Two Projects: English ECD and Russian Production Dictionary

Jurij D. Apresjan (1) & Marina Ja. Glovinskaja (2)

(1) Laboratory of Computational Linguistics, Institute for Information Transmission Problems of RAS; Moscow, Bolshoy Karetny, 19
apr@iitp.ru
(2) Department of Modern Russian, Russian Language Institute of RAS; Moscow, Volkhonka 18/2;
marinaglov@yandex.ru

Abstract

Work on two lexicographic projects is reported. The first project is concerned with compiling a multi-purpose English ECD dictionary counting up to 3500 items; it is close to completion. The second project is aimed at creation of a full-scale Russian production (“active”) dictionary amounting to some 50 000 items; work on it has just been started. Both projects are based on updated versions of the theories of lexical functions, syntactic government, and meaning definition. Emphasis in both of them is on the systematicity of lexicographic description, i.e. on representing the lexica of a language as a system.

Keywords

MTT and its applications, lexical system, ECD, production dictionary, predicate classification, government patterns, lexical functions, meaning definition, collocations description.

1 Introduction

The idea of an explanatory combinatorial dictionary (ECD) was proposed at the turn of the 60-ies of the past century in (Zholkovsky & Mel’čuk, 1966); see also (Mel’čuk, 1974), (Mel’čuk & Zholkovsky, 1984), (Mel’čuk et al., 1984) and the subsequent three volumes of this dictionary; the latest summary is (Iordanskaja & Mel’čuk, 2007).

The idea of ECD is too well known to be reproduced here. Let it be said that it signified a genuine breakthrough in the lexical study of language and made a powerful impact on the thinking of scholars even outside the MTT-paradigm. Yet up till now it has not been implemented in sufficiently representative lexicographic products; in fact, lexicographically it has not gone beyond the experimental stage. The innovative Russian ECD counts 282 entries, the auxiliary entries included. Its French counterpart contains 508 entries. Other experiments in the field of ECD lexicography are still more modest in volume and can serve illustrative rather than informative functions. Even elementary dictionaries for schools are supposed to meet the
minimum of some 3000-3500 entries. Hence the task of compiling ECDs for natural languages formulated forty years ago still remains top priority for present day linguistics.

Work on ECDs nowadays seems to have a better claim to systematicity (which is the keynote of our paper) in view of the following four recent developments in the field of semantics and theoretical lexicography:

(a) The fundamental classification of predicates going back to Ju. Maslov’s and Z. Vendler’s work and further elaborated mostly by Russian linguists; see, for instance, (Bulygina, 1982), (Paducheva, 1996), (Zaliznyak & Shmelev, 2000), (Glovinskaja, 2001), (Apresjan, 2003), (Paducheva, 2004), (Apresjan, 2006b) and some other publications.

(b) The construction of more sophisticated versions of the semantic metalanguage designed for formulating lexicographic definitions and meaning interaction rules (scopes, modifications etc). See (Boguslavsky, 1996), (Apresjan, 2006b).

(c) Advances in the theory of syntactic government: among other things, it has been shown that government properties of large classes of verbal lexemes, namely, the number and nature of their semantic actants, may be predicted to a considerable extent on the basis of their membership in certain classes of the fundamental predicate classification: (Levin, 1993), (Goldberg, 1995), (Kibrik, 2000), (Apresjan, 2003), (Mel’čuk, 2004), (Apresjan, 2006a).

(d) A new understanding of the nature of collocate LFs of the OPER-FUNC family: it has been shown that their values have a meaning of their own (are not semantically void) and, consequently, that the choice of a certain lexical item as value of the given LF with regard to the given argument is semantically quite well motivated (Paducheva, 1991), (Mel’čuk & Wanner, 1996), (Reuther, 1996), (Apresjan, 2004). Interestingly, it turns out to be motivated by the same factor as in the case of syntactic government, namely by membership of a particular item in certain classes and subclasses of the fundamental predicate classification.

This paper will be devoted to two lexicographic projects. The first is concerned with an English ECD which is now being compiled by Ju. Apresjan in the Computational linguistics laboratory of IITP RAS. The second is a production (“active”) dictionary of Russian, research on which is under way at the Theoretical Semantics Section of the Russian Language Institute of RAS. The material for presenting this project is based on M. Ja. Glovinskaja’s research.

