

Published in

Giandomenico SICA (ed.), 2006, Open Problems in Linguistic and Lexicography, Monza (Italy): Polimetrica, 225-355.

Explanatory Combinatorial Dictionary

Igor Mel'čuk

Abstract

The paper presents a new type of dictionary—Explanatory Combinatorial Dictionary [= ECD], which is a formalized semantically-based lexicon designed to be part of a linguistic model of natural language. After stating ECD's main properties (1), we describe an ECD lexical entry (2), then groupings of lexical entries (3), and, finally, principles for compiling ECDs (4); a series of entries for an English ECD is given as an illustration (5).

Table of Contents

Explanatory Combinatorial Dictionary	1
1 General Overview of the ECD	3
1.1 Introductory Remarks	3
1.2 The General Defining Properties of the ECD	5
1.2.1 The Theoretical Nature of the ECD	5
1.2.2 The Formalized Character of the ECD	6
1.2.3 The Internal Exhaustiveness of an ECD Entry	7
1.3 The Specific Defining Properties of the ECD	8
1.3.1 The Active Orientation of the ECD: From Meaning to Text	8
1.3.2 The Semantic Basis of the ECD	10
1.3.3 Cooccurrence as the Main Target of the ECD	11
1.3.4 The ECD Describes All LUs of L Together and in a Similar Way	12
1.3.5 Each Article of the ECD Describes one LU	18
2 The ECD's Microstructure: A Lexical Entry	20
2.1 The Lexicographic Definition in the ECD	22
2.1.1 Substantive Requirements on ECD Definitions	23
2.1.2 Rules for Writing ECD Definitions	27
2.1.2.1 Rule 1: Propositional Form Rule	27
2.1.2.2 Rule 2: Decomposition Rule	28
2.1.2.3 Rule 3: Standardization Rule	32
2.1.2.4 Rule 4: Maximal Block Rule	35
2.1.2.5 Rule 5: Mutual Substitutability Rule	37
2.1.3 Criteria for an ECD Definition: Criteria of Type I	39
2.1.3.1 Criterion I.1: Linguistic relevance of a semantic component	40
2.1.3.2 Criterion I.2a: Cooccurrence with qualifying modifiers	42
2.1.3.3 Criterion I.2b: Cooccurrence with quantifiers	43
2.1.3.4 Criterion I.2c: Cooccurrence with negation	45
2.1.4 The ECD Definition: General Characteristics	47
2.2 The Government Pattern in the ECD	51
2.3 Lexical Functions in the ECD	54
2.3.1 Semantic Derivation	55
2.3.2 Collocation	56
2.3.3 Lexical Function	60
2.4 Illustrative Examples in the ECD	64
3 The ECD's Macrostructure: An ECD's Lexical Super-entry	65
3.1 Basic Notions for the Characterization of Lexical Super-entries	65
3.2 Criteria for the Delimitation of LUs within Vocables: Criteria of Type II	72
3.2.1 Introductory Remarks	72
3.2.2 Criterion II.1: Differentiating Lexicographic Information	74

3.2.3 Criterion II.2: Unifying Cooccurrence (= the Green-Apresjan Criterion)	79
3.2.4 A Comparison of Criteria II.1 and II.2	82
3.3 Organization of an ECD Super-entry [= Vocable]	84
3.3.1 The Vocable Synopsis	84
3.3.2 The Order of the LUs in a Vocable	84
4 Principles for Compiling the ECD	86
4.1 The Formality Principle	87
4.2 Coherence Principles	89
4.2.1 Lexical Unit Internal Coherence Principle	89
4.2.2 Semantic Field Coherence Principle (= Lexical Inheritance)	94
4.3 Uniform Treatment Principles	95
4.3.1 Lexical Unit Uniform Treatment Principle	95
4.3.2 The Vocable Uniform Treatment Principle	97
4.4 The Internal Exhaustivity Principle	98
4.5 Maximal Generalization Principles	99
4.5.1 The Vocable Generalization Principle	99
4.5.2 Semantic Field Generalization Principle	100
4.6 The No Regularly Produced LUs Principle	101
4.6.1 Regular Derivation	101
4.6.2 Regular Polysemy	102
5 An Illustration: A Sample of an English ECD	103
5.1 Some Lexical Entries from an English ECD	103
5.2 Lexico-grammatical Problems Related to Lexical Entries for BAKE	111
Acknowledgments	116
Notes	116
References	127
Index of Terms	137

The paper describes what I consider to be one of the central tools of modern semantics: **the theoretical lexicon of a language**. In the framework of the Meaning-Text approach, this lexicon takes the form of an *Explanatory Combinatorial Dictionary*, or ECD, which is thus the main subject of this text. Its description, of necessity concise and stripped down to a minimum, will be presented in five steps:

1. General overview of the ECD
2. The ECD's microstructure: lexical entry
3. The ECD's macrostructure: lexical super-entry
4. General principles for compiling the ECD
5. Examples: several lexical entries from an English ECD

Given the complexity of my task and natural limitations of space, even a superficial review of the domain cannot be offered here; no parallels will be drawn between the ECD and similar/related approaches in different frameworks.

Since my own lexicographic practice is mostly focused on French, in this paper I will often use French data; in order to simplify the presentation, the lexicographic definitions of French lexical units are formulated here directly in English.¹

Abbreviations and Notations

DSyntA : Deep-Syntactic Actant	SemA : semantic actant
ECD : <i>Explanatory Combinatorial Dictionary</i>	SemR : Semantic Representation
GP : Government Pattern	'X' : the meaning of X
L : a particular language	$_L_1 \dots L_2_$: the idiom $L_1 \dots L_2$
L : a particular lexical unit	\oplus : the operation of linguistic union (p. 00)
LF : Lexical Function	'X Y' : X is the presuppositional part of a meaning (p. 00)
LU : Lexical Unit	*X : ungrammatical expression X
MTM : Meaning-Text Model	'X : grammatically dubious expression X
MTT : Meaning-Text Theory	#X : semantically and grammatically correct, but pragmatically unacceptable expression X

Writing conventions

- Italics* : linguistic examples, i.e., cited linguistic forms
Boldface : highlighting (including titles)
Courier New : technical terms on their first occurrence
CAPITAL LETTERS (10 points) : lexical expressions

1 General Overview of the ECD

I will begin with a few introductory remarks (1.1) and then consider the ECD's defining properties, general (1.2) and specific (1.3).

1.1 Introductory Remarks

The main tenet of this paper can be formulated as follows:

A formalized semantically-oriented and cooccurrence-centered lexicon must be one of the central components of a linguistic description of any language.

It is presumed that a complete linguistic description of language **L** includes two major components: a lexicon and a grammar.

- The lexicon of **L** is the set of all minimal (= elementary and quasi-elementary) lexical signs of **L**: roughly, the set of its Lexical Units [= LUs].²

- The grammar of **L** is composed of two sets:

- The set of all minimal grammatical signs of **L**, that is, of its inflectional and derivational means (affixes, apophonies, conversions).

- The set of all rules of **L**: semantic, syntactic, morphological, and phonological. A rule can be very general or very specific; but it must apply to a set of signs, not an individual sign.

Thus, a lexicon of **L** describes **L**'s individual lexical signs, and the grammar of **L** covers a) **L**'s individual grammatical signs and b) the behavior of sets of **L**'s signs.

The lexicon of any **L**, in combination with **L**'s grammar, ensures meaning-to-text and text-to-meaning transitions. More precisely, it plays a central role in establishing correspondences between a given Semantic Representation [= SemR] and all the Deep-Syntactic Representations [= DSyntRs] of all the utterances that express this SemR. The function of the lexicon in this process is to be a depository of all lexical data—information related to individual LUs of **L** and necessary for a linguistic model to go from a given SemR to all corresponding DSyntRs (and beyond), as well as vice versa.

Here, I consider a particular type of such a lexicon: the *Explanatory Combinatorial Dictionary* [= ECD], on which I have been working, together with several colleagues, over a period of forty years.

The ECD is a monolingual dictionary, proposed in the late 1960s by Alexander Zholkovsky and myself (Žolkovskij & Mel'čuk 1965, 1966, 1967). A little later, Jurij Apresjan joined us, so that the very first versions of ECD's lexical entries for Russian were authored by all three of us. In its present form, the ECD implements many of Apresjan's ideas (see Apresjan 1969a, b and 1974, as well as his more recent lexicographic output: Apresjan 1980, 1988a, b, 1990a, b, 1995, 2002, 2004). Three specimens of the ECD are available in a printed form: Mel'čuk & Zholkovsky 1984 for Russian, as well as Mel'čuk *et al.* 1984-1999 and Mel'čuk & Polguère (to appear) for French; see also Polguère 2000. A dictionary of Spanish collocations—DICE (= *Diccionario de colocaciones del español*)—is being developed: Alonso Ramos 2003 and 2004.

Over the last 40 years, numerous theoretical and descriptive papers on the ECD have been published, but lack of space prevents me from offering the reader a complete bibliography; some more references are given in the course of this exposition.

The ECD is set apart from other dictionaries by its general and specific (= linguistic) properties. They will be presented below in some detail; but first, let me point out that they are ideal objectives rather than actual accomplishments: **they are requirements that define a 'dream' ECD**, while published fragments of the ECD, compiled by ordinary humans, do not always fully live up to these requirements.

1.2 The General Defining Properties of the ECD

Compared to most existing dictionaries, the ECD has three distinctive general properties:

- 1) theoretical orientation,
- 2) formalized character,
- and
- 3) completeness at the level of each entry.

These properties follow exclusively from **our choice of an approach**; in other words, the ECD is—in the above-mentioned respects—the way it is because we want it to be like this, based on our scientific convictions. These properties do not reflect the nature of language or particularities of some specific languages. It is in this sense that I call them *general*; they are opposed to *specific*, or *linguistic*, properties of the ECD, discussed in the next subsection.

1.2.1 The Theoretical Nature of the ECD

An ECD is theory-oriented, in two senses:

- An ECD is compiled within a specific linguistic theory, in this case the Meaning-Text Theory [= MTT] (see Mel'čuk 1973, 1974, 1981, 1988a: 43-101, 1977). MTT presupposes a Meaning-Text Model [= MTM] of the language **L** under discussion—a model that features autonomous Semantic, Syntactic, Morphological, and Phonological linguistic modules and puts strong emphasis on the lexicon; the ECD constitutes an integral part of the Semantic Module of an MTM.

Most current linguistic theories view a linguistic description of a language as a grammar; a lexicon is taken to be an indispensable, but somehow less interesting annex to this grammar, where all the idiosyncrasies and irregularities that cannot be successfully covered by the grammar are stored. By contrast, MTT considers the lexicon as the central, pivotal component of a linguistic description; the grammar is no more than a set of generalization over the lexicon, secondary to it.

With such a viewpoint on the role of the lexicon, small wonder that the ECD is the focus of our approach.

All lexicographic concepts used in an ECD (Lexical Unit, lexeme, phraseme, collocation, Government Pattern, Lexical Function, and a host of others)

are rigorously defined within MTT; taken as a system, they form a well-developed lexicographic metalanguage.

- An ECD is a theoretical lexicon (cf. R. Lakoff 1973: 162-164) rather than a practical or conventional dictionary. More specifically, written within a clearly stated theoretical linguistic framework, an ECD purports to store all lexical knowledge shared by speakers of **L**. It contains all of the LUs of **L**—lexemes and phrasemes, as prescribed by the theory (i.e., by MTT)—and it is designed to dovetail with the grammar (\approx syntax + morphology) of **L**.

Conventional dictionaries are normally not consistent with a particular linguistic theory, and they cannot be: they are developed to satisfy practical needs, not to boost the science of language. Unlike them, the ECD is not limited by commercial, typographical, or even pedagogical constraints, which are inevitable for any conventional dictionary: an ECD is not meant to serve a particular public or to be adapted to a particular level of understanding of its prospective users. It is developed for the sake of linguistics and should satisfy the normal requirements for scientific descriptions.³

NB: 1. Since nothing is as practical as a good theory, an ECD of **L** can be successfully used as a source of several types of practical dictionaries. Once an ECD (or a fragment thereof) is compiled for **L**, it can be simplified, without loss of rigor or systematicity, for various practical uses: as a dictionary for learners, as a reference book for translators and editors, etc. It is easy to foresee bilingual or even multilingual dictionaries based on the ECD ideology. Applications of an ECD in computer text processing are even more obvious. Therefore, an ECD provides the infrastructure for practical products.

2. What just has been said should not be construed as a rejection of pedagogical considerations by ECD lexicography. Of course an ECD must be easily surveyable and consultable product—otherwise even its developers will not be able to deal with it. As a consequence, an ECD needs to be as pedagogical as possible; nevertheless, pedagogical constraints cannot be allowed to prevent an ECD lexicographer from formulating anything that he finds necessary.

