

Pragmatic Components in the Slovene Lexical Database Meaning Descriptions

Mojca Šorli

Trojina, Institute for Applied Slovene Studies
Cesta v Kleče 16, 1000 Ljubljana, Slovenia
E-mail: mojca.sorli@trojina.si

Abstract

The possibility to analyse vast amounts of linguistic data has brought about changes both in methodology as well as in the ways we perceive certain language phenomena. A key insight gained by computational methods in language analysis is undoubtedly the importance of lexical co-occurrence and usage patterns for the description of lexical meaning. Corpus analysis and new methods in the analysis of pragmatic components of meaning have also yielded significant results in areas such as the treatment of semantic prosody. The present paper does not focus on what is traditionally subsumed under connotation or the speaker's attitude (e.g., swear words, pejorative and offensive language, praise, excuses, requests, demands, etc.), but on ways in which the pragmatic (functional) meaning that arises from various contextual features can become an integral part of lexicographic descriptions. This is important for the treatment of all of those lexical items whose meanings reside in their function rather than in their bare lexical-semantic meaning, as this is particularly the case with phraseology and idiomatics. From another perspective, pragmatics turns out to be an effective means of sense discrimination in works of lexical and lexicographic relevance, as will be shown in the continuation.

Keywords: lexicographic description; lexical database; pragmatics

1. Introduction

The Slovene Lexical Database (hereafter SLD)¹ is a monolingual lexical resource that will provide a corpus-driven account of the core vocabulary of the Slovene language, including semantic, syntactic, collocational and phraseological information, supported by illustrative examples. Data from the existing 620-million-word FidaPlus reference corpus of Slovene² is recorded and lexicographically treated using the Sketch Engine corpus query tool. The SLD is similar to other modern electronic databases³ in that it is constructed on the principles of lexicogrammar, but it gives more overall importance to meaning descriptions.⁴ Explanations are compiled drawing on those lexicographical practices that appear to be successful in terms of either the rigour of their lexicogrammatical approach or their user-friendliness, or both. Syntagmatics and meaning descriptions have been given more prominence, perhaps intentionally somewhat at the expense of the description of the inherent features and paradigmatic associations of

words.⁵ In principle, the main goal of a lexical database is not to produce (finalised) meaning descriptions, but to create lexical profiles of words by describing lexicographically relevant information (Atkins; Fillmore; Johnson, 2003); nonetheless, in the SLD, a great deal of thought has been given to the nature and form of lexicographic description as it should be provided in dictionaries, particularly those for upper elementary and intermediate school goers. This is partly due to the specific conditions in the Slovene language community – a small dictionary market and fairly limited human and financial resources in the field of lexicography. With these goals in view, in setting the guidelines for meaning descriptions we have considered corpus-based and pragmatically aware contemporary monolingual dictionaries:⁶ MEDAL, LDOCE, and COBUILD are quoted below in order to provide comparison.

¹ Slovene Lexical Database (2008-2012): The database's operation is co-financed by the European Union, the European Social Fund, and the Ministry of Education and Sport of the Republic of Slovenia. It presently contains 2,500 entries (<http://www.slovenscina.eu/Vsebine/En/Aktivnosti/LeksikalnaBaza.aspx>).

² www.fidaplus.net

³ The SLD is close in scope and methods to the recently compiled DANTE database: <http://www.webdante.com/>.

⁴ *Meaning description* is used with reference mainly to the SLD, while *definitions* are referred to as products of particular (past or future) lexicographic traditions. Throughout the article, *explanation* is used analogously to *meaning description*, in its broadest possible sense.

⁵ According to Atkins, Fillmore and Johnson (2003), a complete description of the lexicographically relevant information required for the proper analysis of a keyword would have to include the word's *inherent features* (part of speech class and subclass, semantic type, etc.), its *pragmatic features* (information about users and user communities, contexts of use, emotional affect, etc.), its *paradigmatic associations* (synonymy, antonymy, meronymy, etc.), and its *syntagmatic or combinatorial features* (information about the context a word creates or satisfies, expressed in terms of grammatical and semantic phrase types and lexical collocations).

⁶ Namely, the Macmillan Dictionary and Thesaurus: Free English Dictionary Online (MEDAL 2010), the Longman Dictionary of Contemporary English (LDOCE 2003) and the Collins COBUILD English Language Dictionary (1995).

1.1 The conceptual framework

Semi-automated electronic databases, such as FrameNet⁷ (Fillmore et al., 2003), Corpus Pattern Analysis (hereafter CPA)⁸ (Hanks, 2004) and the Cobuild Project⁹ (Sinclair, 1987) have all been studied in the construction of the SLD. FrameNet primarily builds ontologies and is concerned with the identification of semantic participants and argument structures by means of predetermined and largely formalised syntactic-semantic categories. It is therefore “more or less limited to recording information about the combinatorial requirements of the words it studies” (Atkins; Fillmore; Johnson, 2003). CPA both records the participant structure of a sentence pattern and provides a schematic explanation of the particular pattern (implicature) that establishes the relationships between the identified participants. The implicature does not pretend to be a dictionary definition, partly because it is ascribed to a particular pattern rather than to a conventional dictionary “sense”. Cobuild explanations, on the other hand, are characterised by the clarity and naturalness of the definition language, achieved by describing the meaning of the headword in terms of its typical syntactic patterns and the immediate context surrounding it, and “[u]nlike classical definitions Cobuild definitions make their headwords an integral part of mentioning them, and so deal with the meanings of the words being defined both as entities and activities” (Barnbrook, 2002: 19-20). Metalinguistic information, which traditionally had no place in dictionary descriptions, is now foregrounded (COBUILD: 495): If you **explain** something, you give details about it or describe it so that it can be understood.

