliť <zahraničným Maďarom> maďarské občianstvo je nebezpečný pre Európu, uviedol predseda vlády SR Robert Fico v rozhovore pre dnešné vydanie rakúskeho denníka Kurier.[5]

A synonym, although much less frequently used, is the Slovak collocation *štatútový Maďar* 'status Hungarian'. The Czech equivalent has not been found in Czech journalism.

The next collocation denoting the Hungarians as Hungarian minority representatives is the Czech and Slovak combination *menšinový Maďar* 'minority Hungarian'.

> Hospodářské noviny, 25.2.2002 Premiér Orbán varoval nepřímo Bratislavu, že pokud nebude zákon o <menšinových Maďarech> respektovat, nemusela by získat souhlas Budapešti se vstupem do NATO.[6]

In a similar sense, the collocation *etnický Maďar* 'ethnic Hungarian' occurs, used in journalism to identify Hungarians living outside of Hungary, although its dictionary meaning has a broader semantic range.

> Texty z www.birdz.sk, 2004 Maďarský parlament schválil zákon, ktorý zaručuje <etnickým Maďarom> v susedných krajinách osobitné možnosti zo zdravotného i vzdelávacieho hľadiska, ako aj z pohľadu pracovných príležitostí. Po novom roku vstúpil zákon do platnosti.[7]

---

[4] The result is assessed solely through the lens of the so-called Beneš' Decrees and a compatriot act on foreign Hungarians which are hot topics with both Bratislava and Budapest in the conflict.

[5] Dual citizenship is a big threat to Europe. The plan of the Hungarian government to grant Hungarian citizenship for foreign Hungarians is dangerous for Europe, said Prime Minister Robert Fico in an interview in today's edition of the Austrian *Kurier* newspaper.

[6] Prime Minister Orbán indirectly warned Bratislava that unless the act treats minority Hungarians with respect, Bratislava will not obtain the consent of Budapest to its accession in the NATO.

[7] The Hungarian Parliament passed a law which guarantees ethnic Hungarians living in neighboring countries the possibility of special health and education benefits, as well as jobs. After the new year the law came into force.

Among the significant collocations that differ in Czech and Slovak journalism, one should mention the word combination *neslovanský Maďar* 'non-Slavic Hungarian' which emphasizes the non-Slavic roots of Hungarians.

> SME 2003/04, 2003-04-25 O Slovensku a Slovákoch môžeme teda hovoriť odvtedy, ako sa slovanskí Nitrania ocitli v kontraste s <neslovanskými Maďarmi>, čiže odvtedy, ako sa ocitli v Uhorsku.[8]

To the other non-analysed collocations belong *rodilý/rodený Maďar* 'a native Hungarian'.

The next group are words connected with the existing Hungarian minority: *asimilace/asimilácia* 'assimilation', *autonomie/autonómia* 'autonomy'.

These assimilation issues in Czech journalism are associated with the Slovak situation.

> Respekt, č. 27/2003 PUB Přitom se dá říct, že zákon je nyní neškodný a navíc stejně zbytečný jako už jeho první verze, protože <asimilaci Maďarů> stěží zabrání, což ví jak Medgyessy, tak Dzurinda.[9]

> Lidové noviny, 3. 12. 2004, PUB "Součástí slovenské státní strategie je latentní <asimilace Maďarů>", řekl na jaře Béla Bugár.[10]

In the Slovak language corpus this collocation is also not very frequent. There are two types of the occurrences, written from the Hungarian and Slovak perspective. The first one represent negative attitude towards this issue, because the occurrences contain lexemes referring to the fears – fear of assimilation, for example; the assimilation is compared to masked tools.

> Hospodárske noviny, 2008-12-09 Obavy z <asimilácie Maďarov> na Slovensku majú reálny základ, toto nemôžeme poprieť.[11]

> SME 1994-08-05 Známe je tiež napr. aj to, že silný odpor k alternatívnym školám je motivovaný podozrením, že sú iba dobre maskovaným nástrojom <asimilácie Maďarov>.[12]

The occurrences, seen from the Slovak point of view, contain an analysis of factors which hold back Hungarians' assimilation.

---

[8] We can therefore speak about Slovakia and Slovaks from the times the Slavic Nitrania were in contrast with non-Slavic Hungarians [...].

[9] It can be said that the law is now harmless, and in addition, redundant in its first version, because assimilation of Hungarians is hardly evitable, as Medgyessy and Dzurinda know.

[10] This aspect of the Slovak government strategy is a latent assimilation of Hungarians, said Béla Bugár.

[11] Fears of Hungarians' assimilation in Slovakia have real basis, a fact we cannot deny.

[12] It is also known, for example, that strong resistance to alternative schools is motivated by suspicion that they are only well-masked tools of Hungarians' assimilation.

SME 2010-08-20 Zákon o dvojakom občianstve vraj zabráni <asimilácii Maďarov> Nový maďarský zákon o dvojakom občianstve, ktorý uľahčuje Maďarom žijúcim za hranicami získanie maďarského občianstva, môže zvrátiť asimilačné trendy v susedných krajinách.[13]

The issue of autonomy in the Slovak language corpus is much more represented than in the Czech language corpus and refers to the Slovak and Romanian situation.

Nové slovo Jeho klamstvo nechtiac odhalil sám Tökes, keď na pôde Európskeho parlamentu vyhlásil, že s predsedom SMK skutočne hovorili o <autonómii Maďarov> na juhu Slovenska.[14]

The next group of words are lexemes denoting a displacement of the Hungarians. This type of connection is highly represented in the Slovak language corpus (*vysídlenie* 'displacement', *vysídľovanie* 'displacement', *vysídlený* 'displaced', *odsunutý* 'moved out', *vysťahovanie* 'displacement', *odsun* 'move out', *vysťahovaný* 'displaced'). It is connected with the issue of compensation for this ethnic group.

Hospodárske noviny 07/04,doc.date=2007-04-11 1948 Časť <vysídlených Maďarov> bola istým spôsobom odškodnená už vtedy.[15]

Hospodárske noviny 07/10,doc.date=2007-10-17 Mečiar je proti odškodňovaniu <Vysídlenie Maďarov> zo Slovenska po druhej svetovej vojne je podľa lídra ľudovcov Vladimíra Mečiara vnútroštátny problém.[16]

This group of collocations is actually limited only to a single one (*odsun Maďarů* 'Hungarians' moving out') in Czech journalism. The collocation is restricted to historical contexts.

The next group consists of the following Slovak lexemes: *urážanie* 'insult', *hanobiť* 'defame', *neubližovať* 'hurt', *strašenie* 'to scare sb', *nenávidět* 'hate'. In the Czech corpus we find only one their equivalent *nenávidět* 'hate'. In Slovak journalism, as opposed to the Czech, these lexemes concern most often Slovak context.

[13] The act regarding dual citizenship will supposedly hold back Hungarians' assimilation. A new Hungarian law on dual citizenship which enables Hungarians living beyond the borders of Hungary to have citizenship can reverse the assimilation trends in neighboring countries.

[14] Tokes unwillingly revealed his lie when, at the European Parliament, he said that the President of SMK really talked about Hungarians' autonomy in southern Slovakia.

[15] Part of displaced Hungarians was compensated in some way.

[16] Mečiar is against the compensation. According to People's Party leader Vladimir Mečiar, Hungarians' displacement from Slovakia after World War II is an intrastate problem.

Hospodárske noviny 2006-08-15 Ján Slota údajne <hanobí Maďarov> a podnecuje... Slotovi hrozí vyšetrovanie Predseda SMK Béla Bugár podal trestné oznámenie na lídra národniarov.[17]

SME 2009/04 2009-04-05 <Strašenie Maďarmi> a autonómiou rezonuje najmä medzi voličmi zo severu krajiny, ktorí s občanmi maďarskej národnosti neprichádzajú do kontaktu.[18]

Hospodárske noviny 2007-10-23 Boli tu rôzne nedorozumenia a nezhody, lebo Slováci <nenávidia Maďarov>, a naopak, a hašteria sa medzi sebou.[19]

The modification of the Hungarian government passwords *Neubližujte Maďarom* 'Do not hurt the Hungarians' is particularly noteworthy. In Slovak journalism we found the following occurrence:

SME 2010/05 2010-05-28 <Neubližujte Maďarom>, ublížia si dosť aj sami...[20]

In Czech journalism, the collocation discussed (*nenávidět Maďary* 'to hate the Hungarians') does not apply to the present situation of Czechs but to a Slovak, Hungarian, or historical situation.

Another interesting collocation we must mention is the connection describing the Hungarian nature, *temperamentní Maďar* 'a temperamental Hungarian'. It is quite highly located in the Czech language corpus.

Mladá fronta DNES, 11.10.2002 PUB Někoho zaujme nový pohled na písňové aranžmá pražského souboru Karmína, někoho rustikální Poláci s muzikou z Horní Oravy, jiného <temperamentní Maďaři>.[21]

To complete this analysis, we would like to focus on the collocations containing the adjective *Hungarian*. In Czech journalism, many collocates (excluding proper names) denote Hungarian food (*tokajské* 'Tokay', *uherák* 'salami', *čabajka* 'salami'), other characteristics of the Hungarian national culture and phenomena (*čardáš* 'czardas', *puszta* 'Puszta', *forint* 'forint', *honvéd* 'Honvéd', *gróf* 'count', *folklor* 'folklore', *temperament* 'temperament'), lexemes connected with the extreme political views (*ultrapravice* 'ultra-right', *ultranacionalista* 'ultra-nationalists', *šovinista* 'chauvinists'), the words related to the nationality and origin (*národnost* 'nationality', *menšina* 'minority', *původ* 'origin', *žid* 'Jew').

[17] Jan Slota allegedly defames Hungarian and encourages ... Slota can be investigated. Chairman Béla Bugár filed a complaint to the leader of the nationalists.

[18] They use Hungarians and autonomy to frighten voters mainly from north Slovakia who are out of touch with citizens of Hungarian nationality.

[19] There have been various misunderstandings and disagreements, because Slovaks hate the Hungarians and vice versa, and they argue among themselves.

[20] Do not hurt the Hungarians; they hurt a lot by themselves.

[21] Anyone would be interested in the new arrangement of the song Karmina, in rustic Polish music from Upper Orava, or in temperamental Hungarians.

Among the Slovak collocates, apart from the lexemes related to food (*čardáš* 'czardas', *tokaj* 'Tokay', *maďarská gulášová polievka* 'Hungarian goulash soup') and words related to politics (*Maďarská Starokonzervatívna strana* 'the Hungarian old-conservative party', *hnutie* 'movement', *maďarská Socialistická strana* 'the Hungarian socialist party', *šovinista* 'chauvinist', *m. kresťanskodemokratický* 'Christian Democratic'), we also find *inforádio* 'inforadio', *hajzel* 'asshole', *forint* 'the forint', *pusta* 'Puszta', *vyžla* 'weed', *garda* 'guard', *menšina* 'minority', *krátkosrstý stavač* 'shorthaired pointer', *repatriant* 'rapatriant', *prihraničie* 'border', *prizvuk* 'accent'.

From the analysis of collocations with the adjective *Hungarian* it was discovered that lexemes denoting foods are more strongly represented in Czech than in Slovak, which allows us to conclude that this area for Slovaks is more familiar than for Czechs. Words related to the policy, however, are more represented in the Slovak than Czech language, because this topic is taken up more often in Slovakia. It is worth noting that in both languages collocates refer to the extreme political views, which may suggest an image of the Hungarian as a politician with extreme political views. Among Slovak collocates, there are more words denoting animals (e.g. dogs) than among the Czech collocates. It may indicate the opposition of a rustic culture (Slovakia) and an urban one (Czech)[22]. We must also note a vulgarism, which co-occurs with the adjective 'Hungarian'. The collocation *hajzel maďarský* 'Hungarian asshole', used in different contexts, is a part of a negative attitude to Hungarians by some Slovaks.

To sum up, the groups of collocations are quite similar in the Czech and Slovak languages, but differ in the severity of the specific characteristics and contexts. In both corpora, words related to the Hungarian minority are very frequent. Among Slovak and Czech co-occurring words we can separate lexemes associated with assimilation and autonomy. Many of the Slovak lexemes also relate to the issue of displacement. This group is also represented in the Czech corpus, but to a much lesser extent. Similarly, the words that can have negative overtones are properly limited to the contexts in Slovakia and they do not represent Czech attitude. In Slovakian journalism, many more occurrences showing a negative image of the Hungarians appear, mostly as a threat and problem for Slovakism.

3. The collocational profile of the Hungarians (Magyars) in the Bulgarian language

The Bulgarian National Corpus (Български национален корпус, referred to here as BNC) according to data given on its website (http://ibl.bas.bg/BGNC_bg.htm) contains 320 million tokens; the number is noted elsewhere as

[22] From the conversation with Anna Kobylińska, PhD (Slovakist, culturologist).

400 million or even 450 million (as of March 4, 2012). The Corpus was created by the employees of the Institute for Bulgarian Language, Bulgarian Academy of Sciences. Its contents developed in the years 2001–2009. "Materials in the Corpus reflect the state of the Bulgarian language (mainly in its written form) from the middle of 20th c. (1945) until present"[23], it cannot, however, be considered fully synchronic as it features numerous quotations from folk literature and texts in archaic pronunciation. The description of BNC does not unfortunately contain assessment of balance and proportions in its constituents. It is assumed that the Corpus consists in separate blocks of "specialized" quotations – from the language of sports, economy, law, offices, parliament shorthand, etc.

Bulgarian ethnonyms for 'Hungarians' are: *унгарец/унгарка* (adj *унгарски*; the partly syncretic adverb *унгарски* 'Hungarian, in the Hungarian way' does not affect the analysis) and *маджарин/маджарка* (adj *маджарски* + adv *унгарски*, similarly). The second variant is closer to the auto-ethnonym 'Magyar' and refers mostly to historical contexts (approx. until the 19th c., as indicated by Corpus material), frequently those of folklore and/or emotionally charged. The lexeme *маджарин* (along with *маджарка* 'Hungarian woman', in sg. as in pl.) is noted in the BNC 764 times (0.000191% of 400 million BNC tokens), the adjective *маджарски* (in any genus and number) – 455 times (0.00011375%), *унгарец* – 1757 times (0.00043925%) and *унгарски* – 6512 (0.001628%). Despite this evident disproportion one should assume complementarity of these names and focus on the specific ways both word groups function: first that of *маджарин/маджарски*, then *унгарец/унгарски*.

While preparing the present article, the BNC was affected by serious technical impairment: lack of response in tools allowing for research in collocations. The research was therefore conducted "manually" (even without using regular expressions), by means of copying portions of Corpus material for further analysis. This could result in imperfect precision of numeric data or pertinence of assessment. The group most easily and precisely examined were the collocates placed before the adjectives *маджарски* and *унгарски* – the mentioned technical problems and lack of means for ample presentation in the present team analysis are reasons to focus on these relations.

3.1. *Маджарин, маджарски*

As the greater part of collocations of the substantive *маджарин* refers to ancient eras (the "pre-Turkish", see remarks closing this section), the collocations with adjective *маджарски* contain mostly words belonging in the domain of history, associated with notions such as hierarchies of power, army, wars, violence, etc., frequently with negative emotional charge: *крал* ('king', 92 occurrences), *земя* ('territory', 16), *копеле* ('bastard', 13), *юнак* ('hero', 7), *войвода*

[23] http://ibl.bas.bg/BGNC_bg.htm, 4.03.2012.

('voivode, leader', 7), *източници* ('sources, documents', 7), *наемници* ('mercenaries', 6), *власт/власти* ('authority/authorities', 4), *конници* ('horsemen', 3), *кралство* ('kingdom', 3), *нашествия* ('attacks', 3), *княгиня* ('duchess', 2), *принцеса* ('princess', 2), *корона* ('crown', 2), *владетел* ('monarch').

Less frequent are words of more universal functions, referring to modern or contemporary history: *вино* ('wine', 3 – apparently, Hungarian wine does not attract Bulgarian attention), *език* ('language', 3), *емиграция* ('emigration', 2), *дума* ('word', 2), *дреха* ('outfit'), *крез* ('richman', on the activity of the Hungarian American businessman and philantropist George Soros), etc.

To complete the analysis of collocations of the adverb *маджарски* some remarks on the founding noun *маджарин* should be added. Almost 2/3 among the 216 occurrences of the lexeme *маджарин* (m sg.) concern the nickname of a historical figure that had served for the Hungarian king, Philippo Scolari (1369–1426) who (just as e.g. Prince Marko) has been changed, from an authentic person of little significance, into a hero of folk epic in which he is a leader (or even a king), apparently cruel towards Bulgarian people.

Other contexts in which the word *маджарин* occurs are related to less distant past: the interference of the Catholic Habsburg monarchy in the Bulgarian life at the decline of Turkish reign, Hungarian historical figures (e.g. Mór Jókai), generalized notions on neighboring nations, ethnic tensions; some quotations account for deeply negative stereotypes.

Wars waged between the former states of Bulgaria and Hungary result in voices reminiscent of battles, alliances, etc. Common collocations for *маджарин* are therefore adjectives *страшен* 'terrible', *черен* 'black' or *див* 'wild'.

The BNC mentions other ancient cultural or economic contacts; there incidentally occurs the interesting issue of conflicting relations between Bulgarians and Hungarians in the Banat territory (in today's Romania).

The word *маджарин* is sporadically met in information concerning the events of the 19th c. (the anti-Austrian uprising in 1848), with the exception of quotations referring to present times, e.g. the situation of Hungarian minority in Romania. The word *маджарин* then gains a historical undertone (parallel to that in today's Polish).

Some (usually unfavorable) generalizations on Magyar customs are observed, e.g. concerning culinary habits strange to Bulgarians or the lack of refinement in their everyday language.

3.2. *Унгарец, унгарски*

The word *унгарец* and the adjective *унгарски* derived thereof are contemporary terms, semantically neutral, meaning '1. member of Hungarian nation; 2. citizen of Hungarian state'. They can also refer to past times, like the usually used *маджарин*, *маджарски* (see above, p.1).

The collocations of the adjective *унгарски* are rather distinctive and can be divided into several main groups.

The most important collocate in the BNC is *език* ('language', occurred 226 times). This is partly due to official documents (e.g. Bulgarian versions of European Union related) being featured in the BNC, and partly to common opinions on the strangeness of Hungarian and difficulties in learning and translation.

As mentioned before, the adjective *унгарски* also refers to past eras, although this domain is usually associated with the (largely prevailing) adj *маджарски*; there are numerous combinations with lexemes *крал* ('king', 108 occurrences), *кралство* ('kingdom', 21).

Collocates referring to recent history of the Hungarian state and people are: *въстание* ('uprising', 32 times), *революция* ('revolution', 41), *събития* (euph. 'events', 45). This concerns the years 1848 (the uprisings of Kossuth and Bem), 1919 (proletarian revolution of Béla Kun) and October 1956 (with hints of shame that Bulgarians did not afford such mobilization, also subsequent expressions of Polish anti-communist resistance are evoked on occasion).

The BNC also contains entries accounting for critical (attentive and anxious) observation of successes and failures of Hungarian economy, usually in more recent times. It seems that Bulgaria whose economy is also rather unstable should mind to avoid economic mistakes that affected Hungary. This is confirmed in high frequency of collocates: *банка* ('bank', 104 entries), *форинт* ('forint', 73), *икономика* ('economics', 31), *компания* ('company, consortium', 52), *фирма* ('firm', 8), *пазар* ('market', 44), etc.

The BNC contains numerous quotations accounting for attentive overwatch of Hungarian politics. Their tone is similar to that concerning economic issues. See collocates: *власти* ('authorities', 32), *столица* ('capital city', 109, also depicted as urban creation and tourist attraction), *демократичен* 'democratic' [*съюз* 'union', *форум* 'forum' in names of parties] (36), *законодателство* ('legislation', 17), *министър* ('minister', 12), *органи* ('organs [executive powers]', 218), *парламент* ('parliament', 34), *партия* ('party', 11), *правителство* ('government', 115), *президент* ('president', 33), *премиер* ('prime minister', 83), et al.

There are also many entries referring to Hungarian culture and science, generally assessed in positive terms, see: frequency of collocates *кино* ('cinema', 18 instances), *културен* 'cultural' [mostly *институт* 'institute'] (59), *учени* ('scientists', 19).

The problem of ethnic minorities, significant for Hungarians (especially of the Hungarian minority in the Romanian Transylvania, the Serbian Vojvodina, in southern Slovakia and in Austria and Slovenia) is also reflected in the BNC material. These problems only potentially affect Bulgarian society, which has to face its own problems concerning ethnic minorities. This is illustrated by

collocates: *малцинство* ('minority', 47 occurences), *население* ('population', 12), *произход* ('origin', 45). To complete these final observations, the collocation of the noun *унгарец – етнически* ('ethnic', 80 occurrences), of considerable quantity, should be included.

In conclusion to specification of collocates for *маджарин/маджарски* and *унгарец/унгарски*, it can be assumed that the BNC contains rich research material which reflects the interest that Bulgarian society sees in Hungarian problems as well as their attentive observation and criticism. The usage of *унгарец/унгарски* (as opposed to *маджарин/маджарски*) does not feature stereotypes or other generalizations, apart from the prevailing positive opinion on Hungarian science and culture. The quotations account for importance of ethnic (minorities') problems in the Hungarian social relations.

Bulgarians have developed expressive images (stereotypes) centered on neighbors (antagonists throughout history): Gypsy-Roma, Albanian-Arnaut, Greek-Byzantine, Turk; or nationalities geographically remote, "exotic", the knowledge of which is commonly unverifiable (Spaniards, the French, Italians, Americans). The image of Hungarians is different. It seems nowadays emotionally uncharged and not stereotypical – the ethnos observed being too close to exoticize, its interests not having collided with those of Bulgarians for a long time; it lacks therefore reasons to e.g. generalize or to "domesticate the enemy". The prevailing position is that of observing an indirect neighbor who shares a common situation in geopolitics and economy, and whose history can bring useful conclusions.

4. The collocational profile of the Hungarians in the Slovene language

The aim of this part of the paper is to analyze the collocational profile of Hungarians in the Slovene language. It is an analysis based on corpora and therefore it has its obvious limitations. However, it can be considered useful while analyzing stereotypes as a whole and their proper parts. It can be also considered useful for those involved in research on the Hungarian minority in Slovenia, which, according to the 2002 census, is the fifth largest in the country (6,243 people living mainly in the northeastern municipality of Lendava and its closest surroundings)[24].