1.2.2 The Formalized Character of the ECD

An ECD is a formalized dictionary—to such an extent that one can say that it is a lexical database. It puts heavy emphasis on explicitness and consistency:

- The ECD's explicitness means that nothing should be left to the user's intuition or logical abilities; nothing should be communicated through analogy or examples; everything has to be stated in an overt and precise way. To achieve this, the lexicographer is obliged to use a pre-established and well-developed lexicographic metalanguage.

- The goal of consistency in the ECD has two implications: first, similar LUs should be described in a similar way, so that the degree of semantic relatedness of two LUs is paralleled by the degree of the similarity of their entries; and second, different aspects of one LU, i.e. its semantic, syntactic, and lexical cooccurrence properties, should be described in conformity with each other.

This again follows from the ECD's general scientific *Einstellung*: to serve as the central component of a theoretical model of **L**. The formalization of the ECD makes it especially appropriate for computer treatment. On the one hand, it can be developed with the help of a computer, that is, by means of special lexicographic editing tools (see Polguère 2000); on the other hand, it can be widely used in an electronic format.

Conventional dictionaries also strive of course for systematicity and consistency, but—precisely because of their practical/commercial/pedagogical orientation—they cannot achieve the level of formalization set for the ECD. The primary concern of an ECD lexicographer is to observe all formal requirements, which allows for easy verification of linguistic facts. For this, a fairly sophisticated formal metalanguage has been developed over the years, and it is pressed into service for writing ECD entries.

1.2.3 The Internal Exhaustiveness of an ECD Entry

An ECD seeks to be exhaustive with respect to each individual LU it describes. That is, whatever a native speaker knows about an LU *L* of **L** must be fully presented in the ECD entry for *L*: no *etc.*'s or outright lacunae are allowed. Once again, this follows from the ECD's scientific conception.

Since the complexity of a lexical entry for an ECD greatly exceeds that of an entry in a conventional dictionary, for the time being an ECD lexicographer cannot envision an ECD as (more or less) exhaustive in the traditional sense—that is, including all or at least most of the lexical stock of **L**. Nowadays we can plan only for substantial fragments of the lexical stock of a language to be covered. However, as far as any one lexical entry is concerned, it must be developed exhaustively: it should contain all of the information needed to use the head LU successfully in all possible contexts.

1.3 The Specific Defining Properties of the ECD

Along with the three above-mentioned general properties, the ECD has five specific, or linguistic, distinctive features:

1) Orientation: it presents the lexicographic data organized in the direction from meaning to text, i.e., in such a way as to enable the user to pass from a given meaning to the corresponding texts

2) Semantic basis: the definition of L forms the basis of the entry for L

3) Coverage of the restricted lexical cooccurrence of each headword (= the LU described in a particular lexical entry, or a dictionary article): all collocations of L are included in the entry for L

4) Word list: it includes lexemes and idioms in the same order and treats them in the same way

5) The target of a lexical entry: each one is a monosemous LU.

These properties are determined by **the nature of human languages** as understood by MTT; that is, in this respect, the ECD is the way it is because the natural language is like this.

1.3.1 The Active Orientation of the ECD: From Meaning to Text

An ECD is an active, or encoding, dictionary: the information about words and expressions it contains is collected and presented exclusively from the viewpoint of text synthesis, i.e., speaking/writing, and not from that of text analysis, i.e., understanding spoken/written text. An ECD aims to provide the user with the maximum of linguistic means that exist in **L** for the expression of a given meaning in a given situation and a given context. This is in accordance with the general approach of MTT, in which the speaking process is considered to be more linguistic than the (language) understanding process, since the latter requires lots of extralinguistic knowledge and common sense (which are not part of linguistic competence as such). Any linguistic MTT-description is therefore organized from the viewpoint of the passage from meaning to text, and this is true of the ECD in particular. It is designed to answer questions not of the type ‘What does such and such an expression mean?’, but rather questions of the type ‘How do you express such and such a meaning?’.

For ease of reference by human users, the printed ECD entries are arranged in alphabetical order by headwords, but a rich system of cross-references and special codes allows the user to find all the LUs he might need in connection with his intentions. Thus, for instance, in the entry for DOG, the user can expect to find:

- the term of any particular breed of dogs—BORZOI, BULLDOG, GREAT DANE, GREYHOUND, etc.;
- the term for a dog of no definable breed—MONGREL, MUTT; CUR;
- the form of the sign warning of an attack dog's presence—BEWARE OF THE DOG;
- the terms for various human activities related to dogs—KEEP [a dog], WALK [a dog], HUNT [with a dog], MUSH, etc.;
- the names of different actions typical of dogs—BARK, SNARL, HOWL, WHINE, YAP, SNIFF, WAG [the tail], BITE, MAUL, etc.;
- the names of such artifacts used for dogs as COLLAR, MUZZLE, LEASH, DOGHOUSE, BASKET, etc.

Practically, any English LU that has the component 'dog' in its definition must be cross-referenced in the entry for DOG. As a result, an ECD entry will be much richer and more complicated than a corresponding entry in a conventional dictionary.

Among other things, an ECD specifies, for any headword L, all of its semantic derivations (see below, 2.3.1, p. 00ff.): for instance, for any L that denotes an event it should give its action noun (DISCOVERY₁ for DISCOVER, FIGHT_N for FIGHT_V, etc.), the names of its actants (DISCOVERER 'who has discovered', DISCOVERY₂ 'what has been discovered', etc.) and of its circumstantials (for instance, 'place of ...': THEATER for HOSTILITIES, HOLSTER for PISTOL, PIGSTY for PIG, etc.)—insofar of course as these exist.

Let it be emphasized that such expressions cannot simply be listed—each of them must be associated with a semantic description. Thus, for an ECD, it is not sufficient to have the expression *walk a dog* in the entry for DOG; there must also be a description of the meaning:

take the D. out so that
it can get exercise
and relieve itself : [to] walk [ART ~]

NB: In some cases, a conventional dictionary supplies for a headword a list of related terms, but without specifying their meaning, so that there is no way to identify the term you need. Thus, the French dictionary PR 2001 has practically all the useful expressions under the entry CHIEN ‘dog’, but with no explanations. Thus, it lists *promener son chien* ‘[to] walk one’s dog’, but nowhere in the dictionary is the meaning of this collocation given.

As a result of this policy, a typical ECD entry contains a series of ‘subentries,’ with mini-definitions and other types of lexicographic information. This point will be developed later.

The consistently synthetic orientation of the ECD does not prevent its use for analysis as well. Synthesis requires much more linguistic information than analysis (because when you try to understand, you can often guess many things from the context, while when you are speaking you have to know exactly how you should say what you want to say—it is far less probable that a guess will be right). Therefore, the information an ECD contains is quite sufficient for analysis, i.e., for language understanding.

1.3.2 The Semantic Basis of the ECD

Because of its orientation, an ECD is a semantically based dictionary: instead of simply listing data about words, it explicitly shows the links between each lexical element and its meanings. (Hence the adjective *explanatory* in the name of the dictionary.) This property of the ECD comes from two sources:

— On the one hand, the ECD’s semantic orientation is determined by the general belief that natural language is, in the first place, a tool for expressing meanings, so that semantic considerations underlie everything else in language and, consequently, in an MT-model of language; an ECD, which is a part of this model, must of course have this characteristic.

— On the other hand, the semantic orientation is inevitable in an active dictionary. Since an ECD is geared to helping the user find the right linguistic expressions for the meaning he wants to convey, it must concentrate on the description of meaning-to-text relations and therefore have a semantic basis.

Practically, this means three things:

- First, lexicographic definitions of LUs are much more rigorous and complex in an ECD than in a conventional dictionary.
- Second, all the information elements in the entry for a given LU L must conform to the definition of L, whereas a conventional dictionary habitually does

not show particular links between the components of the definition and other elements in the entry.

- Third, as was just said above, each LU L' given in the entry for L must be associated with the meaning 'L'. In other words, the entry for L contains a mini-definition of any L' mentioned in it. An example is provided on the previous page: it is not enough simply to put the expression [*to*] *walk one's dog* into the entry for DOG; such an expression must be 'introduced' by its definition, so that what is given is the lexical means L' necessary to express a given meaning 'L' related to DOG.

1.3.3 Cooccurrence as the Main Target of the ECD

An ECD is a cooccurrence-centered dictionary: it puts a strong emphasis on describing the whole of the restricted syntactic and lexical cooccurrence of L in the greatest possible detail. (Hence the adjective *combinatorial* in the name of the dictionary.) This property comes again from two sources:

- On the one hand, the ECD must faithfully reflect the properties of natural language. Restricted cooccurrence of signs, i.e., cooccurrence that is not determined semantically (= by the signified) or phonologically (= by the signifier) is a typical feature of natural languages, so the ECD has to pay close attention to the restricted combinability of LUs. (The phrase *to pay close attention* in the preceding sentence illustrates this point: it is a good example of restricted lexical cooccurrence: *pay + attention, close + attention*.)

- On the other hand, it is the active character of the ECD that brings the restricted cooccurrence to the fore. A conventional dictionary does not have to bother to indicate that ILLNESS can be *grave* or *serious*, but not **heavy* (as it is, e.g., in Russian: *tjažělaja bolezn'* lit. 'heavy illness'). A normal user—not necessarily a native English speaker—when encountering the phrase *grave/serious illness* in text, understands it without difficulty, and **heavy illness* will probably never be found in native speakers' speech. But an active dictionary should deal precisely with cases in which the user is looking for the expression of the meaning 'illness of high intensity' and therefore needs an explicit indication of the appropriate term(s).

The ECD puts forward a system of special techniques and tools for a rigorous description of restricted syntactic and lexical cooccurrence of the

headword **L**—the *Government Pattern* [= GP] and *Lexical Functions* [= LFs] respectively. (These concepts are discussed in Subsections 2.2 and 2.3.) To the best of my knowledge, no conventional dictionary uses such a methodology.

1.3.4 The ECD Describes All LUs of **L** Together and in a Similar Way

Given its formalized character, the ECD aims to cover all LUs of **L** in the same homogeneous way. There are three major types of LUs (cf. Endnote 2):

- A *lexeme*, i.e., a minimal LU—a single word in a single well-defined sense.

This is by far the largest part of the lexical stock of any language.

- A *compounding element*—a radical that can be used only as part of a compound. In English, this type of LU is typically represented by ‘neo-classical’ elements—of the type ANTHROPO- ‘human_N’, ICHTHYO- ‘fish’ (*ichthyophagous*, *ichthyosaur*, etc.), NEURO- ‘nerve’, SINO- ‘China’, or else -OLOGY ‘science’, -PHILE ‘lover’, -PHOBIA ‘fear’. Compounding elements, known in English lexicography as *combining forms*, are typical of some languages and rather marginal in others, as they are in English or French. However, they are rather numerous even in these languages. Thus, the *Webster Illustrated Contemporary Dictionary* (ed. by S. Landau, 1987, Chicago, Ill.: Ferguson, 938-950) lists more than 500 compounding elements of Greek and Latin origin used in English; for more on compounding elements in English, see Prčić 2005.

- An *idiom* or a *quasi-idiom*, i.e., a *multilexemic LU*—roughly, a set phrase (in our examples below a set phrase is shown by the symbols “_ _”):

— The meaning of an idiom does not include the meaning of any of its components, as in *_shoot the breeze_* ‘chat idly’, *_kick the bucket_* ≈ ‘die’, *_ride herd_* [*on N*] ‘control [*N*] by watching [*N*] closely’, *_call the shots_* ‘being in charge, give orders’, or *_once bitten, twice shy_* ‘somebody who has once suffered a misfortune is overcautious not to suffer it again’, etc.

— The meaning of a quasi-idiom includes the meanings of all its components, but none of them in a *communicatively dominant position*⁴, plus an unpredictable semantic configuration, as in *_start a family_* ‘conceive the first child with one’s spouse, thus creating a complete family’, *_fried eggs_* ‘dish made with the liquid contents of chicken eggs fried in a particular way’, or Fr. *_donner le*

sein_ [à N_Y] lit. ‘give the breast [to N_Y]’ = ‘X feeds the baby Y with X’s own milk by putting the nipple of a breast of X’s to Y’s mouth’, etc.

In what follows, idioms and quasi-idioms will not be distinguished—expressions of both types are called simply *idioms* (since in the present context, the distinction between them is irrelevant).

Compounding elements are not common in European languages and, because of this, do not cause serious difficulties for a lexicographer. From time immemorial, they have been entered in dictionaries of these languages along with words. From now on, they will be ignored in this paper. But idioms are a completely different story: while most conventional dictionaries do not systematically include them as regular headwords along with lexemes, the ECD does exactly this.

How original is this move? Including some—although by no means all—idioms in the dictionary on the same footing as monolexemic entries is typical of English lexicography. For other languages, the distinction between dictionaries of words and dictionaries of phrases is, to the best of my knowledge, much stricter. Thus, for French, German, Spanish and Russian we have separate dictionaries of phraseological expressions of different types. Even in English, set phrases are, as a rule, described in separate dictionaries: Rodale 1947, Makkai 1975, Cowie *et al.* 1993, OCD 2002.