1.2 The theoretical background - the Sinclairian lexicographic tradition

The Cobuild definition style is perhaps the most literal transfer into practice of what Halliday (2007: 26) summarises as follows: “In general, it is unwise to assume that meaning is captured in dictionary entries, in the definitions or explanations given against the words. Dictionary definitions can and should be informative and helpful, and, when well written, they provide a paraphrase or explanation of meaning. But the meaning is not necessarily fully contained or exhaustively captured within such a definition. This is not to say that meanings are vague or ethereal. Within the conventions of a particular language, meanings contrast with each other with considerable precision. Words do not mean whatever we want them to mean, but are governed by social convention. Nonetheless, we cannot assume, without qualification, that the wording of a dictionary definition is an ideal representation of what a word means.”

1.2.1 Meaning as event vs. meaning as entity

Within a wide range of reflection on the nature of meaning, there have been various attempts to define its complexities, motivated, among other things, by the need to explain language and the ways in which it is used to an average user. The theoretical framework for some of the modern views on pragmatics can be traced back to Piotrowski's (1989: 73-74) formulation: “Thus, on the one hand meaning can be seen as a sort of entity: concept, notion, prototype, stereotype, or fact of culture. On the other hand, meaning can be seen as a sort of activity: skill, knowledge of how to use a word.” From this understanding, the so-called “use-mention” dichotomy was derived, built on extensively by Sinclair et al. in the Cobuild project. Hanks adds complexity to the “use-mention” pair by claiming that “[d]ictionaries are much concerned with accounting for what it is that an utterer may expect a hearer to believe” (1987: 20). In his tribute to elegance in lexicography, Rundell (2010: 357) points out how the instability of word senses has long been observed by thinkers about words, summing up Hanks's conclusions on the issue as “it makes more sense to think of meanings as events rather than (as their treatment in dictionaries implies) independently existing entities” (ibid.).

1.2.2 Use and meaning – the metalinguistic approach

The analysis of instances of natural text has long shown that some words are more literally “used” to produce a desired effect, such as to convey the intentions of the speaker, than others; or, as Sinclair (1991: 126) puts it, “[t]he statement may be about what people mean when they use a word or phrase, rather than what the word or phrase means.” A well-known example employed by Sinclair to demonstrate how a restatement of meaning becomes a metalinguistic comment on the way the word is used in a context of situation is:

If you call a woman a bitch, you mean that she behaves in a very unpleasant way

vs.

derog. A woman, esp. when unkind or bad-tempered (Barnbrook, 2002: 178).

As Hanks (1987: 203) succinctly put it, “in the most common meaning of this word, what is at stake is the utterer's intention to insult, not the semantic convention associated with the sense. This meaning must be distinguished from the “literal” meaning, which although rather rare is privileged.”

2. The Meaning Descriptions in the Slovene Lexical Database

2.1 General principles

The descriptions are formed as full-sentence definitions,¹⁰ with a view to further lexicographical treatment for the purposes of general, and particularly

⁷ <http://framenet.icsi.berkeley.edu/>.

⁸ <http://nlp.fi.muni.cz/projekty/cpa/>.

⁹ J. M. Sinclair, 1987 – see References.

¹⁰ Pioneered by the COBUILD 1 project (1987).

student (upper elementary and intermediate school), monolingual audiences. The treatment of lexical data in the SLD sets out to describe individual lexical items, their meanings and usage by means of FrameNet-type “scenarios”, which includes defining the range of semantic and syntactic combinatorial possibilities (valencies of each word in each of its senses) (Kocjančič; Zaranšek, 2009). The distinguishing features of the SLD descriptions are: the entry headword is integrated into the definition in its natural context; the syntactic environment of the headword and its semantic participants (semantic roles) are manually annotated: obligatory participants are in block letters for purposes of semi-automated pattern retrieval; to complete the picture, a description of broader circumstances of meaning is provided. Subordinate to the level of argument structure are the levels of grammatical patterns or structures, and collocations; the general rule is not to bend the argument structure too much towards either of these groups of lexical information, but rather to formulate it as a summary of all of them (ibid.).

The SLD meaning descriptions, then, are not yet ready-made dictionary definitions, but rather semantically and pragmatically informed “implicatures” which provide a platform for further work on explanations and definitions tailored to the needs of specific target groups. The policy of identifying the argument structure for each lexical unit¹¹ and annotating the semantic roles within what, at the same time, have to be adequate and comprehensible (and elegant at that) explanations has resulted in a relatively unique definition language. The aim was to bring together the best of what modern lexical descriptions based on, or influenced by, the contributions of computational lexicography, such as FrameNet, CPA and the Cobuild project, have had to offer. Combining formalisation with efforts to produce intelligible and simple meaning descriptions has resulted in an occasional clash of emphasis, which has had to be resolved independently for each situation and with regard to the reference skills of an average end-user. The fact that the database is intended essentially both for dictionary compilers – who will be compiling a student dictionary – as well as for general users who might be interested in querying somewhat raw linguistic data has added to the challenge.