During the research, the FIDAPLUS corpus was used. It is the second largest and the second youngest corpus of all the Slovenian corpora. It was made public in the year 2006 and it consists of 621.15 million tokens, being collected since 1990. As far as sources are concerned, the vast majority of words are taken from press (88.52%), the second largest source is books (8.74%), whereas only 1.24% of tokens have been taken from the Internet, and 1.5% is of other origin.

[24] CIA World Factbook, 2010.

From the technical point of view, two lemmas were analyzed in all their forms – these were the noun *Madžar* 'Hungarian' and the adjective *madžarski* 'Hungarian'. As far as the collocations are concerned, the closest neighborhood of both lemmas was taken into consideration, i.e. one word before and after the lemma. From the statistical point of view, the mutual information (MI and MI3 values) was the most important factor; collocations which had the MI3 value of over 7.00 were considered statistically relevant.

The lemma *Madžar*, in all its forms, had 10778 records, which is 0.0017% of all the lemmas in the corpus, while *madžarski* was found 23691 times (i.e. 0.0385%). The results can be grouped into several categories showing whether there is, or not, any language stereotype of a Hungarian in the Slovene language.

Before analyzing the results, we have to mention that the fact that a large number of collocations are connected with politics, sport and culture, is a result of the proportions of particular kinds of texts included in the corpus. Since over 88% of the material is taken from press, there is over-representation of these thematic areas, which is very frequent in this kind of media. Out of these three, the collocations are mainly names of professions and notions connected with politics and culture as well as with various sports.

Results here can be grouped into two categories. First, is concerning the noun *Madžar*, where four categories are most important. These are collocations connected with geography (23.8% of all the relevant collocations, showing both places of origin of large groups of Hungarians in Slovenia and abroad like Pomurje, Prekmurje or Vojvodina) and sport (15.8%), adjectives showing various features of Hungarians (19.4%) and verbs showing their activities (18.2%). As far as the adjective *madžarski* is concerned, far more groups could be established, the most important of which are shown in the table below.

Group	**Occurrences (%)**
Politics	3038 (12.8%)
Army/military	2425 (10.2%)
Nation/nationality	2311 (9.8%)
Culture	1969 (8.3%)
Sport	1506 (6.4%)
Geography	1000 (4.2%)
Economy	755 (3.2%)
Minority	696 (2.9%)
Immigration	602 (2.5%)
Language	442 (1.9%)
Professions	425 (1.8%)
Cuisine	425 (1.8%)

History	390 (1.6%)
Industry	260 (1.1%)
Tourism	203 (0.9%)
Other nations	175 (0.7%)
Fascism/occupation	173 (0.7%)
Media	144 (0.6%)
Agriculture	118 (0.5%)
Others	4275 (18.1%)

Before a more detailed description is given, it has to be mentioned that out of these groups, one can see that the most important features of a Hungarian for Slovenians are:

- being neighbors; the vast majority of the geographical collocations concerning both the noun and verb are linked with the regions/towns bordering Slovenia, the noun *sosed* 'neighbor' has 44 occurrences itself.
- being militant and belligerent; among the collocations connected with the army/military service one can find such words as *vstaja* 'uprising', *odpor* 'resistance', *robusten* 'robust', *uporen* 'rebellious', *skregan* 'on bad terms with', *bojevit* 'feisty', *vročekrven* 'hot-tempered'.
- there is a large group of historical collocations, which shows both the ties between Hungary and Slovenia throughout history (such as *kralj* 'king', *kraljestvo* 'kingdom', *fevdalen* 'feudal', *plemstvo* 'nobility', *plemič* 'nobleman', *prestol* 'throne') as well as historical events in common history (such as *revolucija* 'revolution', *revolucionar* 'revolutionary' – both concerning the 1848 revolution).
- interestingly, there is a large group among the collocations connected with history that I decided to consider as a separate group, and named it 'fascism/occupation'; these lemmas are connected with the late 30's and early 40's regimes in Hungary which were both pro-fascist (lemmas such as *fašizem* 'fascism', *fašistski* 'fascist', *profašističen* 'pro-fascist'), the Hungarian occupation of a part of Yugoslav territory (*okupator* 'occupier', *okupacija* 'occupation', *okupatorski* 'connected with occupation'), and the revisionist ideology of the so-called 'Great Hungary'[25] (*revizionizem* 'revisionism', *iredenta* 'irredenta', *iredentist* 'irredentist' and *iredentizem* 'irredentism').
- among the concordances connected with the army and military service there are also many historical associations such as *honved* 'soldier of the Hungarian army during the 1848 war with Austria', *pandur*, *žandar* – 'member of military police', *huzar* – 'member of light cavalry unit'; however, the majority of concordances in this group concern the current situation.

[25] i.e. Hungary in its pre-I world war borders.

- as far as cuisine is concerned, there are mainly stereotypical collocations, concerning 'national Hungarian' dishes and specialities, such as *golaž* 'goulash', *paprikaš* 'paprika stew', *tokaj*, *tokajec* 'Tokay wine', and *salama* 'salami sausage'; this is somewhat a stereotypical picture of Hungarian cuisine in all the European countries.
- the "minority" group consists of many records, but the vast majority of them (582 out of 655) is the collocation *madžarska manjšina* 'Hungarian minority'; what is interesting, is that it does not always mean Hungarian minority in Slovenia, but also sometimes refers to Hungarian minority in other countries, mainly in Serbia and Slovakia.

Among the 'others' groups there are many interesting cases which cannot be examined thoroughly due to the limited quantity of the paper. However, the most significant collocations are those connected with assimilation of a minority (15 occurrences altogether for the noun and verb), a small group of words referring to population movement (i.e. *asimilacija* 'assimilation', *asimilirati* 'to assimilate', *begunec* 'fugitive', *izseljenec/izseljenka* 'refugee', *izseljevati* 'to migrate', *avtohton* 'autochtonous') We have to mention, however, that, unlike the collocations from the large groups above, all these are single separate cases in the corpus, which obviously cannot be used to create a rule.

In summary, this research does not give us any expressive stereotypes, which could be a result of several factors. The most important is that although Slovenia and Hungary remained in one country for several centuries, they did not establish strong political or cultural ties and their fields of interest were different. Slovenia has always been pro-Austrian, whereas Hungary was not interested in Slovenia as a field of its influence. The attitude towards Hungarians in contemporary Slovene can be considered rather neutral, which is not typical for a nation bordering Hungary; however, it may come out of the fact that the Hungarian minority in Slovenia is not large and that there are no collisions of its interests and demands with inner politics of the country. We can, however, distinguish what we call current attitude, which is as I mentioned neutral and historical, and can be negative in some aspects.

5. The collocational profile of the Hungarians in the Polish language

For a number of reasons one can expect the collocational profile of Hungary and Hungarians in a corpus of Polish to be a positive one. First, there is the proverb *Polak, Węgier – dwa bratanki* (lit. 'Pole and Hungarian – two nephews'), known well enough to be cited usually without the second part: *I do szabli, i do szklanki* ('Both for a sabre and for a glass'). It is not sure whether the proverb refers to ethnic Hungarians or the inhabitants of the so-called 'Upper Hungary', i.e. Slovaks [see divergent opinions in: Krzyżanowski 1970, Markiewicz i Romanowski 2005, Tazbir 1991, Kaczorowski 1998]. Besides, the

proverb is not as old as one might assume (it probably originated after 1848, see Csapláros 1964). Nevertheless, it works in Poland as a confirmation of the centuries-long Polish-Hungarian friendship.

Second, the positive attitude of Poles to Hungarians is attested in public opinion surveys. About 50 per cent of Poles declare their liking to Hungarians, while the rest remain undecided rather than hostile or unfriendly [Pawlak 2008].

Third, the portraits of Hungarians in the 19th and 20th century Polish literature are likewise positive [Sieroszewski 1995: 60]. The figures portrayed are, with minor exceptions, noble, faithful and loyal, courageous, and sensitive to social injustice.

Can the friendly attitude of Poles to Hungary and Hungarians be confirmed in a collocational analysis in the National Corpus of Polish (Narodowy Korpus Języka Polskiego, henceforth NKJP)? In order to answer this question, one may refer to the collocator built into PELCRA, one of NKJP search engines. It is reasonable to start the analysis from the balanced subcorpus of ca. 300 million tokens (nearly 300 million words). Most of its sources were first published after 1945: half of them come from periodicals, a third from fiction and non-fiction books, the rest from the spoken language (of different levels of formality) and the Internet.

The following lexemes are the subject of the analysis: *Węgry* ('Hungary'), *Węgier* ('Hungarian', noun), *węgierski* ('Hungarian', adjective), *po węgiersku* 'Hungarian-style', as well as *Madziarsko*, *Madziar*, *madziarski*, which are the stylistically marked synonyms of *Węgry*, *Węgier*, *węgierski*, respectively. The relative frequencies in the balanced subcorpus are not easy to measure, because some word-forms are common to *Węgry* and *Węgier*. Approximately, *Węgry* occurs 25 times in every million words, *Węgier* – 9 times, *węgierski* – 14 times, *Madziar* – only once. Other words display a frequency of below one occurrence per million words.

The first conclusion from the analysis, a rather surprising one, is that a 300-million corpus of Polish does not allow for an image of Hungary or Hungarians to be drawn on the basis of the collocations alone. Thirty most significant collocates for the lexeme *Węgry*, ranked according to the chi-square values, are:

> *król, stolica, regent, reprezentacja (drużyna sportowa), ambasador, premier, prymas, reprezentant, południe, królestwo, historyczny, terytorium, dzisiejszy, wyjątek, prezydent, wyzwolić, republika, odbyć się, rządzić, wygrać, przykład, teren, obywatel, zająć, rząd, trener, granica, gospodarczy, los, władza*

> *king, capital (city), regent, team, ambassador, prime minister, Primate, (a) representative, (the) south, kingdom, historical, territory, present-day, exception, president, to liberate, republic, to take place, to govern, to win, example, terrain, citizen, occupy, government, coach, (the) border, economic, fate, authorities*

Proper names have been omitted from the list (we will return to them soon) and so have numerals and function words. The majority of collocates belong to the historico-political vocabulary (e.g. *król* 'king', *stolica* 'capital city', *premier* 'prime minister', *wyzwolić* 'liberate', *rządzić* 'govern', *obywatel* 'citizen', *gospodarczy* 'economic') and only a few words relate to sports (e.g. *reprezentacja* 'team', *trener* 'coach'). As for proper names, they refer mainly to the countries with which Hungary has maintained political or cultural relationships (the first five are the Czech state, Slovakia, Romania, Slovenia, and Bulgaria), as well as people connected with Hungary. A query for the lexeme *Węgier* brings similar results, the only difference being a high rank of the word *bratanek* 'nephew', a component of the aforementioned proverb, which is often quoted and commented on in the Internet.

The collocations of the adjective *węgierski* are somewhat more diversified. Fifty most significant common nouns, placed immediately left or right to the key-word, are:

> *puszta, forint, mniejszość, czardasz, gulasz, tokaj, król, huzar, wino, powstanie, rewolucja, salami, parlament, królewna, ambasada, żandarm, pochodzenie, Cygan, pisarz, koncern, Żyd, nacjonalista, Żydówka, granica, politolog, rząd, magnat, elita, ubezpieczyciel, język, żołnierz, taniec, kuchnia, owczarek, tron, dyplomacja, parlamentarzysta, delegacja, hrabia, indeks (giełdowy), korona, opozycja, narodowość, potrawa, kompozytor, korzenie, socjalista, powstaniec, małżonka, melodia*

> *Puszta, forint, minority, czardas, goulash, Tokay, king, hussar, wine, uprising, revolution, salami, parliament, princess, embassy, military policeman, origin, gypsy, writer, concern (company), Jew, (a) nationalist, Jewess, (the) border, political scientist, government, magnate (nobleman), (the) elite, insurer, language, soldier, (a) dance, cuisine, sheepdog, throne, diplomacy, member of parliament, delegation, count (earl), (share) index, crown, opposition, nationality, dish (food), composer, roots (origin), (a) socialist, insurgent, wife, melody*

Beside historico-political vocabulary a small number of particularly 'Hungarian' words appear, related to local peculiarities, e.g. *puszta*, *forint*, *czardasz* 'czardas', *gulasz* 'goulash', *tokaj* 'Tokay', *huzar* 'hussar', *salami*.

Some words can be particularly useful in a collocational reconstruction of a stereotype, e.g. of an ethnic group. In Polish, the use of the word *prawdziwy* 'true' may prove the very existence of a stereotype, *cf. prawdziwy Polak* 'a true Pole', while nominal phrases ending in *po polsku*, *po węgiersku*, etc. may help capture some of its components, *cf. karp po polsku* 'Polish-style carp'. Unfortunately, in the 300-million balanced corpus of Polish there are just a few occurrences of the phrase *prawdziwy Węgier*, which does not allow for generalizations, and among the nouns preceding the phrase *po węgiersku* 'Hungarian--style' there are only *gulasz* and *placek* (kind of fried pastry).

The findings suggest that either a 300-million corpus of Polish is too small to retrieve a global image of Hungary and Hungarians on the basis of the collocations alone, or such an image does not exist in Polish texts at all, and we have to settle for a fragmentary picture, restricted to a few historical events and some local color. In order to find out which of the guesses is true, one can repeat all the queries in the full corpus, which, albeit worse-balanced, has an impressive number of nearly one billion words.

This time the search for *po węgiersku* brings quite a long list of meals typical of Hungarian cuisine (or thought to be Hungarian), with goulash, carp, coffee, paprika, fried cakes, pork, herring, fish soup etc. As to *prawdziwy Węgier*, the collocation displays both culinary and agricultural contexts, e.g.

> Bela Hamvas jako prawdziwy Węgier wraz z mlekiem matki wyssał miłość do wina.[26]

and political contexts, e.g.

> W nacjonalistyczno-kulinarnym kodzie ironicznie stosowanym przez Vargę oznacza to, że Węgrzy w dalszym ciągu marzą o swojej imperialnej wielkości. Wspominają czasy, gdy Słowacja, kawałek Rumunii i Serbii należały do nich: „To wszystko kiedyś było nasze" – mówi ojciec pisarza podczas wycieczki samochodowej przez Słowację. „Powiedział to raczej z melancholią niż ze złością, bo był prawdziwym Węgrem" – dodaje autor.[27]

New collocates of the lexemes *Węgry* and *Węgier* appear in the full corpus as well, but do not change the former observations, for they fit well into the semantic categories distinguished before, namely history and politics (taken together) and the sports.

However, new nouns collocating with the adjective *węgierski*, once added to the ones found before, allow the historico-political vocabulary to be divided into a few smaller categories. These are, firstly, words related to the monarchical and aristocratic past of Hungary (e.g. *król* 'king', *arystokrata* 'aristocrat', *hrabia* 'count', *tron* 'throne'); secondly, words, relating to the struggle for the nation's independence (e.g. *powstanie* 'uprising', *rewolucja* 'revolution'); thirdly, words concerning ethnic matters (e.g. *emigrant* 'emigrant', *mniejszość* 'minority', *pochodzenie* 'origin', *Cygan* 'gypsy', *Żyd* 'Jew', also *korzenie* 'roots' as a synonym of ethnic origin).

[26] As a true Hungarian, Bela Hamvas took his love to wine in with his mother's milk.

[27] In the national-culinary code, applied ironically by Varga, this means that Hungarians keep dreaming about their imperial power. They recall the times when Slovakia, some parts of Romania and Serbia used to belong to them: 'All that was once ours', says the writer's father during a car trip across Slovakia. 'He said this with melancholy rather than anger, for he was a true Hungarian', adds the author.

As can be seen from above, the collocational analysis in the full Polish corpus adds new elements to the image emerging from the balanced corpus. Still, the progress seems small. A closer inspection of selected contexts is necessary to notice, e.g., the emerging picture of Hungary as a country with bitter historical experience, and the Hungarians longing for the former political power.

To complete the analysis, we need contexts in which *Madziar* and its derivatives appear in the present-day Polish. NKJP shows that they belong to the domains which we already know: history (both distant and recent) and sports competition. Sports journalists use the word *Madziar* instead of *Węgier* in order to add color to the story and to avoid repetitions.

Etymologically, both *Węgier* and *Madziar* relate to the members of the Asian nomadic tribes who settled in central Europe and gave birth to the Hungarian state (Brückner 1957, Kurczabowa 2004). Still, in the 19th century, as the historical dictionaries show, *Madziar* was an indigenous inhabitant of Hungary, while *Węgier* just belonged to the Hungarian homeland of nations (*natio hungarica*). Nowadays the two words are basically synonymous, but stylistically different. Although contemporary dictionaries mark *Madziar* as colloquial or jocular, the corpus analysis suggests that it would be more accurately defined as ornamental.

6. Conclusions

The collocation image of Hungarians is not equally clear in each of the Slavonic languages. This may be due to technical factors (corpus too small or ill-structured, low frequency of a lexeme, malfunctioning tools or their absence) or sociological ones (absence of a clear picture of the nationalities in the culture of a nation, mostly due to the lack of historical or contemporary contact/experience). Another reason may be the lack of a negative image, especially in the press. Perhaps that is why the most difficult thing for us was to reconstruct the image of Hungarians in the Polish language.

In the Bulgarian language a clear collocation image is drawn by one of the two lexemes defining Hungarians, referring to the historical contexts in which Hungary is seen as an aggressor. A neutral image is found in the Slovenian language and in Bulgarian (if we consider the other lexeme identifying the contemporary Hungarian), while in the Slovak language a clearly negative image prevails. Czech collocations suggest a negative attitude towards the Hungarians, but one can see (after placing the collocations in context) that they often occur in the statements reporting Slovak attitudes and problems. It seems that even such occurrences may affect the perception of the Hungarians among Czech people.

In Polish, Slovenian, and Bulgarian, collocations relate primarily to nationality, in the Czech and Slovak languages the national minority image dominates. In

most languages we have studied, words associated with food are frequent. The exception is the Bulgarian language: for Bulgarian people, Hungarian food is not marked, because their cuisine has similar characteristics. In Polish, Bulgarian and Slovenian – but not in Czech or Slovak[28] – an important group of collocations relate to the sphere of the royal and noble. The same situation concerns the words related to revolutions, rebellions, social liberation and struggle: they are visible in collocations of each studied language with the exception of Czech and Slovak. Pejorative lexemes are found primarily among Slovak and Bulgarian collocations (but only in contexts of tokens *маджарин*, *маджарски*) and partly in the Czech collocations. In all languages, a group of words related to politics can be seen: Slovak and Czech collocates are dominated by words defining extreme political views, while the Slovenian collocates allow to create a Hungarian-occupier category. Almost everywhere the characteristic feature of Hungarians is their language and temperament.

Hungarian images depend on historical and contemporary experience and are very different, although the predominant image is a neutral one (with the exception of the Slovak language in which it is clearly negative).

References

Bartmiński, Jerzy [2006], *Językowe podstawy obrazu świata*. Lublin.

Brückner, Aleksander [1957], *Słownik etymologiczny języka polskiego*. Warszawa: Wiedza Powszechna.

Bulgarian National Corpus, http://dcl.bas.bg/en/corpora_en.html, http://ibl.bas.bg/BGNC_bg.htm.

CIA World Factbook [2010], http://www.cia.gov/library/publications/the-world-factbook/, 01.03.2012.

Csapláros, István [1964], *„Polak, Węgier – dwa bratanki"* (*Les Polonais et les Hongrois sont des bons amis*). Slavica. Annales Instituti Philologicae Slavicae Universitatis Debreceniensis de Ludovico Kossuth Nominatae IV, 81–106.

Čermák, František / Šulc, Michal [2006], *Kolokace*. Lidové noviny. Praha.

FIDAPLUS korpus slovenskega jezika, http://www.fidaplus.net/, 02.03.2012.

Hebal-Jezierska, Milena [2011], *Kolokační obrazy některých lexémů patřících do sémantického pole cizinec v českém tisku* (*s metodologickými úvahami)*. In: Čermák, F. (ed.), InterCorp. Lidové nowiny. Praha, 109–121.

Kaczorowski, Aleksander [1998], *Tu nie będzie dyktatury. Wywiad z Ivanem Czicz-manecem*. Gazeta Wyborcza 79, 18, http://szukaj.wyborcza.pl/Archiwum/1,0,386048,19980403RP-DGW,Tu_nie_bedzie_dyktatury,.html.

Keibel, Holger / Belica, Cyril / Kupietz, Marc / Perkuhn, Reiner [2011], *Approaching grammar: Detecting, conceptualizing and generalizing paradigmatic variation*.

[28] In the Czech and Slovak this group is limited to the words related to noble, but not royal.

In: Konopka, M. / Kubczak, J. / Štícha, F. / Waβner, U. (eds.), Grammar und Korpora 2009. Tübingen: Narr Verlag, 329–356.

Kiss, Gy Csaba [1995], W podwójnym zwierciadle: Węgrzy i Słowacy. In: Walas, T. (ed.), *Narody i stereotypy*. Kraków: Międzynarodowe Centrum Kultury, 53–57.

Korpus prim-5.0=inf, Bratislava 2010, *Slovenský národný korpus*. Available at www.korpus.sk.

Korpus SYN2009PUB *Český národní korpus – SYN2009PUB.* Ústav Českého národního korpusu FF UK, Praha 2010. Available at WWW: http://www.korpus.cz.

Krzyżanowski, Julian (ed.) [1970], *Nowa księga przysłów i wyrażeń przysłowiowych polskich*, vol. 2. Warszawa: PIW.

Kurczabowa, Krystyna [2004], *Węgier i Węgry*. Poradnia Językowa PWN, http://poradnia.pwn.pl/lista.php?id=5012.

Markiewicz, Henryk / Romanowski, Andrzej [2005], *Skrzydlate słowa. Wielka księga cytatów polskich i obcych*. Kraków: Wydawnictwo Literackie.

Mc Enery, Tony / Wilson, Andrew, *Corpus Linguistics*, http://www.lancs.ac.uk/fss/courses/ling/corpus/Corpus3/3FRA1.HTM.

NKJP: *Narodowy Korpus Języka Polskiego*, http://www.nkjp.pl.

Pawlak, Agnieszka [2008], *Sympatie i antypatie narodowe? – fakty i mity*. Realia i Co Dalej... 2, 163–178, http://realia.com.pl/dzial_7/artykul_84.html.

Sieroszewski, Andrzej [1995], *Stereotyp Węgra w literaturze polskiej XIX i XX wieku*. In: Walas, Teresa (ed.), Narody i stereotypy. Kraków: Międzynarodowe Centrum Kultury, 58–62.

Tazbir, Janusz [1991], *Węgry jako symbol i przestroga w literaturze Staropolskiej*. Odrodzenie i Reformacja w Polsce XXXVI, 147–161.

