

PHONOLOGICAL AND SEMANTIC MARKERS IN THE SUBSTANDARD LEXICON

DANKO ŠIPKA

Adam Mickiewicz University, Poznań

1. Introduction

This paper is an attempt to identify both phonological and semantic markers of a substandard lexicon, as well as to propose both a more general taxonomy of these markers and a model of their lexicographic treatment. The materials used are derived from the 25,000-entry SerboCroatian-English Colloquial Dictionary Project completed in July 2000 at the Language Research Center of the Washington DC metro area company called MRM/McNeil. The resulting dictionary is available at: <http://www.dunwoodypress.com>.

2. Methodology and Conceptual Map

The methodological framework of this analysis rests on both cognitive-linguistic methodology (see Lakoff and Johnson 1980, Langacker, 1991 and 1998) and a cross-cultural approach as represented in Pike (1996) and Wierzbicka (1992). In particular, the following elements of the cognitive-linguistic and cross-cultural methodology have been used in the dictionary and in this analysis:

- a. The binary, phonological-semantic model of the linguistic sign and the negation of grammatical autonomy as presented in Langacker (1998: 11): “Both lexicon and grammar are claimed to be fully and properly describable as assemblies of symbolic structures [...] In cognitive grammar, a *symbolic structure* is defined as consisting in the association between a *semantic structure* and a *phonological structure* – its semantic and phonological *poles*. Every lexical and grammatical element is thus attributed some kind of semantic and phonological value, whether specific or schematic.”
- b. The negation of semantic autonomy: “Rather than imposing artificial boundaries, it [cognitive semantics] posits a gradation between semantics and

pragmatics, and also between linguistic and general knowledge” (Langacker 1998: 3),

- c. The role of metaphor in language, as presented in Lakoff and Johnson (1999), and earlier in Lakoff and Johnson (1980),
- d. The idea that “In natural language meaning consists in human interpretation of the world. It is subjective, it is anthropocentric, it reflects predominant cultural concerns and culture-specific modes of social interaction as much as any objective features of the world ‘as such’”, Wierzbicka (1988:2), reiterated in Wierzbicka (1992).

Linguistic units and processes modelled in terms of cognitive and cross-cultural linguistics, it is argued here, are the markers of certain identities in the substandard sphere. We therefore need to elaborate both on the notion of linguistic identity and on the concept of the substandard.

3. Linguistic Identities and the Substandard

It is a well-known phenomenon that the use of language builds an identity. The levels of identity relevant for this discussion can be summarized as follows:

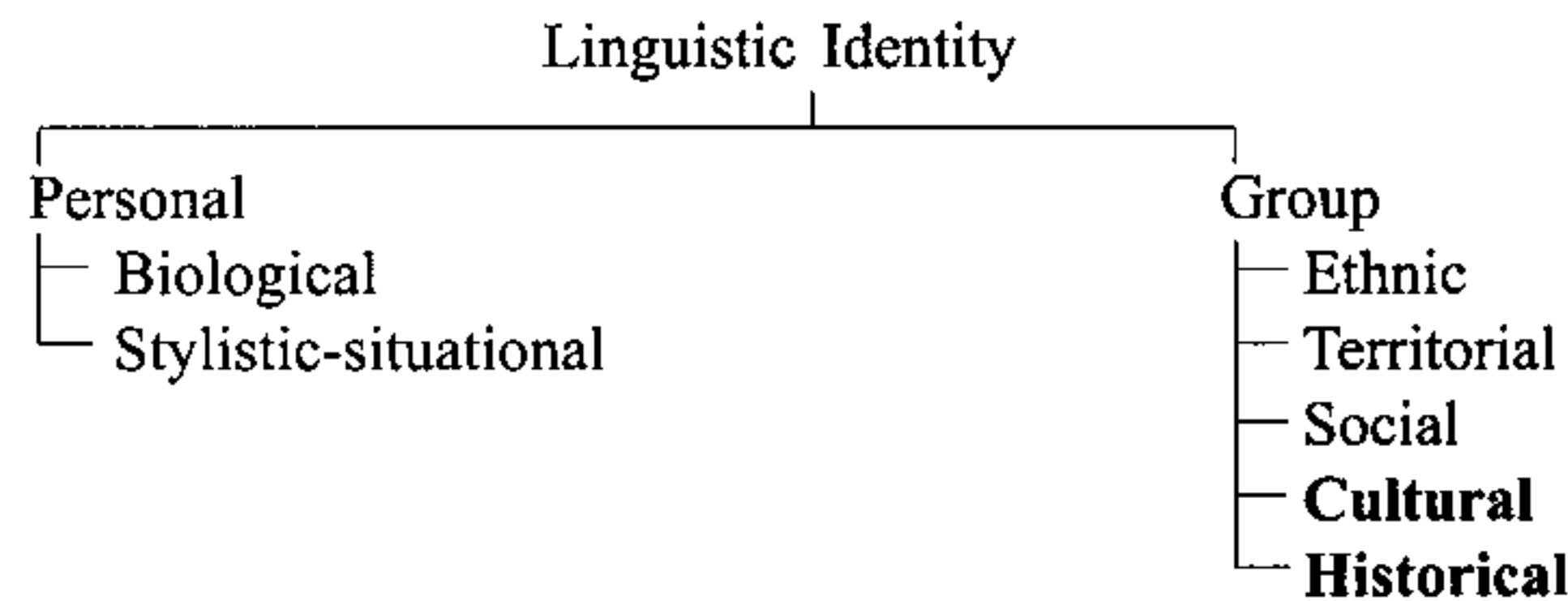


Diagram 1. Types of Linguistic Identity

Biological identity comprises the features such as sex, age, etc. which are independent of a person’s will and which can exist independently of specific groups of people. Stylistic-situational identity pertains to the behavioral patterns of a person, in general or in particular situations, which are also independent of groups.

The first three categories of group identity are well elaborated in literature (even in such introductory texts as Crystal 1987). However, the last two categories are seldom mentioned. Cultural identity comprises features that are dependent upon a cultural sphere (for example “European”, “Slav”, “inhabitant of the Balkans”, etc.). Historical identity pertains to features conditioned by a specific historical period, with its attendant fashion or ruling ideology (for example “Eastern European”).

All categories of identity can be marked with an array of linguistic means. For example, territorial identity can be marked by use of a phoneme, morpheme, syntactic construction, or lexeme which cannot be found in the standard language.

Substandard, as defined here, pertains to the repertoire of linguistic units, which because of either its semantic or phonological component, is incompatible with the standard-language discourse. The substandard is thus defined in a negative manner, relative to the standard. The following two quotations from AHD (1992: s.v.) express other possible terms and approaches that can be used to refer to this field or its subfields.

nonstandard [...] *Usage Note:* The term *nonstandard* was introduced by linguists and lexicographers to describe usages and language varieties that had previously been labeled with terms such as *vulgar* and *illiterate*. *Nonstandard* is not simply a euphemism but reflects the empirical discovery that the varieties used by low-prestige groups have rich and systematic grammatical structures and that their stigmatization more often reflects a judgment about their speakers rather than any inherent deficiencies in logic or expressive power. Note, however, that the use of nonstandard forms is not necessarily restricted to the communities with which they are associated in the public mind. Many educated speakers freely use forms such as *can't hardly* or *ain't I* to set a popular or informal tone. Some dictionaries use the term *substandard* to describe forms, such as *ain't*, associated with uneducated speech, while reserving *nonstandard* for forms such as *irregardless*, which are common in writing but are still regarded by many as uneducated. But *substandard* is itself susceptible of disparaging interpretation, and most linguists and lexicographers now use only *nonstandard*, the practice followed in this Dictionary.

dialect [...] *Synonyms:* *dialect, vernacular, jargon, cant, argot, lingo, patois*. These nouns denote forms of language that vary from the standard. *Dialect* applies to the vocabulary, grammar, and pronunciation characteristic of a specific geographic area or locality. The *vernacular* is the everyday language spoken by a people as distinguished from the literary language. *Jargon* is the specialized language used by a social or occupational group but not understood by the general public. *Cant* now usually refers to the specialized vocabulary of a group or trade and is often marked by the use of stock phrases. *Argot* applies especially to the language of the underworld or, by extension, to that of any specific group. *Lingo* is applied, often humorously or contemptuously, to language that is unfamiliar or so specialized that it is difficult to understand. *Patois* refers especially to a regional dialect without a literary tradition or to a creole.