2 The English ECD

At present the Dictionary\(^1\) counts over 2500 entries; in the coming two years it will be expanded to the size of some 3500 entries.

---

\(^1\) This dictionary is an offspring of a project aimed at constructing a computer-aided device for teaching and learning lexica in the mode of linguistic games with the computer. The project was conceived at the beginning of the 90-ies in the Computational Linguistics Laboratory of the Institute for Information Transmission Problems of RAS. The linguistic data – ECDs of English and Russian counting up to a thousand entries each and lists of LFs of both languages – were elaborated by Ju. D. Apresjan, the algorithms – by L. L. Tsinman; the set of games was devised jointly by those two authors; the programs
Apart from the Dictionary proper a dictionary of English LFs counting 116 functions has been compiled. Each LF is supplied with a definition and an extensive list of examples, in record cases (like MAGN or OPER1) exceeding 300.

It should be noted that the English ECD is a multi-purpose dictionary designed above all for all sorts of computer applications. In particular, its material is used:

(a) in the bidirectional English-to-Russian and Russian-to-English machine translation system, known as ETAP-3, – for resolving syntactic and lexical ambiguities in parsing and overcoming lexical mismatches between languages in translation;

(b) in a computer-implemented system of paraphrasing Russian sentences;

(c) in computerized lexical textbooks which provide a foundation for learning various aspects of vocabulary (mostly word meanings, synonyms, antonyms, converse terms, all sorts of semantic derivatives, and lexically constrained collocations) in the mode of playing games with the computer (see (Apresjan, 1996), (Apresjan & Tsinman, 2002), (Apresjan et al., 2003), (Boguslavsky et al., 2004), (Apresjan et al., 2007a), (Apresjan et al., 2007b)).

Since the English ECD was aimed at a variety of computer applications it had to be designed along slightly different lines than its Russian and French predecessors. On the whole it is less sophisticated. Its definitions and government patterns are simpler, and the number of LFs is smaller. On the other hand it incorporates recent findings in the theory of syntactic government, the theory of LFs and lexical semantics, which provide for greater systematicity in the lexicographic treatment of these phenomena.

Below the apparatus of LFs and meaning definition will be taken up in greater detail.

2.1 LF apparatus in the English ECD

2.1.1 The Make-up of LFs

In computer-implemented systems part of the original LF apparatus had to be ignored because some LFs could not be identified with formal devices at the disposal of ETAP-3 or required the use of disproportionately complicated algorithms. This pertains to non-standard LFs, the composite LFs like FUNC0 + MAGN (The storm was raging), two-argument LFs of the type His teeth were clattering with fear due to (Iordanskaja, 1972) and some others.

On the other hand, the LF apparatus of the English ECD was substantially expanded as compared with its classical prototype. Several new families of LFs were introduced. One of
them, namely, the family REALi-M – FACTi-M, will be briefly considered below. It emerged as a result of splitting up the old family of REALi – LABREALij – FACTi and was described in detail in (Apresjan, 2001). Therefore it will suffice to list the basic arguments which led to this decision.

In the classic version of the MTT the LF family at issue was composed of three classes: REALi – LABREALij – FACTi. Syntactically they were completely likened to three other classes of LFs, namely, OPERi – LABORij – FUNCi but semantically were opposed to them. “Unlike the “empty” LFs Operi, Funci and Laborij, however, the functions we are now concerned with correspond to a specific content – ‘to fulfill a demand contained in the meaning of W0’” (Mel’čuk & Zholkovsky, 1984: 56). Examples: FACT0(knife) = cut, FACT0(gun) = shoot, FACT0(ship) = sail, FACT0(dream) = come true, FACT0(doubt) = prove true, be confirmed, REAL1(accusation) = prove, REAL2(accusation) = agree with.

As these examples show, the classic version of the LF-theory assigned the LFs of the REALi – LABREALij – FACTi family to two semantic classes of words. The first class are names of artefacts like knife, gun, ship. The second class are predicate words like accusation, bet, dream, doubt, temptation whose meanings contain various modal components. In (Apresjan, 2001) it was argued that the values of the respective LFs (like, for instance, cut = FACT0(knife) and come true = FACT0(dream)) could not be subsumed under the same semantically based definition.