Alicja Witalisz*

DID THE ANCIENT ROMANS USE COMPUTER SOFTWARE? LEXICOGRAPHERS IN THE FACE OF SEMANTIC CHANGE

1. Introduction

Part of my research on English semantic loans in Polish, carried out before 2006[1] and after, involved the examination of selected dictionaries of contemporary Polish to confirm the potential institutionalization of particular semantic loans. Some of the semantic loans in question, although relatively new as a linguistic phenomenon, are already included in the dictionaries of contemporary Polish published after 2000. This study led to some observations of a lexicographer's work.

The inclusion of a semantic loan in a dictionary entry means equipping an already existing Polish lexeme or a well-established loanword with an additional sense or senses that have appeared and been used in a variety of contexts in post-1989 Polish[2]. Since a large portion of my research on semantic loans from English concentrated on establishing a set of identification criteria of such loans, the lack of etymological information in the extended dictionary entries was very much noticeable. The more so that in those cases where the semantic loan names a very recent phenomenon, e.g. a computer-related object or activity, and the semantic extension happens to operate on an earlier Latin loanword, an average dictionary user, or a future user analyzing the Polish of

* Institute of Neophilology, Pedagogical University of Cracow.

[1] Findings of the research published in Witalisz [2007a].

[2] The terms 'loanword' and 'semantic loan' have received various interpretations in the literature. To avoid ambiguity, a 'loanword' is a foreign word whose form and meaning (or at least one sense) have been imported by the receiving language; a loanword may be adapted on the graphic, phonological, morphological and semantic levels, e.g. P. *jeansy*/*dżinsy* (< E. *jeans*), P. *leasing* (< E. *leasing*). In the case of a 'semantic loan', it is only part of the meaning (one or more senses) of a polysemous foreign lexeme that is imported by the receiving language and acquired by a native lexeme or a well-established loanword, e.g. P. *mysz* (< E. *mouse* 'computer device'), P. *definiować* (< E. *to define* 'to describe the nature or qualities of'; < L. *definio*) [*cf.* Haugen 1950: 214, 219; Hope 1960; Mańczak-Wohlfeld 1995: 44–77; examples from Witalisz, 2007a].

the early 2000s, might have the impression that for instance *application* (*software*) had already been known to the ancient Romans (see examples 6a, 6b, 6c and 7a, 7b below). This results from two facts. The lexicographer retains the etymological information of the source language of this 16th-century Latin loanword but resigns from providing the source language of the 20th-century semantic loan[3].

2. Position of a new sense in a dictionary entry

The newly-developed or borrowed senses usually receive the last position in a dictionary entry, and are placed under the traditional senses of the headword, as in the entries below:

(1a) **aplikacja** 1. ('...'); 2. ('...')[4]; 3. 'Aplikacje to programy komputerowe …' (E. Applications are computer software…); 4. 'Aplikacja o pracę lub o stypendium …' (E. job or scholarship Application …) [ISJP, I: 36]

(1b) **aplikacja** 1. ('...'); 2. ('...'); 3. *w informatyce* (E. *inform. tech.*) 'komputerowy program użytkowy ...' (E. computer application program ...); 4. *w środowisku biznesowym* (E. *in business*) 'podanie, prośba ...' (E. application, request) [SWOTP, 40]

(1c) **aplikacja** <łac. *applicatio*> (E. from Latin) 1. ('…'); 2. ('…'); 3. *inform*. (E. *inform. tech*.) 'program użytkowy …' (E. application program …) [USJP, I: 107].

Their inclusion in a dictionary confirms their institutionalisation[5] in the receiving language. An average user of Polish comprehends and uses the new sense actively in a variety of contexts. In this particular example, none of the dictionaries though lists two other new senses of P. *aplikacja*, i.e. 'the act of applying for', e.g. *membership application*, and 'the act of applying a cosmetic' [*cf.* Witalisz 2007a: 226].

In individual cases, the lexicographer may decide to move the borrowed sense up or even to the first position in the dictionary entry, which reflects the high frequency of use of a particular new sense, e.g.:

(2a) **aplikacja** 1. ('…'); 2. 'komputerowy program użytkowy' (E. computer application program); 3. ('…'); 4. ('…'); 5. ('…'); 6. 'podanie, np. o pracę' (E. job application) [STPS, 24]

[3] This problem has already been remarked on in Witalisz [2007b].
[4] ('...') marks the traditional sense of a lexeme.
[5] The term has been adopted from Leech [1974: 227].

(2b) **promocja** 1. 'w handlu, spopularyzowanie produktu' (E. in trade, popularizing a product); 2 ('…'); 3 ('…') [STPS, 639]

(2c) **promocja** 1. 'promocja jakiegoś towaru …' [E. a product promotion]; 2. 'promocja jakichś działań, idei, pomysłów …' [E. promotion of ideas…]; 3. ('…'); 4. ('…') [ISJP, II: 270]

P. *promocja* appears a particularly problematic case. One of the dictionaries lists its two new senses, of which one is classified as a semantic loan, the other as part of a separate loanword:

(3a) **promocja I** 1. 'działania zmierzające do spopularyzowania czegoś' (E. popularizing a product); 2. ('…'); 3. ('…')
promocja II 1. 'wzmożona działalność reklamowa …' (E. intensive advertising of a product) [WSJPP, 665]

Generally, the position of the new sense added to the entry may vary among dictionaries or even among the subsequent editions of the same dictionary, as in the case of P. *artykułować* 'to express thoughts, opinions' (< E. *to articulate*) [*cf.* ISJP, I: 44 and USJP, I: 128][6].

3. Possible reasons for lack of etymological information

A lexicographer who decides to extend a dictionary definition by the addition of a new sense which is a semantic loan faces the problem of identifying the source language of the loan. At least, in theory. In practice, this problem is not addressed and hardly any dictionary of contemporary Polish undertakes the challenge. There may be several factors which motivate this strategy of avoidance.

3a. Semantic loan identification problems

As for loanwords, the etymological information is often provided in dictionary entries, though this is one of the lexicographic practices that proves a certain degree of, to put it euphemistically, subjectivity in dictionary writing[7]. In the case of semantic loans several problems come to the fore. One of them is the confirmation that we indeed deal with a contact-induced semantic innovation, i.e. that the new sense is a result of borrowing the polysemy of a foreign lexeme rather than an independent semantic development, such as for instance P.

[6] On the hierarchy of senses in polysemous lexemes see Kurkowska [1965], Jadacka [1990].
[7] *Cf.* the criticism of etymological information in Polish dictionaries by Walczak [1989].

architekt 'creator', as in *architekt pokoju* (E. 'peace architect'), or P. *platforma* 'a formal declaration of principles', e.g. *platforma współpracy* (E. 'cooperation platform').

Secondly, establishing the source language of the new sense may prove an uneasy task in view of the fact that many of the Polish lexemes that have recently acquired new senses under English influence are in fact earlier Latin loanwords whose Latin etymons are also semantically broader, e.g. P. *definiować* ('to describe the nature or qualities of' < E. *to define*) < L. *definio*, and also because the semantic extension may be of international character. However, in the case of the recent semantic loans which acquire international status it is easy to detect the source language since many of them are metaphorical extensions and belong to either culture-specific or computer-related terminology, e.g. Polish *mysz*, Spanish *ratón*, German *Maus*, Czech *myš*, French *souris*, Hungarian *egér*, etc. (< E. *mouse* 'computer device').

Also, a lexeme that has acquired a new sense may be a reduced form of a larger expression that itself is a calque of a foreign phraseological unit, as in P. *szczyt* (< E. *summit*) which has adopted the meaning of the whole phrase P. *konferencja na szczycie* (< E. *summit conference*) [Witalisz in press].

One other problem, indirectly connected with the above, is deciding on the status of the innovation that has appeared under foreign influence and classifying it as either a semantic loan or a loanword that happens to be homonymous to the already existing Polish word, which is the case with for example P. *kursor*, an old Latin loanword (L. *cursor*), and P. *kursor*, a recent English loanword meaning 'an indicator used to show the position on a *computer monitor*' (< E. *cursor*)[8].

3b. Subjectivity in dictionary writing

Even though an average language user reaches for a dictionary as an ultimate and infallible source of linguistic information, a careful reader notices a certain degree of subjectivity, which takes its source in several factors. We shall refrain from commenting on the obvious differences in the contents of dictionary entries that define the same lexical item.

Discrepancies are also present in providing the etymology of lexemes in dictionaries, not excluding dictionaries of loanwords or etymological dictionaries. This lack of agreement among lexicographers as to the origins of the same lexeme may result from the different understating of the 'source' of a loanword. Does Polish *tomahawk* come from English *tomahawk* or from Algonquian *tāmāhāk(an)*? In other words, the source of a loanword may either be perceived as the direct source of a loanword (P. *tomahawk* < E. *tomahawk*) or as

[8] For differentiation criteria see Kurkowska [1976: 101ff], Markowski [2000: 108], Witalisz [2007a: 78ff].

the ultimate etymological source (P. *tomahawk* < Al. *tāmāhāk(an)*). The etymological information in dictionary entries may then include either 1) the direct source of a loanword or 2) the ultimate etymological source, or 3) both of them plus the possible intermediary links [Walczak 1989: 76]. This leads to deriving words such as for instance Polish *tajfun* (E. *typhoon*) from either German *Taifun* [SJPD, SWOR, SWOT] or English *typhoon* [SWOK] or Chinese *taj fun* [SJPSz, MSJP] or Greek *typhón* [SJPSz]. Some of the dictionaries also mention Arabic *tūfan* as a possible intermediary link, though in various configurations [SWOK, SWOT]. If one of the purposes of dictionary writing is recording facts from the history of a language, choosing the first of the three possible options seems sufficient. The least justifiable would be listing only the ultimate etymological source, as this might suggest non-existing language contacts. We cannot obviously exclude cases in which the direct source equals the ultimate etymological source [for a thorough discussion and exemplification of this problem see Fisiak [1962], Walczak [1989]].

As for semantic loans, unfortunately, the problem may appear in the case of some anglicized Latinisms, i.e. new senses which are outcomes of English linguistic influence and which are acquired by earlier Latin (or French) loanwords in Polish. The appearance of the new sense of P. *definiować* 'to describe the nature or qualities of' results from copying the polysemy of E. *to define*, which, just as its Latin etymon, is broader semantically than its Polish counterpart. The same is true of the new senses of P. *adresować coś do* 'to communicate directly to' (< E. *to address sth to*), P. *dedykować* 'to commit to a goal' (< E. *to dedicate*), P. *generować* 'to bring into existence' (< E. *to generate*). The lexicographer might decide to follow the third of the above options and include both the direct source of the semantic loan and its ultimate etymological source, especially in view of the fact that the Latin or French loanwords in Polish used to be semantically broader (just like their etymons) when they were first borrowed in the 16th-18th centuries [Witalisz 2007a: 150ff]. This brings us exactly to what is advocated in the present paper, i.e. providing the etymological information for the earlier loanword and also for the recent semantic loan, especially if the latter is a result of the influence of a different language than the former.

For semantic loans that are culture- or technology-related, the problem of deciding on the type of the source of a loan disappears, as the newly-acquired senses were coined in English, e.g. P. *administracja* 'government' (< AmE *administration*) (< L. *administratio*). Thus the direct source of this American English semantic loan equals its etymological source and there are no intermediary links.

The subjective approach to dictionary writing also manifests itself in the use of divergent criteria for deciding on the polysemy or homonymy of lexical

items. The much quoted example of homonymy illustrated by two nouns, **bank I** and **bank II** [in CIDE, NODE], or even **bank III** [in LDCE1, LDELC, OALED, OCDEE, WCDIE], turns out to be captured as one polysemous lexeme in CCELD, LAAD and LDCE2. These contradictory categorizations of words like *bank* stem from the fact that the lexicographer, in deciding on the polysemy or homonymy of a lexical item, may choose to give priority to for instance the etymology criterion rather than to any of the other two criteria commonly used for this purpose, i.e. the formal identity/distinctness criterion or the close semantic relatedness criterion[9]. Lipka [2002: 156] argues that polysemy and homonymy are not "absolute opposites, and total homonymy must be seen as the end-point of a continuum" [see also Jadacka 1990: 438]. Also, some linguists may change their understanding of the two phenomena over the years [*cf.* Lyons 1968: 406, and 1977: 554 qtd. in Lipka 2002: 153–154].

This last observation is also true for the perception and treatment of loans. Firstly, as already illustrated with the example of P. *kursor*, lexicographers may differ in deciding on whether an innovation is semantic or lexical, in other words, whether they deal with a semantic loan or a loanword, which in practice means that the innovation will either be listed as an extension of the already existing entry or will receive a new entry of its own. Secondly, a loan classified as a semantic loan may happen to be reclassified as a loanword, and *vice versa* [*cf.* e.g. P. *elektorat* 'the body of persons entitled to vote' in Markowski [2000: 108] and [2005: 218] or P. *promocja* 'popularizing a product', 'promotion of ideas' in ŁSTS [2001: 210] and Markowski [2005: 217]]. This different treatment of loans is also dictated by the use of distinct differentiating criteria, or simply by new developments in linguistic studies.

3c. Unsettled status of semantic innovations

At every stage of language development there appear innovations, including semantic innovations, of which only a few are likely to become institutionalized. The fate of a particular innovation cannot really be foreseen. Whether a semantic neologism will enrich the lexical system or become obsolete and come out of use depends on its nominal or expressive character. The names of new objects, phenomena or activities are much more likely to settle in the lexical system of language than innovations expressing emotions. These are liable to be replaced by still newer means of expression.

In the case of neologisms derived in word-formation processes, the likelihood of the institutionalization of a new word is directly proportional to the number of derivatives it forms and to the number of words that are coined

[9] A detailed discussion of the differentiating criteria used by various linguists is offered in Lipka [1986].

with the use of the same base [Jadacka 2001: 152]. As for semantic innovation, the language user is not faced with a new lexeme whose new form and new sense have to be remembered; it is only the new sense that is attached to a familiar word, or, in other words, a familiar word is used in an untypical context. Although this may be seen as a factor facilitating the institutionalization of semantic innovations in a lexical system, Polish dictionary writers represent a rather cautious attitude and refrain from extending dictionary entries prematurely. This is not a satisfactory solution to the researcher of language change. Storing a semantic innovation, even if it proves to be of ephemeral nature, would contribute to the more precise recording of the history of a language and of language change.

3d. Lack of space

One other practical reason for not providing the etymological information for the new semantic loans is the limited space in printed dictionaries. This, however, should no longer be the issue in the era of online information storage.

4. Ways of introducing semantic loans in dictionaries

Polish lexicographers choose one of the two ways of treating contact-induced semantic innovations. One of them is listing the innovation as a separate dictionary entry, thus assuming it is a loanword homonymous to the already existing lexeme in the receiving language, as in the following cases:

(4a) **promocja I** (łac. *promotio*) (E. from Latin) 1. 'przyznanie uczniowi prawa przejścia do następnej klasy' (E. putting ahead to the next higher grade at school); 2. 'nadanie tytułu naukowego lub stopnia oficerskiego' (E. advancement in rank or position) [traditional senses]
promocja II (ang. *promotion*) (E. from English) 1. 'ogół działań służących wylansowaniu jakiegoś towaru' (E. popularizing a product); 2. 'czasowe obniżenie ceny jakiegoś towaru' (E. lowering the price of a product) [NSWO, 509]

(4b) **promocja I** 1. 'nadanie tytułu naukowego lub stopnia oficerskiego' (E. advancement in rank or position); 2. 'przyznanie uczniowi prawa przejścia do następnej lasy' (E. putting ahead to the next higher grade at school) <łac.> (E. from Latin) [traditional senses]
promocja II 1. 'wzmożona działalność reklamowa …' (E. intensive advertising of a product); 2. 'sprzedaż nowych towarów … po celowo obniżonej cenie dla zwiększenia ich popularności' (E. selling a product at a reduced price to popularise it) <ang.> (E. from English) [WSJP, 1395]

(5) **I mysz** 'niewielki gryzoń wszystkożerny ...' (E. a small rodent…)
II mysz <ang. *mouse*> (E. from English) *inform*. 'urządzenie eletroniczne ….' (E. an electronic device …) [USJP, II: 752; also in STPS, 420; WSJPP, 478].

As for (4a) and (4b), it is a lexicographer's decision to regard **promocja II** used in its new senses as a separate loanword from English, rather than a semantic loan. In the case of (5), however, listing **II mysz** separately seems an awkward practice in view of the fact that *mysz* is a native Polish lexeme. Although the positive aspect of such a decision is providing the etymological information (presumably of the new sense), placing **II mysz** in a separate entry unavoidably and quite mistakenly suggests that it is a loanword from English.

The other strategy is listing the new sense (or senses) under the traditional sense (senses) of the existing headword. In the majority of the dictionaries used for this study no etymological information for the new sense is provided, which is a source of two types of misleading information. In the case of an earlier Latin or French loanword which today acquires a new sense under the influence of English, the new sense, whether listed in the final or middle position in the definition of a polysemous word, happens to be either preceded or followed by the etymological information concerning the loanword itself. If no other etymological information is provided for the new sense, the implication is that all the senses listed come from the same source as the original loanword, i.e. either from Latin or French, as in the examples that follow:

(6a) **aplikacja** 1. ('...'); 2. *inf*. (E. *inform. tech*.) 'częsta nazwa użytkowych programów ...' (E. a frequent name of application programs); 3. ('…') – łac. (E. from Latin) *applicatūra*, *applicatio* [SWOZOA, 42]

(6b) **aplikacja** (łac. (E. From Latin) *applicatio*) 1. ('…'); 2. ('…'); 3. *inform*. (E. *inform. tech*.) 'program użytkowy …' (E. application program) [USJP, 2003, I: 107]

(6c) **aplikacja** 1. *praw*. (E. *law*) ('...'); 2. ('...'); 3. *inf*. (E. *inform. tech*.) 'komputerowy program użytkowy' (E. computer application program)' <łac.> (E. from Latin) [WSJP, 34)]

(7a) **molestować** 1. ('…'); 2. 'wykorzystywać seksualnie' (E. to harrass sexually) z łac. *molestare* (E. from Latin) [NSEJP, 359; also in WSEH, 436]

(7b) **molestować** 1. ('…'); 2. 'zwykle w połączeniu z wyrazem *seksualnie*: nakłaniać kogoś do kontaktów seksualnych' (E. in a collocation with the word *sexually*: to harrass sexually) <łac.> (E. from Latin) [WSJP, 868]

Dictionary users are misled also in the case of Polish native lexemes which are extended semantically under English influence. Not providing the etymological information for the new sense in an extended dictionary entry implies that it is a semantic innovation independent of any foreign influence, e.g.:

(8) **mysz** 1. ('…'); 2. 'Mysz to małe urządzenie elektroniczne …' (E. Mouse is a small electronic device …) [ISJP, I: 907; also in WSJP, 886; WSJPP, 478]

There are very few instances in which the dictionary entry includes an annotation explaining that a particular sense of a lexeme is used in reference to, say, the American reality, as in the case of P. *administracja* (E. *administration* 'government') [USJP, 2003, I: 14].

Among the many dictionaries of contemporary Polish that have appeared after 2000, only two, ŁSTS and SWOTP (an extended version of the former), undertake the task of providing etymological information for the new contact-induced senses, though the information is not quite explicit. For instance, sense number 4. of the noun *aplikacja* is remarked on in the following way:

(9) **aplikacja** 1. ('...'); 2. ('...'); 3. *w informatyce* (E. *inform. tech.*) 'komputerowy program użytkowy ...' (E. computer application program ...); 4. *w środowisku biznesowym* (E. *in business*) 'podanie, prośba ...' (E. application, request) UWAGA: używanie wyrazu w tym znaczeniu jest przejawem bezkrytycznej mody na angielszczyznę … (E. ATTENTION: using this word in this [4th] sense is an indication of an uncritical fashion to use English) [SWOTP, 40]

Similarly implicit etymological information is found in NSEJP:

(10) **molestować** 1. ('…'); 2. za sprawą skandalu obyczajowego spowodowanego przez prezydenta światowego mocarstwa pojawiło się w 1995 r. nowe znaczenie czasownika *molestować* 'wykorzystywać seksualnie' (E. the new sense of the verb *to molest* 'harrass sb sexually' appeared [in Polish] in 1995 due to a moral scandal caused by the president of a world superpower) [NSEJP, 359].

Stating explicitly and neutrally the source language of the new sense would be more informative and certainly less space-consuming. Regardless of the unsatisfactory manner in which the etymological information is provided in the quoted examples (9, 10), this practice is not performed consistently throughout these dictionaries.

The problem of providing etymological information for semantic loans is avoided in the case of earlier English loanwords, e.g. P. *film* (< E. *film*) whose new sense 'coat, layer' is also English-sourced, or P. *surfować* (< E. *to surf*) and its new sense 'to look up information on the Internet'. Here, no additional etymological information is necessary.

Finally, it ought to be added, without hope or agenda, that an ideal dictionary entry that lists a semantic loan should also include information in which of the varieties of English, British, American or any other, the new sense is sourced. As much as it seems mission impossible, it may be argued that in some cases of recent English semantic loans in Polish this task appears feasible, as in the new senses of the following well-established Latin loanwords: P. *segregacja* (*rasowa*) 'separation of humans into racial groups' (< E. (*racial*) *segregation*), P. *administracja* 'government' (< E. *administration*), P. *konwencja* 'a formal gathering of people for political purposes' (E. *convention* (*of a political party*)). All the new senses quoted here were imported from American English and may be termed 'cultural semantic loans' or 'culture-specific semantic loans' [Witalisz 2006][10].

5. Conclusion

The examples quoted above illustrate two major problems in the rendering of semantic loans in contemporary Polish dictionaries. Firstly, semantic innovations that appear in Polish as a result of English linguistic influence are differently interpreted, which is manifested in contradictory dictionary entries defining the same innovation. Generally, there is lack of agreement among lexicographers as to whether a particular semantic innovation is a semantic loan, a loanword, or an independent semantic development (e.g. *cf.* P. *promocja* in its new senses: in (2b) and (2c) – semantic loan; in (4a) and (4b) – a separate loanword; in (3a) – both). Secondly, avoiding etymological information is a common practice.