4. Substandard Phonological and Semantic Markers

In addition to marking a linguistic unit as substandard, substandard phonological and semantic elements also mark one’s identity.

There are two possible relations of the substandard markers to the standard language. It is possible that the substandard phonological or semantic unit is simply different from its standard counterpart, the substandard nature of the unit or any given identity is marked by that difference. On the other hand, the substandard unit can be derived by modifying the standard one. In this case the substandard unit presupposes the existence of the standard one; i.e., the standard form is both temporally and logically precedent to the substandard one. The first relation can be called parallel or concurrent, and the second is known as sequential marking of the substandard. The following table demonstrates these points:

Table 1. Parallel and sequential relationships between substandard and standard.

Parallel		Sequential		relation
↑ standard	↑ substandard	substandard ↑ standard		
standard	substandard	standard	substandard	
četiri 'four'	ćetri 'four'	zdravo 'hi' zdra-vo (1-2)	vozdra 'hi' vo-zdra (2-1)	examples – phonological
upasti 'fall into'	upasti 'fall'	mrkva 'carrot'	mrkva 'redhead'	examples – semantic
		<i>Redheads look like carrots.</i>		

Obviously, lexical borrowing in the substandard is a subclass of the parallel relation type.

As we can see, the relationships between standard and colloquial lexemes are not merely ones of neighboring and exclusion. Many nonstandard lexemes assume the existence of the standard ones and vice versa. This fact is particularly true for the semantic extensions, which have been identified in 66% of cases in this dictionary. Building identities using phonological markers is rather simple, both lexicologically and lexicographically, and a comprehensive list of these markers can be found in Šipka (2000: Introduction). I will therefore focus on the semantic markers in the remainder of this section.

The semantic extensions identified in this project point to the three relevant distinctions presented in Table 2; each will be explained in turn:

Table 2. Types of semantic markers.

1. internal nonstandard	standard-nonstandard
2. internal semantic	phonetic-semantic
3. identity-bound	more general

The first category above points to the fact that a nonstandard semantic extension can have its source domain in nonstandard or in standard lexicon. The first type of extension can be seen in *drkati* 'obscene for: to masturbate' with its destination domain 'fiddle with, touch or play with without any reason or purpose' (both meanings are nonstandard); or *eki* 'pragmatic operator used to express surprise' and 'question tag' (both meanings are nonstandard). The second situation can be illustrated by:

devojka 'young female' (standard) and 'prostitute' (nonstandard) or standard *taster* 'key (e.g., on a keyboard)' and nonstandard 'informer'.

The first value within the second category in Table 2 has been thoroughly elaborated in numerous studies. Two examples will suffice to demonstrate the point. First, the metaphor "A remote place is like a hole" in an internal semantic extension *rupa* 'hole' – 'remote place, hellhole'. Second, "Redheads look like carrots" which is behind the extension *mrkva* 'carrot' and 'redhead'.

The second value is seldom mentioned in the literature. This concept covers the situations where the partial similarity of forms triggers semantic extension. Thus, nonstandard Serbo-Croatian has a series of lexemes to denote a homosexual with the following source domains: *pedala* 'pedal', *pedikir* 'pedicure', *pešak* 'pedestrian', *peškir* 'towel', *Peđa* 'proper name', *Pežo* 'Peugeot, a car brand'. The only link between these meanings and 'male homosexual' is that they are phonologically similar to *peder*, the most common nonstandard word for a male homosexual.