Indeed, with regard to artefacts, no matter whether they are tools like knife and gun, or means of transportation like train and ship, the formulation ‘to fulfill a demand contained in the meaning of the keyword’ defies any literal understanding. The meanings of such nouns do not include the component ‘demand’. Their only common property is the use for which they are normally intended. So the definitions of the respective LFs should contain formulations like ‘X functions or is functioning according to its destination’ (for LF FACT0, cf. the gun shoots, the movie is on, the ship sails); ‘to use X according to its destination (for LF REAL1, cf. read a book, take medicine, listen in to the radio); ‘to be accessible for use according to X’s destination’ (for LF REAL2, cf. be on the market, be on at the theatre); ‘to act upon somebody or something with the help of X according to the latter’s destination’ (for LF LABREAL1-2, cf. drink something from a cup, treat somebody in the hospital).

As concerns words of the second semantic class, that is, nouns like accusation, bet, dream, hypothesis, opportunity, temptation and so on, the only common elements in their lexical meanings (and, hence, their definitions) are various modal senses. But then the common semantic core of such verbs as prove (the accusation), win (the bet), realize (one’s dream), confirm (the hypothesis), avail oneself of (the opportunity), resist <repel> (temptation) etc. which are presumably the values of LF REAL1 from the respective arguments cannot be captured by such formulations as ‘to fulfill the demand contained in X’ or ‘to use X according to its destination’.

As a matter of fact the situation with such words is even more complicated. Some of them, for instance, accusation and dream, have no common semantic components at all. The only

---

2 X stands for the keyword; Pi in the definitions below stands for the i-th participant of X.
common property of all such words is a certain pragmatic expectation associated with the
referent of the keyword. Of a hypothesis we expect that sooner or later it will be proved, of a
debt – that it will be paid, of a dream – that it may come true and so on. It is only through
the notion of a pragmatically natural, or, in a sense, normal expectation that the definitions of the
LFs under consideration can be formulated.

This was the reason why the new family of LFs was introduced, – the family REALi-M and
FACTi-M, where M means ‘modality’. All the LFs of this family were defined through the
notion of normal expectations. Consider the following examples.

FACT0-M = ‘X is in the state corresponding to normal expectations in the respective situation
(a collocate verb taking X as its grammatical subject)’; cf. The attack succeeds, Hopes come
true, Illusions dissolve, Justice triumphs.

FACT1-M = ‘X affects P1 in a way which is normally expected of X in the respective situation
(a collocate verb taking X as its grammatical subject and P1 as its primary object)’; cf.
Conscience tortures somebody, Duty demands of somebody, Promise binds somebody.

REAL1-M = ‘To do with regard to X that which is normally expected of P1 in the respective
situation (a collocate verb taking P1 as its grammatical subject and X as its primary object); cf.
attain one’s aim, win a combat, pay a debt, keep <fulfill> one’s promise.

REAL2-M = ‘To do with regard to X that which is normally expected of P2 in the respective
situation (a collocate verb taking P2 as its grammatical subject and X as its primary object); cf.
repulse aggression, run <break> the blockade, overcome a difficulty, accept the invitation

REAL3-M = ‘To do with regard to X that which is normally expected of P3 in the respective
situation (a collocate verb taking P2 as its grammatical subject and X as its primary object); cf.
take the hint, fulfill the order, comply with the request.

This crucial semantic difference between two classes of arguments was the main reason for
splitting the former REAL-FACT family into two families. However, there were other
arguments as well.

1. The simple REAL – FACT family features a richly represented LF LABREAL1-2, cf. treat
somebody in hospital, examine something under the microscope and so on. This is readily
understandable. It is natural for a language to have verbs meaning ‘to act upon the object with
an artifact which is intended for such use’. The modal family REAL-M – FACT-M has no such
LF.

2. The LFs of the simple REAL-FACT family are closely connected with two other LFs –
DEGRAD = ‘X becomes worse or unfit for use though the loss of its vital function’ and
PREPAR = ‘to prepare X for use according to its destination’. Cf. Beer stales, A bulb fuses,
Metal rusts, Milk turns <goes sour> (DEGRAD); undo the bed, set the clock, serve dinner,
load the gun (PREPAR). Neither of them is represented in the REAL-M – FACT-M family.