Based in part on the Cobuild definition taxonomy – which was primarily designed to serve the purposes of the computer processing and formalised accordingly – and on an analysis of the early SLD entries, a new definition taxonomy was built taking into account the specifics of the Slovene language. The main purpose of the taxonomy was as much to provide guidance within the broad spectrum of defining possibilities as to homogenise the choices the compilers were making.

¹¹ *Lexical unit* is used throughout as “a unit of meaning”, unless otherwise indicated.

2.1.1 Conciseness and simplicity

The descriptions follow the maxim that the words used in them will be either more precise or easier to understand than the headword is by itself (Barnbrook, 2002: 49). This generally means that rare, polysemous or difficult words, as well as figurative expressions, are avoided in the descriptions, which essentially aim to tell what we already know about the meaning. The concept of “exactness” is secondary in importance to intelligibility, which subsumes brevity, conciseness and simplicity:

a HUMAN **breathes** by taking AIR into his/her lungs and pushing it out again¹²
ČLOVEK ali ŽIVAL **dih**a tako, da potegne ZRAK v pljuča in ga nato spet potisne ven

Where there are indications in the corpus data that (pragmatic) circumstances contribute decisively to a particular sense, the headword may require a more extensive description:

an **argument** is a logically derived reason used in a debate to persuade the listeners or the opponents to support you

argument je logična utemeljitev stališča v razpravi, s katero skuša človek pridobiti naklonjenost poslušalcev ali prepričati nasprotnike

2.1.2 The syntactic-semantic description

Meaning descriptions in the SLD are schematically divided into two parts or levels:

- a) The participant structure: all of the identified participants and circumstances are assigned semantic types or semantic roles. Syntactic and semantic information is overtly marked in order to enable automatic retrieval of patterns of usage. The assumption is that each meaning is realised within a syntactic pattern consisting of all of the words, expressions and situations in the co-text that contribute decisively to the meaning of a lexical unit.
- b) “The scenario” is the level of description that states the general situation of meaning, the relationships between the participants and other sense-discriminating, particularly pragmatic, components of meaning (Gantar et al., 2009: 108).

Obligatory vs. optional elements

Each meaning description includes all of the participants and circumstances that are needed to construct a particular meaning. In the process of identifying obligatory elements, we have also dealt with cases of

¹² In the paper all of the examples from the SLD are translated into English to aid understanding and listed first. Although sometimes awkward, the translations are intentionally as literal as possible, so that the organisation of the original descriptions remains evident.

null instantiation, but we will not further elaborate on their treatment here. Block letters are used to mark the obligatory participants, while the remaining information appears in lower case (see below) and is considered as part of the “scenario” (in italics). Participants are identified as obligatory if in at least some contexts they are syntactically (or contextually) expressed, i.e., their instantiations are to be found in the corpus data:

if a HUMAN **beseeches** another HUMAN, s/he begs them to help him/her out of a DIFFICULT SITUATION or to do something that means a lot to him/her
 če ČLOVEK **roti** drugega ČLOVEKA, ga obupano prosi, da mu pomaga iz STISKE ali stori nekaj, kar mu veliko pomeni

- “Please,” he **besought** me, “give me a chance to meet my son ...”
- Even as a young woman she was **besought** by some not to confine herself to the convent, with all that energy, passion and glitter in her eyes.
- I burst into tears **beseeking** and begging her to be more understanding and gentle.
- He **besought** all his friends not to betray his secret to anyone.

The semantic roles of obligatory participants are annotated in all of the descriptions of verbs as well as of those nouns and adjectives that are “verbal” in nature and therefore construct their meanings with analogous valency patterns:

a **reproach** is a critical expression of dissatisfaction or disappointment, usually in a quarrel, that a HUMAN has endured by another HUMAN

očitek je povzetek nezadovoljstva ali razočaranja, ki ga ČLOVEK izrazi nad ravnanjem drugega ČLOVEKA, navadno med prepiranjem

- A common **reproach** to Anna was that she did not show enough interest in the learning skills of her child.
- Their children are growing up in a hostile environment, often saturated with mutual **reproaches** and conflicts.
- A severe source of conflict can be mentally or physically handicapped children, especially if the parents are full of **reproach** for each other.

This is especially the case with the adjectival and nominal meanings typically activated in the predicative position:

if a HUMAN is **frivolous** s/he does not think enough about the CONSEQUENCES of his/her actions, or does not care about them

ČLOVEK je **lahkomiseln**, če ne razmišlja dovolj O POSLEDICAH svojih dejanj ali mu zanje ni mar

In some situations, however, participants, and especially circumstances, are typically expressed but are not decisive for the realisation of a particular meaning. Such elements are identified as optional and viewed as part of the “scenario”. They are, in principle, introduced by the

adverb “usually”, i.e., a hedge, providing a wider context of situation:

if a HUMAN **flours** FOOD s/he sprinkles it with FLOUR, usually in the process of cooking
 če ČLOVEK **pomoka** ŽIVILO, ga potrese z MOKO, navadno v postopku priprave jedi

- They are then cut to pieces, which we **flour** with the rest of the flour and place in buttered cookie moulds.
- We **flour** them with buckwheat flour and add stock.
- The dough is then **floured**, covered and left in a warm place to rise.