The main objective of this paper was to advocate the inclusion of (more) reliable etymological information in dictionary entries, which in the case of the linguistic material discussed, is not impossible to obtain. Dictionaries indirectly record the history of a language, which is a witness to and a recorder of the historical and social events of the whole community. It is the lexicographer's obligation to record those changes in the form of accurate dictionary information. Providing detailed etymological information would not be a new practice in lexicography. If, in pursue of precision, some lexicographers undertake the task of showing the whole way a loanword has travelled before entering Polish by providing both its direct and etymological sources (plus any inter-

[10] *Cf.* Bloomfield [1933: 444] on the notion of 'cultural borrowings'.

mediary links) [MSJP, SJPSz, SWOR], why not add etymological information concerning semantic neologisms which appear due to foreign influence. Anyway, providing separate etymological information for the (loan)word and its new sense(s) (in cases in which there are two different sources) would be more beneficial for students of language than enumerating all the languages a word has gone through before entering Polish. In this way, a dictionary would register the acts of borrowing of a foreign word or sense, which result from foreign linguistic influence, which, in turn, reflects cultural contacts [*cf.* Walczak 1989: 90]. Depriving semantic loans of etymological informational creates gaps in dictionary entries which in a few years' time may be hard to fulfill in view of the processes of linguistic internationalization and globalization.

The example of contact-induced semantic innovations in Polish shows just one of many possibilities of how theoretical studies may be at the service of practical applicability. Thus the expected outcome of linguistic research including the theoretical description and the empirical analysis of semantic loans, is, in this very case, the possibility of practical application of its theoretical findings in lexicographic works, i.e. providing appropriate, research-based etymological information in dictionaries.

References

Bloomfield, Leonard [1933], *Language*. New York: Holt, Rinehart and Winston.

Fisiak, Jacek [1962], *Złożony kontakt językowy w procesie zapożyczania z języka angielskiego do polskiego*. Język Polski XLII(4), 286–294.

Hope, T. E. [1960], *The analysis of semantic borrowing*. In: Dubois, E. T. *et al.* (eds.), Essays Presented to C. M. Girdlestone. Newcastle upon Tyne: University of Durham, King's College, 125–141.

Haugen, Einar [1950], *The analysis of linguistic borrowing*. Language 26(2), 210–231.

Jadacka, Hanna [1990], *Z problemów opisu leksykograficznego haseł wieloznacznych*. Poradnik Językowy 1990(6), 436–442.

Jadacka, Hanna [2001], *System słowotwórczy współczesnej polszczyzny (1945–2000)*. Warszawa: Wydawnictwo PWN.

Kurkowska, Halina [1965], *O typach zróżnicowania znaczeniowego wyrazów*. Prace Filologiczne XVIII(4), 65–71.

Kurkowska, Halina [1976], *Zapożyczenia semantyczne we współczesnej polszczyźnie*. In: Magnuszewski, Józef (ed.), Z problemów współczesnych języków i literatur słowiańskich. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego, 99–109.

Leech, G. [1974], *Semantics*. Harmondsworth: Penguin.

Lipka, Leonhard [1986], *Homonymie, Polysemie oder Ableitung im heutigen Englisch*. Zeitschrift für Anglistik und Amerikanistik: ZAA; A Quarterly of Language, Literature and Culture 34, 128–138.

Lipka, Leonhard [2002], *English Lexicology. Lexical Structure, Word Semantics & Word-formation*. Tübingen: Gunter Narr Verlag Tübingen.

Lyons, John [1968], *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Lyons, John [1977], *Semantics*. Volume 2. Cambridge: Cambridge University Press.

Mańczak-Wohlfeld, Elżbieta [1995], *Tendencje rozwojowe współczesnych zapożyczeń angielskich w języku polskim*. Kraków: Universitas.

Markowski, Andrzej [2000], *Jawne i ukryte nowsze zapożyczenia leksykalne w mediach*. In: Bralczyk, Jerzy / Mosiołek-Kłosińska, Katarzyna (eds.), Język w mediach masowych. Warszawa: Upowszechnianie Nauki – Oświata "UN-O", 96–111.

Markowski, Andrzej [2005], *Kultura języka polskiego. Teoria. Zagadnienia leksykalne*. Warszawa: Wydawnictwo Naukowe PWN.

Walczak, Bogdan [1984], *Rola desygnatów i realiów w badaniach etymologicznych nad zapożyczeniami*. In: Język – Teoria – Dydaktyka. Materiały VI Konferencji Młodych Językoznawców-Dydaktyków, Katowice 12–14.11. 1981. Kielce: Wyższa Szkoła Pedagogiczna, 149–164.

Walczak, Bogdan [1989], *Objaśnienia etymologiczne w powojenych słownikach polskich*. In: Saloni, Zygmunt (ed.), Studia z polskiej lesykografii współczesnej III. Białystok: Dział Wydawnictw Filii UW *w Białymstoku*, 73–98.

Witalisz, Alicja [2006], *English linguistic influence on Polish and other Slavonic languages*. Quaderni del Centro di Studi Linguistico-Culturali. Available at: http://amsacta.cib.unibo.it/ 918/1/OCCASIONALp.pdf.

Witalisz, Alicja [2007a], *Anglosemantyzmy w języku polskim – ze słownikiem* [English Semantic Loans in Polish – with a Dictionary]. Kraków: Tertium.

Witalisz, Alicja [2007b], *Some remarks on hidden Anglicisms*. Paper presented at 38th Poznań Linguistic Meeting 2007. Available at: http://ifa.amu.edu.pl/pl/files/Presentations/PLM2007_Presentation_ Witalisz.pdf/.

Witalisz, Alicja [in press], *A reconsideration of the notions of semantic loan and loan translation in the analysis of English linguistic influence on Polish*.

Dictionaries

CCELD – *Collins Cobuild English Language Dictionary*, 1987.

CIDE – *Cambridge International Dictionary of English*, 1995.

ISJP – *Inny słownik języka polskiego*, M. Bańko (ed.), vol. I–II, 2000.

LAAD – *Longman Advanced American Dictionary*, 2002.

LDCE1 – *Longman Dictionary of Contemporary English*, 1990.

LDCE2 – *Longman Dictionary of Contemporary English*, 1995.

LDELC – *Longman Dictionary of English Language and Culture*, 2000.

ŁSTS – *Łatwy słownik trudnych słów*, A. Markowski (ed.), 2001.

MSJP – *Mały słownik języka polskiego*, S. Skorupka, H. Auderska, Z. Łempicka (eds.), 1968.

NODE – *The New Oxford Dictionary of English*, 1998.

NSEJP – *Nowy słownik etymologiczny języka polskiego*, K. Długosz-Kurczabowa, 2003.

NSWO – *Nowy słownik wyrazów obcych*, A. Latusek (ed.), 2003.

OALED – *Oxford Advanced Learner's Encyclopedic Dictionary*, 1998.
OCDEE – *Oxford Concise Dictionary of English Etymology*, 1993.
SJPD – *Słownik języka polskiego*, W. Doroszewski (ed.), vol. I–XI, 1958–1969.
SJPSz – *Słownik języka polskiego*, M. Szymczak (ed.), vol. I–III, 1978–1981.
STPS – *Słownik 100 tysięcy potrzebnych słów*, J. Bralczyk, (ed.) 2005.
SWOK – *Słownik wyrazów obcych i zwrotów obcojęzycznych*, W. Kopaliński (ed.), 4th ed., 1968.
SWOR – *Słownik wyrazów obcych*, Z. Rysiewicz (ed.), 12th ed., 1967.
SWOT – *Słownik wyrazów obcych PWN*, J. Tokarski (ed.), 1971.
SWOTP – *Słownik wyrazów obcych i trudnych*, R. Pawelec (ed.), 2003.
SWOZOA – *Słownik wyrazów obcych i zwrotów obcojęzycznych z almanachem*, W. Kopaliński, 2000
USJP – *Uniwersalny słownik języka polskiego*, S. Dubisz (ed.), vol. I–IV, 2003.
WCDIE – *Webster Comprehensive Dictionary*. International Edition, 1998.
WSEH – *Wielki słownik etymologiczno-historyczny języka polskiego*, K. Długosz-Kurczabowa, 2008.
WSJP – *Współczesny słownik języka polskiego*, B. Dunaj (ed.), 2007.
WSJPP – *Wielki słownik języka polskiego*, E. Polański (ed.), 2009.

Solomija Buk*

THE EPITHETIZATION INDEX IN A WORK OF FICTION (ON THE BASIS OF THE TEXT CORPUS OF IVAN FRANKO'S LONG PROSE FICTION)

1. Introduction

Due to modern trends of applied linguistics, statistical linguistics in particular, every text has its own statistical parameters [Köhler, Altmann 2005: 16], such as the number of word-forms and words, text richness, index of nominalization, index of verbal definitions etc.. Linguists consider proportions of parts of speech (PoS) in text, especially the correlation of nouns and adjectives, as relevant style characteristics.

Special attention to nouns in text was given even in ancient times by Plato and Aristotle. In the tractate *Poetics,* two main PoS were distinguished: noun and verb. A noun can occur as the main word in a clause, phrase or sentence. Currently psychologists and sociologists take it into consideration when characterizing cultures, nations, and languages. The adjective plays an important role in text and style too. It gives an attribute and description to the noun, providing new information about it. In the literature it is one of the most important stylistic tropes of poetic speech. Often in the special literature, one can find statements like “This author is more poetic, s/he uses more epithets”, “In this text there are more epithets”, “This writer likes epithets”, etc. Such opinions are not objective unless scrupulous study is made. To measure this notion, the epithetization index was introduced. It is the relation between the number of nouns and the number of adjectives in a given text. This index shows how many nouns occur per one adjective; the lower the value, the more epithets are in the text.

Of course from this point of view the “epithet” is understood in a wide semantic (not syntactic) meaning, because not all adjectives play the role of epithets in the strictest understanding of this term. But all of them give the infor-

* Department for General Linguistics, Ivan Franko National University of Lviv.

mation which an epithet provides: characteristics of an object. For instance, in the sentence *ukr*. `Divchyna bula rozumna` [The girl was wise], "rozumna" syntactically is a part of the predicate, but semantically "rozumna" characterizes the girl.

Epithetization index is one of the most interesting characteristics of text. According to [Ruszkowski 2004: 52], the value of this parameter has a huge significance for stylistic identification of style and level of its officiality. It can distinguish the style of separate texts, of writers, of literature types and genres as well as of functional styles. It can also suggest the age of the person, whose speech is under consideration.

The article presents, for the first time, a study of the epithetization index for the complete works of long prose fiction by Ivan Franko. The presented analysis is a part of a larger project of complex liguostatistical and quantitative study of the long prose fiction by Ivan Franko. The main aim is to investigate author's style peculiarities by quantitative methods using text corpus [Buk 2011b].

Ivan Franko (1856–1916) lived in Western Ukraine during the Austro--Hungarian Empire. He is the national Ukrainian figure of encyclopedic knowledge: a poet, writer, translator, ethnographer, social and literary critic, journalist, economist, political activist, doctor of philosophy, as well as author of the first detective novels and modern poetry in the Ukrainian language. Along with Taras Shevchenko, he has had a tremendous impact on modern literary and political thought in Ukraine. The modern study of his heritage from different aspects is an important part of Ukrainian national research.

It this work, the epithetization index is calculated for all nine of Franko's Ukrainian novels: *Boa constrictor* (1884), *Boryslav smijet'sja / Boryslav Laughs* (1881), *Zakhar Berkut* (1882), *Ne spytavshy brodu / Without Asking a Wade* (1886), *Dlja Domashnjoho Ohnyshcha / For the Hearth* (1892), *Osnovy suspil'nosti / Pillars of Society* (1895), *Perekhresni stezhky / The Cross-Paths* (1900), *Velykyj Shum / Great Noise* (1908), *Petriji j Dovbushchuky / Petrijs and Dovbushchuks* (1912, second edition).

The present article has the following structure. In Section 2, the proportions of PoS in linguistic studies are overviewed. Section 3 describes the Ivan Franko electronic text corpus and its markup organization, and analysis of direct and author's speech, together with morphological characteristics of words. In Section 4 the statistical features of PoS in Franko's novels are presented, with special attention paid to nouns and adjectives. The epithetization index in Franko's works, with comparison to average values for Ukrainian prose fiction, is interpreted in Section 5. Section 6 contains conclusions and research prospects.

2. Proportions of parts of speech in linguistics studies: short overview

The quantity of PoS, with respect to proportion in different texts poses an interesting problem in modern linguistics. It is a multidimensional notion. Every PoS has its own generalized meaning; it denominates the human categorization of the world. That is why the number of PoS, their distribution and correlation can be unique characteristics of culture, language, styles, genres, authors, chronological time, etc.

American social psychologist R. Nisbett compared dominant PoS among the Western and Eastern civilizations, to explore the relationships between language and thought. He concluded that in West-European languages, nouns are more frequently used than in Eastern languages (Japanese, Chinese, Indian, etc.), where verbs prevail [Nisbett 2009: 69–88, 105–122]. "This difference starts out in childhood, where studies prove that Western children learn nouns faster than verbs, and Asian children learn verbs faster than Western children" [Eigenauer 2006: 7].

The quantitative analysis of PoS and their percentage for different styles and genres was made on the material of different languages, in particular English [Potapenko 1999; Bowen 2005; Barrett *et al.* 2007; Leech *et al.* 2001], Japanese [Masamitsu 2005], Polish [Kamińska-Szmaj 1988, 1989; Laskowska 1989; Rachwałowa 1986; Ruszkowski 2004; Skubalanka 1986: 141–155; Zarębina 1983, 1985; Zgółkowa 1987], Russian [Klimova 2004, Shajkievich *et al.* 2003], and Ukrainian [Perebyjnis 1967; Tyshchenko 1970]. For example, Masamitsu cited two formulations for estimating the proportion of PoS. Both laws related the occurrence of nouns and other PoS in Japanese text, and demonstrated the existence of a linear dependence between them [Masamitsu 2005: 90]. A fragment in [Barrett *et al*., Greenberg, Schwartz 2007] was devoted to a comparison of PoS tag densities in seven English documents hand-selected from four different domains (medical, financial, military, and narrative). S. Krzyśka in his work [Krzyśka 2008] explored the qualitative and quantitative distribution of different PoS in children's literature in comparison to oral literature (folklore song, dialect story), written language (press, modern poetry), and oral language (school, children). The author approved the statement that nouns require other PoS in text, especially adjectives.

Many linguists have pointed out the importance of quantitative and statistical methods in the analysis of particular writers (see, e.g. [Puzynina 1988: 445–454; Korpysz 2010: 55]). Many studies of writers' dictionaries contain specific information about PoS, concluding that PoS play a dominant role in stylistics and determination of the author's style: K. Čapek [Čermák 2007], M. Pavić [Vasić 1998], F. Dostojevskij [Shajkievich *et al.* 2003], J. Przyboś [Zębaty-Michałakowa 1982], C. Norwid [Puzynina, Korpysz 2008–2012], I. Franko [Buk 2011a], etc.

In light of the studies cited above, the proportions of PoS, and in particular nouns and adjectives, are current topics in different fields of linguistics. They can be considered properties of an individual author's style, as well as a characteristics of a specific work [Perebyjnis *et al.* 1985: 152].

3. The Ivan Franko Text Corpus and its markup

Classes of text corpora include not only the corpora of national languages but also those of specific writers. Writers of national importance often have such a corpus: W. Shakespeare, J. Joyce, K. Čapek, M. Pavić, F. Dostojevskij, H. Skovoroda, etc.

The Corpus of Ivan Franko's long prose fiction is currently being created at the Ivan Franko National University of Lviv. It is the first stage of a larger project of the I. Franko text corpus. A comprehensive statistical description of Franko's PoS proportion is planned as one of the project outcomes [Buk 2007, 2011b]. Due to the classical approach [see McEnery *et al.* 2006: 23; Meyer 2002: 81; Aston, Burnard 1998: 24; Shyrokov *et al.* 2005: 207–208; Dems'ka--Kul'chyc'ka 2005: 152–168], the Franko text corpus has three levels of markup: external, structural and internal.

Information about the bibliographic description of works, the education and origin of the author, the time, duration, and place of writing, etc. is coded in the external markup of the texts.

The structural annotation contains the markup of text divisions (sections, paragraphs, sentences, author's definition of genre, dedications, epigraphs), author's notes about remaking, citations, author's signature of work (date and/or place of the work creation, original name or pseudonym), author's and editor's footnotes, poetic insertions, information about the direct and author's speech etc. [Buk 2009]. Comparison of direct and author's speech is a useful instrument in text studies. These two types of speech have diverse rations among works of Franko which correlates with the observed variation in PoS.

The internal markup of the Franko corpus includes semantic, morphological, syntactic, and anaphoric data. In this work the morphological markup is the most important for the topic under consideration. Lemmatization as part of a morphological markup is necessary for all Slavic languages because of the multiplicity of inflectional forms. In the process of lemmatization, every token (word-form) was given a standard form. In the Franko corpus, the standard approach to PoS separation for Slavic languages was followed. So there are 10 PoS: noun, verb, adjective, adverb, pronoun, numeral, conjunction, particle, preposition, and interjection. They function differently in text and in the dictionary.

The homonyms are disambiguated in the corpus by applying contextual analysis. For instance, *або* (conj., particle), *біг* (noun, verb), *волів* (noun, verb),

горі (noun, adv.), дні (noun Dat. from *дно*, noun pl. Nom. *день*), *де* (adv., particle), *доктор* (medical doctor and academic degree), *зимою* (noun, adv.), *її, його, їх* (pers. and posess. pron.), *коло* (noun, prep.), *корч* (a plant, a spasm), *коса* (a braid, a scythe), *круг* (noun, adv.), *лютий* (noun, adj.), *ніж* (noun, particle), *палати* (noun, verb), *перед* (noun, prep.), *поверх* (noun, adv.), *попасти* (to get and to shepherd), *раз* (noun, adj., conj.), *святий* (noun, adj.), *чому* (adv., pron.), *шкода* (noun, predic. word), and others. In the corpus, the phonetic variants (being mostly euphonic alternations) are joined with the basic forms: *б/би/би-м*; *в/у*; *бачитися/бачитись*; *вбити/убити*; *вважатись/уважатися*; *весь/ввесь/увесь*; *лякатися/лякатись,* etc. This process is required to obtain a clear picture of PoS in the text.

The morphological markup provides the opportunity to obtain rich information about word-forms, lexemes (lemmas), PoS, etc. In this manner, all nine Ukrainian novels by Franko are described: *Boa constrictor, Boryslav smijetsja* [*Boryslav Laughs*], *Zakhar Berkut*, *Ne spytavshy brodu* [*Without Asking a Wade*], *Dlja domašnjoho ohnyšča* [*For the Hearth*], *Osnovy suspil'nosty* [*Pillars of Society*], Velykyj *šum* [*The Great Noise*], *Petriji j Dovbushchuky* [Buk 2010a–f], as well as *Perekhresni stežky* [*The Cross-Paths*] [Buk, Rovenchak 2007a,b, 2006–2012]. There are over 370 500 tokens and 22 600 lemmas in total.

4. Nouns vs. Adjectives in the Long Prose by Ivan Franko

For quantitative processing of the long prose fiction by Franko [Buk 2011b], a scheme for the statistical description of text properties was developed. The following parameters were calculated: text size (the number of tokens, N), the total number of lemmas (V) in the vocabulary (i.e., the list of different lemmas), the index of variety (V/N), the mean repetition of a word in the text (N/V), the number of hapax legomena (i.e. words occurring only once in the text), the indicator for vocabulary variability, i.e. exclusiveness index for the text (V_1/N) and for the dictionary (V_1/V), the number of words with frequency higher than 9 in the text ($V_{10,\mathrm{T}}$) and in the vocabulary (V_{10}), and concentration indexes for the text ($V_{10,\mathrm{T}}/N$) and vocabulary (V_{10}/V). The proportions of PoS, especially nouns and adjectives, have significant potential to characterize the author's style.

In an average Ukrainian text, nouns are the most frequently used semantic PoS; they constitute 25,8% of the text. The next most common are verbs (18,3%), followed by adjectives (8,6%) [Perebyjnis 1985: 157].

Given the morphological markup of the Franko text corpus, the frequency of nouns and adjectives could be automatically determined in the vocabulary and text. It was previously shown [Buk 2011a] on the example of six novels by I. Franko, that all PoS form three conditional groups: 1) with higher activity (frequency) in the text than in the dictionary (synsemantic PoS, pronouns),

2) vice versa: with higher activity in the dictionary than in the text (nouns, verbs, adjectives) and 3) with approximately the same activity in both text and dictionary (adverbs, numerals). In this classification nouns and adjectives have the same type of behavior and demonstrate the highest variety they belong to the same group. For these PoS, the relative numbers in the vocabulary exceed the proportions in the text. Nouns occupy on average about 22% of the text and 33% of the vocabulary, whether the proportion of adjectives is three times higher in the vocabulary than in the text: 19% versus 8%. Thus, the latter group is primarily responsible for the relative richness of the vocabulary.

It is thanks for this second group of PoS wins.

Table 1 shows noun and adjective distribution in the text and vocabulary of Franko's long prose. It is supplemented by three more novels (compared to the one proposed in [Buk 2011a]): *Boa constrictor, Ne spytavshy brodu [Without Asking a Wade], Petriji j Dovbushchuky* and represents the complete works of long prose by I. Franko. It demonstrates the stated correlations.

NOVELS	**NOUNS**		**ADJECTIVES**	
	text	**dictionary**	**text**	**dictionary**
Boa constrictor	24.64%	31.52%	9.35%	18.95%
***Boryslav smijetsja* [*Boryslav Laughs*]**	**21.30**%	**30.80**%	7.00%	18.40%
Zakhar Berkut	**26.30**%	31.00%	**9.80**%	20.80%
Ne spytavshy brodu [*Without Asking a Wade*]	22.29%	33.38%	7.91%	19.08%
***Dlja domašnjoho ohnyšča* [*For the Hearth*]**	22.80%	33.30%	**7.80**%	20.30%
***Osnovy suspil'nosty* [*Pillars of Society*]**	23.50%	33.90%	7.20%	18.70%
***Perekhresni stežky* [*The Cross-Paths*]**	23.60%	**36.60**%	6.90%	**17.90**%
***Velykyj šum* [*The Great Noise*]**	24.40%	**36.90**%	7.50%	**17.60**%
Petriji j Dovbushchuky	24.50%	32.20%	8.90%	**18.00**%
Long Prose Fiction by Ivan Franko	23.70%	33.90%	8.04%	16.90%
Ukrainian Long Prose Fiction	25.8%		8.6%	

Table 1: Noun and adjective distribution in the text and dictionary of Franko's works, Franko's long prose, and Ukrainian mid-twentieth century fiction[1]

[1] The data are from [Perebyjnis 1985: 157].