---

3 In some work on LFs collocations like repulse aggression, run the blockade, overcome a difficulty are
treated as values of ANTIREAL2. Broader research is necessary to arrive at a definitive solution.
3. In its turn, the family REAL-M – FACT-M features a richly represented LF REAL3-M which is practically absent in the family REAL-FACT. This LF is defined for three-argument nouns denoting mostly speech acts: fulfill the order, comply with the request etc.

4. Besides, the same family has an abundance of antonyms of the type ANTIFACT0-M, ANTIREAL1-M, ANTIREAL2-M and ANTIREAL3-M: An attempt fails, A candidate loses, A dream dissolves, Negotiations collapse (ANTIFACT0-M); lose a battle, miss <lose, let slip> a chance, repudiate a debt, shirk one’s duty, break a promise (ANTIREAL1-M); defy somebody’s authority, fail an examination, swallow an insult, give in to an obstacle (ANTIREAL2-M); ignore advice, disregard a ban, reject a claim, reject <decline> an offer, break <disobey> an order (ANTIREAL3-M). This can be accounted for by the fact that part of the meanings of the argument nouns are modal senses with antonymous counterparts: ‘possible’ – ‘impossible’, ‘permission’ – ‘ban’. Since in most instances it is the antonymy of the ‘P’ – ‘non-P’ type, the LFs of this family are capable of paraphrases by the rule X = not ANTI2(X); cf. The attack failed – The attack did not succeed, to fail the exam – not to pass the exam. The family REAL-FACT has no antonymous LFs, and such paraphrases are ruled out.

2.1.2 Semantic motivation of LFs

As has been stated above, it is assumed in the present paper that all the values of all collocate LFs have lexical meanings of their own and, consequently, that the choice of a particular value in a particular case is semantically quite well motivated. This assumption is corroborated by empirical evidence and theoretical considerations.

The most important piece of empirical evidence is the fact that when the given LF has two or more values with regard to the given keyword, they are normally slightly different semantically. Consider the verbs make, commit and perpetrate in the function of OPER1 from the noun blunder. To make (a blunder) has obviously the most general meaning. To commit (a blunder) implies negative evaluation by the speaker of the mental digression at issue, and to perpetrate (a blunder) suggests both, a greater mistake and its negative evaluation.

Error also differs from mistake in suggesting a greater digression from the truth. On the other hand,
it differs from both, mistake and blunder, in that it denotes not only a single act of erroneous judgment but also a mental state. That is why error freely collocates not only with the actional OPER1 verbs like make, commit and perpetrate (an error) but also with the stative OPER1 be in (error). Be in a mistake does occur in the corpora but sounds rather odd, and *be in a blunder is downright impossible. Semantic agreement tells.

To be able to discuss semantic motivation on a more general plane it is necessary to recall the familiar notion of predicate classes. For the subsequent discussion we shall need five classes of predicates – actions (talk), activities (war), processes (growth), states (need), and properties (courage).

Actions, in the present context, are defined as predicates which meet the following two conditions: (a) the assertive part of their explication at the level of semantic primitives displays the component ‘do’ as the top node; (b) the situation they denote falls within one round of observation, i.e. can be observed from beginning to end without interrupting the observation.

Activities are predicates which denote a sequence of heterogeneous and non-synchronous actions directed at attaining the same final aim and requiring many rounds of observation to be observed from beginning to end.

Processes are predicates whose assertive part at some point in their semantic reduction displays the component ‘become’ as the top node. Put a bit less technically, processes denote change, i.e. a situation developing of itself, without human intervention.

A state is a semantic quark, i.e. an element smaller than a prime; a property is a prime.

Let us turn now to the problem at hand. It has already been stressed above that there are no semantically void LF. The choice of a particular value V of a certain LF with regard to a particular keyword L is semantically quite well (though not completely) motivated.