In sharp contrast to such an attitude, an ECD includes as separate full-fledged entries of the same type the lexemes and the idioms of **L**. It is thus a dictionary of words **and** (set) phrases, and the phrases numerically prevail: on the one hand, all idioms appear as separate entries, on the other, all collocations—and they number in millions!—are included in the entries for their bases. To make a long story short, an ECD is, in a sense, a phrasal dictionary.

The massive inclusion of phrases in the ECD’s word list forces the lexicographer to consider problems typical of phrases, which are not encountered when one deals with words only. To better see the complexity of these problems, consider the French phrase *comme un âne* lit. ‘as an ass’. It has two senses, discussed in (1):

- (1) **a.** In the expression *têtu comme un âne* ‘stubborn as a mule’ the phrase *comme un âne* is an intensifier: it means roughly ‘very’ and is an element of the value of the Lexical Function⁵ Magn: $\text{Magn}(\textit{têtu}) = \textit{comme un âne}$. In

this sense, it is absolutely synonymous with the phrase *comme une mule* ‘as a mule’, which is also used as an intensifier of TÊTU.

The whole expression *têtu comme un âne* is a collocation of TÊTU.

b. In all other combinations *comme un âne* means ‘foolishly’ = ‘like a fool’:

(i) *Comme un âne, il a dit non* ‘Like a fool, he said no’.

(ii) *Il s’est comporté comme un âne* ‘He behaved like a fool’.

Note that in (1)a) no foolishness is implied:

c. *Paul est très intelligent mais têtu comme un âne*
‘Paul is very intelligent, but stubborn as a mule’.

In (1)a) the phrase *comme un âne* ‘very’ includes the noun ÂNE1 ‘he-ass’ and is an idiom, i.e., an LU of French. Theoretically, this idiom should be entered in an ECD as an independent entry. Practically, however, this is not very useful: since the phrase occurs mainly with the adjective TÊTU ‘stubborn’, being its collocata (= intensifier), it is sufficient to enter it in the lexical entry for TÊTU; as a result, we obtain:

TÊTU

...

Magn : *comme un âne, comme une mule*

In point of fact the phrase *comme un âne* ‘very’ cooccurs as an intensifier also with several verbs of shouting: *crier* ‘shout’ (<*brailler* ‘bellow’, *gueuler* ‘scream’, *hurler* ‘shriek’). Because of this, the lexicographer may want for it an independent entry. This, however, is not very relevant from our viewpoint: At any rate, the phrase must be stored in the dictionary articles for the verbs it intensifies, and this is sufficient for the linguistic model.

In (1)b) the phrase *comme un âne* ‘like a fool’ includes the noun ÂNE2 ‘fool’ and represents a free expression: it means exactly what it says (cf. *Quel âne, ce type-là !* ‘What a fool, this guy!’). Therefore, this phrase should not be stored in a dictionary at all; it is sufficient to store the lexeme ÂNE2.

As for the phrase *le coup de pied de l’âne* lit. ‘the blow of foot of the he-ass’ = ‘a mean and cowardly verbal attack on a beaten adversary’, it is an idiom and constitutes an autonomous entry in a French ECD. This idiom is given its own Government Pattern, its Lexical Functions (such as Oper₁: *allonger, asséner*, neutral *donner*, and Oper₂: *recevoir*), etc. The phrases *passer* (<*sauter*) *du coq à l’âne* lit. ‘pass (jump) from the rooster to the donkey’ = ‘change the subject in a sudden and

illogical way’ are both also idioms and also constitute separate entries of an ECD. At the same time, the phrase *l’âne de Buridan* ‘Buridan’s ass’ = ‘a legendary ass, who belonged to somebody called Buridan and who died of starvation between two identical bales of hay, being unable to choose between them’ is a collocation of ÂNE1 ‘ass’ = ‘donkey’ (since its meaning includes the meaning ‘âne1’ in the communicatively dominant position), and it is entered in the article for this noun: it does not need a separate entry.

As was just indicated, the ECD does not describe all phrasemes⁶ in the same way: idioms and quasi-idioms, on the one hand, and collocations, on the other, are given different treatment.

- Idioms, that is, full phrasemes (the phrases of the type Fr. *_ARRIVER <VENIR> COMME UN CHEVEU SUR LA SOUPE_* lit. ‘arrive like a hair on the soup’ ≈ ‘[of a remark, etc.] be incoherent and out of place’), and quasi-idioms, that is, quasi-phrasemes (phrases of the type Fr. *_DONNER LE SEIN_ [à N_V]* ‘give the breast [to N_V’]) cannot be effectively described as a function of just one of their components; they each receive a separate lexical entry—just like lexemes. A lexical entry for an idiom/ quasi-idiom has (almost) the same structure and gives the same type of information as one for a lexeme. Since from the perspective that interest us here idioms and quasi-idioms behave similarly, I will not distinguish them from now on, calling all of them indiscriminately *idioms*.⁷

- Collocations, that is, semi-phrasemes (the phrases of the type *accept/decline an invitation, pay a visit* or *keen interest*) can be and are described as a function of one of their components, known as the *base* (shown here in boldface); therefore, a collocation is presented only in the article for its base (as a *sui generis* subentry, as mentioned above).

As a result, an ECD lists, in the same alphabetical order, both lexemes and idioms. Thus, the French phrase *_RAT DE CAVE_* lit. ‘cave rat’ ≈ ‘very thin long candle’ = ‘wax taper’ is an obvious idiom having a proper definition and particular lexical cooccurrence. In practice, all conventional French dictionaries recognize this fact, but, instead of explicitly reflecting the autonomy and lexical character of the expression, they put *_RAT DE CAVE_* in the article for *CAVE*. However, describing an idiom in the entry for one of its components is unacceptable in the ECD framework:

1) It would mean embedding one full-fledged lexicographic entry within another. Embedding of this type is not a good solution from either the practical or the logical viewpoint.

2) This treatment is never consistent. Thus, PR 2001 stores *_CHEMIN DE FER_* lit. ‘iron road’ = ‘railroad’ and *_POMME DE TERRE_* lit. ‘earth apple’ = ‘potato’ as single LUs and puts them into general alphabetic order, while *_RAT DE CAVE_* OR *CORDON-BLEU* lit. ‘blue ribbon’ = ‘good cook’ (even if the latter is spelled with a hyphen) are hidden inside entries for *CAVE* and *CORDON* respectively.

3) Embedding of articles for idioms raises an artificial, but difficult problem: how to determine the ‘host’ for an idiom? Thus, should *_RAT DE CAVE_* be stored under *CAVE* or under *RAT*? There are no obvious reasons in favor of the choice either; the lexicographers of PR 2001 did not see any rationale, since they have put *_RAT D’HÔTEL_* lit. ‘hotel rat’ = ‘hotel thief’ under *RAT*, but *_RAT DE BIBLIOTHÈQUE_* lit. ‘library rat’ = ‘bookworm’, under *BIBLIOTHÈQUE*. (Semantically, *_RAT D’HÔTEL_* is closer to *HÔTEL* than *_RAT DE BIBLIOTHÈQUE_* is to *BIBLIOTHÈQUE*, so that, logically, one would expect just the opposite.)

To avoid unnecessary complications, the ECD presents idioms, as has been said, in the general alphabetic order of entries. Thus, a French ECD lists as its separate entries, together with single lexemes, all the idioms:

— noun phrases such as *_POT DE CHAMBRE_* ‘chamber pot’ and *_VASE DE NUIT_* lit. ‘night vase’, both meaning ‘chamber pot’, *_RAT DE BIBLIOTHÈQUE_* ‘bookworm’, *_RAT DE CAVE_* ‘wax taper’, *_RAT D’HÔTEL_* ‘hotel thief’, *_TABLE DE NUIT_* ‘night table’, etc.;

— verb phrases such as *_APPORTER DE L’EAU À [A_{poss-Y} = *mon, ton, son, ...*] MOULIN_* lit. ‘bring some water to A_{poss-Y} mill’ ≈ ‘give an argument in favor of N’s_Y viewpoint’, *_BRISER LA GLACE_* ‘break the ice’, *_COUCHER EN JOUE_ [N_Y]* lit. ‘lay [N_Y] in cheek’ = ‘take aim at N_Y [with a rifle]’, etc.;

— adjective phrases such as *_COUSU DU FIL BLANC_* lit. ‘sewn with white thread’ ≈ ‘that sticks out a mile/as a sore thumb’ or *[N_Y] _TOUT CRACHÉ_* lit. ‘[N_Y] all spit’ ≈ ‘someone who is the spitting image [of N_Y]’, etc.;

— prepositional phrases such as *_EN MIETTES_* lit. ‘in crumbs’ ≈ ‘destroyed’, *_PAR DESSUS LE MARCHÉ_* lit. ‘by over the market’ = ‘in addition, to boot’, etc.;

— full clauses (proverbs and sayings) such as *_AUSSITÔT DIT, AUSSITÔT FAIT_* ‘No sooner said than done’, *_L’ARGENT NE FAIT PAS LE BONHEUR_* lit. ‘The money does not make the happiness’ = ‘Money does not buy happiness’, *_TEL PÈRE, TEL FILS_* ‘Like father, like son’, etc.

To ensure a better and easier access to idioms for linguistic synthesis (= linguistic production), the ECD systematically uses semantic cross-references. Thus, the idiom *_COUCHER EN JOUE_* ‘take aim [with a rifle]’ is cross-referenced under *FUSIL* ‘gun, rifle’, *VISER* ‘take aim’ and *TIRER* ‘fire_v’; *_APPORTER DE L’EAU À [A_{poss-Y}] MOULIN_* is referenced from *ARGUMENT*. And for analysis needs—that is, for the user who is looking up an idiom—the ECD has cross-references to the idiom under all its constituents (minus, of course, structural words): *_COUCHER EN JOUE_* is cross-referenced also under *COUCHER* and *JOUE*, and so forth. (These, however, are simple references, with no definition or other types of information: such information is to be found in the lexical entry for the idiom.)

At the same time, the ECD aims at giving, for every LU L, all of L’s restricted lexical cooccurrence, so that the entry for L contains a great number of phrasemes of the third type—*collocations*: multilexemic expressions in which L fully retains its own meaning and syntactic-morphological properties. Since all collocations of L are described in the entry for L, collocations do not need lexical entries of their own. For instance, *pot shot* is described only under *SHOT*, *keep close watch* under *WATCH_N*, and *lodge a complaint* under *COMPLAINT_N*. This is done by using Lexical Functions [= LFs] (both standard and non-standard), a major innovation of the ECD.

The number of collocations in a dictionary averages roughly a couple of dozen per lexical entry, so they actually outnumber the LUs. (The distribution of collocations per LUs is extremely heterogeneous: some LUs have scores of collocations, while some have none.)

Until now, only a few lexicographers have attempted to cover restricted lexical cooccurrence systematically: A. Reum, at the beginning of the last century, whose dictionaries were reprinted as Reum 1953 (French) and Reum 1955 (English), J. Rodale (1947: English), W. Beinhauer (1978: Spanish), M. Benson *et al.* (1986: English), Ilgenfritz *et al.* (1989: French), Binon *et al.* (2000: French), and OCD (English); there are also some specialized dictionaries of individual types of

lexical cooccurrence, e.g., Deribas 1975 for Russian, Günther & Förster 1987 for Russian and German, and Bosque 2004 for Spanish. Yet such descriptions have never been carried out on a serious scale; besides, they are presented as separate lexicographic products: dictionaries of collocations. In sharp contrast, the ECD aims to describe all the collocations of any LU L in it within the lexical entry for L. As a result of this policy, plus the inclusion of all the idioms in the ECD word list, the ECD appears, as we said above, as a phrasal dictionary. I believe that the ECD corresponds fairly closely to the type of phrasal dictionary J. Becker (1975) and A. Pawley (1985) were so vigorously campaigning for.⁸

What is proposed here with respect to mono- and multi-lexemic expressions boils down to a systematic lexicographic description of all the lexemes and idioms of **L** in parallel, and the inclusion of all collocations in the entries for LUs. This proposal stems from the conviction that text synthesis is best done in terms of phrases (mostly collocations) rather than individual words.

NB: (Quasi-)idioms may, and sometimes must, appear as well in other lexical entries: *_shoot the breeze_* under CHAT, *_black belt_* under KARATE, and *_donner le sein_* ‘give the breast’, under BÉBÉ ‘baby’, ALLAITER ‘nurse’, etc. But these are only semantically motivated cross-references; the complete lexicographic descriptions of these idioms are found in their own lexical entries.

1.3.5 Each Article of the ECD Describes one LU

In contrast to conventional dictionaries, the basic unit of lexicographic description in the ECD is a *lexical unit* [= LU]: a lexeme, a compounding element or a (quasi-)idiom, i.e., a word, a part of a word or a set phrase taken in one well-specified sense—rather than a polysemous word, polysemous compounding element or polysemous idiom. In the ECD, each LU has its own lexical entry, and each lexical entry corresponds to one LU. All relevant lexicographic information is, strictly speaking, attached to an individual LU.