After adding the abovementioned three novels to the text corpus we observed a clearer picture of these morphological classes. Nouns make up 23.7% of the text and almost 34% of the dictionary, whereas adjectives cover 8.04% of the text and almost 17% of the dictionary.

The largest number of nouns is in the text of *Zakhar Berkut* (26.3%), which is even more than in average Ukrainian prose (25.8 %). The smallest sum of nouns is in *Boryslav smijetsja* (21.3%). In the dictionary, the highest abundance of nouns is exhibited in *Velykyj šum* (36.9%) and *Perekhresni stežky* (36.6%), and the lowest in *Boryslav smijetsja* (30.8%).

Adjectives are most prevalent in the text of *Zakhar Berkut* (9.8%) and *Boa constirctor* (9.35%). The lowest proportion is in *Perekhresni stežky* (6.9%). In the dictionary *Zakhar Berkut* (20.8%) and *Dlja domašnjoho ohnyšča* (20.3%) have the most adjectives, while *Perekhresni stežky* (17.9%) and *Velykyj šum* (17.6%) have the least.

In summery, *Perekhresni stežky* and *Velykyj šum* display the greatest variety of both nouns and adjectives in the dictionary; *Zakhar Berkut* has the most in the text. *Boryslav smijetsja* has the lowest proportion of nouns in the text and dictionary. On average the long prose fiction by Ivan Franko has a lower percentage of both nouns (23.7%) and adjectives (8.04%) than Ukrainian long prose fiction in general (25.8% and 8.6% respectively).

5. Index of epithetization in Franko's works

The quantitative relations between PoS are known as an important element of statistical text characteristics. Apart from index of epithetization, there are the index of verbal definitions (relation of the total adverb occurrences to the total adjective occurrences), and the level of nominalization (relation of the total noun occurrences to the total verb occurrences) [Ruszkowski 2004: 50; Kamińska-Szmaj 1988: 128]. In Table 2, the epithetization index of each Franko novel is compared to average indexes for Franko's long prose, as well as for Ukrainian fiction of the mid-twentieth century.

NOVELS	Epithetization index in text	Epith. index in author's speech	Epith. index in direct speech
Boa constrictor	2.66	2.51	5.03
***Boryslav smijetsja* [*Boryslav Laughs*]**	3.04	2.79	5.04
Zakhar Berkut	2.69	2.51	3.31
Ne spytavshy brodu [*Without Asking a Wade*]	2.82	2.71	3.53

***Dlja domašnjoho ohnyšča* [*For the Hearth*]**	2.92	2.66	3.63
***Osnovy suspil'nosty* [*Pillars of Society*]**	3.28	3.21	3.48
***Perekhresni stežky* [*The Cross-Paths*]**	3.45	3.20	3.81
***Velykyj šum* [*The Great Noise*]**	3.25	3.15	3.43
Petriji j Dovbushchuky	2.74	2.60	3.36
Long Prose Fiction by Ivan Franko	3.05	2.82	3.72
Ukrainian Long Prose Fiction	3.00		

Table 2: Quantitative relationship between nouns and adjectives (Epithetization Index) in Franko's works, Franko's long prose, and Ukrainian mid-twentieth century fiction

The highest indexes are observed in *The Cross-Paths* (3.45), *Osnovy suspil'nosty* (3.28), and *Velykyj šum* (3.25). These works have the largest ratio of nouns per epithet/adjective. The lowest values are in *Boa constrictor* (2.66), *Zakhar Berkut* (2.69), and *Petriji j Dovbushchuky* (2.74). In these novels, the author used the largest number of epithets/adjectives per noun. In long prose fiction by Ivan Franko overall, the average epithetization index is 3.05 (it is obtained from bulk data and is thus somewhat different from the average which might be obtained from Table 2, due to differences in text size of the novels). It is a bit higher than in Ukrainian long prose fiction in general (3.00); thus Franko used slightly less epithets than are normally found in Ukrainian literature.

Another aspect is to compare these figures in author's and direct speech. From the observed data, it is clear that the direct speech has significantly more nouns per adjective than author's speech: 3.72 versus 2.82. The greatest difference between them is observed in *Boa constrictor* (5.03 : 2.51) and *Boryslav smijetsja* [*Boryslav Laughs*] (5.04 : 2.79). The most homogeneous allocation of nouns and epithets is in *Osnovy suspil'nosty* [*Pillars of Society*] (3.48 : 3.21), *Perekhresni stežky* [*The Cross-Paths*] (3.81 : 3.20) and *Velykyj šum* [*The Great Noise*] (3.43 : 3.15). This indication can be attributed to the varying ratio of authors to direct speech, but not in every case: *Osnovy suspil'nosty* [*Pillars of Society*], *Perekhresni stežky* [*The Cross-Paths*], and *Velykyj šum* are highly dialogical novels, but roughly the same correspondence is found in *Dlja domašnjoho ohnyšča* [*For the Hearth*], Ne spytavshy brodu [*Without Asking a Wade*], *Boryslav smijetsja* [*Boryslav Laughs*], as is shown in Figure 1.

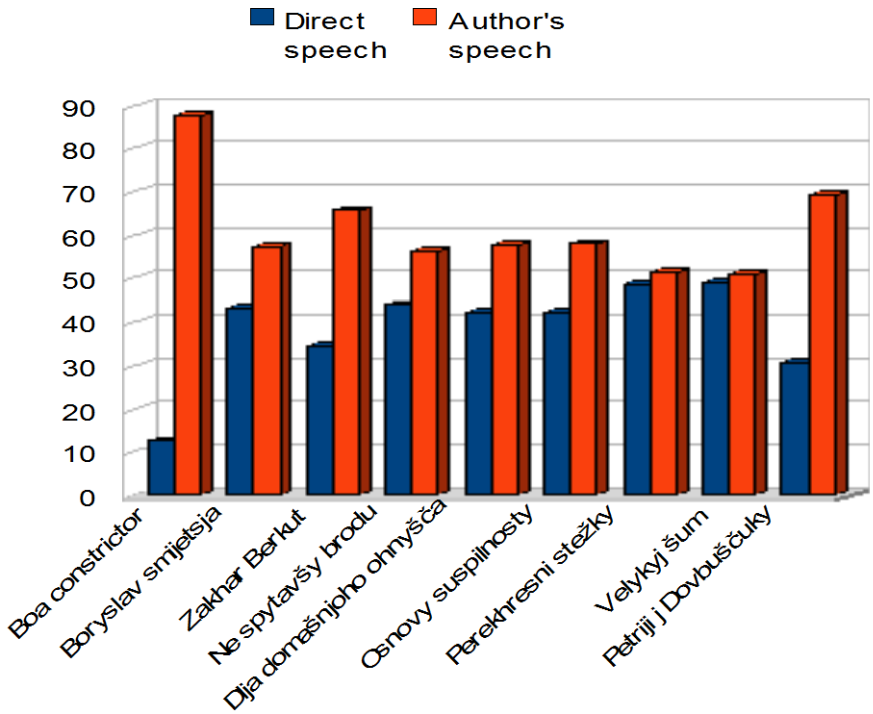

Fig. 1. The correlation between direct and author's speech in the novels by Ivan Franko

Boa constrictor and *Petriji j Dovbushchuky* have comparatively less dialogues, but the epithetization index relation in their direct and indirect speech is different: 5.03 : 2.51 and 3.36 : 2.60, respectively. So, these dependences need a deeper study.

In Ukrainian long prose fiction in general, the fraction of direct speech is 28.2% of tokens [Perebyjnis 1981: 10], while in Franko's prose it is 40.6% (more than 1.5 times more). The epithetization index is approximately the same for Franko's and Ukrainian prose (close to 3.0), the writer still used more epithets per noun in author's speech than normally found in Ukrainian literature. Indeed, as the index is much higher in the direct speech; author's speech must balance it to obtain nearly the same mean value. From this point of view, Franko's prose is more poetic; he gives more attributions and descriptions to the object of description, more information about it.

Although the index of epithetization is only one of many instruments used for the stylistic analysis of text, it can be considered as a parameter to complement qualitative text analysis.

6. Conclusions and research perspectives

Thanks to the morphological markup in the text corpus of Ivan Franko's long prose fiction (total size over 370 500 tokens and 22 600 lemmas) we had the possibility to investigate important aspect of the author's style peculiarity – proportion of PoS separately for every novel, as well as for Franko's long prose fiction in general. For these PoS, the relative numbers in the vocabulary exceed the proportions in the text. Nouns occupy on average about 24% of the text and 34% of the vocabulary, the proportion of adjectives is almost three times higher in the vocabulary than in the text: 17% versus 8%. So, the vocabulary richness of Franko's works largely depends on morphological classes.

At the same time, these PoS have unique behavior in each Franko's works: *Perekhresni stežky* and *Velykyj šum* display the highest variety of both nouns (36.6% and 36.9%, respectively) and adjectives (17.6% and 17.9%, respectively) in the dictionary in the comparison to other novels, whereas *Zakhar Berkut* has the same in the text (26.3% and 9.8%). *Boryslav smijetsja* has the least use of nouns in the text and dictionary (21.3% and 30.8%). Summarizing, on average, long prose fiction by Ivan Franko has a lower percentage of both nouns (23.7%) and adjectives (8.04%) than Ukrainian long prose fiction in general (25.8% and 8.6%).

The epithetization index reflects this picture. Being the relation between the number of nouns and the number of adjectives in the text, this index shows how many nouns are present per adjective, the lower the value the more epithets are in the text. The lowest numbers of nouns per adjectives/epithets are in *Boa constrictor* there (2.66), *Zakhar Berkut* (2.69), and *Petriji j Dovbushchuky* (2.74), i.e. in these novels the author used the largest amount of epithets per noun. The highest index is in *The Cross-Paths* (3.45), *Osnovy suspil'nosty* (3.28), and *Velykyj šum* (3.25). In these works the largest quantity of nouns occur per one epithet/adjective.

In Franko’s novels, direct speech has significantly more nouns per adjective than author’s speech: 3.72 against 2.82. But there was no precise correlation found between the size of authors-direct speech correspondence and epithetization index.

The average value of the epithetization index in long prose fiction by Ivan Franko is 3.05. It is marginally higher than the respective parameter of texts written by other Ukrainian authors (the average values of Ukrainian prose fiction is 3.00). At first glance, it seems that Franko used slightly less epithets than is typical in Ukrainian literature. The difference however is not significant. Considering the fact that Franko's novels have 1.5 times more direct speech, he still used more epithets in author's descriptions than other Ukrainian writers.

This method of epithetization index comparison could also be applied to genre, register or authorship analysis. In future, it will be interesting to obtain this index also for Polish novels by Ivan Franko and compare the data. It will also be important to compare the obtained results with Polish language of Franko's times.

References

Aston, Guy / Burnard, Lou [1998], *The BNC Handbook. Exploring the British National Corpus with SARA.* Cambridge: Edinburgh University Press.

Barrett, Leslie / Greenberg, David / Schwartz, Mark [2007], *A Syntactic Feature Counting Methods of Selecting Machine Translation Training Corpora*. Language and Computers: Studies in Practical Linguistics 60(1), 1–19.

Bowen, Rhowen [2005], *Noun complementation in English: a corpus-based study of structural types and patterns*. Göteborg: Göteborg University.

Buk, Solomija [2007], *Ivan Franko Text Corpus: an Attempt to Determine Main Parameters*. In: Applied Linguistics and Linguistic Technologies: MegaLing 2006. Kyiv: Dovira, 72–82.

Buk, Solomija [2009], *Strukturne anotuvannja u korpusi tekstiv* (*na prykladi prozy Ivana Franka*). Ukrajins'ka mova 3, 59–71.

Buk, Solomija [2010a], *"Zakhar Berkut" Ivana Franka u svitli statystychnoji lingvistyky*. Naukovyj chasopys nacional'noho pedahohichnogo universytetu imeni M. P. Drahomanova. Serija 10. Problemy gramatyky i leksykolohiji ukrajins'koji movy 6, 172–176.

Buk, Solomija [2010b], *Kvantytatyvna parametryzacija tekstiv Ivana Franka*. In: Ivan Franko: Studiji ta materialy. Lviv: Vydavnychyj centr LNU imeni Ivana Franka (*in press*), http://arxiv.org/abs/1005.5466 .

Buk, Solomija [2010c], *Statystychni kharakterystyky romanu Ivana Franka "Osnovy suspil'nosti"* (*na osnovi chastotnogo slovnyka tvoru*). Visnyk: Problemy ukrajins'koji terminologiji 676, 90–93.

Buk, Solomija [2010d], *Interpretacija leksyky romanu Ivana Franka "Velykyj Shum" u kil'kisno-statystychnomu aspekti*. Studia Methodologica 30, 145–153.

Buk, Solomija [2010e], *Roman Ivana Franka "Dlja domashnjogo ognyshcha" kriz' pryzmu chastotnogo slovnyka.* Movoznavstvo. No. 4, 56–66; arXiv: 1006.0153v1 [cs.CL], http://arxiv.org/abs/1006.0153.

Buk, Solomija [2010f], *Statystychna struktura romanu I. Franka "Boryslav smijet'sja"*. Uchenyje zapiski Tavricheskogo nacional'nogo universiteta im. V. I. Vernadskogo. Serija "Filologija" 23(62), No. 3, 114–118.

Buk, Solomija [2011a], *Chastyny movy u slovnyku i teksti Ivana Franka* (*na materiali velykoji prozy*). Lingvistychni studiji 22, 62–66.

Buk, Solomija [2011b], *Lingvostatystychnyj opys "Ne spytavshy brodu" Ivana Franka*. Visnyk L'vivs'kogo nacional'nogo universytetu. Serija filologichna. Frankoznavstvo 55, 230–242.

Buk, Solomija [2012], *Chastotnyj slovnyk romanu Ivana Franka "Boa constrictor"*. In: Batsevych, F. S. (ed.), Stežkamy Frankovogo tekstu (komunikatyvni, ligvosemiotychni, kognityvni ta lingvostatystychni vymiry prozy). Lviv: Vydavnychyj centr LNU imeni Ivana Franka (*in press*).

Buk, Solomija / Rovenchak, Andrij [2007a], *Chastotnyj slovnyk povisti I. Franka "Perekhresni stežky"*. In: Stežkamy Frankovogo tekstu (komunikatyvni, stylistychni ta leksychni vymiry romanu "Perekhresni stežky"). Lviv, 138–369.

Buk, Solomija / Rovenchak, Andrij [2007b], *Statistical Parameters of Ivan Franko's Novel "Perekhresni stežky" ("The Cross-Paths")*. In: Grzybek, P. / Köhler, R. (eds.), Quantitative Linguistics. Vol. 62: Exact Methods in the Study of Language and Text: dedicated to Professor Gabriel Altmann on the occasion of his 75th birthday. Berlin–New York: Mouton de Gruyter, 39–48.

Buk, Solomija / Rovenchak, Andrij [2006–2012], *On-lajn konkordans romanu Ivana Franka "Perekhresni stežky"*, http://www.ktf.franko.lviv.ua/~andrij/science/Franko/concordance.html.

Čermák, František [2007], *Slovník Karla Čapka*. Praha: Nakladatelství Lidové noviny; Ústav Českého národního korpusu.

Dems'ka-Kul'chyc'ka, Orysja [2005], *Osnovy nacional'nogo korpusu ukrajins'koji movy.* Kyjiv: Instytut ukrajins'koji movy NAN Ukrajiny.

Eigenauer, John [2006], *Summary of The Geography of Thought*, http://www.eigenauer.com/GeographyOfThought.pdf.

Kamińska-Szmaj, Irena [1988], *Części mowy w słowniku i tekście pięciu stylów funkcjonalnych polszczyzny pisanej (na materiale słownika frekwencyjnego)*. Biuletyn Polskiego Towarzystwa Językoznawczego XLI: 127–136.

Kamińska-Szmaj, Irena [1989], *Słownictwo tekstów popularnonaukowych w ujęciu statystycznym*. Rozprawy Komisji Językowej. Wrocławskie Towarzystwo Naukowe XVI, 69–87.

Klimova, Larisa [2004], *Vysokochastotnyje imena sushchestvitel'nyje, prilagatel'nyje i chislitel'nyje v sovremennom russkom jazykie*. Dis. ... kand. filol. nauk: 10.02.01, Ekaterinburg.

Köhler, Reinhard / Altmann, Gabriel [2005], *Aims and Methods of Quantitative Linguistics*. In: Problems of Quantitative Linguistics: A Collection of Papers. Chernivtsi: Ruta, 12–42.

Korpysz, Tomasz [2010], *Słowniki języka autorów jako typ opracowań leksykograficznych*. Poradnik Językowy 4, 51–72.

Krzyśka, Sławomir [2008], *Statystyka języka wierszy dla dzieci*. In: Krasoń, Katarzyna / Mazepa-Domagała, Beata (eds.), Wyrazić i odnaleźć siebie, czyli o sztuce, ekspresji, edukacji i arteterapii, Wydawnictwo Górnośląskiej Wyższej Szkoły Pedagogicznej im. Kardynała Augusta Hlonda. Mysłowice, 271–279, http://www.staff.amu.edu.pl/~slawekk/statystyka_wierszy.pdf.

Laskowska, Elżbieta [1989], *Charakterystyka słownictwa nauczycieli i kandydatów na nauczycieli klas początkowych*. Bydgoszcz.

Leech, Geoffrey / Raygon, Paul / Wilson, Andrew (eds.) [2001], *Word Frequency in Written and Spoken English. Based on the British National Corpus.* Pearson Educational Ltd.

Masamitsu, Ilto [2005], *Quantitative linguistics in Japan*. In: Köhler, Reinhard / Altmann, Gabriel / Piotrowski, Rajmund (eds.), Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch – An International Handbook. Berlin–New York: de Gruyter, 82–95.

McEnery, Tony / Xiao, Richard / Tono, Yukio [2006], *Corpus-based language studies: an advanced resource book*. London: Routledge.

Meyer, Charles F. [2002], *English Corpus Linguistics: An Introduction.* Cambridge: Cambridge University Press.

Nisbett, Richard [2009], *Geografia myślenia*, przekł. E. Wojtych. Sopot: Smak Słowa.

Perebyjnis, Valentyna [1967], *Statystychni peremetry styliv*. Kyjiv: Naukova dumka.

Perebyjnis, Valentyna (ed.) [1981], *Chastotnyj slovnyk suchasnoji ukrajins'koji khudozhnjoji prozy*. In 2 vol. Kyjiv: Naukova dumka.

Perebyjnis, Valentyna / Muravyc'ka, Margaryta / Darchuk, Natalija [1985], *Chastotni slovnyky ta jikh vykorystannja.* Kyjiv: Naukova dumka.

Potapenko, Alla [1999], *Stylemetrychni harakterystyky prykmetnykiv riznoji morfemnoji struktury v suchasnij anglijs'kij movi*: Avtoref. dys... kand. filol. nauk: 10.02.04; Kyjiv. derzh. lingv.un-t. Kyjiv.

Puzynina, Jadwiga [1988], *O celach i metodach pracy nad słownikiem języka pisarzy*. Z polskich studiów slawistycznych. Seria: VII „Literatura, folklorystyka, problematyka historyczna", Warszawa.

Puzynina, Jadwiga / Korpysz, Tomasz [2008–2012], *Internetowy słownik języka Cypriana Norwida*. Pracownia Słownika Języka Norwida, Wydział Polonistyki, Uniwersytet Warszawski, Warszawa, http://www.slownikjezykanorwida.uw.edu.pl/index.php.

Rachwałowa, Maria [1986], *Słownictwo tekstów naukowych.* Wrocław: Ossolineum.

Ruszkowski, Mariusz [2004], *Wskaźnik epitetyzacji w badaniach stylistycznych*. Respectus Philologicus 5(10), 48–53.

Shajkievich, Anatolij / Andrjushchenko, V. M. / Rebeckaja, N. A. [2003], *Statisticheskij slovar' Dostojevskogo*. Moskva: Jaz. slav. kul'tury.

Shyrokov, V. / Buhakov, O. / Grjaznuhina T. (eds.) [2005], *Korpusna lingwistyka*. Kyjiv: Dovira.

Skubalanka, Teresa [1986], *Założenia gramatyki stylistycznej*. In: Typy opisów gramatycznych języka. Wrocław, 141–155.

Tyshchenko, V. [1970], *Chastota chastyn movy v riznych funkcional'nych styljah suchasnoji ukrajins'koji movy*. Pytannja strukturnoji leksykolohiji. Kyjiv: Naukova dumka, 215–224.

Vasić, Smiljka [1998], *Polazne osnove novije srpske proze. Knjiga II. Hazarski rečnik Milorada Pavića: frekvencijski rečnik*. Beograd: Kultura.

Zarębina, Maria [1973], *Statystyczna struktura tekstu gwarowego* (*na przykładzie wsi Ząb w powiecie nowotarskim*). Język Polski LIII(1), 5–16.

Zarębina, Maria [1985], *Próba statystycznej analizy słownictwa polszczyzny mówionej* (*synteza danych liczbowych*). Wrocław.

Zembaty-Michalakowa, Maria [1982], *Poezja Juliana Przybosia w świetle badań statystyczno-językowych na tle porównawczym*, Wrocław: Wydawn. Uniwersytetu Wrocławskiego.

Zgółkowa, Halina [1987], *Ilościowa charakterystyka słownictwa współczesnej polszczyzny. Wybrane zagadnienia*. Poznań: UAM.

Natalia Zborowska*

THE LANGUAGE OF POLISH COMPUTER GAME PLAYERS (FOR EXAMPLE OF 'WORLD OF WARCRAFT' USERS). RECONNAISSANCE

Games become the context for many other media and various cultural activities.
Wiesław Godzic [2006: 11]

The very appearance of the medium does not change the world.
The media are communication – that's all. And it's a lot.
Mirosław Filiciak [2006: 194]

Games as a subject of study in Poland

Computer games, created in the 1970s were initially regarded as an unsophisticated entertainment for children and the youth. However, as it was noted in the opening sentence of this text, they are becoming an important medium not only for their users, but also for researchers, including these from Poland, who for several years have dealt with ludology as a matter of their scientific research. The Games Research Association of Poland, founded in 2004, tries to promote this different perspective of perceiving games [the history, objectives: Surdyk 2007]. Its members emphasize the importance of this phenomenon, for example one of its founders says:

> do not underestimate the phenomenon of computer games [...] because of low public awareness and incomplete, or even rudimentary knowledge about the advantages (but also potential risks) of computer games, or more broadly speaking electronics. They are underestimated and categorized by laymen as a phenomenon of little relevance, frivolous or even malicious or evil. The controversy surrounding various titles, which due to their medially attractive character distract attention from the less shocking information about the use of educational and didactic games, contributes to the spread of this view on games. [...]. Rarely, and rather in

* Doctoral Studies, Institute of Polish Philology, Pedagogical University of Cracow.