Especially interesting from this point of view is the allegedly “void” family OPERi-LABORij-FUNCi. Within it the degree of semantic motivation grows with the transition from OPER to LABOR and FUNC, and inside each of these classes – with the transition from smaller to larger numbers. In all these cases motivation is sufficiently great to provide a foundation for stating the general trends and forming useful lexicographic expectations, although it is insufficient for formulating rules.

Usually semantic motivation manifests itself at two-levels. First of all, the choice of a particular value V of a certain LF with regard to the keyword L is determined by membership of V and L in a certain semantic class or classes of the fundamental predicate classification. For example, LABOR12 from actions typically collocates with an actional verb subject (somebody or something to something): subject to analysis, to censorship, to a check-up, to criticism, to an examination, to inspection, to an interrogation, to punishment, to torture. LABOR12 from

---

4 Some (though not all) English and American explanatory dictionaries consider the senses of ‘act’ and ‘state’ to constitute two distinct lexical meanings of the noun error, but even if this treatment is accepted, our basic argument still holds true.
states typically collocates with a stative verb hold (somebody or something in something): hold in abhorrence, in contempt, in esteem, in respect.

Secondly, at a deeper level it is determined by a narrower semantic subclass of predicates to which the keyword belongs.

The point will be illustrated in more detail below with the material of OPER1 and OPER2.

**OPER1** from actions are verbs whose lexical meanings either include the sense ‘do’ as their basic component or boil down to it. In the accumulated sufficiently representative material OPER1 from actions displays the greatest variety of expression: of the total number of 77 verbs in our material it has 52 at its disposal. Here is the list of those verbs: administer (a blow), ask (a question), bear (evidence), bring (a charge against somebody), carry out (an operation), cast (a glance), come to (a conclusion), commit (a crime), crack (a joke), deal (a blow), deliver (a blow), do (harm), draw (a comparison), drop (a hint), effect (marriage), enact (murder), exercise (control), exhibit (a charge), fetch (a groan), fight (a battle), furnish (an answer), give (advice), impose (punishment), incur (expenses), inflict (a blow), issue (an order), launch (a challenge), lay (an accusation), lend (assistance), let out (a cry), lodge (a claim), make (an attempt), participate in (a competition), pay (a call), perform (an experiment), perpetrate (a blunder), pose (a question), proffer (a complaint), put forth (an effort), put forward (a proposal), raise (a claim), read (a paper = ‘report’), render (assistance), speak (with an accent), stage (an experiment), strike (a blow), take (care), tell (a lie), tender (an apology), throw (a scene), wear (mourning), win (a victory).

**OPER1** from activities is mostly expressed by the verbs carry on (correspondence, research, work), conduct (a campaign, a ceremony, an inspection, a war), wage (a campaign, a contest, a war). All of them imply a greater time interval occupied by the respective situation; in contradistinction to such action OPER1 verbs as cast (a glance), crack (a joke), drop (a hint), give (advice), let out (a cry), strike (a blow), throw (a scene), none of them can be thought of as an instant situation.

**OPER1** from processes is almost uniformly expressed by the verb undergo (a change, decay, decomposition, deformation, disintegration, mutation). The only current alternative to undergo is the composite expression be subject to (change, decay, decomposition, deformation, disintegration, mutation). It is to be noted that the semantic role of the first actant of a process is that of PATIENT, i.e. somebody or something who or which are not doing anything of their own will but have things happening to them, are under the impact of an outer force. The verb undergo as an exponent of this meaning has no rivals among verbs.

**OPER1** from states has up to three main exponents, the most universal being be in (confusion, despair, ecstasy, error, a faint, fashion, a fury, horror, hysteries, mourning, need, a trance, vogue). The other two are experience and feel; a little more on them will be said below. It will be remembered that there is a subtle temporal difference between states and properties. A prototypical property is a constant characteristic of a person or thing, and in a language with the ‘have’ – ‘be’ opposition it is usually predicated to the POSSESSOR by the ‘have’ type of verb. A prototypical state is a transient characteristic of a person or thing. People and things get into certain states, are in them for some time and sooner or later go out of them. It is therefore natural to predicate a state to its EXPERIENCER with the verb be in.
So much for semantic motivation of the choice of a particular value of OPER1 at the level of the largest classes of the fundamental predicate classification. Semantic motivation (preferences) for OPER1 at the second level may be illustrated with the following instances:

*Commit* is preferred as the value of OPER1 from nouns denoting negatively evaluated acts: *commit aggression, a blunder, a crime, an error, murder, a sin, suicide, treachery*.