Related LUs having an identical signifier and sharing non-trivial semantic components in their signifieds are grouped into *vocables*, so that this grouping reflects polysemy (see below, Section 3, where a rigorous definition of the concept is given: Definition 7, p. 00). Thus, the English vocable IMPROVE includes six LUs—in this case, lexemes—and each one has a separate lexical entry; in the illustration below, lexemes are identified by their definitions and simple examples, but their lexical entries are not cited.

IMPROVE, verb

IMPROVEI.1a *X improves* ≡ ‘The value of the quality of X becomes higher’

[*The weather suddenly improved; The system will improve over time*]

IMPROVEI.1b *X improves Y* ≡ ‘X causes₁ that Y improvesI.1a’⁹

[*The most recent changes drastically improved the system*]

IMPROVEI.2 *X improves* ≡ ‘The health of a sick person X improvesI.1a’

[*Jim is steadily improving*]

IMPROVEI.3 *X improves at Y* ≡ ‘X’s execution of Y improvesI.1a, which is caused₁ by X’s having practiced or practicing Y’

[*Jim is steadily improving at algebra*]

IMPROVEII *X improves Y by Z-ing* ≡ ‘X voluntarily causes₂ that the market value of a piece of real estate Y becomes higher by doing Z-ing to Y’

[*Jim improved his house by installing indoor plumbing*]

IMPROVEIII *X improves upon Y* ≡ ‘X creates a new Y’ by improvingI.1b Y’

[*Jim has drastically improved upon Patrick’s translation*]

(The phrase [*to*] *improve oneself* is a set expression—actually a quasi-idiom, meaning ‘to educate oneself, to improve one’s culture’,—and should be considered here.)

The sense-distinguishing (= lexicographic) numbers, in accordance with existing dictionaries, purport to capture semantic distances between LUs within a vocable: Roman numerals express the larger distances, Arabic numerals, smaller ones, and letters, the smallest distances (for lexicographic numbers see 3.1, p. 00). Distances themselves are measured 1) by the proportion of shared semantic material and 2) by the regularity of the difference in question. Thus, the four lexemes grouped under IMPROVEI are considered to be closer to each other than to IMPROVEII and IMPROVEIII, because all of the former, but not the latter, include the semantic component ‘improveI.1a’. IMPROVEI.1a and IMPROVEI.1b are especially close since their semantic difference—’P’ ~ ‘cause₁ to P’ —is very regular in English. (For more details on lexicographic numbers, see 3.2, p. 00ff. The semantic component ‘cause₁’ represents non-agentive causation, as in *The falling tree* ⟨*The bullet*⟩ **killed the dog**; ‘cause₂’ stands for the agentive causation: *John* **killed the dog**.)

The policy of having separate LUs as separate entries has the advantage of making it possible to indicate for each sense its morphological peculiarities (if any), its special derivation range, its specific syntactic and lexical restricted cooccurrence, and so on—that is, to ensure the internal coherence of lexical entries. In ordinary dictionaries, after different senses have been lumped together under one head entry, there is not enough flexibility to indicate what properties go with which sense. The entry becomes cluttered with additional markings, and not really all individual properties of the senses are or even can be explicitly indicated.

All the pieces of information that are shared by all the LUs of the same vocable are extracted from individual LUs and ‘raised’ to the level of the vocable, so as to avoid unnecessary repetitions (this was done with the part-of-speech indication ‘verb’ in our mini-example). The policy of extracting common features and stating them only once, under a superior unit, is followed throughout the ECD, i.e., not only in vocables, but with respect to semantic and lexical fields as well (for the corresponding notions, see below, Subsection 3.1, pp. 00ff, and 4.3.2, p. 00).

In a nutshell, the ECD is an active phrasal dictionary, based on the semantics of the LUs treated and stressing their restricted cooccurrence; its unit of description is a Lexical Unit, that is, roughly, a word or a set phrase taken in a particular sense (rather than a polysemous word, as in all current dictionaries).

2 The ECD’s Microstructure: A Lexical Entry

An ECD entry is a full description of an LU L—a lexeme or an idiom. It comprises three major divisions, which correspond to the three components of the linguistic sign, as it is understood in MTT. A linguistic sign *s* is a triplet:

$$s = \langle \text{‘Signified’} ; / \text{Signifier} / ; \Sigma(\text{syntactics}) \rangle,$$

where the signified (*signifié*) and the signifier (*signifiant*) are taken in their Saussurean sense, and syntactics is the set of properties that control the cooccurrence of *s* with other signs (see Mel’čuk 1993: 109ff and 2006: 384ff).

An LU is presented in the ECD as a linguistic sign. Each component of the triplet is described in a ‘zone’ of an ECD lexical entry; each zone is, in its turn, subdivided into sub-zones (indicated in boldface).

1) The *Semantic Zone* describes the signified of L, i.e., it contains all of the data concerning L's semantic properties. It consists of two sub-zones: the **definition** of L, which fully specifies L's meaning, and L's **connotations** (meanings that the language **L** associates with L, but that are not part of its definition; see Iordanskaja & Mel'čuk 1984 and 2006).

2) The *Phonological/Graphematic Zone* deals with the signifier of L, i.e., it gives all of the data concerning its phonological properties. Here again we find two sub-zones: L's **pronunciation**, including its syllabification, and any non-standard **prosodic properties** (see, in connection with this, the detailed study Apresjan 1990a), as well as L's spelling variants (if any).

3) The *Cooccurrence Zone* presents the syntactics of L—i.e., all of the data concerning its combinatorial properties: with what inflectional elements L combines and how it behaves in such combinations, what derivations L admits, what are L's possible syntactic actants, etc. It is organized into five sub-zones, where L's morphological, syntactic, lexical, stylistic, and pragmatic cooccurrence is specified.

The morphological sub-zone: **Inflection Data**

The **Inflection Data** (conjugation/declension class, irregular forms, missing forms, permitted alternations, etc.—cf. Apresjan 1988a) cover the morphological cooccurrence of the stem of L from the viewpoint of its inflectional paradigm.

The syntactic sub-zone, with two sub-sub-zones

The **Government Pattern** deals with the active syntactic valence of L: its cooccurrence with its syntactic actants (Deep- and Surface-), while **Part of Speech and Syntactic Features** describe L's passive syntactic valence: its participation in specific constructions as a dependent.

The lexical sub-zone: **Lexical Functions**

Here are presented all lexical links between L and other LUs of **L**—in terms of LFs. LFs cover both semantic derivations and collocations of L with individual LUs or very small and irregular groups of LUs.

The stylistic sub-zone: **Usage Labels**

Usage Labels specify, for the headword L, its speech register (**informal, colloquial, vulgar, poetic**, etc.), temporal (**obsolescent, archaic**) and geographical (**British, Indian, Australian**) variability, and the like.

The pragmatic sub-zone

Pragmatic Clues pinpoint the real-life situations in which a particular expression is appropriate/inappropriate: e.g., in English, a street sign features the expression *No parking* rather than the fully understandable and syntactically well-formed (but non-conventional) [#]*Parking forbidden* or [#]*Prohibition against parking*, while in French the opposite is true: the correct formula is *Défense de stationner* lit. ‘Prohibition to park’, rather than [#]*Aucun stationnement* ‘No parking’ (the symbol ‘#’ indicates pragmatic inappropriateness). The expression NO PARKING, given as a value of a non-standard LF in the lexical entry for the verb PARK_V, must be supplied with the indication “on a street sign”; the expression BEST BEFORE ..., given in the entry for such nouns as CAN_N [tinplate container] or PACK_N, has the indication “on a container of canned or packed food”; etc.

4) To the above three major zones, the ECD adds a fourth, the *Illustrative Zone*. Completely redundant from a strictly scientific viewpoint, it is useful for the human user of the ECD: linguistic illustrations not only facilitate the understanding of a lexicographic description, but also serve to substantiate the claims about possible/impossible expressions made in the corresponding entry.

Only the following five of the sub-zones mentioned above will be considered here: from the *Semantic Zone*, **Definition** (to the exclusion of connotations); from the *Cooccurrence Zone*, **Government Pattern** and **Lexical Functions** (to the exclusion of morphological data, syntactic features, etc.); and from the *Illustrative Zone*, **Examples**. Only these sub-zones appear in the lexical entries for the verb BAKE given in Section 5.

2.1 The Lexicographic Definition in the ECD

Since the definition of the LU L in an ECD entry for L in language **L** is actually the central component to which all the other aspects and elements of L’s lexicographic description are geared, it deserves a more detailed characterization. It is also the most difficult part of a lexical entry to write: as Anna Wierzbicka has put it, ‘The process of constructing a lexicographic definition is—or should be—a search for truth’ (Wierzbicka 1996: 264).

The ECD-type definition of an LU L is, theoretically speaking, its Semantic Representation [= SemR]; however, it is not presented in the standard

form of a labeled network supplied with the specification of Semantic-Communicative areas (Rheme ~ Theme, Given ~ New, Foregrounded ~ Backgrounded ~ Neutral, etc.), as should be done in the framework of Meaning-Text semantics. An ECD definition is written ‘verbally’—i.e., in a special semantic metalanguage, which is basically **L** itself, but submitted to particular constraints. (The verbal definition of L and its standard SemR are of course in one-to-one correspondence.)

An ECD definition has the following tripartite form, as all lexicographic definitions do:

definiendum ≡ definiens,

where the *definiendum* is the presentation of the LU L to be defined, and *definiens*, the presentation of its meaning ‘L’ (i.e., the definition in the narrow sense); ≡ is the symbol of equivalence. I will discuss the ECD definition in four sub-subsections:

- Substantive requirements for ECD definitions
- Rules for writing ECD definitions
- Criteria to be used for writing ECD definitions
- General characteristics of ECD definitions

2.1.1 Substantive Requirements on ECD Definitions

There are three linguistic conditions that the definition of an LU L in an ECD must satisfy; namely, it should ensure an accurate description of

- L’s links with the extralinguistic world—L’s denotational potential;
- L’s semantic links with related LUs in the lexicon—L’s paradigmatic potential;
- L’s syntagmatic links with other LUs in the sentence—L’s syntagmatic potential.

Let us consider these conditions in turn.

1) L’s denotational potential is the whole set **D** of extralinguistic world entities or facts to which L (or, more precisely, the meaning of L) can be applied—i.e., to which L refers. The definition of L must include all the components necessary and sufficient to allow the use of L when appropriate—i.e., when the speaker wants to refer to an element of **D**—and to disallow it when inappropriate. It should not, however, embody the entire amount of knowledge a speaker possesses

in connection with **D**: much, or even most, of this pertains to the things and/or situations denoted by L, and not to L itself.

An ECD lexicographic definition must by all means avoid including any information about the real world (i.e., encyclopedic information), beyond what is strictly necessary to distinguish the meanings of LUs being described.

The aim of semantic decomposition of lexical meanings is restricted to determining the applicability of LUs to things and situations already identified, while it should not try to contribute to the identification of the thing or the situation itself. The biological definition of the scientific concept “CAT” aims at distinguishing real cats from other mammals, that is, at identifying a cat; the semantic definition of the English lexeme CAT, on the other hand, should be concerned only with the use of the lexeme as applied to the previously already identified animal (cf. the remarks in McCawley 1986). In other words, an ECD definition does not reflect the naive concepts speakers have of the corresponding object or event, but tries to encapsulate the subconscious speakers’ knowledge of the LU and its uses. For instance, an English speaker knows that cups normally come with saucers, that they are breakable, that they should be washed, etc., but this knowledge concerns the object and is irrelevant for the lexicographic definition of the noun CUP: a cup will still be called *a cup* by an English speaker, even if (say, in a particular culture) it is not put on a saucer, never washed and is made of unbreakable plastic. (Lexicographic definitions of such English lexemes as CUP, MUG, SAUCER and the like are thoroughly discussed, although in a different perspective, in Wierzbicka 1985: 10ff; cf. also Wierzbicka 1991, especially pp. 87-91, where the lexicographic definitions of CAT and BOTTLE are reproduced.) It is by no means easy to draw a line between what is encyclopedic knowledge and what pertains to the linguistic meaning of an LU, but a linguist must do his best in barring the data about objects and situations from the definitions of corresponding LU.¹⁰

2) L’s paradigmatic potential is the whole set of LUs in the lexicon of **L** with which L shares important semantic material (= has semantic bridges, see **3.1**, Definition 5, p. 00). From this viewpoint, the definition of L must allow for the following two things:

• On the one hand, the definition of L must ensure a correct selection of L and/or of its semantic partners in possible paraphrases; it must state clearly the semantic similarities and differences between L and its potential substitutes. Thus, consider the ECD-style definition of ESCAPE1.1c (*He escaped from Cambodia/from his captors*):

X escapes from Y by way of Z to W ≡ ‘X being in place Y¹ where Y² is, such that (something related to) Y² threatens X and such that it is possible that X will not be able to move away from Y¹ before the threat by Y² to X is realized,|| X succeeds in intentionally moving away from Y¹ via Z to place W, which causes¹ that the threat by Y² is not realized’

Explanations

1. Y¹ and Y² represent what is known as a *split variable*: Y = Y¹/Y². The technique of split variables allows us to cover by the same definition such different expressions as *escape from Cambodia/from the prison camp* [= Y¹] and *escape from one’s captors/from the execution* [= Y²]. The variable Y corresponds to two different participants of the situation ‘[to] escape1.1c’, whose expressions, however, are incompatible in one clause and so need not to be represented by two different variables, i.e., to correspond to two different SemA-slots.
2. The symbol ‘||’ indicates that everything preceding it is a presupposition; see below, 2.1.3.4, p. 00.
3. Recall that ‘cause¹’ stands for non-agentive causation.