> a form of curiosity, do the media provide information about the positive effects of games (not just computer ones) on the intellectual and manual skills of players, or about their application e.g. in medicine and rehabilitation [Surdyk 2007: 10].

Another founder of the association, Mirosław Filiciak concludes that nowadays there can be observed an appreciation of the role of computer games as an important element of contemporary culture, not just a childish entertainment. He points out that the study of computer games initiated research on video games in the 1990s [Filiciak 2006: 17], and he rightly treats games in his academic research as:

> a communication tool between individuals, but also as a communication platform for entire groups of people [Filiciak 2006: 24].

These opinions are confirmed by numerous media, cultural and psychological studies on the games and their users, which were conducted in our country [Gądek 2009: 221–222]. Among them there appear the works examining the impact of playing on the psyche, the increase in aggression among players, the history of games and their use in teaching, as well as the works placing the games in the context of the 'culture of participation', described by Jenkins [Filiciak 2007], which strongly emphasize one of its components – the figure of active customers (related to interpersonal communication). The nature of the work on computer games is not only the domain of these specializations; the literary scholars also actively participate in the ludology discourse. The examples include the works of Jan Stasieńko [who writes about the most important monographic works on games, Stasieńko 2005: 44–48], Joanna Wrycza [2008] or of Monika Górska-Olesińska [Górska-Olesińska 2009].

Linguistic study of games

The above are just some of the names of Polish researchers involved in the subject of computer games. How, then, does a linguistic investigation on this field look like? A large part of it, showing the methodology of games, mentioning researchers and studies carried out (until 2008) was presented by Kamila Gądek [2009]. It includes works on the computer users' language, which is seen as a starting point for the study of language games. The researcher rightly states that there is no purely academic linguistics, but this does not mean that the language of the players was not described from different perspectives. The use of terminology taken from areas that have already been used in games research is postulated in these studies.

Some works dealing with the language of so-called computer geeks or the language of the Internet players describe certain aspects of the players' language. A good example is the publication of Jan Grzenia [2006], which draws

attention to such aspects as: types of communication, interrelationships between language and technology, and the most important features of the Internet language; noting that a full description requires long-term studies in the face of constant change. Kamila Gądek draws attention to the presence of the quotes and their meaning in the statements concerning the games [2009, this subject is elaborated on in her later article]. Already mentioned Joanna Wrycza [2008] examines the emoticon as a challenge to the existing punctuation. She states, however, like Grzenia, that:

> witnessing changes within the written word under the influence of the new electronic medium is both privilege and challenge. It is worth to analyze these changes, even though it is hard to judge on their importance [...] [Wrycza 2008: 9].

Among the works about culture and media, in his work from 2006, Mirosław Filiciak presents a players' slang [Filiciak 2006: 122–133]. This work, like the one of Dominika Urbańska-Galanciak, which deals with the characteristics of language players, was published in a series titled 'Cyberculture, the Internet, the Society'. The author in her cultural studies work presents a large section about, as she calls it, the language of cybertaiment [Urbańska-Galanciak 2009: 106–136]. In it she includes a summary of the extensive research, pertaining to the games, on online forum users' language. This work, however, is not a linguistic treaty.

'World of Warcraft' – a short presentation

The video game 'World of Warcraft' is a MMORPG (Massive Multiplayer Online Role-Play Game) made by Blizzard company, that means, it is an RPG (Role-Playing Game) game type, but without single player mode [the categories of gaming: Filiciak 2006: 69–74] . It is now the most popular game of this genre in the world. Currently, about 10 million players use it (this is the number of active accounts), not so long ago there were two millions more [http://www.world-of-warcraft.pl; the popularity of such games in recent times is described in: Filiciak 2006: 77–78]. When a player purchases the game and its additions on the available servers, he or she has to register and create a character (or several ones) having a unique look and abilities, and being a member of one of the two factions in the game, then he or she plays, achieving next levels, fighting enemies with the help of other players.

The need to cooperate with other players to be successful (you cannot achieve certain stages of the game otherwise) is just this aspect of this type of game, which makes the communication between the players so important. They create informal groups and the formal ones, called guilds [Filiciak 2006: 113–121]. The strengthening of social ties, often starting in the real world

(friends from school, work, family) or continuing in it (meetings of members of the guild) is reinforced by communication channels, created by the players

> each group has its own communication channel in which its members talk. Although such a specific text messaging mode, narrowed only to members of the group, is obviously aiming at the improvement of the exchange of information when performing the tasks, players will use it not only to talk about game and to exchange tips [Filiciak 2006: 113].

Such channel is built into the interface of the game, because, as Filiciak rightly notes:

> social interaction is the effect of a natural players' desire to meet the need of having contact with others [...] and of creators of the program's concern to maintain balance in the virtual universe [Filiciak 2006: 118],

however, communication through voice messaging is much more popular, as

> verbal communication in real time is the primary means of players' communication [Filiciak 2006: 123].

Because of that, the fans have produced a special language to communicate with each other (they call their language 133T or elite speak). The possibilities of communication in the game environment are discussed in Filiciak [Filiciak 2006: 85].

Linguistic observations

Unfortunately, a researcher encounters certain difficulties in the analysis of the language. One of them is the lack of a representative corpus for this language [Gądek 2009: 122], for this reason I will use examples taken from the two biggest 'World of Warcraft' fans' fora [http://hellground.pl/, http://www.world-of-warcraft.pl/forum/index.php]. Thus, I will try to describe the main local, individual features of the language used by only these game players, and I will indicate, using the available literature, which place it occupies in the general players' language (for the reference material I will use Mrs. Urbańska-Galanciak's assertions). She writes:

> Internet spread has brought into the Polish language certain properties of speech acts that conflict with established knowledge of the linguistic and stylistic norms. Features specific to the Web, like ennoblement of colloquialisms, phrasal compounds, proverbs, sayings, and the expansion of euphemism, expressiveness, profanity, para-words, exclamations and abbreviations often violate the criteria of standard language. The Internet offers strong patterns to create new rules of communication for users, resulting from the conviction that the prevalence of the

> use of language certifies its accuracy. In this way, the lame, often incorrect or offensive words, which, according to linguistic purists reduce the culture of the language, become widely used [Urbańska-Galanciak 2009: 132–133].

However, the investigator's duty is, after all, to describe even such structures of language, which do not belong to grammatically and linguistically correct ones.

The words used most often by the 'World of Warcraft' players include the abbreviations and acronyms in English, well known to players of different nationalities. The very name of the game is sometimes shortened to 'WoW' ('what would you say, if one of the additions to 'WoW' introduced the 'subrases'?', 'Did you try the voice chat in WoW?'). It is read and pronounced phonetically in Polish, and is associated with the exclamation point 'Wow!' that is the English expression of some positive surprise (just now abbreviating is becoming more popular in other games of this type, e.g. 'League of Legends' has an abbreviation 'LoL', in the English meaning 'laughing out loud').

Most frequently used abbreviations (acronyms), are typical of the Internet language (in general), e.g. 'BRB – be right back', 'Thx, ty – thanks, thank you', 'of – of course', 'cya – see you', 'Np (or Nop) – no problem', 'lvl – level (in the game)' Other equally often used acronyms are unique to this game, associated with actual in-game characters, places, having specific names: names of locations ('AB – Arathi Basin', 'WSG – Warsong Gulch', 'EOTS – Eye of the Storm', 'BT – Black Temple') or the characters in the game ('DK – Death Knight'). The players are inclined to use acronyms because of their economy of language when players need the rapid exchange of information, especially when playing on-line [to read about the language of acronyms, check: Urbańska-Galanciak 2009: 133]. However, as Filiciak writes, the sole passive knowledge of these acronyms allows the non-speaking English players to move freely in the game world [Filiciak 2006].

In addition, there are expression well known for Internet users: 'PL' as Polish (language), e.g. 'the WoW in PL?', 'nick' (name of character controlled by the player) e.g. 'You see, I moved to RPPVP and it was ok, but yesterday I had to change nick (for free) because they did not like my first: Elite', 'post' – comment added by the user ('posts that do not contain any arguments will be ruthlessly deleted'), 'fejspalm' (here in a different version of spelling), expressing the feeling of helplessness against the behavior/statements of another Internet user: ('I just do facepalm, when I see these 'fraction X is better than Y' woofs'), 'kumać' (understand something) or 'kozaczyć, przykozaczyć' – to do something to arouse the admiration, respect ('I play only after work so I didn't przykozaczyłem my lvl'). Common English borrowings are recorded phonetically, e.g. 'screen' – abbreviated in this case, meaning the use of the Print Screen button on your keyboard, or to copy the screen ('insert skrin with all the b because it's something that I don't want to believe in'). Polish abbreviations of these expressions, for example 'OMB – O Mój Boże', as the calque of the

English language) do not occur. Instead, emoticons are present [Wrycza 2008: 29–39], which, being an expression of player's emotions while writing, help in socializing and creating the illusion of real contact.

Filiciak, analyzing a different game of the same type, writes:

> the dominance of English and the lack of native vocabulary makes the Poles playing in MMOs use language calques, usually inflecting the English terms [Filiciak 2006: 128].

The players themselves create their own dictionaries, of abbreviations, but also of the concepts for the use of their less experienced mates [http://hellground.pl/]. By analyzing them, we can see that large part of vocabulary (except specialist words, occurring only in this game and related to the specificity of the game) is common to other games of this type. These words include for example: 'farming' (doing something monotonous like: raising money, gaining experience, constantly repeating the same actions), 'ninja' (the player who takes something during a joint mission, without consulting the others), 'pet' (monster controlled by the player), 'tank' (a player who attacks an enemy; there is also a verb 'tankować', Filiciak 2006). Of course, the players use the words in different ways e.g. 'nob' – the player with little experience is more pejorative term than neutral 'newb/newbie' ('Will I find some fellows there? Friendly for noobków?' – this is also a good example of mechanisms of creating diminutives among players).

The English words used by them are instinctively inflected by them, according to the rules of Polish conjugation, e.g. 'expić' meaning 'achieving experience' ('expienie with mage without Frost Armoru is a little annoying', or with 'ks': 'I ekspiłem very well'), 'raidować/rajdować' as walking the raids, the journey to defeat the enemies with the other players ('[add] allows raidowanie level 85'), or declination of the variety of names such as professions, selectable by the player ('it is difficult to choose the wrong profession for the class. With the exception of Tailoring, Leatherworkingu and Blacksmithingu') or the 'life' in the game universe: 'realm' – land appearing in the game, 'instancja' as a court, not in the meaning of the degree in the hierarchy of power, but of the game (dungeon, desert, forest): 'In each realmie there are the insty, raidy, etc and there is no difference between them'.

Players tend to use diminutives in situations when they talk about the issues closest to them, about their favorite games or their avatars (characters at the helm in the game) for example 'podstawka' is a basic version of the game ('which CDs of the game disc must I have installed, if I have podstawka, 'The Burning Crusade', 'Wrath of the Lich King' and 'Cataclysm'?'), 'hordziaki' are the players belonging to the 'Horde' (characterized in a positive way, a player using this name belonged to this group), 'szamek' a shaman, one of the characters available in 'WoW' ('play what you could play, for example professing

szamka professing elements'), and 'skilki' – their skills ('some have a similar dilemma with skilkami'), 'profki' - professions, which may be choosen ('Which profka is giving more money?'), and 'itemki' – stuffs you need/owned ('do you know how well itemki drew?').

The most commonly used prefixes are 'wy-'. 'po-' and 'z-'for example: 'if the whiz wyhealował YS I thought, that he is good', 'I'm sure you wydropisz something for you', 'from major rajdujące guild is welcome and appropriate wyenchantowany wygemowany gear', 'powbijam repe Also, 'I pogrindować, you'll have to wait 15 minutes until zrespawnuje!', 'sometimes they can zkickować even 10 even in raids'.

Quotation marks are rare [for frequent use of quotation marks, as the expression of uncertainty: Gądek 2009], more often, apostrophes are used, even if you do not need them. They separate the core of English words from suffixes, for example, 'do not waste your time on gank'owanie low lvl'owych players', you also started once, you know how it discourages. Being in a group, do not need'uj items that are redundant, or would serve you for the second embodiment'.

Within the area of syntax, most Polish words are mixed with English, with Polish rules of grammar (declination or conjugation), for example: 'quick guide get Golda', 'Master transmutacji/potionów', 'Where to buy bagi [bags for the herbalist, the profession in the game] to herbaslimu?' The players themselves sometimes say that 'the essence of wow-owego language is 'omfg omfg rofl lolz u fckin nub', which can be translated to 'hahaha, you're a funny Czesiek', or 'hihi, oh this man is ridiculous, hoho', so they mainly use acronyms (often repeated).

There was also a tendency to stylization [Urbańska-Galanciak 2009: 134], both in the use of so-called 'zkaraibszczony' English (using the 'crippled' version of the language-specific spelling of its 'speech' for example, 'see ya' ('see you'), 'mon' ('come on'), which is related to the phenomena of speech literalization and writing orality [Michael Sandbothem's thesis, referred to by Stasieńko 2005: 39] and in the use of the fans' works like prayers to the game producers or the Decalogue of players behavior. Stylizing [about stylizing look: Urbańska-Galanciak 2009: 113, 128] pertains also to the characters' names, assuming a certain, specific knowledge of popular culture (Tolkien's trilogy, 'Star Wars' universe), and the knowledge of the popular online video clips, for example, often parodied priest Natanek ('Do you drink? No. Do you smoke? No. But maybe 'World of Warcraft?').

Analyzing language of a forum for all gamers, Urbańska-Galanciak said:

> Internet environment for players keeps consistency through a kind of linguistic communication, by deliberately emphasizing linguistic diversity [...]. The players use a language rich in borrowings, neologisms, specialized terminology, apply numerous lexical, syntactic and morphological innovations [Urbańska-Galanciak 2009: 107].

The researcher also mentions features most characteristic of the language of computer gamers in general. These include, according to her, the abundance of borrowings, transcriptions of English words, duplicated words, creating synonyms, adjectives in the original spelling or phonetically polonized, the use of Polish suffixes added to English words, a tendency to parody, stylization and references, the use of archaisms, neologisms, euphemisms and metaphors [Urbańska-Galanciak 2009: 115–125].

However, some of the linguistic phenomena observed by her are not present in 'WoW' players language in such a large amount, as in the general language of players. The most important features of the language of players, according to her, are secrecy and using colloquialisms, undermining the principles of grammar and spelling which causes difficulty in understanding the messages for the public who does not know this unfamiliar language [Urbańska-Galanciak 2009: 115], however in the case of 'WoW' the players, have a lot of good will, and they are friendly to the new fans of this game, often helping them to understand the specific language they use. This is particularly understandable, given the fact that

> living in [...] this type of game worlds [...] would be boring if it were not possible to communicate with other players, which is the absolute axis of the game. It is not indeed the accident – developers deliberately design the 'downtime' that somehow forces the establishment of dialogue between the players [Filiciak 2006: 79].

So without this help and patient explaining the rules by the players of this particular language game, they will not achieve the success of communication and thus, measurable success in the game, because

> communicating with others is one of the primary things the game [...] the game is primarily mass communication [Filiciak 2006: 83].

Conclusions and research perspectives

Nicholas Ostler, a linguist, said in 2011 that new technologies will replace English as the 'lingua franca'. Kamila Gądek few years ago wrote about the language of the computer players, that it is a

> language still remaining in the stage of formation and continuous transformations, in which changes occur faster than normally. It is therefore a kind of miniature, in which we can observe some changes and linguistic phenomena. The development of players' language is also important as a part of research on the Polish language, especially as more and more people have at least occasional contact with the games and, increasingly, they are also recognized as an important element of contemporary culture. Perhaps with time, the development of computer games market and the progressive increase in the average age of players in Poland will increase the impact of vocabulary related to computer games in a general

> language, as was the case with the vocabulary associated with the use of computers [Gądek 2009: 225].

Undoubtedly, especially among the younger generation, we observe changes in the way we use language. Often, however, as it is observed, these are seasonal fashion for specific words or expressions. Language of the Internet, like the news posted in it every second, is sometimes elusive. Today, only PC players, faithful to the older types of games, like classic adventure games, use expressions such as 'go two screens to the front', which a few years ago were frequent, and which were described by Gądek [2009].

As it can be seen, although these expressions are still understood, during the short term they got old and are used by fewer people. The strength of the impact of the language of this social group, in this case the computer players, seems to relate only to certain aspects of their language. They will, as it seems, dominate first in the language of players minimally associated with them (e.g. playing browser games available on social networking sites), and then become popular in the spoken colloquial Polish. Some are expected to remain fashionable and used only by specific user groups, like so called 'pokemon language', existing only in written form, for example the use of alternating large and small letters ('writing wave': Śliwa 2012), which is a secret language that connects that specific community, because

> slang is just another factor which integrates community and emphasizes the fact that users depend on the cooperation and agreement [Filiciak 2006: 124].

He calls it 'quilt' (term by Elizabeth Byrd), a creation or:

> the result of joint work of the various participants of communication [Filiciak 2006: 124].

The same researcher reporting on the thesis of Miriam Eladhari [Filiciak 2006: 131–132] says that when she looked at a big collection of video games, she found that the number of determinants that can be used to describe them is small. Perhaps the same situation occurs when we try to describe language of computer players – there are not too many determinants of this language, at least those that occur in communication between players of all games. It would therefore be interesting to study the overall turnout and the study of the language of each group of players, fans of specific games. An interesting aspect would be to gain access to talks with (very rarely) one's own voice chat, but more often analysis of programs used by the players, such as 'Team Speak', 'Ventrilo', 'Mumble', and even 'Skype', conducted by the players and their analysis. This could help to build vocabulary corps of players, because they often overlap among those who are fans of games belonging to one type. Studies

conducted to date will be the most valuable because, as it can be seen, this language is changing rapidly, but on the other hand, its core seems to be already established. For recent years, with more and better knowledge of English and the presence of more advanced players (as opponents or companions for fun) on foreign servers, young (but not only young) Poles have begun their adventure with network games and contact with peers from abroad. Perhaps in a few years an interesting task for applied linguistics will be a description of the acquisition of new language skills by the players and the impact of their favorite online games for learning correct grammatical structures, vocabulary growth (surveyors have already conducted research on the personal beliefs of players, about how to improve their language skills through games like 'WoW', see Greloch, http://www.ankieter.pl/wyniki/index/id/5854). Such tests could be used for practical purpose - to create new activation techniques and foreign languages teaching.

Bibliography

Filiciak, Mirosław [2006], *A virtual playground. Network Games and the transformation of the contemporary culture*. Warszawa: Wydawnictwa Akademickie i Profesjonalne.

Filiciak, Mirosław [2007], *Computer games as a manifestation of cultural participation*. In: Augustyn, Surdyk / Szeja, Jerzy (ed.), Culture-Generative Function of Games. The Game as a medium, text and ritual. Volume II. Language Culture Communication. Series No. 2. Poznań: Wydawnictwo Naukowe UAM, 107–113.

Gądek, Kamila [2009], *The methodological basis for the description of the language of computer gamers*. In: Skarżyński, Mirosław / Czelakowska, Anna. Volume 3. Library of LingVaria. Kraków: Księgarnia Akademicka, 221–227.

Godzic, Wiesław [2006], *Introduction. The games – about the contemporary culture*. In: Filiciak, Mirosław, A virtual playground. Network Games and the transformation of contemporary culture. Warszawa: Wydawnictwa Akademickie i Profesjonalne, 7–11.

Górska-Olesińska, Monika [2009], *The word in the web. Electronic discourses*. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Grzenia, Jan [2006], *Linguistic communication in the Internet*. Warszawa: PWN.

Ostler, Nicholas [2011], *New Lingua Franca*. Tygodnik Forum 26, 54–56.

Stasieńko, Jan [2005], *Alien Vs. Predator – computer games and literary studies*. Wrocław: Wydawnictwo Naukowe Dolnośląskiej Szkoły Wyższej Edukacji TWP we Wrocławiu.

Surdyk, Augustyn [2007], The *nature and objectives of the Polish Society of Games Research and the perspective of the study of games in Poland*. In: Surdyk, Augustyn / Szeja, Jerzy (eds.), Culture-Generative Function of Games. *Culture-Generative Function of Games. The Game as a medium, text and ritual*. Volume

II. Language Culture Communication. Series No. 2. Poznań: Wydawnictwo Naukowe UAM, 7–14.
Śliwa, Marta [2012], The *analysis of comments of users Sweet Brokacik forum*. Typescript.
Urbańska-Galanciak, Dominika [2009], *Homo players. Strategies of reception of computer games.* Warszawa: Wydawnictwa Akademickie i Profesjonalne.
Wrycza, Joanna [2008], *Galaxy of the internet language*. Gdynia: Novae Res – Wydawnictwo Innowacyjne.

Websites (source examples)

Greloch, Mateusz, *Computer games as a modern method of developing language skills [Poll]*, http://www.ankieter.pl/wyniki/index/id/5854.
Forum World of Warcraft – HG HellGround.pl http://hellground.pl/.
Forum World of Warcraft – Polish Official Fan.
http://www.world-of-warcraft.pl/forum/index.php Service.

Ewa Maciejewska-Stępień*

PRACTICAL APPLICATIONS OF ETHNOLINGUISTIC RESEARCH

Due to the growing multilingualism and multiculturalism of the European Union, it has become crucial to promote the intercultural dimension in language pedagogy [*cf.* European Commission 1995]. This need has been clearly reflected in the European policy on language, culture and education [Council of Europe 2001] as well as in the literature on applied linguistics [e.g. Byram 1997, Kramsch 1998, Bandura 2007]. Now, more often than ever, language learners expect and are expected to develop intercultural communicative competence (ICC) [Byram 1997], which is supposed to be achieved with the help of a variety of teaching procedures devised to attain this goal. It seems that one of the main concepts explored by Polish ethnolinguists, i.e. linguistic worldview [Bartmiński 2009a], along with pertinent ethnolinguistic research results might significantly enrich the existing procedures. Thus, the purpose of this paper is to suggest ways in which the notion of linguistic worldview, supported with the research results, might be used to improve ICC in foreign language learners. The paper consists of four parts: first, it defines ICC; second, it discusses the most common teaching techniques for improving ICC in foreign language learners; third, it presents the concept of linguistic worldview; and, finally, it suggests ways in which the notion might be adapted to develop some of the existing techniques.