*Draw* is preferred as the value of OPER1 from nouns denoting a mental operation, especially various operations of comparison: *draw an analogy, a comparison, a conclusion, a distinction, a parallel*.

*Enjoy* is preferred as the value of OPER1 from nouns denoting those social attributes of people which can be seen as privileges: *enjoy freedom, privileges, a good reputation, rights (VS. enjoy good health)*.

*Experience* and especially *feel* often collocate with nouns denoting emotional states and attitudes: *experience grief, joy, pleasure (VS. experience need); feel fear, hatred, pity*.

*Meet with* is preferred as the value of OPER1 from nouns denoting bad states of affairs: *meet with an accident, a calamity, a loss, bad luck, failure, a reverse (VS. meet with success)*.

*Suffer* and *sustain* are also preferred as the value of OPER1 from nouns denoting bad states of affairs, but on a larger scale: *suffer a catastrophe, casualties, damage, a defeat, an injury, losses, a shipwreck; sustain casualties, a great damage, a defeat, injuries, a loss*.

**OPER2** are verbs which are similar to the passive form of a verb in that typically their grammatical subjects play the semantic role of PATIENT. Now, as is clear from our discussion of processes, PATIENT is a role assigned to someone or something that is acted upon. ‘Being acted upon’ is a “passive” sense, and indeed, the most frequent values of OPER2 are verbs and collocations with a passive meaning. Cf. *be under (control), undergo (censorship), meet with (a rebuff) and receive (a beating)*.

Out of these four values *being under* is the least specific and does not convey any meanings apart from the general idea of being controlled by somebody. It serves as OPER2 from actions, activities, states, and states of affairs, both, good and bad: *be under arrest, an attack, a ban, the care of somebody, the charge of somebody, consideration, control, discussion, editorship, the fire, somebody’s influence, somebody’s observation, somebody’s power, somebody’s pressure, somebody’s protection, repair, the sentence (= verdict), somebody’s supervision, somebody’s suspicion, treatment*.

At the second level of motivation the following preferences in the choice of OPER2 values can be observed:

*Undergo* as value of OPER2 is preferred with three groups of nouns: (a) nouns denoting hostile actions or activities: *undergo criticism, an interrogation, punishment, torture*; (b) a closely related group of nouns denoting various forms of inspection: *undergo censorship, a checkup, an examination, inspection, a test, sea trials (of a ship)*; (c) a group of nouns denoting interference: *undergo an operation, reform, repair, surgery*. 
Meet with as value of OPER2 is preferred with nouns denoting various forms of approval, disapproval and opposition: meet with applause, approval, opposition, a rebuff, a good reception, a refusal, a repulse, resistance.

2.1.3 A new strategy of LF definitions

The new system of LF definitions reproduces the basic pattern of definitions from Mel’čuk’s classic version of MTT. Yet all the definitions are modified to do justice to the fact that the values of LFs of the OPER-FUNC family are meaningful lexical units. Therefore, apart from the syntactic properties of the respective LFs, minimal semantic information on their content in the form of several elementary senses is introduced. For instance, the definition of OPER1 includes the senses ‘do’, ‘be in’ and ‘have’ – to reflect the ability of this class of verbs to collocate with nouns denoting actions, states and properties respectively. Below several sample definitions constructed along these lines will be adduced.

OPER1 = ‘To do X, to have X or to be in the state X (a collocate verb taking P1 as its grammatical subject and X as its primary object)’: fight (a battle), deal (a blow), give (consent), have (a debt), be in (a fury).

INCEPOPER1 = ‘To start to do X, to have X or to be in the state X (a collocate verb taking P1 as its grammatical subject and X as its primary object)’: join (battle) / enter (a battle), incur (a debt), fall into (despair) / sink in (despair).

FINOPER1 = ‘To cease to do X, to have X or to be in the state X (a collocate verb taking P1 as its grammatical subject and X as its primary object)’: drop (the charge), lose (courage), go out of (existence).