This definition allows one to replace ESCAPE1.1c with FLEE in some contexts (*Dith Pran escaped Cambodia* ≈ *Dith Pran fled Cambodia*), but not in others (*Dith Pran barely escaped* < **fled* > *Cambodia*); this is ensured by a careful choice of the central component in the corresponding definitions. Thus, the combinability of ESCAPE1.1c with BARELY ‘with a small margin of probability’ is guaranteed by the central component ‘succeed’ in the definition of ESCAPE1.1c (‘barely succeeded’ is semantically correct), while the central component in the definition of FLEE is ‘move away from ...’, which does not combine with ‘barely’: *‘barely moved away from ...’ is semantically incorrect. (See Apresjan 1992b on a program for a Russian dictionary of synonyms, in which the lexicographic data should allow one to formulate definitions that satisfy this requirement; for the dictionary itself, see Apresjan 2004.)

- On the other hand, the definition of L must make explicit all intuitively felt semantic links between L and all LUs L'_i related to L (in the same vocable or in other vocables) in **L**'s lexicon. We cannot accept for verbs ESCAPE1.1a (*He escaped from the state penitentiary*) and ESCAPE1.1c the definitions proposed in AHD 1981: 'break loose from confinement' and 'succeed in avoiding (capture, danger, or harm)', since they do not explicitly show the semantic relatedness of these two senses of ESCAPE. In an ECD, the definitions of ESCAPE1.1a and of ESCAPE1.1c must share a central component: for instance, '(intentionally) move away from ...'.

Another example of the need for explicit presentation of the semantic links between LUs is the definition of the noun SNOW: in an ECD, it should include the component 'white'—not because snow is white (this is a piece of encyclopedic information about the substance «snow», rather than about the lexeme SNOW), but because English has expressions like SNOWY 'pure white', [*white*] _AS SNOW_, SNOW-[*white*], and SNOW3 '[*white*] cocaine powder', which are perceived as semantically related to SNOW; the definition of SNOW must make this speakers' perception explicit. At the same time, the definition of SALT, even though salt is also white, should not include the component 'white'—since there are no supporting English expressions (no **salty white*, **salt-white* or something similar; *white as salt* might be found, but it is not as clichéd as *white as snow*). I will return to the discussion of this case in connection with the Criterion of linguistic relevance, Subsection 2.1.3, p. 00.

3) The syntagmatic potential of L is the whole set of L's lexical 'partners'—LUs that coccur with L. The definition of L must account for two types of lexical cooccurrence:

- In the first place, the definition of L must ensure the proper combinability of L with L's non-restricted lexical partners, that is, all LUs that can/cannot coccur with L on the basis of semantic considerations only (restricted lexical cooccurrence is covered by Lexical Functions, mentioned above and discussed below). Thus, GRAFT² [*corruption*] cannot be defined as 'the practice of obtaining money unlawfully or unfairly...' (LDoCE 1978), since *practice* and *graft* differ in their morphology (*practice* is countable, while *graft* is uncountable) and, as a result, in their free lexical cooccurrence: *these various practices* vs. **these various grafts*. A definition of GRAFT², which is better in this respect, could be 'obtaining money

unlawfully or unfairly ...’, since the morphosyntactic behavior of gerunds is closer to that of *graft* than that of *practice*. To sum up, the definition of L must fully account for L’s free (i.e., semantically motivated) combinability with other LUs in the text.

- At the same time, the definition of L must contain components that can accept meanings expressed by L’s LFs. Thus, since English has the expression *pass an exam* [*pass* = $\text{Real}_2(\text{exam})$, a standard LF, see below, 2.3.3, p. 00] the definition of EXAM must include the component ‘... Y’s goal being to show the necessary level of knowledge or skills ...’; [*to*] *pass* expresses the achievement of this goal. (For more on the links between the definition of L and its LFs, see below, Subsubsection 2.1.3.2, p. 00.)

So that is what we want from a lexicographic definition in an ECD. It must:

- ensure L’s correct denotation (by specifying the applicability of L to previously identified entities of the extralinguistic world);
- make explicit all L’s paradigmatic links with other LUs in the lexicon (allowing, among other things, for correct choices between paradigmatic relatives of L); and
- cover all L’s syntagmatic links with other LUs in the sentence (i.e., predict L’s semantically controlled lexical combinability, both free and restricted).

2.1.2 Rules for Writing ECD Definitions

To make an ECD definition conform to the three above requirements, five rules are proposed that control the formal correctness of definitions: Rule 1 concerns the form of the definiendum, Rules 2 through 4, the form of the definiens, and Rule 5, the equivalence ‘definiens = definiendum’ itself.

As for the factual (i.e., linguistic) correctness of ECD definitions, it must be checked by several lexicographic criteria, see Subsection 2.1.3.

2.1.2.1 Rule 1: Propositional Form Rule

Let the LU L to be defined have a signified that is a semantic predicate or a quasi-predicate:¹¹

$$\text{‘L}(X_1 ; \dots ; X_n)\text{’}.$$

The meanings that correspond to the arguments X_1, \dots, X_n are known as Semantic Actants [= SemAs] of L (for SemAs, see Mel'čuk 2004a).

L's definiendum must be a propositional form—an expression constituted by L supplied with variables X, Y, Z, ... that represent L's SemAs (and with structural elements such as *with*, *out of*, ..., syntactically relating the variables to L).

Unlike all conventional dictionaries, a definiendum in an ECD cannot be, generally speaking, simply the name of the LU being defined. For a predicative LU L (i.e., L whose meaning is a predicate or a quasi-predicate), the definiendum is the name of L supplied with variables which stand for L's Semantic Actants. (These variables are symbolized by uppercase Latin letters X, Y, Z, etc.) Thus, in order to define REPROACH_V (as in *The Senate leader reproaches him for recent remarks*), the ECD actually defines the expression *X reproaches Y for Z*; for CHANGE_V (in *Life has changed into a endless succession of shocks*), the definiendum is *X changes into Y*; and for SKIRMISH_N you have to define *skirmish between X and Y over Z*.¹²

Only a non-predicative LU, semantically close to a proper name, can have a definiendum that is not a propositional form: the name of a natural substance (*water, sand, air*), of a wild animal or of a plant species (*giraffe, oak, wheat*), or else of a unique natural object (*Sun, Moon*).

2.1.2.2 Rule 2: Decomposition Rule

The definiens of an LU L must be written in terms of meanings of two or more LUs L_1, L_2, \dots, L_n such that 1) ' L ' = ' L_1 ' \oplus ' L_2 ' \oplus ... \oplus ' L_n ' and 2) each ' L_i ' is semantically simpler than ' L '; in other words, the lexicographic definition of the meaning ' L ' must be its decomposition.

The symbol \oplus stands for the operation of linguistic union—in this particular case, semantic amalgamation or the uniting of two meanings.

Comments on Rule 2

1. Semantically simpler. The central point in the formulation of Rule 2 is the requirement of defining ' L ' in terms of simpler meanings. The expression *be semantically simpler than* is used here in a technical sense: 'be included in the definition of'. Therefore, the semantic relation 'be semantically simpler than', as

understood here, does not apply to any two meanings; it is applicable only to such a pair of meanings that the one must be defined in terms of the other. It makes no sense to ask, for instance, what is semantically simpler—‘abhorrence’ or ‘invite’.

The meaning ‘L’ is semantically simpler than the meaning ‘L’ if and only if [= iff] ‘L’ can be defined in terms of ‘L’ but not vice versa:

$$‘L’ = ‘L’ \oplus \dots \oplus ‘L’_n’, \text{ while } ‘L’ \neq ‘L’ \oplus \dots \oplus ‘L’_m’.$$

Let me illustrate the notion of semantic simplicity by an example. Which is simpler: ‘manII’ (= ‘adult male human’) or ‘woman’ (= ‘adult female human’)? Following A. Wierzbicka (1972: 44ff), we say that the meaning ‘woman’ can—and must—be defined without mentioning ‘manII’, exclusively by women’s capacity of giving birth; but ‘manII’ is impossible to define without mentioning ‘woman’, since ‘manII’ has to be defined by the capacity of causing that a woman gives birth.¹³ As a result, the meaning of the LU MANII ‘adult male person’ includes the meaning of the LU WOMAN ‘adult female person’, but not vice versa; WOMAN is thus semantically simpler than MANII.

Generally speaking, the possibility of defining ‘L’ in terms of ‘L’ but not vice versa can be determined in the process of subsequent decompositions and substitutions. However, this is not always immediately obvious. Consider the French nouns ASTRONOME /astronɔm/ ‘astronom+er’ and ASTRONOMIE /astronɔm+i/ ‘astronom+y’. One can say that ‘astronome’ is \approx ‘person who does astronomy’, defining ASTRONOME via ASTRONOMIE. Inversely, it is also possible to say that ‘astronomie’ is \approx ‘science done by astronomers’, and then ASTRONOMIE is defined via ASTRONOME. Both statements are factually true. However, from a lexicographic viewpoint, only the first is acceptable as an ECD definition; here is why.

- If ‘astronome’ = ‘person who does astronomy’, then at the next stage of decomposition, ‘astronomie’ (= ‘astronomy’) is defined as \approx ‘science of celestial bodies’; we do not need to return to ‘astronome’, and a vicious circle is avoided.

- If, on the other hand, ‘astronomie’ = ‘science done by astronomers’, then, while continuing the decomposition, one is forced to define ‘astronome’ (= ‘astronomer’) as ‘person who studies celestial bodies’. But in this case, the substitution gives us

‘astronomie’ = ‘science done by people who study celestial bodies’.

By an obvious reduction, ‘science done by people who study X’ is simply ‘science of X’. Consequently, we can write

‘astronomie’ = ‘science of celestial bodies’.

In this way, we come back to the first statement: ‘astronome’ must be defined as ‘person who does astronomy’, but not the other way around.

An additional problem in this case is created by the fact that the morphological relation between LUs ASTRONOME and ASTRONOMIE is inverse with respect to their semantic relation: *astronome* \subset *astronom+ie* [morphologically ASTRONOME is simpler], while ‘astronome’ \supset ‘astronomie’, because ‘astronome’ = ‘**person who does astronomy**’ [semantically, ASTRONOME is more complex]. Here, morphological complexity contradicts semantic complexity. Note that in English the situation with the equivalents of ASTRONOME and ASTRONOMIE is different: in English, ASTRONOM+ER and ASTRONOM+Y are formally of equal complexity (morphological intersection instead of inclusion), so there is no immediate conflict with semantic inclusion.

There exists no foolproof method that would allow one to automatically determine relative semantic simplicity of two LUs. Only by several successive attempts can the linguist reach a good formulation of a definition. Nonetheless, in spite of the lack of easy techniques for its application, the notion ‘be semantically simpler than’ is absolutely crucial for the ECD.

2. Semantic decomposition. Rule 2 means that a definition of ‘L’ must of necessity be ‘L’'s decomposition; the scientific methodology of semantic decomposition, launched in the late 1960s by A. Wierzbicka, is central to the Meaning-Text approach. This methodology has three important consequences.

- In an ECD, it is forbidden to define by synonyms: a synonym of L is by no means a decomposition of the meaning of L. Thus, one cannot define the French verb RIPOSTER ‘[to] retort, react, talk back’ [*Elle riposta en éclatant d’un rire fou* ‘She reacted [to this] by bursting into hysterical laughter’] simply by *répondre* ‘[to] answer’, *répliquer* ‘[to] rejoin’ ou *réagir* ‘[to] react’: such definitions are by no means decompositions.

However, the ban on synonyms as definitions does not preclude the use of a poorer synonym of L as the central (= generic) component in The definition of L. Let me illustrate using RIPOSTER:

X riposte à Y par Z ≡ ‘Person X *répond*_{1.2} [= reacts] to words or gestures of person Y that X believes to be addressed by Y to X and harming X—by using words or gestures Z that X addresses to Y with the purpose to harm Y’.

This definition covers at least the following cases:

(2) **a.** *À leur proposition de se rendre, le maréchal riposta par un mot devenu célèbre*

‘To their proposal to surrender, the marshal retorted with the word that became famous’. [This was the exclamation *Merde !* ‘Shit!’, with which the commander of the Old Guard, Marshal Cambronne, answered the British proposal to surrender during the Battle of Waterloo, 18 June 1815.]

b. *Va ranger ta chambre, dit maman, et ne riposte pas !*

‘Go and clean your room, Mother said, and don’t talk back!’

c. *Riposter à un policier par un bras d’honneur, c’est de l’audace !*

‘To reply to a policeman by giving him the finger is really audacious!’