2.2 Metalanguage in full-sentence meaning descriptions

2.2.1 If/when-sentences

The if/when-sentence puts the description into a metalinguistic mode in which the natural usage of the headword is “encoded implicitly within the description text itself rather than explicitly as a separate, densely encoded abbreviation which the user may well ignore” (Barnbrook, 2002: 7-9). In principle, if/when-sentences make the description more explicit, thus facilitating the inclusion of pragmatic components of meaning. The if/when-sentence is also a typical definition type, particularly for verbs, in the SLD:

if a HUMAN or CIRCUMSTANCES **degrade** a HUMAN, his/her EFFORTS, or his/her KNOWLEDGE, they destroy his/her sense of value or importance, or diminish his/her role

če ČLOVEK ali OKOLIŠČINE **degradirajo** ČLOVEKA, njegove NAPORE ali ZNANJE, ga razvrednotijo, mu vzamejo veljavo ali zmanjšajo njegovo vlogo

COBUILD: Something that **degrades** someone causes people to have less respect for them. (...the notion that pornography degrades women... //When I asked him if he had ever been to a prostitute he said he wouldn't degrade himself like that).

There is enough flexibility in the guidelines to prevent a forced and inappropriate use of this type – if/when-sentences can be cumbersome and therefore inappropriate for some meanings – however, we abide by the rule on full-sentence definition. While acknowledging the advantages of the Cobuild strategy in describing words or phrases that typically occur in quite limited contexts, Rundell (2010: 361) is critical of its application where contextual features are not especially salient, as it “can sometimes lead to definitions which mislead the reader by overspecifying typical contexts of use (Rundell, 2006: 330-331).” On providing arguments in favour of full-sentence definition, Barnbrook (2002: 55) also acknowledges the fact that “[t]he adequacy of the contents of any individual dictionary is a separate consideration.” These objections are indeed more relevant for dictionaries. If/when-sentences often create the need to use more pronouns and anaphoric

expressions, which can be at the expense of elegance. Each solution in the SLD is subject to consideration from the perspective of the potential proliferation of anaphoric and deictic elements.

2.3 Pragmatics and the definition strategies

Various aspects of lexicographic description have been studied in the SLD, including the potential for rendering pragmatic components an integral part of meaning description, for which so-called “projection” proves to be extremely useful.

2.3.1 The “projection” principle

This description is characterised by reported speech or by the so-called “report” element of the co-text in the left side of the definition (Sinclair, 1991: 126-127), which re-establishes the traditional lexicographic equation as a comment on usage rather than as a description of the intrinsic meaning of the headword. The label “projection” was taken from Halliday (Barnbrook, 2002: 151-152). Hanks (1987: 204) links the projection principle – and the use of a strategy such as “If you say that ...” or “If you call someone a ...” – directly to the description of figurative senses and phraseology or idiomatic expressions (see section 2.4):

if we say that TREES and BUILDINGS **soar** we mean that they rise up very high
če rečemo, da DREVESA ali ZGRADBE **silijo** navzgor, menimo, da segajo zelo visoko

Some headwords and their meanings thus need to be treated with special attention to what, in actual fact, “people mean” when they use them in writing or speech. The SLD meaning descriptions attempt to make this sometimes very subtle layer of meaning as evident as possible within the explanation itself, rather than using labels, usage notes and so on to convey comments on usage. It is understood that this information is an inseparable part of meaning. An alternative strategy is to use a “something is an expression for something” formula which generally has been avoided:

a **consumer** is an expression used for someone who regularly buys and uses goods or services, especially with regard to his/her rights

potrošnik je izraz za posameznika, ki redno kupuje in uporablja trgovske izdelke in storitve, zlasti kadar so v ospredju njegove pravice

COBUILD: A **consumer** is a person who buys things or uses services. (*...claims that tobacco companies failed to warn consumers about the dangers of smoking.// ...improving public services and consumer rights.*)

To express specific circumstances of usage in this case, MEDAL, for example, uses the label Economics, the second part of the description specifying usage in relation to the grammatical feature “singular with plural meaning”, which indicates that “the consumer” is viewed as a category of people (economics. “someone

who buys and uses goods and services. The expression the consumer is often used for referring to consumers as a group”). To the definition “someone who buys and uses products and services”, LDOCE adds a usage note: “A consumer is anyone who pays for goods and services. This word is used especially when you are talking about people’s rights (*Consumers have a right to know what they are buying*)”.