OPER2 = ‘To undergo X, to experience X, or to be subject to X (a collocate verb taking P2 as its grammatical subject and X as its primary object)’: undergo (a checkup), have (somebody’s confidence) / enjoy (somebody’s confidence), be under (somebody’s influence).

INCEPOPER2 = ‘To start to undergo X, to experience X, or to be subject to X (a collocate verb taking P2 as its grammatical subject and X as its primary object)’: fall into (an ambush), get under (control), meet with (resistance).

FINOPER2 = ‘To cease to undergo X, to experience X, or to be subject to X (a collocate verb taking P2 as its grammatical subject and X as its primary object)’: get out of (control) / go out of (control), fall out of (fashion) / go out of (fashion), pass out of (sight) / disappear from (sight).

2.2 A Sample Entry of the ECD of English

WOUND 1.1 ‘an object on the body part P2 of a living being P1 which is a result of rather serious injury to the skin and tissues of P2 caused by an outer force or object P3, usually in the form of a hole, a cut or a tear, which is bleeding or has bled’ [This explication takes into account M. Ja. Glovinskaja’s definition of the Russian word rana; in the rest of this entry we have also profited from Mantha, S. & I. Mel’čuk, 1999].
Two ECD Projects: English and Russian

3 The Russian Production Dictionary

Work on this project started a year ago in the Theoretical Semantics Section of the Vinogradov Russian Language Institute of RAS. In 2006 we compiled a word-list for the future dictionary counting over 50,000 items and are now working on the general layout for a description of certain semantic classes of highly polysemous words. Unlike the English ECD, this dictionary is intended for the lay public, with no particular knowledge of linguistic theories or terminology. Therefore, although we aim at compiling a contemporary dictionary meeting the basic theoretical requirements of systemic lexicography, we do not employ any technical terms or the LF notation. All the formalisms are used mostly as background tools, that is, as means to collect lexical material and arrange it systematically, while the actual language of description is kept maximally close to the object language.

To bring out the difference between the treatment of lexical items in the English ECD and in the Russian production dictionary we shall adduce a sample entry for the lexeme *рана* ‘wound’ compiled for the Russian dictionary by M. Ja. Glovinskaja. However, not to go into the laborious task of glossing every Russian word and expression, we shall exemplify it with the respective English lexeme *wound*. In the present paper the entry is broken into three broad zones of definition, government pattern and collocations; each zone is briefly commented on.

---

WOUND 1.1

Definition: ‘an object on the body part P2 of a living being P1 which is a result of rather serious injury to the skin and tissues of P2 caused by an outer force or object P3, usually in the form of a hole, a cut or a tear, which is bleeding or has bled’.

This definition is a result of the following considerations: wound is one of the lexemes expressing the idea of an injury on the body of a human or some other living being. The class comprises such words as bruise, shiner, sore, suffusion; abrasion, graze, scrape, scratch; cut, gash, laceration, rent, slash, tear; dislocation (of the hip), rick (of the wrist), (shoulder) slip, spraining, strain, twist, wrench; cleft, fissure, fracture, rupture; contusion; and some others.

Every semantic class in the Russian Production Dictionary is described according to a certain unified layout, as was first proposed independently in (Iordanskaja, 1970) and (Wierzbicka, 1972) and later pursued in all of A. Wierzbicka’s semantic research. For the class of injuries to which wound belongs the general layout is due to M. Glovinskaja; it includes the following parameters to be specified in the definition:

(a) GENUS PROXIMUM. A wound in English, as well as rana in Russian, is conceived of above all as a material OBJECT, although in a number of collocations this sense tends to shift to senses ‘place’ and ‘result’. In this respect wounds are similar to bruises, shiners, sores etc. and are distinct from dislocations, sprainings, twists, fissures and fractures which cannot be thought of as objects.

(b) TYPE OF INJURY. A wound is an injury to the skin AND tissues of a body part; cf. a scratch, which is an injury to the skin, and a shiner, which is an injury to the tissues. Cf. also sprainings which are injuries to the ligaments, and fractures which are injuries to the bones. Besides that, a wound usually implies bleeding at the moment of observation or prior to it, while some other injuries, especially all sorts of twists and sprainings, have no such implication.