It is clear that ‘riposter’ is a particular case of ‘répondre_{1.2}’, so that the latter is a poorer synonym of the former; therefore, using ‘répondre_{1.2}’ in the definition of ‘riposter’ is more than justified: it is the only possibility. What Rule 2 forbids is the use of a synonym as a whole definition.

- In an ECD, it is forbidden to define using semantically empty LUs. Thus, to continue our example with the French verb RIPOSTER, an ECD lexicographer cannot accept the definition given in PR 2001: RIPOSTER ≡ ‘faire une riposte’ [= ‘make a ripost’]. Although the equality ‘riposter’ = ‘faire une riposte’ is factually true, it cannot serve as a definition, because ‘faire une riposte’ is not a decomposition of ‘riposter’: the verb FAIRE ‘make’ is empty here (it is a light, or support, verb), and ‘riposte’ is semantically equal to ‘riposter’.

- Vicious circles, the current plague of practically all existing dictionaries, are thus successfully avoided in the system of lexicographic definitions in an ECD.¹⁴

3. Semantic primitives. Requiring that a definition of L should be a decomposition of the meaning of L guarantees that, by carrying out subsequent decompositions of lexical meanings of language **L** as far as possible, one will

inevitably arrive at a set of LUs $\{L_i^{\text{PRIM}}\}$, whose meanings cannot be decomposed any further in terms of the meanings of other LUs of **L**. The LUs $\{L_i^{\text{PRIM}}\}$ are the semantic primitives OF **L**. The modern study of semantic primitives has been vigorously developed by A. Wierzbicka (the reader is referred to her work: Wierzbicka 1972, 1980, 1996 and Goddard & Wierzbicka 1994). But, however interesting and important, the question of semantic primitives cannot be discussed here. I will limit myself to the following two remarks.

- It is not the case that semantic primitives are not decomposable at all: they are not decomposable, i.e., not definable, only in terms of other lexical meanings of **L**. In principle, they are definable, but in terms of extralinguistic notions—logical, psychological, mathematical, or physical ones (see Mel'čuk 1989). Thus, the meaning 'no/not' (= negation) is a semantic primitive: it seems impossible to define the meaning of the LUs NO or NOT in terms of semantically simpler LUs of English. Yet in logic the expression NO/NOT = \neg is defined easily:

Negation \neg is an operation such that **if** A is a true proposition, **then** \neg A is a false proposition, and vice versa.

This is a perfect definition, yet it cannot appear in a dictionary of English along with the definitions that have been presented above. Other examples of the same type of lexicographically inadmissible definitions include WATER = 'H₂O', LIGHT = 'electromagnetic waves of a given frequency', or CAT = '*felis felis*' (the meanings of these LUs are not semantic primitives: they can and should be properly defined in the ECD-style). Such definitions characterize the thing denoted by the LU in question, its referent, rather than the linguistic meaning of the LU itself.

- Semantic primitives in our perspective are language-specific—unlike universal primitives of human thought introduced by Wierzbicka. In point of fact, in MTT, we should speak of the semantic primitives of English, Chinese, Swahili, Totonac, etc. This does not preclude (near-)identity of the sets of semantic primitives for different languages; this is, however, a serious problem that must be passed over in the present context.

2.1.2.3 Rule 3: Standardization Rule

Given the formal nature and rigorous logic of an ECD, the ECD-style definition of an LU must be made in a uniform semantic metalanguage subject to

explicit constraints, which are applicable in a homogeneous way to the whole lexicon. In the case of an English ECD this semantic metalanguage is ‘processed’ English, and the constraints in question can be expressed by the Standardization Rule:

The lexicographic definitions in the ECD should contain neither 1) ambiguous expressions (= carrying different meanings) nor 2) synonymous expressions (= carrying the same meaning).

Comments on Rule 3

1. Non-ambiguity of defining elements. The first constraint means that each expression used in a definiens must always have one and the same meaning: it is a *semanteme*—i.e., the meaning of a well-defined LU. In order for this requirement to be satisfied, all lexical items of **L** have to be disambiguated by means of distinctive lexicographic numbers that specify the sense under discussion. Lexicographic numbers are the current practice of all existing dictionaries—which, however, never use them in their own definitions; as a result, the definitions are quite often highly ambiguous. Most of the time, for a human user this ambiguity is successfully resolved by the context of the definition and, especially, by the examples; however, for a non-native speaker it can create problems, and from a scientific viewpoint it is completely unacceptable.

To illustrate the problem of ambiguous expressions in definitions, consider the definition of the French lexeme HAUTEUR_{1.1} ‘height [e.g., of a tower]’ as presented in PR 2001:

HAUTEUR_{1.1} ≡ ‘dimension dans le sens vertical, de la base au sommet’

[‘dimension, in the vertical direction, from the base to the summit’]¹⁵

This definition is multiply ambiguous, given that DIMENSION has, in the same dictionary, 6 lexicographic senses, SENS has 3, BASE, 11, and SOMMET, 3 (the adjective VERTICAL is monosemous); as a result, the definition of HAUTEUR is formally interpretable in 594 (= 6 × 3 × 11 × 3) different ways! Such a state of affairs cannot be tolerated in an ECD. The definition has to be rewritten as follows:

HAUTEUR_{1.1} ≡ ‘dimension_{1.2} dans le sens² vertical, de la base_{1.1} au sommet₁’

Lexicographic distinctive numbers here are borrowed from the same dictionary, which has them, but does not use them in its own definitions (for the discussion of this problem, see Rey 1990: 52).

Using disambiguated defining elements in the definitions allows for automatic substitution of these elements by their own definitions, and this enables formal verifications of consistency. Even for a human user, definitions with distinctive numbers of their elements are of great help: such a definition ensures that the user gets the exact meaning of the LU he is looking up. Moreover, for any formal treatment, including computer processing, this is the only option. From now on, the English definitions proposed in this paper will use the disambiguated elements, with lexicographic numbers borrowed from LDoCE 1978. (Structural words, used in the definitions for readability, are of course not disambiguated. Since the sense differentiation in LDoCE 1978 is not always satisfactory from our viewpoint, the disambiguation shown in this paper is, on many occasions, quite approximate—which is not too serious in the present context.)

2. **Non-synonymy of defining elements.** The second constraint in Rule 3 means that each meaning to be expressed in a definiens is always—i.e., in all definitions appearing in an ECD—expressed by the same LU. To respect this constraint, it is necessary to determine, for each meaning to be expressed in a definition, one and only one LU that will express it in all definitions. However, it is extremely difficult to recognize identical meanings expressed by different LUs—at any rate, much more difficult than to recognize formally identical LUs carrying different meanings; therefore, the second constraint poses even more problems than the first one. Consider a simple example involving the names of artifacts:

WATCH is often defined as ‘device allowing one to know the time’,

HAMMER—as ‘tool for striking’,

KNIFE—as ‘instrument servng to cut with’, and

SPOON—as ‘utensil used to carry food to the mouth’.

It is not immediately obvious that the LUs *allowing* [to], *for*, *servng* [to] and *used* [to] express (in the context of the definitions sketched out here) the same meaning. But let us suppose, at least for the sake of our discussion, that they do. Then this meaning must always be expressed, in ECD definitions, by one and the

same lexical expression whose choice is determined by the following four conditions:

1) The expression selected must be the least ambiguous possible—in order to allow for a better functioning of the linguist’s intuition. Thus, although the preposition *for* in the definition of HAMMER expresses the necessary meaning quite well, it is too ambiguous and therefore must be rejected. (Even disambiguated with a lexical number it remains difficult to process.) *Allowing*, *serving* and *used* are lexically less ambiguous, but still create polyvalent expressions: each means simultaneously ‘actually serving’, ‘which can serve’, and ‘designed to serve’.

2) The expression selected must be the least idiomatic possible; that is, it should not carry additional nuances difficult to filter out.

3) The expression selected must have the greatest syntactic flexibility possible: it should be usable in the widest variety of contexts.

4) And last but not least, the expression selected must be semantically the most precise possible.

Taking these conditions into account, we shall choose the expression ‘designed for’ for the four LUs above; then the definitions (of course incomplete) can be formulated as follows:

WATCH : ‘device₁ designed₂ for showing₃ the time₁₂ ...’

HAMMER : ‘tool₁ designed₂ for striking₂ ...’

KNIFE : ‘instrument₁ designed₂ for cutting₂ ...’

SPOON : ‘utensil₂ designed₂ for carrying₁ food₁ to the mouth₁ ...’

The expression *designed₂ for* is clearly less ambiguous than its competitors: it is also less idiomatic and/or idiosyncratic, it is syntactically quite flexible, and semantically, it is the most appropriate for artifacts.

2.1.2.4 Rule 4: Maximal Block Rule

If the lexicographic definition of L contains a semantic configuration composed of the meanings ‘L₁’ ⊕ ‘L₂’ ⊕ ... ⊕ ‘L_n’ such that this configuration is semantically equivalent to the meaning of the LU L’ that exists in **L**, so that

$$‘L_1’ \oplus ‘L_2’ \oplus \dots \oplus ‘L_n’ = ‘L’ ,$$

then ‘L’’, and not the above configuration of meanings, must appear in the definition.

The semanteme ‘L’ is the maximal block with respect to the semanteme configuration ‘L₁’ ⊕ ‘L₂’ ⊕ ... ⊕ ‘L_n’.

In other words, Rule 4 (first formulated by Ju. Apresjan; see Apresjan 1969a: 14, 1969b: 421, 1974: 95) requires obligatory semantic reduction; this means that the semantic decomposition of the meaning under description must be minimal, or the shallowest possible. Such an approach ensures gradual decomposition (into ‘semantic immediate constituents’) and thus makes lexicographic definitions more manageable and surveyable; otherwise they would be too long and cumbersome.

In contrast to the three preceding rules, Rule 4 is not logical—in the sense that a good definition which follows this rule is absolutely equivalent to a good one which does not. However, it is very important methodologically: it guarantees that every semantic decomposition is the shallowest possible and thus allows the linguist to avoid arbitrary decisions as to where to stop decomposing in a lexicographic definition. Logically, a definition can be either the shallowest or the deepest possible. But the deepest decomposition will of necessity be constructed in terms of semantic primitives; and writing the definitions only in terms of semantic primitives suffers from at least three shortcomings:

- A linguist writing the definitions for LUs of **L** directly in terms of semantic primitives must first have at his disposal a well-established set of semantic primitives of **L**, and such a set is not yet available (Wierzbicka herself has modified her starting hypotheses several times and is still developing the inventory of semantic primitives, which has grown in about 30 years from 13 to over 60). Before constructing a dictionary, the linguist is forced to accomplish a preliminary enormous task: to establish the set of semantic primitives for **L**.

- A definition written only in terms of semantic primitives is very long and extremely complex, which makes it unwieldy: not only the dictionary user, but the lexicographer too will find it difficult to work with it. Worse still, the speaker’s intuition balks at the evaluation of such definitions. For instance, certain definitions in Wierzbicka 1985 reach two printed pages—even with the use of many intermediate semantic components, that is, components not reduced to primitives. Thus,

the definition of BIRDS takes up two pages: 180-181, as does that of LEMONS: 310-311, and many others.

Wierzbicka herself proposes the use of intermediate semantic components. The idea of the maximal block is in harmony with this proposal, adding only the requirement that the intermediate components be used obligatorily wherever this is possible.

- With the deepest semantic decomposition possible, the semantic links between LUs are not directly visible in their definitions. Thus, in the definition of PROFESSOR one will not find ‘teach’, since ‘teach’ will be replaced by a configuration of semantic primitives.

It is to avoid these three shortcomings that the notion of the maximal block is introduced.

2.1.2.5 Rule 5: Mutual Substitutability Rule

What do we want from a lexicographic definition of L in an ECD from the viewpoint of its relation to L itself? Clearly, it should reflect the linguistic intuition that native speakers have as to the meaning of L as closely as possible; however, this informal requirement is difficult to check. More formally, the definition of L should satisfy the following general condition:

An ECD definition of an LU L should guarantee absolute mutual substitutability with L in text: L must be replaceable by its definition and the definition of L must be replaceable by L in any imaginable context (with the exclusion of meta-linguistic ones)—*salva significatione* (i.e., stylistic elegance or even normal lexical cooccurrence may be violated).

Let me emphasize that the substitutability of L and its definition required here is not substitutability *salva veritate*, i.e., with the preservation of the truth value—expected in many philosophical approaches to synonymy and hence to the theory of defining—but the substitutability *salva significatione*, i.e., with the preservation of the same meaning. The notion of the ‘same meaning’ is basic in the Meaning-Text semantics.

Mutual substitutability of the definition (= definiens) and the unit defined (= definiendum) with preservation of meaning, as well as the greater semantic simplicity of the defining elements in the definiens with respect to the unit defined,

are major requirements in Wierzbicka's semantic approach, repeatedly stated and defended in her publications since 1960s. We faithfully follow these requirements here.