Finally, the third category deals with identity. The extension can be identity-bound or more general. Table 3 provides several examples of identity-bound metaphoric extensions:

Table 3. Identity-bound semantic extensions

	lexeme	source domain	link	target domain
1.	<i>krmak</i>	'boar' (the animal)	non-Muslim is like a boar	'non-Muslim'
2.	<i>leš</i>	'corpse'	senior citizen is already a corpse	'senior citizen'
3.	<i>pop</i>	'priest'	homosexual is like a priest (no wife)	'homosexual'
4.	<i>pička</i>	'vagina'	what matters to men about a girl is her vagina	'girl'

The first example elucidates an identity determined by one's religious/ethnic affiliation (e.g., "Bosnian Muslim") and religious practices (e.g., the Muslim diet excludes pork). The second example focuses on identity determined by one's inclusion in an age group, and the third describes an identity formed by the Catholic religion (i.e., celibacy is compulsory for Catholic priests), and finally (hetero)sexuality is the basis for the fourth example. These lexemes in their extended senses have been used within their groups of people with specific connotative values. For example, the lexeme may be used in an offensive sense, as in *Ne mogu tog krmka na oči!* 'I cannot even look at that boar/non-Muslim'; or in a facetious one, as in *On je pop pa ga zato ne zanimaju pičke.* 'He's a priest/homosexual and that's why he is not interested in vaginas/girls'.

On the other hand, words such as *rasturiti*, ‘destroy’, which extends into ‘win’; or *plakati*, ‘cry’, which extends into ‘complain’, have a much wider cross-cultural distribution owing to the simple fact that they are based on certain (almost) universal cognitive and affective faculties of the human species.

The (sub)cultural webs which determine any particular semantic extension can be very intricate and non-apparent for the individuals outside that given (sub)cultural milieu. The lexeme *fataljist* ‘person who cleans parks’, for example, is determined by the following linguistic facts: a) the existence of a nonstandard verb *fatati* ‘to catch’, b) the existence of the noun *list* ‘leaf’ with Albanian interference, and, c) the existence of the noun *fatalist* ‘fatalist’. On the other hand, there is a common perception that Albanians have performed such jobs in the former Yugoslavia (therefore “b”, above). The lexeme *fatalist* is changed into “Albanian” *fataljist*, re-segmented into: *fata-ljist* ‘catches leaves’ but still remains present (i.e., a person with that social-economic status is perceived as a fatalist). The corpus of this project contains numerous examples with such intricate networks. This example is also a marker of specific historical identity (the existence of the former Yugoslavia and various ethnic groups living together in it). Facetious contaminations for automobiles from the former Soviet-block countries like *varburđini* (Wartburg + Lamborghini) ‘Wartburg (a brand of car)’, *škodilak* (Škoda + Cadillac) ‘Škoda (a brand of car)’ are fully understandable only in the context of pre-1989 realities and the corresponding identities.

As we can see from the categories presented, the semantic units have been interwoven with complex networks of knowledge and identity. At the same time, the semantic side of the sign is not only coexistent with its phonological counterpart, but in some situations the phonological side can trigger the semantic change. This brings us back to the cognitive-linguistic negation of both grammatical and semantic autonomy mentioned at the outset of this paper.

The above-mentioned categories constituted a challenge to the lexicographic treatment in this dictionary and required an adequate response. The solutions employed in this dictionary project will be discussed in the next section.

4. A Proposal of Lexicographic Treatment

The lexicographic treatment of the substandard phonological and semantic markers should provide solutions to the following:

- a. all identities marked by either phonological or semantic side of the lexeme should be identified,
- b. all markers should be identified,
- c. the links between standard and substandard phonological units should be accounted for, in particular for the sequential substandard markers,
- d. the markers should be elucidated for the users of the dictionary who may have different identities (primarily English-speaking ones).