2.3.2 Circumstances of meaning – hedging

Pragmatic information is often located in the circumstances of meaning. In view of the fact that a high degree of granularity is presupposed in the SLD meaning descriptions, pragmatic components, along with the semantic-syntactic behaviour of words, become prominent indicators of meaning (nuances) and often play an important role in the process of sense discrimination. The above examples show that – seemingly for reasons of the structure of natural discourse – the pragmatic background will often fit naturally into the end part of the description. Early attempts to annotate pragmatic elements in the descriptions for purposes of automatic retrieval were abandoned due to seeming inconsistencies in the lexicographic treatment of pragmatic information. It turned out that this type of information is quite naturally and consistently located in the semi-formalised parts of descriptions beginning in “usually” (also “especially”), which provide typical situations of meaning. Where relevant, the definitions are given below from MEDAL and LDOCE in order to provide comparison of the ways in which they describe (pragmatic) circumstances:

if a HUMAN **distorts** INFORMATION, FACTS or someone’s STATEMENT, s/he intentionally presents them in a way that is no longer accurate or true, usually because s/he wants to hide something or to harm someone¹³

če ČLOVEK **izkrivlja** PODATKE, DEJSTVA ali IZJAVE, jih namenoma prireja ali navaja neresnično stanje, navadno zato, ker hoče kaj prikriti ali komu škodovati

• *Such comments **distort** the truth in an insidious way and are covertly destructive.*

• *During the interrogation he presumably **distorted** the truth, gave false information, but above all he openly lied about how he’d dealt with the problem himself.*

• *A while ago a group of renowned Danish scientists accused him of **distorting** facts to make them fit with his own theories thus violating the research community ethic.*

MEDAL: to change something such as information so that it is no longer true or accurate (*The paper was accused of **distorting** the truth// The programme presented a distorted picture of her life.*)

¹³ Examples of usage would be needed to adequately support the existing meaning descriptions throughout the text, but they could not, unfortunately, always be listed for reasons of limited space.

LDOCE: to report something in a way that is not completely true or correct (*His account was badly distorted by the press*)

The section introduced by “usually” typically contains information on cause, reason, intention, manner or other circumstances of meaning – this strategy is adopted analogously in many monolingual dictionaries, including COBUILD, MEDAL and LDOCE. In the SLD, the question often arises whether a circumstance is to be interpreted as typical or obligatory given that participants and circumstances must be explicitly identified as obligatory (set in block letters) or optional (in lower case). In order to be either one or the other, the element must recur in the corpus data; however, the evidence is not always unequivocal, especially as we move away from evident syntactic patterning to more elusive contextual clues. Sometimes components of meaning have to be inferred from the co-text, sometimes even based on our general knowledge of the world. Below are some examples of unclear situations, which are illustrated by the descriptions in the EFL dictionaries. Here we touch upon the issue of semantic prosody (more on this in section 2.5).

The example below shows that the “for pleasure” part is interpreted in the SLD as a non-obligatory participant, while MEDAL presents it as a “core” element of the definition, adding instead as an extension “especially to a series of places”, which provides more detailed information on the type of experience in store:

a HUMAN **cruises** around the SEA or a RIVER by BOAT, usually for pleasure or as part of a holiday
ČLOVEK s PLOVILOM **križari** po MORJU ali REKI, navadno za zabavo ali preživljanje prostega časa

MEDAL: to sail in a ship for pleasure, especially to a series of places (*The first week was spent cruising around the Baltic// They'll be going cruising the Greek islands next week.*)

LDOCE: to sail along slowly, especially for pleasure (*We were cruising in the Caribbean all winter.// an evening spent cruising the River Seine*)

COBUILD: If you **cruise** a sea, river, or canal, you travel around it or along it on a cruise. (*She wants to cruise the canals of France in a barge.// During their summer holidays they cruised further afield to Normandy and Brittany. (A cruise is a holiday during which you travel on a ship or boat and visit a number of places).*)

Another example of discrepancy follows where the “by force” element is recognised in the SLD and in LDOCE as an added circumstance, while in MEDAL and COBUILD this is a central component of the “scenario”:

if an INDIVIDUAL or a group of PEOPLE **overthrows** a HUMAN or a REGIME they take away his power, usually by using force

če POSAMEZNIK ali SKUPINA LJUDI **strmoglavi** DIKTATORJA,

PREDSEDNIKA ali VLADO, mu navadno z uporabo sile odzame moč

MEDAL: to force a leader or government out of their position of power (*Her father was overthrown in a military coup in the seventies.*)

LDOCE: to remove a leader or government from power, especially by force

COBUILD: When a government or leader **is overthrown**, they are removed from power by force. (*That government was overthrown in a military coup three years ago.// ...an attempt to overthrow the president.*)

2.4 Phraseology – figurative meanings and idiomatic expressions

Many, but not all, figurative meanings in the SLD fall under the phraseology section. Figurativeness is neither necessary nor a sufficient, but merely a typical condition for listing a lexical unit under phraseology. Often, in order to describe (figurative) meaning and idiomatic expressions, there is a need to adopt what Hanks (1987: 203) has called a “displacement strategy” while referring to the Cobuild tradition. This is particularly important where a phraseological unit allows for a literal interpretation, which “must be guarded against” (ibid.). Hanks quotes the example: If you twist someone round your little finger, they will do anything that you want them to do. For Hanks, “that is open to the objection that it is either a false statement about English or a false statement about the world, or both.” Certain notions go with a conventional interpretation of figurative meaning, and this must be “indicated on both the left-hand and the right-hand side of the explanatory equation (ibid.):” If you say that you can twist someone round your little finger, you mean that they will do anything that you want them to do.