Mutual substitutability includes, of course, substitutability within the definitions themselves. Thus, consider the following equivalences:

(3) a. *X reveals Y to Z* ≡ 'information «Y» being hitherto hidden¹ from Z by someone and X believing that Z would like to know «Y»,|| X directly¹ causes² that «Y» becomes¹ known³ by Z'

(cf. Wierzbicka 1987: 308-309)

b. *Y becomes W* ≡ 'Y begins¹ to be² W'

and

c. *Y is [= BE²3] known by Z* ≡ 'information «Y» is² in Z's mind¹'

By substituting 'become' and 'known' in (3)a) by their definitions (= semantic decompositions) given in (3)b) and (3)c), we obtain ((3)d):

d. *X reveals Y to Z* ≡ 'information «Y» being hitherto hidden¹ from Z,|| X directly¹ causes² that «Y» begins¹ to be² in Z's mind¹'.

Indeed, *John revealed to all his colleagues that he had traveled to Coruña* does mean 'John directly caused [by saying, writing, or showing something] that the information «John had traveled to Coruña», which was hitherto hidden from all John's colleagues, began to be in the mind of all his colleagues'.

Absolute mutual substitutability of the definiendum and the definiens *salva significatione* is the central methodological requirement of the MTT approach to theoretical semantics and the lexicon; other features of the ECD follow from a strict observance of this requirement. Without substitutability, we cannot claim that the meaning of the definiendum is equal to that of the definiens, and thus the concept of definition itself collapses: it loses all positive content. A scientific approach to lexicographic definitions is impossible if we do not advance, following Anna Wierzbicka's effort, the requirement of absolute substitutability of the definiens and the definiendum.

NB: The inclusion of mutual substitutability of the definiendum and the definiens among the rules for the formal correctness of definitions can be questioned: is it not rather a substantial requirement? I believe it is formal, in the sense that it does not involve any specific property of natural language. On the other hand, where this rule is classified is not that important—provided it is observed.

2.1.3 Criteria for an ECD Definition: Criteria of Type I

The rules for ECD definitions ensure their formal correctness, or well-formedness, as imposed by Meaning-Text Theory. An ill-formed lexicographic definition cannot be properly justified, criticized, or improved. If it is well-formed, it satisfies the condition of being usable, but not, as yet, the condition of sufficiency; “to be sufficient” it must also be:

- explicitly linked with all semantically related definitions in the dictionary, and
- factually true—i.e., it must correspond to the facts of **L**.

The goal of any dictionary is of course to have true definitions; formally correct but factually false definitions are good for nothing.

Therefore, along with rules for formal correctness of lexicographic definitions, we need substantive criteria that target the relation linking the definition under consideration 1) to other semantically related definitions and 2) to actual semantic facts of **L**—more specifically, to the meaning of the LU **L** under description. These criteria help the linguist select the semantic components to be included in, or excluded from, the definition of **L**. I will refer to them as Type I Criteria, or Criteria I, since later I will introduce another type of lexicographic criteria, Criteria II—for distinguishing the LUs within a vocable (see Subsection 3.2, p. 00).

Let me start with the criterion that is necessary for ensuring the internal semantic coherence of the dictionary—namely, the explicit links between the definition of **L** and semantically related definitions. Consider an LU **L** that denotes a physical phenomenon/object/substance **P**. What properties of this **P**, which is the referent of **L**, must be reflected in the definition of **L**? Some of these properties are constitutive—if they are not included in the definition of **L**, **L** becomes applicable to other referents which are not **P**s. Thus, ‘being solid’ is a constitutive property of ice, as ‘being invisible’ is a constitutive property of air: something liquid cannot be called ICE in English, and a visible gas is not AIR (even if some fantastic beings in a science fiction novel use it for breathing). Such properties control the correct usage of the corresponding LUs and have to be represented in their lexicographic definitions. The difficulty appears when the property in question is not constitutive, but still quite typical of **P**. Take, for instance, the white color of snow, sugar, salt

and rice. Should we put the component ‘white’ in the definitions of the English nouns SNOW, SUGAR, SALT and RICE? (This question—with respect to SALT—was discussed above: see 2.1.1, p. 00.) Criterion I.1 helps to find the answer.

2.1.3.1 CrIterIon I.1: Linguistic relevance of a semantic component

Let there be ‘ σ ’, a semanteme or a configuration of semantemes that is a candidate for the inclusion in the definition of L; ‘ σ ’ reflects a non-constitutive, but typical property of the referent of L, so that one feels tempted to have ‘ σ ’ as a component in the definition of L.

The definition of L must include the component ‘ σ ’ if and only if language **L** has at least one other LU L’ that is formally linked to L and has ‘ σ ’ in its meaning.

The existence of the LU L’ with the indicated characteristics demonstrates the linguistic relevance of ‘ σ ’ in the definition of L.

L’ can be formally linked to L in one of the following three ways, involving three important linguistic phenomena:

- L’ is another LU of the same vocable to which L belongs: polysemy.
- L’ is an LU derived from L: derivation.
- L’ is a phraseme—more precisely, an idiom—that contains L: phraseology.

In these three cases, the inclusion of ‘ σ ’ in the definition of L ensures an explicit specification of the semantic link perceived by speakers between L and L’ (formally speaking, ‘ σ ’ represents a semantic bridge between L and L’).

Examples

Polysemy

(4) L = CLOUD_I ‘accumulation₂ of white_{1a} or grayish₁ substance₁ ... (that partially₁ blocks₂ the sun_{2a})’;

‘ σ ’ = ‘... that partially₁ blocks₂ the sun_{2a}’;

L’ = CLOUD_{III} [*on* N_Y] ‘fact X ... that (partially₁) spoils₂ the positive₁ character₂ of the fact Y [*_as if_1* X were a cloud_I that partially₁ blocks₂ the sun_{2a}]’ [*as in This sad news was the only cloud on this otherwise excellent vacation or Chandra casts a cloud on the anti-matter theory*].

The existence of CLOUD_{III} shows the linguistic relevance of the component ‘σ’ in the definition of CLOUD_I. The semantic link between CLOUD_{III} and CLOUD_I is immediately obvious to an English speaker—it is a comparison with CLOUD_I [a live, even if conventional metaphor]; it has to be shown in the definitions of both lexemes. On the other hand, according to the concept of vocable (see Def. 3 below, p. 00), two lexemes of the same vocable should explicitly manifest their semantic bridge. As a result, we have to include the component ‘σ’ in the definition of CLOUD_I, which allows us to have the component ‘[_as if_1 X were a cloud_I that (partially₁) hides₁ the sky ...]’ in the definition of CLOUD_{III}, and the semantic link—a semantic bridge—is ensured.

Derivation

- (5) L = STORM_I ‘meteorological phenomenon₁—violent₂ winds₁ and rain₂ ...’;
 ‘σ’ = ‘violent₂’;
 L’ = STORMY_{II} ‘accompanied₂ by violent₂ expression₁ of feelings₁ [_as if_1 it were a storm_I]’ (*stormy quarrel* ⟨*stormy tears*⟩).

The component ‘violent₂’ is necessary in the definition of STORM_I to ensure a semantic bridge with STORMY_{II}; it indicates the basis of an obvious comparison.

Phraseology

- (6) L = SNOW_I ‘white_{1a} cold₁ substance₁...’;
 ‘σ’ = ‘white_{1a}’;
 L’ = [WHITE] _AS SNOW_, SNOW-[WHITE]

Let it be added that the presence of the component ‘white_{1a}’ in the lexicographic definition of SNOW_I is confirmed by the polysemy and the derivation tests:

- English has SNOW_{II} ‘cocaine in powder form—WHITE as snow_I’.
- English has SNOWY_{II} ‘pure WHITE’ (*snowy hair*).

NB: 1. Now I can answer the question asked above: The definitions of SUGAR, SALT and RICE do not mention ‘white color’, even if these substances are factually white, because English has no expression in which these lexemes are involved to express whiteness: **white as sugar*, **sugar-white*, **salty white*, **rice whiteness*, ... But, for instance, Russian says *saxarnye zubki* lit. ‘sugary nice.little.teeth’ = ‘very white nice little teeth [of a child or a young woman]’, so that the definition of the Russian lexeme SAXAR ‘sugar’ must include the component ‘white’. Thus, our approach is strictly lexicological, not at all encyclopedic.

2. The semantic bridge between L and L’ can be a connotation of L rather than a component in the definition of L. Thus, STONE_N has a connotation ‘cruel indifference’, which is justified by the existence of the adjective STONY, as in a *stony heart* ⟨*stony stare*⟩.

Thus, Criterion I.1 (= Criterion of Linguistic Relevance) allows the linguist to make a linguistically justified decision in a case where the necessity of a semantic component in a lexicographic definition is not straightforward.

Criterion I.1 is aimed at enhancing the coherence of our lexicographic description—making explicit all semantic links between related LUs. The next criterion, I.2, or rather a group of three criteria—I.2a, I.2b and I.2c, targets the factual truth of the definition. It addresses the combinatorial possibilities of L—L’s cooccurrence with qualifying modifiers, quantifying modifiers and negation.

2.1.3.2 Criterion I.2a: Cooccurrence with qualifying modifiers

The definition of L must explicitly reflect L’s cooccurrence with qualifying modifiers—that is, it must include a semantic component ‘ σ ’ capable of ‘accepting’ the meaning of the modifier in question—technically, of being the argument of the corresponding predicate.

Examples

Adjectival modifier

(7) APPLAUSE: this noun readily accepts adjectival modifiers of the type Magn/AntiMagn: *deafening* ⟨*frenetic, frenzied, thunderous*⟩ or *scattered* ⟨*subdued, thin*⟩, which express intensification/attenuation; therefore, the definition of APPLAUSE must include a semantic component ‘ σ ’ that admits this kind of qualification. Here is a tentative definition (‘ σ ’ is shown in small caps):

X’s applause to Y for Z \equiv ‘Repeated¹ clapping¹ by X as a sign⁴ of approval¹ by X of Y’s Z, THE FORCE¹ AND/OR RATE² of the clapping¹ being² proportional² to the degree¹ of the approval¹’.

Adverbial modifier

(8) Fr. BATTREII ‘beat, defeat’ (as in *Jean a battu Pierre au tennis* ‘J. beat P. at tennis’): the verb *battreII* can take adverbial intensifiers of the type Magn, such as *à plate couture* lit. ‘to flat seam’ \approx ‘soundly’ and *complètement* ‘completely’; consequently, its definition must include an intensifiable component. An ECD-style definition of BATTREII cannot use as the central component ‘avoir le dessus’ = ‘have the upper hand’, defining BATTREII as ‘avoir le dessus sur ...’ = ‘have the upper hand over ...’ (as does, e.g., PR 2001), because AVOIR LE DESSUS is an idiom—i.e., one LU, so that it

does not represent the decomposition of the meaning ‘battre Π ’. (Moreover, the meaning ‘avoir le dessus’ is not easily intensifiable: *avoir le dessus ?complètement* ⟨*?totalement*⟩ ‘have completely ⟨totally⟩ upper hand’.) Here is what can be proposed as a better definition:

X bat Π Y dans Z pour W \equiv ‘X and Y being²₃ opposed₁ in struggle₁ Z over W, || X causes₂ such damage¹₁ to Y that Y is²₃ unable to continue₁ Z, _as a result_ of which Y does not obtain₁ W’.¹⁶

The choice of the above examples is not fortuitous: the modifiers shown are restricted lexical cooccurents of L—elements of the values of the Lexical Function Magn. Such cooccurents, known as *collocates*, have intimate semantic links with components in the definition of L and facilitate the linguist’s job in establishing the semantic content of ‘L’. Systematically accounting for the correspondence between the definition and the restricted lexical cooccurrence of L (= L’s LFs) is one of guiding principles in the developing and presenting ECD entries.

2.1.3.3 Criterion I.2b: Cooccurrence with quantifiers

|| The definition of L must explicitly reflect L’s cooccurrence with quantifiers—
|| especially with plural markers and numerals.

Example

Consider four semantically related French nouns—roughly speaking, names of edible plants:

AIL ‘garlic’, OIGNON ‘onion’, CAROTTE ‘carrot’, CHOU ‘cabbage’

Pluralization shows the first division between them: one can say *Apporte-moi des oignons* ‘Bring me [a few] onions’/*des carottes* ‘[a few] carrots’/*des choux* ‘[a few] cabbages’, but no **Apporte-moi des ails* ⟨*des aulx*⟩ ‘Bring me [a few] garlics’. (AIL is pluralizable, but in a completely different sense: *Les ails* ⟨*Les aulx*⟩ *du Mexique sont très diversifiés* lit. ‘Mexican garlics are highly diversified’; the plural with AIL can mean only ‘different sorts of ...’ rather than ‘several units of ...’.) Moreover, unlike the three other nouns, AIL does not combine with numerals:

(9) *Apporte-moi un* ⟨*trois*⟩ *oignon*⟨*s*⟩ *une* ⟨*trois*⟩ *carotte*⟨*s*⟩/*un* ⟨*trois*⟩ *chou*⟨*x*⟩ !
but
**Apporte-moi un ail* ⟨*trois ails/aulx*⟩ !