1. Intercultural communicative competence

The best known discussion of interculturalism in language pedagogy stems from the work of Byram [1997], who describes intercultural communicative competence (ICC) as a reformulated and extended version of the widely accepted concept of communicative competence [Canale & Swain 1980; Hymes 1972; van Ek 1986].

* Institute of Applied Linguistics, University of Warsaw.

The notion of communicative competence has lately been criticized, as it implies that the native speaker should function as a model for foreign language learners. As Kramsch [1998] remarks, although the idea of the native speaker as a model seemed fairly straightforward about thirty years ago, it has recently become quite controversial. She points out that scholars have questioned the "identity" of the native speaker [e.g. Davis 1991; Paikeday 1985] as well as his/her seemingly "unquestioned authority" [Byram & Zarate 1994; Phillipson 1992] and even the "appropriateness of the one native speaker norm in a time of large-scale migrations, cross-national and cross-cultural encounters, and increasing linguistic and pragmatic differences among speakers of the same language" [Kramsch 1998: 16]. Similarly, Byram argues that the native speaker model is inappropriate because it creates an "impossible target and consequently inevitable failure" [1997: 11]. He adds that this model implies a "schizophrenic" kind of competence:

> It would imply that a learner should be linguistically schizophrenic, abandoning one language in order to blend into another linguistic environment, becoming accepted as a native speaker by other native speakers. This linguistic schizophrenia also suggests separation from one's own culture and the acquisition of a native sociocultural competence, and a new sociocultural identity [Byram 1997: 11–12].

Thus, Byram [1997] rejects the native speaker model. Instead, he suggests replacing it with the intercultural speaker model, which assumes that successful intercultural communication depends on the ability "to interact with 'others', to accept other perspectives and perceptions of the world, to mediate between different perspectives [and] to be conscious of their evaluations of difference" [Byram *et al.* 2001: 5]. Similarly, Kramsch proposes it is the intercultural speaker that language learners should aspire to, not the "untroubled mythical" native speaker [1998: 27].

Byram defines ICC as an ability "to interact with people from another country and culture in a foreign language" [1997: 71]. He argues that ICC consists of four competences: linguistic, sociolinguistic, discourse and intercultural. The first three are based on van Ek's [1986] model of communicative ability; however, they have been reformulated to account for replacing the native speaker with the intercultural speaker:

> **Linguistic competence:** the ability to apply knowledge of the rules of a standard version of the language to produce and interpret spoken and written language;
>
> **Sociolinguistic competence:** the ability to give to the language produced by an interlocutor – whether native speaker or not – meanings which are taken for granted by the interlocutor or which are negotiated and made explicit with the interlocutor;

> **Discourse competence:** the ability to use, discover and negotiate strategies for the production and interpretation of monologue or dialogue texts which follow the conventions of the culture of an interlocutor or are negotiated as intercultural texts for particular purposes [Byram 1997: 48].

Thus, linguistic, sociolinguistic and discourse components share their main characteristics with the concept of communicative competence. What makes them different is the idea of the intercultural speaker which they imply and an additional component which Byram [1997] refers to as intercultural competence.

Intercultural competence (IC) encompasses several interdependent *savoirs*, i.e. attitudes, knowledge and skills connected with intercultural communication. Attitudes (*savoir être*) form the foundation of IC. Byram claims that successful intercultural communication depends on attitudes of curiosity, openness and readiness to discover different perspectives on familiar and unfamiliar phenomena [1997: 50]. This involves readiness to "decenter" [Kohlberg *et al.* 1983, after Byram 1997: 34] and to relativize one's own values, beliefs and behaviours [Byram *et al.* 2001].

Another important factor is knowledge (*savoirs*) of one's own and one's interlocutor's social groups, their products and practices, and of "the general processes of societal and individual interaction" [Byram 1997: 51]. Thus, such knowledge consists of two major categories: (i) knowledge about groups, products and processes, which is always partially present as a result of socialization, and (ii) knowledge of the processes of interaction, which is crucial in successful intercultural communication but "not acquired automatically" [Byram 1997: 35].

Apart from appropriate attitudes and knowledge, there are also important skills that intercultural speakers need to possess. These are skills of interpreting and relating (*savoir comprendre*), and skills of discovery and interaction (*savoir apprendre/faire*). The former refer to the ability to interpret phenomena from another culture and relate them to one's own culture, whereas the latter describe the ability to acquire knowledge of a culture and use it in real-life interaction. By comparing, interpreting, relating and discovering values and beliefs in ideas, events, documents and real-time interaction, intercultural speakers are able to identify ethnocentric perspectives, and as a consequence, they are able to see how misunderstandings can arise and how they might possibly resolve them [Byram 1997: 52].

Byram claims that developing these four aspects of IC in an educational context should be enriched with developing critical cultural awareness (*savoir s'engager*), i.e. "an ability to evaluate, critically and on the basis of explicit criteria, perspectives, practices and products in one's own and other cultures and countries" [1997: 53]. This implies the ability to identify either explicit or

implicit values inherent in one's own and other cultures. Such critical cultural awareness, along with knowledge, skills and attitudes, enables the intercultural speaker to act as a cultural mediator, border-crosser or negotiator of meaning [*cf.* Byram 2003: 60; Council of Europe 2001: 105; Roberts *et al.* 2001: 3].

2. Procedures in developing ICC

As mentioned previously, the concept of ICC is a reformulated and extended version of the widely accepted notion of communicative competences [Byram 1997]. Although these competences are associated with different aims and objectives, they are developed with the help of the same teaching techniques [Bandura 2007: 55]. The interest in cultural education over the last forty years has given rise to a wide variety of culture teaching techniques. For the sake of clarity, Stern [1992: 224–232] suggests dividing them into the following "approaches" to culture teaching: creating an authentic classroom environment, providing cultural information, cultural problem solving, behavioural and affective aspects, cognitive approaches, approaches based on literature, real-life exposure to the C2 and making use of cultural community resources. Below, there is a sampling of ideas representing the most common techniques that have been suggested since the early 1970s and that are associated with the above approaches.

An authentic classroom environment refers to the so-called *culture island*, i.e. classroom decorations, displays of posters, pictures, maps, newspaper clippings as well as expositions of realia such as schedules, tickets, menus, questionnaires or programmes. Such authentic objects from the target culture are intended to provide students with "a dose of reality" [Lund 1992] and in this way increase the comprehensibility of linguistic and cultural input [*cf.* Hughes 1984; Chastain 1988]. They help students relate classroom teaching to the real world through facilitating "the simulation of experience in the target culture" [Berwald 1987].

Providing cultural information encompasses such techniques as culture asides, cultural connotations, culture capsules and culture clusters. *Culture asides* are pieces of cultural information or "slices of life" [Taylor 1970], such as a recording of a popular song or a news item, the main purpose of which is to "help the learner to create a network of mental associations similar to those which the items evoke in the native speaker" [Stern 1992: 224]. Similarly, through the use of visual aids and/or word-association activities, techniques which aim at deriving *cultural connotations* make students aware that words and phrases both in the L1 and in the L2 are associated with culture-specific images [Omaggio 1986; Seelye 1984]. C*ulture capsules*, which represent one of the earliest techniques of culture teaching, are brief descriptions of one difference between a native and a foreign custom, accompanied by photos, slides

and realia, and followed by a series of related questions and activities [Taylor & Sorenson 1961]. *Culture clusters*, also a fairly early technique, consist of three or more capsules concerning similar topics and one thirty-minute simulation that integrates the information in the capsule and dramatizes it through a role-play or skit [Meade & Morain 1973].

Culture assimilators epitomize cultural problem solving. They do not simply present learners with information, but make them face a culturally troublesome situation. Typically they consist of three parts: (i) an act of miscommunication between a member of the C1 and a member of the C2, presented in a narrative or dramatic form, (ii) one or more multiple-choice question(s) dealing with the explanation of the conflict and (iii) feedback paragraphs related to the multiple-choice options, explaining whether the choice is likely and, if necessary, providing additional cultural information [Hendon 1980; Lafayette 1978; Omaggio 1986]. Another example of cultural problem solving is *hypothesis refinement*, consisting of research-oriented activities the main function of which is to refine learners' initial perceptions of an aspect of the target culture [Jorstad 1981, after Omaggio 1986].

Behavioral and affective approaches include the auto-motor unit, dramatization, mini-drama, role-play and simulation. *The auto-motor unit* contains a series of oral commands which are based on a cultural theme and to which students are supposed to react physically. Learners observe as the teacher pantomimes actions, for example culture-specific nonverbal messages, and then follow the commands themselves [*cf.* Chastain 1988; Lafayette 1978]. *Dramatization*, *role-play* and *simulation* involve learners in observing and/or improvising verbal and nonverbal exchanges and actions that illustrate culture-specific issues [Byram & Fleming 1998; Stern 1992]. *The culture mini-drama* includes between three and five skits or episodes, which are read, acted out or watched on video, and which present one or more examples of miscommunication. After each episode learners discuss the scenes and obtain additional information until they are able to identify the precise reason of the miscommunication [*cf.* Chastain 1988; Omaggio 1986].

In contrast to most of the above techniques which are fairly experiential, cognitive approaches represent activities which are "academic or systematic in nature" [Stern 1992: 228]. They include: *lectures*, *readings*, *discussions* and *debates* as well as *research techniques* encouraging learner autonomy [*ibid.*]. Their focus may be on areas such as daily life, work and study in connection with their social underpinnings [Buttjes & Byram 1991; Sercu 1995; van Ek & Trim 1991]. Readings may include, for instance, literary works. Some scholars have advocated an integration of the above approaches, which to a large extent have been influenced by social sciences, with the humanistic approaches, which regard *literature* as a source of culture teaching [Fenner 2001; Kramsch 1993].

The remaining two approaches, which make use of cultural community resources and real-life exposure to the C2, combine communicative language teaching with cultural awareness [Stern 1992]. The former refers to language and culture teaching in the target-language milieu where "the everyday environment constitutes a vast resource [...] for culture teaching" [Stern 1992: 231–232]. The latter additionally applies to the foreign language situation, in which real-life contacts may include the following forms: *pen-pals*, *tape-pals*, *computer-mediated encounters* [Carel 2001; Ware 2003], *interviews with native informants* or *tandem partners* [Woodin 2001], *audiotaped interviews*, *videotaped interviews/films* [Burnett & Thomson 1985, after Omaggio 1986] and *visits to other countries* [Parker & Rouxeville 1995; Roberts 1997], which may be connected to *ethnographic projects* [Byram & Fleming 1998; Roberts 1995; Roberts *et al.* 2001].

3. Linguistic worldview

The concept of linguistic worldview is strictly semantic in nature. It "corresponds well with the cognitive and communicative revolution in contemporary linguistics, which emphasises the symbolic (cognitive) and communicative (interpersonal) functions of language" [Bartmiński 2009a: 24]. It is rightly claimed that "the basic communicative function of language is strongly linked to its cognitive function, i.e. the perception and conceptualisation of reality along with an underlying system of values"; therefore, "*the efficiency of communication largely depends on the compatibility of the values and worldviews that are entrenched in language* and thus readily available to the speaker" [Bartmiński 2009a: 222; emphasis added EMS]. It follows that teachers interested in developing ICC in foreign language learners should pay attention to the worldviews and their underlying values entrenched in their students' first language (L1) and in the foreign language studied in the classroom (L2).

The notion clearly relates to the Sapir-Whorf hypothesis, additionally known as the principle of linguistic relativity, according to which language determines (or at least influences) the way people think [Sapir 1978; Whorf 1982]. The term worldview is also associated with Wilhelm von Humboldt's *Weltansicht* understood as "the way the language system shapes the perspective and conception we have of the world and to a large extent shapes the way we negotiate our way through the course of life on a day-to-day basis as we converse with others" [Trabant 1992, after Underhill 2009: 17]. In Poland, the concept of linguistic worldview was introduced to "a wider audience" by Walery Pisarek [1999, after Bartmiński 2009a: 22] and developed by Jerzy Bartmiński and his team, who have devised, over the last 30–40 years, "an integral theory of language, with the human speaking subject (*homo loquens*) at

the centre" [Bartmiński 2009a: 222]. At present, linguistic worldview constitutes one of the key concepts in Polish ethnolinguistic research.

According to Bartmiński, linguistic worldview is "a language-entrenched interpretation of reality, which can be expressed in the form of judgements about the world, people, things or events", which "unites people in a given social environment, creates a community of thoughts, feelings and values" and which "influences [...] the perception and understanding of the social situation by a member of the community" [2009a: 23]. These judgements can be manifested in language, in its lexis, phraseological units, collocations, metaphors, grammar and texts [2009a: 26–34], but they can be also "presupposed", i.e. implied by language forms, based on the socially entrenched knowledge of the world, convictions, beliefs and conventional behavior relating to obligations and prohibitions [2009a: 34–35]. Linguistic worldview is an interpretation of reality from the perspective of an average speaker of a national language, reflecting the speaker's mentality and needs. Polish ethnolinguistic research tends to focus on the most common variant of a national language, its colloquial style, which is considered anthropocentric and ethnocentric, as well as practical and sufficient to meet the requirements of everyday communication.

In the theory of language advocated by Bartmiński and his team, linguistic worldview relates to the following key concepts: values, stereotypes, cognitive definitions, viewpoint, perspective, profiling and the speaking subject.

First of all, worldviews are created on the basis of values assumed by speakers. Values as such permeate the whole language because "language is a *tool* of valuation", "a *source of information* about the values accepted by its speakers" and "a *career (substrate)* of values" [Bartmiński 2009a: 40]. Ethnolinguists tend to regard language mainly as a source of knowledge about values, ideals and judgments important for a given community of speakers. The most useful source of this kind are words referring to concepts which play a crucial role in public discourse. These words are known as "value terms" [Bartmiński 2009a] or "cultural keywords" [Wierzbicka 1997]. They include: *key political concepts*, such as freedom/liberty, independence, democracy, totalitarianism, left and right wing, socialism, communism, revolution, internationalism, globalism, terrorism, state, authority, religion, Europe, America, the world, the East, the West; *social concepts*, such as home, family, nation, homeland, patriotism, equality, honor, tolerance, hospitality, society, humankind; *moral values*, such as justice, brotherhood, solidarity, courage, heroism, responsibility; *personal virtues*, such as dignity, fidelity; *general concepts*, such as human being, human dignity, human rights, love, friendship, education, knowledge, beauty, progress, work, diligence, career, conscience, faith; *concepts of knowledge and cognition*, such as truth, falsehood, evil, lie [Bartmiński 2009a: 220]. Research shows that the sociopolitical and ethical lexis contains numerous

language- and culture-specific elements in different languages and cultures, especially with regard to connotations. Undoubtedly, it is of special importance and urgency for ethnolinguists, as it has the potential to "contribute to a better coexistence of nations" [Bartmiński 2009a: 219–221].

Linguistic worldviews are composed of stereotypes understood as schematic "pictures in the head" [Lippmann 1922], connecting typical descriptive, emotional and valuational features. Just like language, they are based on simplifications, generalizations and valuations; hence, they are perceived as "living in language" [Bartmiński 2009b]. They have two basic functions: minimalizing the effort of getting to know the world and defending one's social position [Lippmann 1922]. Likely to influence behavior, stereotypes relate closely to the system of norms, values, needs and attitudes advocated by a given community of speakers. Ethnolinguists describe them in terms of cognitive definitions.

To explain the meaning of words, ethnolinguists use cognitive definitions, aiming to present the ways in which objects are conceptualized by the speakers of a language, "to represent socio-culturally established and linguistically entrenched knowledge, its categorisation and valuation" [Bartmiński 2009a: 67]. Such definitions differ significantly from traditional lexicographic definitions, as they need to meet the requirements of "content adequacy" and "structural adequacy". The former refers to including common knowledge in definitions, which implies that descriptive terms must be "the colloquial variant of [a] standard [language], with no elevated, bookish or scientific expressions" [2009a: 71]. The latter means that, on the basis of common knowledge, definitions intend to "reconstruct the relationships between the components present in the collective consciousness of speakers" and report on "all the positive features established in the linguistic worldview [of a community of speakers]" [*ibid.*]. For example, on the basis of folk texts, peasant dialects and ethnographic documentation, Polish *deszcz* (rain) has been described in the following manner:

i) SUPERORDINATE CATEGORY: *deszcz to woda* (rain is water)
ii) APPEARANCE: *w postaci kropel* (it appears in the form of drops)
iii) CHARACTERISTICS: *jest drobny* (it is fine)
iv) ACTION: *deszcz pada* (it falls); it falls with alternate strength: *deszcz mży* (drizzles), *kropi* (spits), *roni* (sheds drops), *pryska* (splutters), *leje* (pours), *bije* (strikes), *popaduje* (falls a little), *pokrapuje* (spits a little), *porania* (sheds a little), *popryskuje* (splutters a little); *cf.* also nouns *mżawka* (drizzle), *kapuśniaczek* (Scotch mist), *ulewa, leja* (downpour). According to traditional beliefs, spring rain impregnates the earth and brings about good harvest. *Deszcz powoduje błoto* (it causes mud, as in the proverb *kwarta deszczu – korzec błota*)

v) TIME OF ACTION: in songs there are frequent references to Sunday morning, when the rain sheds an occasional drop (and lovers meet), as opposed to the evening, when it pours (and lovers part)

vi) EVENTS COOCURRING WITH *DESZCZ*:
a) in the world of humans: love and tears, e.g. (in songs) *deszczyk pada, dziewczę biada* (it rains, a girl laments);
b) in nature: storm, lightning; clouds, dew, e.g. (in songs) *deszczyk pada, rosa siada* (it rains, dew sits down);
c) in the supernatural world: (in a song) the deceased father or mother come with the rain to the orphaned child; (in folk beliefs) frogs and fish fall with the rain; (prohibitions and orders) when it rains, one must not plough or sow; it is good to plant potatoes, sow peas or to scythe; *cf.* the proverb: 'He who scythes in the rain, dries in fine weather'.

vii) PLACE OF ACTION: (in a song) *deszcz pada po drobnej leszczynie, po białej brzezinie* (it rains on fine hazel and on white birch – the hazel and the birch are folk symbols of a girl; *cf.* the impregnating function of the rain above).

viii) AUGURY CONCERNING *DESZCZ:* (in common beliefs) it will rain if: (a) one can feel pain in the bones; (b) fire does not burn well, cattle yawn, a dog eats grass, fish thrash about, bees swarm to the hive, chickens bathe in sand; (c) kites (birds) want to drink, crows cry, swallows fly low; (d) the sun beats down, the sun is pale and sad, stars twinkle, the mist rises in the morning

ix) AUGURY BASED ON *DESZCZ:* the rain on the wedding day foreshadows unsuccessful married life [Bartmiński 2009a: 70–71]

Linguistic worldviews depend on the speaking subject's viewpoint and perspective. Bartmiński [2009a: 77] defines the former as "a subjective-cultural factor, decisive for the way an object is referred to", functioning as "a set of directives shaping the content of words and whole utterances". In other words, the point of view adopted by the speaking subject is the position from which he or she considers something. The concept may be illustrated with the following example:

> *Bławatek* 'cornflower', categorised from the point of view of an average nature lover as a 'flower', must be defined with respect to its [appearance], [scent], [blooming season], [distribution], and maybe [use]. From the point of view of a farmer it is categorised as a 'weed' and defined relative to [the place where it grows], its [appearance], [harmful effects on human endeavours], and maybe [methods of uprooting]. From the point of view of a natural scientist it is categorised as a 'plant' and defined relative to the plant definitional schema. As can be seen, points of view are interwoven with human activity in various domains [Bartmiński 2009a: 81].

Viewpoint influences the perspective from which an object is presented in linguistic worldviews. Bartmiński [2009a: 78] understands perspective as "a set of properties of the semantic structure of words, correlated with and, at least to a certain extent, resulting from a point of view". To be more precise, by this "set of properties" he means, firstly, "a set of 'aspects' ('sides') of the object, which are taken into account by speakers in a non-arbitrary order" and which "constitute a profile of the concept in question" and, secondly, "the body of information provided about the object when viewing it from a given 'side'" [2009a: 82]. Speakers choose those aspects of the object which they consider important, i.e. "those which constitute a (socially) established, internally coherent structure of the concept" [*ibid.*]. For instance, a definition of Polish *gwiazda* (star), based on the colloquial knowledge of the world and on the colloquial structuring of that knowledge, should contain aspects such as [appearance], [number], [place where it is seen], [time when it is seen] and [action]. These should be filled with content reflecting the colloquial viewpoint such as "*gwiazda* is 'one of the many small lights in the sky, visible at night, which group into what are called *constellations*, and about which it is said: they shine, twinkle, light up, fade away, fall …'" [Bartmiński 2009a: 82–83].

Linguistic worldviews function in discourse in the form of profiles, i.e. "variant[s] of the image of a given object" [Bartmiński 2009a: 91]. These variants result from the process of profiling, which includes: the categorization of an object, the selection of aspects within the chosen categorization and the description of characteristics within the chosen aspects. Profiling depends on the speaking subject's viewpoint, knowledge, type of rationality and system of values [Bartmiński 2009b: 94]. Thus, differences in profiling are differences in the subjective conceptualization of an object.

As previously mentioned, linguistic worldview is manifested in language. Its reconstruction requires the analysis of lexis, phraseological units, collocations, metaphors, grammar and texts. Lexis is the most "obvious, unquestionable basis" for reconstructing linguistic worldview [Bartmiński 2009a: 26]. As it has already been pointed out, researchers are especially interested in analyzing "value terms" [Bartmiński 2009a] or "cultural keywords" [Wierzbicka 1997]. Linguistic worldview can be also reconstructed on the basis of phraseological units, collocations and metaphors. Although very interesting for ethnolinguists, phraseological units tend to be based on historical knowledge, which requires etymological analyses. As suggested by Bartmiński [2009a: 31], "more informative for the contemporary interpretation of the world are living collocations" and metaphors. For instance, the analysis of words collocating with *tekst* (text) in Polish shows several ways of conceptualizing the term metaphorically [Niebrzegowska-Bartmińska 2007: 23–31]. Texts in Polish are perceived as:

> a result of the work of craftsman, such as a weaver (*snuć opowieść/ wątek/nić opowiadania,* lit. 'spin a story/motif/the thread of a story'); a cook (a text may be *pikantny* 'spicy', *smakowity* 'delicious', *niestrawny* 'unpalatable'); a builder (*plan/szkic tekstu* 'a plan/sketch of a text', *piętra tekstu* 'several storeys of a text', *budować/klecić tekst* 'to build/botch up a text'); a farmer or gardener (*tekst dojrzewa* 'a text ripens', *pomysł kiełkuje* 'an idea sprouts', *tekst jest owocem pracy* 'a text is the fruit of one's work'); an artist or a musician (*rama tekstu* 'the frame of a text', *drugi plan tekstu* 'the background of a text', *autor maluje coś w jakichś barwach* 'the author paints something in such and such colours', *autor porusza czułą strunę* 'the author pulls a sensitive string'). A text may *ranić* 'hurt' and be *ostry* 'sharp' or *cięty* 'biting' – thus it may be conceptualised as an instrument or a weapon. [Niebrzegowska-Bartmińska 2007, after Bartmiński 2009a: 32]

Yet another basis for reconstructing linguistic worldview is grammar. As Bartmiński [2009a: 33] suggests, "[t]he categories of person, number, gender, tense, mood or case are relatively easily observable and comparable (on the cross-cultural scale) mechanisms of the conceptualisation of reality". For example, in Polish, nouns referring to men occur with the endings *-i* and *-owie* in nominative singular (*mężczyźn-i* 'men', *pan-owie* 'gentlemen'), whereas nouns referring to women and things occur with the same ending *-y* (*kobiet-y* 'women', *rzecz-y* 'things'). Similarly, in the past tense, verbs referring to men appear with the ending *-i* (*byl-i* 'they (men) were'), while verbs referring to women and things with *-y* (*był-y* 'they (women, things) were'). This privileged reference to men tends to be interpreted as "a remnant of male dominance in public life" [*ibid.*].