Attempts have been made in the SLD to identify those figurative meanings that require displacement by using the strategy “If you say that ...” as opposed to those which presumably do not. Hard and fast rules in this matter cannot be applied, given that the interpretation of metaphors and figurative language is largely subjective. This is potentially a point where the intuitions of a lexicographer must be exploited, which, of course, does not guarantee the best of outcomes:

if we say that a MUSICAL INSTRUMENT **squeaks**, we find the sound coming from it unpleasant, such as if somebody cannot play properly

če rečemo, da GLASBILO **cvili**, se nam zdi zvok, ki prihaja iz njega, neprijeten, na primer, ker nekdo ne zna igrati

if someone says the time flies s/he has a feeling that time passes quickly¹⁴

¹⁴ In the descriptions of phraseology, we have not explicated obligatory participants (not in block letters). Also, the description techniques have not been prescribed to the same

če kdo izreče, da **čas beži**, ima občutek, da čas hitro mineva

MEDAL: used for saying that time seems to be passing very quickly

LDOCE: used to say that time seems to pass very quickly (*Time flies when you're having fun*)

COBUILD: If you say that **time flies**, you mean that it seems to pass very quickly. (*Time flies when you're having fun.*)

The distinction is easier to grasp if displacement is paralleled with the actual literal usage (the second example is treated in the phraseology section (no block letters)):

if a HUMAN is **following** another HUMAN s/he is walking or driving closely behind to see where they are going

če ČLOVEK **sledi** KOMU, hodi ali se vozi za njim, da bi ugotovil, kam gre

vs.

if somebody says that s/he is not **following** somebody s/he wants to convey that s/he cannot understand what they are saying

če kdo reče komu, da mu ne **sledi**, sporoča, da ne razume, kaj želi povedati

if NATURE or CIRCUMSTANCES **conspire** against a HUMAN s/he has a feeling that they are operating in disaccord with his or her interests or wishes, usually in a critical moment

če se NARAVA ali OKOLIŠČINE **zarotijo** proti ČLOVEKU, ima ta občutek, da

delujejo v nasprotju z njegovimi interesi ali hotenji, nava dno v odločilnem trenutku

The following expression is highly idiomatic, but does not seem to require displacement:

if somebody is always **a step ahead of their time**, s/he has new ideas or does things long before other people do making them show disapproval, distrust or, rarely, enthusiasm

če je kdo vedno **korak pred časom**, pomeni, da je z dejanji ali v mislih pred svojimi sodobniki, kar izzove sodbo okolja, bodisi tako da zbuja nezaupanje, nejevoljo ali redkeje navdušenje

COBUILD: If someone is **ahead of their time** or **before their time**, they have new ideas a long time before other people start to think in the same way. (*He was indeed ahead of his time in employing women, ex-convicts, and the handicapped.//His only fundamental mistake, he insists, is that he was 20 years before his time.*)

The criteria of (non-)compositionality considerably affect the way in which pragmatics comes into play in phraseology. It has been recognised that “units of meaning associated with metaphors – *metaphoremes* –

must obligatorily have a pragmatic function” (Cameron and Deignan, 2006, in Philip, 2009). In corpus linguistics terminology, this pragmatic element is the semantic prosody (ibid.).

2.5 Semantic prosody

The analysis of concordance lines enables the retrieval of typical patterns of meaning not only in the immediate co-text but also in the wider co-text of the headword, where subtle, on the surface less obviously recurring, elements of meaning may be identified. According to Sinclair (1996a: 34),¹⁵ a semantic prosody expresses attitudinal meaning and is on the pragmatic side of the semantics/pragmatics continuum: “it shows how the rest of the item is to be interpreted functionally.” An ethereal, but perhaps even more frequently cited, definition of semantic prosody comes from Louw (1993: 157): “a consistent aura of meaning with which a form is imbued by its collocates.” Semantic prosodies add meaning that goes beyond the meaning already expressed by word-semantics, requiring a close examination of contexts of use and components of meaning that are not always detectable in the immediate surroundings of the headword, or, as Philip (2009) puts it, “[c]orpus texts facilitate the retrieval of recurrent patterns, but they do so at the expense of the context of situation in which the language under study was originally uttered. Semantic prosodies, therefore, have to be inferred by extracting information from the cotext which allows a picture of the context of situation to be built up.” The fact that semantic prosody is somewhat elusive and not always present has given some linguists reason to discard it as “a figment of corpus linguists’ imaginations” (ibid.), similar to the scepticism that permeates pragmatic meaning in general. Nonetheless, corpus evidence shows that semantic prosody, like meaning on the whole, cannot be identified purely with introspection. As Louw explicitly states: “semantic prosodies are a collocational phenomenon and one which is preferably to be regarded as recoverable computationally from large language corpora rather than intuitively” (2000: 48).

From the above, and particularly from the practical analysis of corpus data, it emerges that semantic prosodies are often difficult to describe “clearly and succinctly, and this may well explain the widespread tendency to speak loosely of positive/negative prosodies rather than attempt to articulate the semantic prosody more precisely” (Philip, 2009). In view of this fact, the question arises as to how, and to what extent, to include semantic prosodies in the descriptions of meaning, as such inclusion will inevitably increase their length and complexity:

extent, thus allowing the compilers more individual freedom in deciding which strategy best suits a lexical unit.