(c) ATTENDING CIRCUMSTANCES: A wound usually implies bleeding at the moment of observation or prior to it. In this respect it is different even from sores, let alone all other injuries.

(d) PLACE OF INJURY. A wound may be located anywhere on the human body whereas contusions involve above all the brain. Cf. also the slang sense of mouse meaning a sore under the eye, like the Russian slangy words fonar’ and fingal.

(e) CAUSE OF INJURY. A rick, a sprain, a strain, or a twist may be a result of excessive tension of some part of the human body, while a wound is always a result of impact of some outer force, even when it is received accidentally.

(f) FORM. Cuts and scratches have elongated forms, while the form of a wound is less specified: it may be a hole, a cut or a tear.

Government pattern. As follows from the definition, a wound has three actants; below we specify the governed forms in which they are subordinated to the keyword.
P1, whose: Peter’s wound;
P2, where: on the back, in the stomach;
Two ECD Projects: English and Russian

P3, from what: from a gunshot.

**Collocations.** As has already been stated above, the apparatus of LFs is used as a background tool for collecting the material. However, there remains a mass of material falling outside LF collocations proper which is lexically constrained and therefore requires lexicographic representation (free collocations like *examine the wound, the wound is getting better* <worse> are left out of account).

To present it in a unified format we order collocations by semantic classes which reflect certain pragmatic aspects of the causation, existence, development and cessation of wounds; this is close to the principles of ordering used in (Iordanskaja & Paperno, 1996). For injuries ten semantic classes of collocations were proposed, some of them intersecting with definition parameters. They are listed and exemplified below.

(a) **COMMENCEMENT & CAUSATION:** green wound; receive *sustain* a wound, suffer a wound; inflict a wound, rare make a wound;

(b) **INSTRUMENT OR IMMEDIATE CAUSE:** a bullet *knife, bayonet, shrapnel* wound; a stab *gunshot, operation* wound;

(c) **CIRCUMSTANCES:** a battle wound;

(d) **INJURED BODY PART:** an abdominal *stomach* wound; a wound in the shoulder *in the head*; a flesh wound;

(e) **FORM, TYPE & CHARACTER:** lacerated *punctured, incised* wound, jagged wound, perforating wound; open *gaping* wound, contused wound; wound of entry *of exit*.

(f) **DEGREE:** deep *superficial* wound; slight *light, minor* wound; dangerous *cruel, grievous, serious* wound; mortal *fatal* wound;

(g) **STATE:** painful *agonizing* wound; cured *uncured* wound;

(h) **PROCESSES IN THE WOUND:** the wound is healing *healing up, healing over*, the wound is overcast, the wound closed *cicatrized, repaired*; the wound is bleeding *fester, is infected*, the wound opened, the wound still rankles;

(i) **TREATMENT:** cleanse *wash* the wound; reopen the wound; nurse the wound; dress a wound, bind (up) a wound, cure *heal* the wounds; neglect the wound;

(j) **CONSEQUENCE FOR THE WOUNDED:** recover from the wounds, die from the wounds.

**Acknowledgements**

The present paper was supported in part with grants from the Russian Research Foundation for the Humanities Nos. 06-04-00289a and 06-04-00090a, the Program of the President of the Russian Federation for Supporting Research of the Leading Scientific Schools of Russia No. НШ-5611.2006.6 and the Program for Fundamental Research “Russian Culture in World History” of the Historical and Philological Department of the Russian Academy of Sciences.
Bibliography


Paducheva, E. V. 1991. Predikatnye imena v leksikographicheskom aspekte. In Nauchno-

Paducheva, E. V. 1996. Semanticheskie issledovaniya. Semantika vremen i vida v russkom
jazyke. Semantika narrativa. Moskva: Jazyki russkoj kul'tury.

cul'tury.

VESTI, PROVODIT’ and PROIZVODIT’. In Wanner, L. (ed.). Lexical Functions in
Lexicography and Natural Language Processing. Amsterdam-Philadelphia, 181–208

Reuther, Tilmann. 1997. Das CALLex-Proekt (Computer-Aided Learning of Lexical
Functions). In Kosta, P. & E. Mann (eds.). Slavistische Linguistik 1996. Referate des XXII


Jazyki russkoj kul'tury.