As far as marking the identities is concerned, ethnic, territorial, and social identities have been identified using pertinent labels (tags), such as: *Ser* – Serbian (ethnic identity), *Vojv* – Vojvodinian (territorial identity), and *sl* – slang (social identity). A comprehensive list of these labels is available in Šipka (2000: Introduction). Marking cultural and historical identities required additional explanation in the dictionary entry.

For example, the phrase *ko kamen za kupus* ‘big in size’ (literally: “like the sauerkraut stone”) is a marker of cultural identity. It points to a certain procedure and artifact within the culture of the Balkans. The following explanation was required for the English-speaking user of the dictionary:

ko kamen za kupus PP (lit. like the sauerkraut stone, i.e., the stone traditionally put on the top of the heads of cabbage in a barrel with salty water: sauerkraut is not only shredded but is also made from whole heads of cabbage) [...]

Similarly, the entry for historically marked entries, such as *Škodilak*, required the following explanation:

Škodilak ... (lit. Škoda: a simple Eastern European car + Cadillac)...

The tasks of identifying the markers, relating them to the standard, and elucidating them for the English users of the dictionary have been divided between the introductory text to the dictionary and its main body. It is obvious that not all instances of phonological markers could be included in the dictionary. Practically any lexeme can have altered phonological form (different phonemic value, missing phonemes, additional phonemes, permutations, etc.) These markers are therefore listed in the introductory text. The most frequent lexemes have been included in the dictionary, and their entries also contain the standard language form, for example:

ćega ... [std. gaće] underpants, underwear <Slav!>

From this example we can see that the substandard form is a permutation of the standard form (a common permutation whereby the first syllable has been moved to the final position in the word). The etymological label (here *Slav* – common Slavic inherited into SerboCroatian) contains the exclamation sign, which signals the transformation of the form.

Semantic markers are treated in a similar manner. In the example

dugotalasan ... (lit. long-wave) slow in understanding <Slav*1>

the standard language meaning is provided in the entry and the semantic extension is marked in the etymological label (the meaning has one semantic extension).

The distribution for both phonological and semantic markers has been computed using the data from the etymological label; descriptive statistic data are available in Šipka (2000: Introduction). These data were also the basis for qualitative accounts of

the markers, such as identifying most common metaphorical links, which are also presented in the introductory text.

To summarize, the proposed lexicographic treatment advocates the following policies to cope with the challenge of marking the substandard phonological and semantic markers:

- a. Introductory text to the dictionary should be given a more prominent role. It should become a substantial part of the lexicographic treatment.
- b. Various labels (ethnic, regional, social, etc.) should be used amply.
- c. An additional entry field 'standard' should be introduced and the standard form should be provided in all instances of phonological substandard markers.
- d. An additional entry field 'literal' should be introduced to provide the source domain for any standard-substandard metaphors. Additional explanations for any identity-based semantic extensions should also be included.
- e. Etymological labels should also register phonological and semantic modifications.

As can be seen, the lexicographic treatment in the substandard projects requires an entry structure which differs from that found in general bilingual dictionaries.

REFERENCES

- AHD. 1992. *The American heritage dictionary of the English language* (3rd ed.). Boston: Houghton Mifflin Company.
- Crystal, D. 1987. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Lakoff, G. and M. Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G. and M. Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Langacker, R. W. 1991. *Concept, image, and symbol. The cognitive basis of grammar*. Berlin: Mouton de Gruyter.
- Langacker, R. W. 1998. "Conceptualization, Symbolization, and Grammar". In Tomasello, M. (ed.). 1-39.
- Pike, K. L. 1996. *Talk, thought, and thing*, SIL. <http://www.sil.org>.
- Šipka, D. 2000. *SerboCroatian-English colloquial dictionary: An exercise in cross-cultural cognitive linguistics*. Springfield: Dunwoody Press.
- Tomasello, M. (ed.). 1998. *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Leipzig: LEA
- Wierzbicka, A. 1988. *The semantics of grammar*. Amsterdam: John Benjamins.
- Wierzbicka, A. 1992. *Semantics, culture and cognition: Universal human concepts in culture-specific configurations*. Berlin: Mouton de Gruyter.