¹⁵ Semantic prosody was first used and presented to the research community by Sinclair (1996b) and Louw (1993).

if a HUMAN **equates** SOMETHING with a PHENOMENON, CONCEPT, or CHARACTERISTIC s/he thinks that they are the same things, usually failing to see the difference either as a result of ignorance or intentionally, due to prejudice

če ČLOVEK **enači** KAJ s POJAVOM, POJMOM ali LASTNOSTJO, meni, da gre za enake stvari, pri tem pa navadno spregleda bistvene razlike, bodisi zaradi nevednosti ali namerno, zaradi predsodkov

• *In the process of compiling a draft of the final document the Arabic countries renounced the demand to **equate** Zionism with racism.*

• *In Western countries Muslim faith is **equated** with terrorism and all the Arabian nations are treated as potential suspects who have to prove their innocence.*

• *Because she **equates** good sex with love she persists with her partner even if nothing but sex is good in the relationship.*

Pre-corpus lexicographic descriptions generally included little or no pragmatic information. Electronic text corpora have made a considerable difference in the selection of illustrative material which tends to show typical usage. In their definitions, most contemporary dictionaries as yet fail to convey the complexities of semantic prosodies (or avoid them), but typically imply them in the examples of (typical) usage. This strategy works well on the presumption that implicit information is lexicographically sufficient:

MEDAL: to consider something to be the same as something else (*These people seem to **equate** honesty with weakness.//Don't make the mistake of equating high test scores and intelligence.*)

LDOCE: to consider that two things are similar or connected (*Most people **equate** wealth with success.*)

COBUILD: If you **equate** one thing with another, or if you say that one thing **equates** with another, you believe that they are strongly connected

Each of the examples listed contains either colligational or collocational information about the semantic prosody," i.e., "seem to equate", "don't make the mistake", and "most people". The semantic prosody could be summarised as: "(people) give equal importance to things that are not the same because they cannot, or will not, see the difference." Collocationally, juxtapositions of "honesty" and "weakness", and "wealth" and "success" also contribute to the construction of meaning, based on the conventional associations of these word patterns. These, of course, are subtle indicators that can only be identified as such in the context of the whole situation, and, particularly, when analysed against a vast collection of data.

While semantic prosodies are often equated with the so-called "semantic preference", some studies (Philip, 2009) have shown that the term semantic prosody can be used loosely incorporating what, in effect, are two different levels of meaning analysis, of which the first is

word-centric and the second delexical, functional, phraseological or contextual. By bringing the associations back into a real context of situation, the latter facet of semantic prosody is inextricably pragmatic in nature: *where, when, why* and *to whom* something means what it does (ibid.).

3. Conclusions

Semantic prosody builds along the semantics/pragmatics continuum, and, unlike "communication background" (Verschuere, 1999: 47), is not a pragmatic backdrop on which we could look for infinite implicit meanings; on the contrary, it is a result of empirically identifiable elements of the meaning structure, albeit on the furthest boundaries of a lexical unit of meaning. This has serious implications for the analysis of corpus data and for the selection of the default span of concordance lines, as well as, and not least, for the way lexicographic descriptions of meaning are constructed. Pragmatic information is an integral part of an (extended) unit of meaning, identifiable only by examining its repeated occurrences in corpus data. Although in the present paper the focus is on pragmatic components that can be abstracted from contextual features, in the SLD we have also addressed the questions of word connotation and emotive and attitudinal meaning that can be associated with words per se. Regardless of the theoretical stance, what is obvious in the process of lexical analysis is the difference in the difficulty of describing semantic prosody as opposed to allocating collocations or attributing semantic roles. In our view, in a lexical database it is vital to provide all of the information on the headword that is retrievable from the corpus data – semantic prosody may be difficult to describe lexicographically, but when present it is an integral part of the wording that cannot be severed from the co-text or context. The question remains how to integrate it into dictionary meaning descriptions for the benefit of their users.

The Slovene Lexical Database is still in the process of compilation. Apart from providing information on the inherent features, grammatical patterns and syntagmatics of words it has also explored the possibilities for constructing meaningful lexicographic descriptions that can serve as a basis for dictionaries targeted at younger audiences. Special consideration has been given to ways in which pragmatics, as a component of meaning, can become an integral part of the definition.