Apart from words, phrases, collocations, metaphors and grammar, researchers of linguistic worldview analyze "stereotyped texts", which describe typical motifs with judgments about people, things and events [Bartmiński 2009a: 33-34]. Such texts include: proverbs, legends, myths, fables, anecdotes, songs, riddles and puzzles. As explained by Bartmiński [2009a: 34], they "contain whole repetitive scenarios of behaviour and action, which correlate with the socially accepted roles, with what is expected and what may be breached". On the basis of these scenarios, ethnolinguists record stereotypical motifs and judgments to reconstruct worldviews and to compile cognitive definitions.

A very valuable source of linguistic worldview data are questionnaires containing either open or closed questions about the meaning of words. Although more difficult to interpret, open-ended questions are regarded as relatively more valid because they do not suggest answers to respondents. The simplest but at the same time one of the most important questions asks respondents to describe the meaning of a word, for instance 'Explain the meaning of a family'. Examples of frequent directional questions include: 'What are the features of a typical family?' and 'What are the features of a true family?'.

The modifiers 'typical' and 'true' tend to influence quite significantly the type of the features provided: in the former case respondents supply purely descriptive characteristics (referring to what a family *is* really like), while in the latter they record idealized characteristics (referring to what a family *should be* like). This distinction allows researchers to examine linguistic worldviews in terms of different modalities. Additional questions may include: 'Write down one word that best describes a family' or 'Complete the following sentence: They are a family, but…' [for more examples see Bartmiński 2006; 2009a: 132–148, 149–161, 162–177, 178–198]. In their analysis of responses to such questions, researchers aim to discover which components of meaning are widely recognized, in which order they are listed, how they are evaluated and what configurations of components can be identified [Bartmiński 2009a: 137].

Bartmiński stresses the importance of comparing and contrasting linguistic worldviews in different languages and cultures, as he deeply believes that "such research may today have practical application and help in intercultural communication" [2009a: 219]. He recommends investigating at least six issues, "in accordance with the basic model of linguistic communication: WHO is communicating, WITH WHOM, WHERE, WHEN, WHAT FOR and HOW?" [*ibid.*]. These include:

(i) the expression of collective identity – who are 'we' (self-stereotypes);
(ii) the perception and linguistic construal of others – who are 'they' (heterostereotypes of one's neighbours);
(iii) the conceptualisation of one's place, location in the world;
(iv) the conceptualisation of the communal time in which we live;
(v) the values, declared and actually functioning, in 'our' social world (communal, social, national);
(vi) the means and types of expression, according to the criteria of styles and genre [Bartmiński 2009a: 219–220].

4. Linguistic worldview in procedures for developing ICC

It seems that the concept of linguistic worldview and the linguistic worldview research results have the potential to enrich the existing procedures for developing ICC. Below there is a sampling of suggestions for incorporating the concept into the most common teaching techniques for developing ICC in foreign language learners. Preference will be given to "value terms" [Bartmiński 2009a] because examining them closely is of special importance and urgency in successful intercultural communication. Since discussing the pertinent research results is beyond the scope of this paper, the references in parentheses listed below will guide the reader to the sources which are most likely to be inspiring for language teachers, materials writers and scholars interested in the techniques. As previously, the techniques will be divided in accordance

with the following "approaches" [Stern 1992] to culture teaching: creating an authentic classroom environment, providing cultural information, cultural problem solving, behavioral and affective aspects, cognitive approaches, approaches based on literature and making use of cultural community resources.

An authentic classroom environment, also referred to as the culture island, may serve as a very useful tool for presenting similarities and differences between the linguistic worldviews entrenched in students' first language and in the foreign language studied in the classroom. Carefully chosen posters, pictures and advertisements may perfectly illustrate values, ideals and judgments important for the speakers of the languages. Newspaper clippings, song lyrics or literature excerpts may provide precious insights into the conceptualizations of words such as, for example, friendship, freedom [*cf.* Wierzbicka 1997], homeland [*cf.* Wierzbicka 1997; Bartmiński 2009a, 2009b], mother, house/home, right and left wing [*cf.* Bartmiński 2009a, 2009b], heaven or hell [*cf.* Bartmiński 2009c].

Culture asides, cultural connotations, culture capsules and culture clusters are all recommended for providing learners with linguistic worldview information. Culture asides, such as song recordings, news items or movie scenes, focusing on for instance the perception of different nationalities [*cf.* Bartmiński 2009a; Bartmiński *et al.* 2002; Brzozowska 2007; Mikoś & Tieszen 2005; Skibińska 2005], races [*cf.* Wysocka 2002; Zarzycka 2005] or countries [*cf.* Bartmiński 2009b; Chlebda 2002; Żuk 2004] may help learners to create associations similar to those which certain terms evoke in native speakers. Likewise, visual aids and word-association activities which aim at deriving cultural connotations can make students aware that words, phrases, collocations, metaphors and proverbs both in the L1 and in the L2 are associated with culture-specific images. Culture capsules may be also very helpful in this respect, if they briefly describe a difference, for instance, between a native and a foreign understanding of democracy [*cf.* Bartmiński 2006; Pietrucha 2003], progress [*cf.* Buława 2003] or fairness [*cf.* Borowiecka 2003], illustrated with photos and realia, accompanied by pertinent questions and activities. Culture clusters, encompassing several capsules, should focus on related concepts such as a native and a foreign perception of 'us' and 'others' [*cf.* Bartmiński 2007; Chlebda 2007; Grzegorczykowa 2008; Kołodziejek 2007; Zinken 2000] and end with a simulation which integrates the information in the capsule and dramatizes it through a skit or role-play.

Apart from presenting cultural information, culture assimilators encourage learners to solve a cultural problem by facing a challenge of miscommunication between representatives of different cultures. The reasons for the misunderstanding should be differences in the interlocutors' linguistic worldviews, for instance their divergent conceptualizations of terms such as fate,

courage [*cf.* Wierzbicka 1997], democracy, human rights or justice [see Bartmiński 2006 for an analysis of these and additional terms]. An activity should begin with an act of miscommunication focusing on a concept similar to those mentioned above, presented in a narrative or dramatic form. Next, students should be provided with multiple-choice questions showing the potential causes for the misunderstanding and then with feedback paragraphs giving additional information about the interlocutors' linguistic worldviews. Another type of cultural problem solving, hypothesis refinement, which consists of research-oriented activities, may help learners to refine their initial understanding of an aspect of the linguistic worldview typical of the target culture, such as self-stereotypes, heterostereotypes or one's place in the world [*cf.* Bartmiński 2009b: 208–225].

Behavioral and affective approaches, i.e. the auto-motor unit, dramatization, role-play, simulation and mini-drama, are very helpful techniques when it comes to sensitizing learners to culture-specific verbal and nonverbal exchanges and actions. The auto-motor unit, containing oral commands to which students react physically, may be used to pantomime for example greetings or farewells [*cf.* Jarząbek 1994]. Dramatization, role-play and simulation may encourage learners to explore culture-bound gestures, facial expressions and postures that tend to accompany certain phraseological units or metaphors in the L2 [*cf.* Kozak 2007]. The mini-drama, consisting of several skits or episodes presenting an act of miscommunication between interlocutors who represent different cultures, is an excellent way to make learners aware of discrepancies between conceptualizations of "cultural keywords" [Wierzbicka 1997]. These may include any of the above mentioned value terms [see also Bartmiński 2006].

Cognitive approaches, including lectures, readings, discussions, debates and research techniques, represent "academic" [Stern 1992] teaching methods. To acquaint students with different linguistic worldviews, teachers are advised to draw inspiration for lectures, discussions and debates from the current ethnolinguistic research [see Bartmiński 2009a, 2009b for an introduction to ethnolinguistics]. Readings may include newspaper articles [*cf.* Zarzycka 2005; Zinken 2000], literature [*cf.* Trześniowski 2006; Winiarska 2008; Wysocka 2002] and ethnolinguistic books [e.g. Wierzbicka 1997], adjusted to students' language proficiency and needs. Research techniques should focus on linguistic worldview reconstruction by examining dictionaries, newspaper articles, corpus data and questionnaire responses [see Bartmiński & Chlebda 2008]. They may encourage students, for example, to compile their own cognitive definitions of chosen terms [Bartmiński 2009a: 67–75].

To obtain questionnaire responses about the meaning of words, which tend to be considered the most valuable source of linguistic worldview infor-

mation, learners are strongly recommended to take advantage of real-life contacts in the form of pen-pals, tape-pals, computer-mediated encounters, interviews with native informants/tandem partners and audiotaped/videotaped interviews [*cf.* Bartmiński 2006]. A similar aim may be achieved during visits to other countries, which may be connected to ethnographic projects. Such projects ideally involve "extended participation in the everyday lives of a group; a detailed description of some aspects of these lives [...] and an analysis and interpretation of them through the eyes and ears of the 'outsider'" [Roberts *et al.* 2001: 30]. Thus, students play the double role of participants and observers, which enables them to be close and, at the same time, to keep a distance. Typically, doing ethnography involves a combination of the following stages: (i) a preparation phase, i.e. acquiring background knowledge in the classroom, (ii) a fieldwork phase with time for observation and data collection and (iii) a presentation phase with time for reflection as well as analyzing and presenting data from a newly acquired perspective [*cf.* Parsons & Junge 2001: 205].

It must be stressed that recent studies [e.g. Bandura 2007; Maciejewska-Stępień 2010] provide evidence that developing ICC in the foreign language classroom is perceived as truly relevant to students' current needs. Therefore, it follows that the European policy on language, culture and education should continue strengthening the intercultural dimension to stay in line with one of the fundamental principles defined by the Council of Europe with regard to modern languages and cultures "that the rich heritage of diverse languages and cultures in Europe is a valuable common resource to be protected and developed" and for this reason "a major educational effort is needed to convert that diversity from a barrier to communication into a source of mutual enrichment and understanding" [Council of Europe 2001: 2]. In this respect, the European policy should inspire the educational policies of particular countries to encourage educational institutions, teachers and materials writers to integrate the intercultural component into their curricula. It is hoped that the above suggestions for enriching the teaching procedures for ICC development will serve as a useful contribution in such undertakings.

References

Bandura, Ewa [2007], *Nauczyciel jako mediator kulturowy*. Kraków: Tertium.

Bartmiński, Jerzy (ed.) [2006], *Język – wartości – polityka. Zmiany rozumienia nazw wartości w okresie transformacji ustrojowej w Polsce. Raport z badań empirycznych.* Lublin: Wydawnictwo UMCS.

Bartmiński, Jerzy [2007], *Opozycja „swój/obcy" a problem językowego obrazu świata*. Etnolingwistyka 19, 35–59.

Bartmiński, Jerzy [2009a], *Aspects of Cognitive Ethnolinguistics.* London: Equinox Publishing Ltd.

Bartmiński, Jerzy [2009b], *Językowe podstawy obrazu świata*. Lublin: Wydawnictwo UMCS.

Bartmiński, Jerzy [2009c], *Stereotypy mieszkają w języku: Studia etnolingwistyczne*. Lublin: Wydawnictwo UMCS.

Bartmiński, Jerzy / Chlebda, Wojciech [2008], *Jak badać językowo-kulturowy obraz świata Słowian i ich sąsiadów?* Etnolingwistyka 20, 11–27.

Bartmiński, Jerzy / Lappo, Irina / Majer-Baranowska, Urszula [2002], *Stereotyp Rosjanina i jego profilowanie we współczesnej polszczyźnie*. Etnolingwistyka 14, 105–151.

Berwald, Jean-Pierre [1987], Teaching foreign languages with realia and other authentic materials. *ERIC Clearinghouse on Languages*, 1–6.

Borowiecka, Ewa [2003], *Uczciwość.* In: Bartmiński, Jerzy (ed.), *Język w kręgu wartości. Studia semantyczne.* Lublin: Wydawnictwo UMCS.

Brzozowska, Dorota [2007], *Polska, Ameryka i Europa. Dowcipy o wojnie w Iraku*. Etnolingwistyka 19, 225–238.

Buława, Monika [2003], *Postęp i sprawiedliwość społeczna w tygodniku konserwatywno-liberalnym „Najwyższy Czas!"*. In: Bartmiński, Jerzy (ed.), Język w kręgu wartości. Studia semantyczne. Lublin: Wydawnictwo UMCS.

Byram, Michael [1997], *Teaching and Assessing Intercultural Communicative Competence.* Clevedon: Multilingual Matters.

Byram, Michael / Fleming, Micheal (eds.) [1998], *Language Learning in Intercultural Perspective: Approaches through Drama and Ethnography.* Cambridge: Cambridge University Press.

Byram, Michael / Nichols, Adam / Stevens, David (eds.) [2001], *Developing Intercultural Competence in Practice*. Clevedon: Multilingual Matters.

Byram, Michael / Zarate, Geneviève [1994], *Definitions, Objectives and Assessment of Socio-cultural Competence.* Strasbourg: Council of Europe.

Canale, Michael / Swain, Merrill [1980], *Theoretical bases of communicative approaches to second language teaching and testing*. Applied Linguistics 1, 1–47.

Carel, Sheila [2001], *Students as virtual ethnographers: Exploring the language-culture connection*. In: Byram, Michael / Nichols, Adam / Stevens, David (eds.), Developing Intercultural Competence in Practice. Clevedon: Multilingual Matters, 146–161.

Chastain, Kenneth [1988], *Developing Second-language Skills: Theory and Practice.* San Diego, CA: Harcourt Brace Jovanovich, Publishers.

Chlebda, Wojciech [2002], *Polak przed mentalną mapą świata*. Etnolingwistyka 14, 9–26.

Chlebda, Wojciech [2007], *Kiedy swój staje się obcym*. Etnolingwistyka 19, 89–98.

Council of Europe [2001], *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Davis, Alan [1991], *The Native Speaker in Applied Linguistics.* Edinburgh: Edinburgh University Press.

Ek, Jan A. van [1986], *Objectives for Foreign Language Learning*, vol. 1: Scope. Strasbourg: Council of Europe.

European Commission [1995], *White Paper on Education and Training – Teaching and Learning – Towards the Learning Society* [online] [cited: 30 January 2012]. Access: http://ec.europa.eu/languages/documents/doc409_en.pdf.

Grzegorczykowa, Renata [2008], *Od wspólnoty do obcości. Rozwój znaczeniowy polskiego przymiotnika „obcy"*. Etnolingwistyka 20, 39–50.

Hendon, Ursula S. [1980], *Introducing culture in the high school foreign language class*. Foreign Language Annals 13/3, 191–199.

Hughes, George H. [1984], *An argument for culture analysis in the second language classroom*. In: Valdes, Joyce Merrill (ed.), Culture Bound: Bridging the Cultural Gap in Language Teaching. Cambridge: Cambridge University Press, 162–169.

Hymes, Dell [1972], *On communicative competence*. In: Pride, John B. / Holmes, Janet (eds.), Sociolinguistics. Harmondsworth: Penguin Books, 269–293.

Jarząbek, Krystyna [1994], *Kinetyczne formy powitań i pożegnań*. Etnolingwistyka 6, 65–81.

Kołodziejek, Ewa [2007], *Swoi i obcy w społecznej przestrzeni subkultur*. Etnolingwistyka 19, 133–141.

Kozak, Katarzyna [2007], *Polska frazeologia gestu: od informacji somatycznej do idiomu*. Etnolingwistyka 19, 253–265.

Kramsch, Claire [1993], *Context and Culture in Language Teaching.* Oxford: Oxford University Press.

Kramsch, Claire [1998], *The privilege of the intercultural speaker*. In: Byram, Michael / Fleming, Michael (eds.), *Language Learning in Intercultural Perspective: Approaches through Drama and Ethnography.* Cambridge: Cambridge University Press, 16–31.

Lafayette, Robert C. [1978], *Teaching Culture: Strategies and Techniques.* Arlington, VA: Center for Applied Linguistics.

Lippmann, Walter [1922], *Public Opinion*. New York: The Macmillan Company.

Lund, Steven [1992], *Giving your courses a dose of reality*. ELT Forum 30/3, 10–15.

Maciejewska-Stępień, Ewa [2010], *Intercultural communication training in the European context: An evaluation of a Socrates project*. In: Lewandowska-Tomaszczyk, Barbara / Pułaczewska, Hanna (eds.), Intercultural Europe: Arenas of Difference, Communication and Mediation. Stuttgart: ibidem-Verlag, 323–345.

Meade, Betsy / Morain, Genelle [1973], *The culture cluster*. Foreign Language Annals 6/3, 331–338.

Mikoś, Michał Jacek / Tieszen, Bożena [2005], *Polak w oczach studentów amerykańskich*, Etnolingwistyka 17, 233–240.

Niebrzegowska-Bartmińska, Stanisława [2007], *Wzorce tekstów ustnych w perspektywie etnolingwistycznej*. Lublin: UMCS.

Omaggio, Alice C. [1986], *Teaching Language in Context: Proficiency-oriented Instruction.* Boston, MA: Heinle & Heinle Publishers, Inc.

Paikeday, Thomas M. [1985], *The Native Speaker is Dead!* Toronto: Paikeday Publishing Inc.

Parker, Gabrielle / Rouxeville, Annie (eds.) [1995], *The Year Abroad: Preparation, Monitoring, Evaluation.* London: AFLS/CILT.

Parsons, Judith / Junge, Peter [2001], *"Why do Danes put their elderly in nursing homes?" – Working outside the classroom with adult second language learners*. In: Byram, Michael / Nichols, Adam / Stevens, David (eds.), Developing Intercultural Competence in Practice. Clevedon: Multilingual Matters, 203–216.

Phillipson, Robert [1992], *Linguistic Imperialism.* Oxford: Oxford University Press.

Pietrucha, Monika [2003], *Profile pojęcia demokracji we współczesnym języku polskim*. In: Bartmiński, Jerzy (ed.), *Język w kręgu wartości. Studia semantyczne.* Lublin: Wydawnictwo UMCS.

Roberts, Celia [1995], *Language and cultural learning: An ethnographic approach*. In: Jensen, Annie Aarup / Jäger, Kirsten / Lorentsen, Annette (eds.), *Intercultural Competence.* Vol. 2: *The Adult Learner.* Aarbolg: Aarlborg University Press, 53–69.

Roberts, Celia [1997], *The year abroad as an ethnographic experience*. In: Byram, Michael (ed.), Face to Face: Learning 'Language-and-culture' through Visits and Exchanges. London: CILT, 62–76.

Roberts, Celia / Byram, Michael / Barro, Ana / Jordan, Shirley / Street, Brian [2001], *Language Learners as Ethnographers.* Clevedon: Multilingual Matters.

Sapir, Edward [1978], *Kultura, język, osobowość. Wybrane eseje.* Warszawa: PIW.

Seelye, H. Ned [1984], *Teaching Culture.* Lincolnwood, IL: National Textbook Company.

Skibińska, Elżbieta [2005], *Obraz Polaka i Rosjanina w języku francuskim i w świadomości francuskiej młodzieży*. Etnolingwistyka 17, 213–232.

Stern, Hans Heinrich [1992], *Issues and Options in Language Teaching.* Oxford: Oxford University Press.

Taylor, J. S. [1970], *Direct classroom teaching of cultural concepts*. In: Seelye, H. Ned (ed.), Perspectives for Teachers of Latin American Culture. Springfield, IL: State Superintendent of Public Instruction, 224.

Taylor, H. Darrel / Sorenson / John L. [1961], *Culture capsules*. Modern Language Journal 45, 350–354.

Trześniowski, Dariusz [2006], *Doświadczenie kobiecości w tekstach młodopolskich poetek*. Etnolingwistyka 18, 329–348.

Underhill, James W. [2009], *Humboldt, Worldview and Language.* Edinburgh: Edinburgh University Press Ltd.

Ware, Paige Daniel [2003], *From Involvement to Engagement in Online Communication: Promoting Intercultural Competence in Foreign Language Education*. Unpublished Ph.D. dissertation, University of California, Berkeley.

Whorf, Benjamin Lee [1982], *Język, myśl i rzeczywistość.* Warszawa: PIW.

Wierzbicka, Anna [1997], *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese.* Oxford: Oxford University Press.

Winiarska, Justyna [2008], *Konceptualizacje JA i DRUGIEGO w filozoficznym i poetyckim dyskursie miłosnym*. Etnolingwistyka 20, 179–189.

Woodin, Jane [2001], *Tandem learning as an intercultural activity*. In: Byram, Michael / Nichols, Adam / Stevens, David (eds.), *Developing Intercultural Competence in Practice.* Clevedon: Multilingual Matters, 189–202.

Wysocka, Aneta [2002], *Językowy obraz Afrykanina*. Etnolingwistyka 14, 175–196.

Zarzycka, Grażyna [2005], *Sposoby określania osób czarnoskórych w prasie lokalnej*. Etnolingwistyka 17, 241–257.

Zinken, Jőrg [2000], *„Nie mam nic przeciwko tym obcym, ale ci obcy nie są stąd". Konceptualizacja obcości w dyskursie prasowym w Polsce i w Niemczech*. Etnolingwistyka 12, 53–62.

Żuk, Grzegorz [2004], *Polska i Unia Europejska w tekstach humorystycznych*. Etnolingwistyka 16, 149–172.