4. References

4.1 Dictionaries and Databases

Collins COBUILD English Language Dictionary. (Second Edition). Sinclair, J. (ed.) (1995) London: HarperCollins. (COBUILD)
The DANTE Database. Accessed at: <http://www.webdante.com/>

Longman Dictionary of Contemporary English, Longman: Accessed at: <http://www.ldoceonline.com/>. (LDOCE)
Macmillan Dictionary and Thesaurus: Free English Dictionary Online. Oxford: Macmillan Education. Accessed at: <http://www.macmillandictionary.com/>. (MEDAL)

4.2 Other Literature

- Atkins, S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Atkins, S. (2010). The DANTE Database: Its Contribution to English Lexical Research, and in Particular to Complementing the FrameNet Data. In G.-M. de Schryver (ed.) *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers, pp. 267-297.
- Atkins, S., Fillmore, C. & Johnson, C.R. (2003). Lexicographic Relevance. Selecting information from corpus evidence. *International Journal of Lexicography*, 16(3), pp. 251-280.
- Barnbrook, G. (2002). *Defining language: A local grammar of definition sentences. Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company.
- Cameron, L., Deignan, A. (2006). The emergence of metaphor in discourse. *Applied Linguistics* 27, pp. 671-690.
- Fillmore, C., Johnson, C.R. & Petrucci, R.L. (2003). Background to Framenet. *International Journal of Lexicography*, 16(3), pp. 235-250.
- Gantar, P., Krek, S. (2009). Dictionary Definitions - A Different View: to Describe, to Explain or to Paraphrase? (Drugačen pogled na slovarske definicije: opisati, pojasniti, razložiti?). In M. Stabej (ed.), *Obdobja 28: Infrastruktura slovenščine in slovenistike*. Znanstvena založba FF UL in Center za slovenščino kot drugi/tuji jezik, pp. 151-159.
- Gantar, P., Grabnar, K., Kocjančič, P., Krek, S., Pobirk, O., Rejc, R., Šorli, M., Šuster, S. & Zaranšek, P. (2009). *Standard for the compilation of a lexical unit (Indicator 6): Communication in Slovene Project: Style Guide: the Slovene Lexical Database*. Kamnik. Slovenia. Available at: http://www.slovenscina.eu/Media/Kazalniki/Kazalnik_6/SSJ_Kazalnik_6_Specifikacije-leksikalna-baza_v1.pdf.
- Halliday, M.A.K., Yallop, C. (2007). *Lexicology. A Short Introduction*. London & New York: Continuum.
- Hanks, P. (1987). Definitions and explanations. In J.M. Sinclair (ed.) *Looking up. An Account of the COBUILD Project*. London and Glasgow: Collins ELT.
- Hanks, P. (1994). Linguistics Norms and Pragmatic Exploitations, or Why Lexicographers Need Prototype Theory, and Vice Versa. In: F. Kiefer, G. Kiss & J. Pajsz (eds.) *Papers in Computational Lexicography: Complex '94*. Budapest: Hungarian Academy of Sciences, pp. 89-113.
- Hanks, P. (2004). Corpus Pattern Analysis. In G. Williams, S. Vessier (eds.) *Proceedings of the 11th EURALEX International Congress, Volume I*. Lorient, Université de Bretagne Sud, pp. 87-97.
- Kocjančič, P., Zaranšek, P. (2009). Slovene Lexical Database: the organizing principles of the argument structure. In P. Cantos Gómez, A. Sánchez (eds.) *A Survey on Corpus-based Research*. Murcia: AELINCO, Asociación Española de Lingüística de Corpus, pp. 293-306.
- Louw, B. (2000). Contextual Prosodic Theory: bringing semantic prosodies to life. In C. Heffer, H. Sauntson (eds.) *Words in Context: A tribute to John Sinclair on his retirement*. Birmingham: University of Birmingham, pp. 48-94.
- Louw, W. E. (1993). Irony in the Text or Insincerity in the Writer?: The Diagnostic Potential of Semantic Prosodies. In M. Baker, G. Francis & E. Tognini Bonelli (eds.) *Text and technology: in honour of John Sinclair*. Amsterdam: Benjamins, pp. 157-176.
- Philip, G. (2009). Why prosodies aren't always: Insights into the idiom principle. *Corpus Linguistics Conference. Liverpool. Great Britain*. <http://ucrel.lancs.ac.uk/publications/cl2009/> (Accessed 17 October 2011).
- Piotrowski, T. (1989). Monolingual and Bilingual Dictionaries, Fundamental Differences, In M.L. Tickoo (ed.) *Learners' Dictionaries: State of the Art*. Singapore: SEAMEO Regional Language Centre, pp. 72-83.
- Rundell, M. (2006). More than One Way to Skin a Cat: Why Full-Sentence Definitions have not been Universally Adopted. In E. Corino, E. C. Marelllo & C. Onesti (eds.) *Proceedings XII Euralex International Congress. Torino, Italy, September 6-9 2006*. Alessandria: Edizioni dell'Orso, pp. 323-338.
- Rundell, M. (2010). Defining Elegance. In G.-M. de Schryver (ed.) *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers, pp. 349-375.
- Sinclair, J.M. (1987). *Looking Up: Account of the Cobuild Project in Lexical Computing (Collins Cobuild dictionaries)*. London and Glasgow: Collins ELT.
- Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: OUP.
- Sinclair, J.M. (1996a). The search for a unit of meaning. TEXTUS IX.
- Sinclair, J.M. (1996b). The Empty Lexicon. *International Journal of Corpus Linguistics* 1, pp. 99-119.
- Tickoo, M.L. (ed.) (1989). *Learners' Dictionaries: State of the Art*. Singapore: SEAMEO Regional Language Centre.
- Verschuere, J. (2000). *Razumeti pragmatiko*. (The original: *Understanding Pragmatics*, 1999). Ljubljana: Založba/*cf.