The only way to say this is to use the ‘counter’ TÊTE ‘head’: *Achète une tête*

d'ail/trois têtes d'ail 'Buy a head/three heads of garlic'.

Example (9) shows that the four nouns cannot be defined in the same way. AIL is 'assaisonnement₁ qui ...' = 'seasoning that ...', that is, AIL is a substance, not a unit; ASSAISONNEMENT₁ 'seasoning' admits pluralization only with the meaning 'different sorts of ...' and cannot be quantified by numerals, so that as the central component in the definition of AIL it will ensure the desired morphosyntactic behavior. (AIL has another sense: 'domestic plant whose bulbs produce ail₁', see below.) OIGNON, CAROTTE and CHOU in (9) can be tentatively defined as 'unité de légume qui ...' = 'unit₁ of vegetable¹₁ that ...'.

However, OIGNON, CAROTTE and CHOU, in spite of their obvious semantic relatedness, show differences in quantified contexts:

- (10) a. *manger de l'oignon* <des oignons> / <*de la carotte> *des carottes* / *du chou*
 <des choux> 'eat onions/carrots/cabbage'
 b. *aimer bien l'oignon* <²les oignons> / <²la carotte> *les carottes* / *le chou*
 <²les choux> '[to] like onions/carrots/cabbage'
 c. *L'oignon* <Les oignons>/*La carotte* <Les carottes>/*Le chou* <Les choux>
 /*L'ail pousse(nt) bien dans cette région*
 'Onions/carrots/cabbages/garlic grow(s) well in this region'.

The above differences force us to isolate, for the names of vegetables, the following three types of lexicographic senses that correspond to examples (9) and (10):

1. 'Unit₁ of vegetable¹₁ U that ... [size, form, color, consistency, taste, ...]' [= (9)]
2. 'Edible substance₁ of U ...' [= (10)a-b); this sense is also valid for AIL, since 'spice₁' = 'edible substance₁ that ...']
3. '(Class₃ of) plant(s)²₁ that produce₂ U ...' [= (10)c)]

These definitional schemata should be systematically applied to all vegetable names, in order to distinguish different lexemes in a 'vegetable' vocable. For each lexeme, the possibility of pluralization has to be explicitly indicated: thus, for Type 2 senses ('edible substance₁ of U ...'): OIGNON has both numbers, while CAROTTE has only the plural and is thus—in this sense—a *plurale tantum*; CHOU, on the contrary, is in this sense rather a *singulare tantum*.

Criterion I.2b not only facilitates the differentiation of the lexemes of a vocable, but also helps choose the central [= generic] component of a lexeme's definition. Thus, in a Type 1 sense ('unit₁ of vegetable¹₁ U that ...'), the generic

component ‘unit₁’ allows for the use of the article UN ‘a’ with the corresponding lexeme, the pluralization of it, and its cooccurrence with numerals (*une unité, des unités, trois unités*).

2.1.3.4 Criterion I.2c: Cooccurrence with negation

|| The definition of L must explicitly reflect the way L combines with negation.

In some cases the meaning of the expression *not L* is not a simple negation of ‘L’; then, a close analysis of the meaning of *not L* can throw interesting light on the contents and the organization of the definition of L.

Example 1

As a first approximation, the definition of the noun WIDOW₁ (characterization of the family status of a woman) can be formulated like this (adapted from LDoCE 1978):

(11) a. *X is a widow* ≡ ‘X is a woman₁ who₂ has lost₆ her husband₁ and₂ has not remarried₁’.

NB: In *John’s widow later married my brother* we have a different LU—WIDOW₂: *X is the widow of Y*.

With the definition in (11)a), the sentence *Zhu is not a widow* would mean ‘Zhu is not a woman₁ ...’. But in reality this sentence negates the facts that 1) Zhu has lost her husband and 2) Zhu has not remarried, but affirms that Zhu is an (adult) woman: according to the meaning of (11)a), Zhu cannot be a male or a little girl. To reflect this property of the definition of WIDOW₁, the component ‘woman’ must have a status different from that of the component ‘has lost₆ her husband₁ and has not remarried₁’: only the latter can be negated. More specifically, ‘woman’ is a semantic taxonomic restriction on a Semantic Actant (its semantic type), which functions as a presupposition, while ‘has lost₆ her husband₁ and has not remarried₁’ constitutes the assertion. One of the ways to show the presupposed character of a semantic component in a verbal lexicographic definition is to put its expression into a modifier position:

b. *X is a widow* ≡ ‘X, who is a woman₁, has lost₆ her husband₁ and has not remarried₁’.

The presuppositions can also be indicated in an equivalent way: by the symbol ‘||’, put after presuppositions and thus separating them from the assertion part of the definition:

‘X being a woman₁,|| X has lost₆ her husband₁ and has not remarried₁’.

In what follows, the latter notation is used.

In the reformulated form of (11)b), the definition ensures the correct description of the combination with negation:

c. *X is not a widow* = ‘X being a woman₁,|| X has not lost₆ her husband₁ or has remarried₁’.¹⁷

Example 2

The verb [*to*] PERMIT₁ (≈ ‘say yes’¹⁸) represents a more complex case.

(12) *Leo [= X] did not permit me [= Z] to go [= Y] to France.*

Firstly, *X permits* only those *Z* to do *Y* who WANT to do *Y*, and *X* must know about this wish; we see that sentence (12), in spite of negation, affirms that *I* wanted to go to France and that *Leo* knew this. Therefore, ‘*X* knows that *Z* wants to *Y*’ is a presupposition in the meaning of [*to*] PERMIT₁.

Secondly, *X permits Z* to do *Y* only if *Z* IS SUPPOSED NOT TO DO *Y* AGAINST *X*’S WILL; and the relation between *X* and *Z* that reflects this is affirmed in (12): again in spite of the negation, the sentence affirms that *Leo*’s will is important for me in this respect. Therefore, a possible definition can read as follows:

X permits Y to Z = ‘Knowing¹₁ that *Z* wants¹₁ to do²₁ *Y*, which *Z* is²₃ not supposed¹_{3a} to do²₁ against *X*’s will²₃,|| *X* communicates₁ to *Z* that, _according to _1 some reasons₂, *Z*’s *Y* is²₃ not against *X*’s will²₃’ (cf. Wierzbicka 1987: 108-111)

But this is not all as yet. While *X did not read/eat/sleep/go* etc. are simple negations of *X read/ate/slept/went*, the expression *X did not permit Y to Z ...*, according to the above definition, is not a negation of *X permitted Y to Z*:

X permitted Y to Z ≈ ‘*X* communicated₁ to *Z* that *Z*’s *Y* is²₃ not against *X*’s will²₃’

vs.

X did not permit Y to Z ≠ ‘*X* did not communicate₁ to *Z* that *Z*’s *Y* is²₃ not against *X*’s will²₃’

In point of fact,

X did not permit Y to Z ≈ ‘*X* communicated₁ to *Z* that *Z*’s *Y* is²₃ against *X*’s will²₃’, so that the negation that syntactically attaches to the verb bears semantically not on the central (= generic) component of its meaning, but on an embedded component ‘be²₃ not against *X*’s will²₃’, giving ‘not be²₃ not against’ = ‘be against’. And a

negation bearing on an embedded component of L's meaning rather than on the central one represents a case of antonymy! We have to conclude that *do not permit* is not a 'normal' grammatical negation, but an antonym of [*to*] PERMIT₁—a separate LU of English. This has to be stated explicitly in the lexical entry for PERMIT₁:

PERMIT₁_V

...
 Anti_['not contrary'] : *do not permit*

[Anti—'antonym'—is another Lexical Function, see 2.3 below.]

Hopefully, the above discussion is sufficient to give the reader an idea of the importance and the complexity of the definition in an ECD dictionary article. We can now sum up the discussion and then move on—to other parts of an ECD lexical entry.

2.1.4 The ECD Definition: General Characteristics

To round out the discussion of the definition in the ECD, two points must be addressed: 1) an ECD-style definition *vs.* a SemR of L; 2) the internal structure of ECD definitions.

ECD-style verbal definitions *vs.* SemRs of LUs

As mentioned above, the definition of an LU L in an ECD must in principle be L's SemR, i.e., a semantic network with an indication of communicative sub-areas (indication of the division of the meaning into sub-networks marked 'Rheme ~ Theme', 'Given ~ New', 'Focalized ~ Non-Focalized', 'Presupposed ~ Asserted', etc.); however, actual ECD definitions are sentences in a natural language—they are verbal and linear. Thus, in an English ECD definitions are written in a 'processed' English which has undergone many amputations and some additions, is subject to special constraints on its syntax, and can violate the standard norms of cooccurrence. Nevertheless, it is still a natural language. There are two reasons for the use of verbal formulations instead of semantic networks.

- The first reason is rather practical: greater convenience and ease from the viewpoint of typography and human users. Thus, in spite of proclaimed freedom from pedagogical and commercial considerations, the ECD has to make compromises...

- The second reason is more profound: a formulation in an English-based lexicographic semantic metalanguage is much more readily accessible to the

linguistic intuition of speakers, including the lexicographer himself. (A. Polguère drew my attention to this point; cf. Polguère 1992: 134-135.) A description of a lexical meaning in the form of a network, i.e., a ‘genuine’ SemR, is more explicit and precise. It is well-suited for logical analysis, for ensuring consistency and for all similar formal, computer-like manipulations. Yet in order to check the acceptability of a proposed substitution of a definition for an LU, a speaker has to use the full strength of his linguistic intuition, and for this he needs a linear, language-like text: linguistic intuition balks at technical formalisms. Because of this, ECD-type verbal definitions have a very important role to play.

It goes without saying that the verbal ECD-style definition of L and its corresponding SemR must be equivalent and in one-to-one correspondence. It would be ideal to have in the ECD both types of representations for the meaning of an LU: a canonical network SemR and a linear verbal definition, plus an algorithm for transforming one into another. However, for the time being, this is still a dream; thus, in the ECD, we stick to linear definitions.

Internal structure of ECD definitions

An ECD definition must be structured. This structuring involves, in the first place, at least the following three facets of a definition: 1) the communicative and/or logical status of its components, 2) the different structural roles these play, and 3) the inheritance of Semantic Actants. Let me consider each in turn.

Communicative and/or logical status of a component in an ECD definition

An ECD-type definition must reflect the different communicative statuses of its components. Thus, it explicitly indicates the presuppositions (to the left of the symbol ||). A presupposition in a meaning remains affirmed under negation of the whole meaning: *Jack **does not** help Mary to finish her studies* still implies that Mary is finishing (or at least is trying to finish) her studies, although Jack does not add his resources to Mary’s efforts. A presupposition remains unaffected by interrogation as well: in the question *Is Jack helping Mary to finish her studies?* the proposition ‘Mary finishes (or tries to finish) her studies’ is not questioned but rather affirmed. An ECD-type definition also indicates the *communicatively dominant node* of the meaning represented and may indicate its division into Rheme (= Comment) vs. Theme (= Topic), etc., not shown in our examples.

NB: On the distinction of different logico-communicative layers in ECD definitions, see, in particular, Apresjan 1980: 49ff. E. Padučeva proposes ‘formatted definitions’ as the main tool in lexicographic work: see, for instance, Padučeva 2002 and 2004: 525ff, where four types of lexicographic parameters for organizing the definitions are put forward: taxonomic category of L (‘action’ ~ ‘state’ ~ ...), semantic field (‘speech’ ~ ‘mental act’ ~ ...), set of semantic roles (‘Agent’ ~ ‘Direction’ ~ ‘Obstacle’ ~...), and taxonomic class of the actant (‘liquid’ ~ ‘person’ ~ ...). For a general discussion of the problem, see Lordanskaja & Mel’čuk 1990 and Apresjan 1988b.

In a different vein, some semantic components are, so to speak, default components: a default component is present in the given meaning, if nothing in the discourse contradicts it, but it can be easily suppressed by a contradicting semantic element in the context—without giving rise to a contradiction. Such components are called weak; they are shown in a definition by parentheses. Thus, the meaning of the Russian verb *OPOZDAT* ‘be late’, as in *Ja opozdal na poezd* ‘I was late for the train’, includes the component ‘in spite of X’s intentions’: that is how the above Russian sentence is to be understood. However, one can say *Ja naročno opozdal na poezd* ‘I intentionally was late for the train’, where the meaning of *naročno* ‘intentionally’ neutralizes or suppresses this component (Anna Zaliznjak 1987: 138ff).

Different structural roles played by the components in an ECD definition

Each semantic component ‘ σ ’ within a definition of LU L plays one of the three major roles with respect to the organization of the definition. Roughly speaking, a component ‘ σ ’ can:

- 1) specify a fact about one or several Semantic Actants [= SemAs] of L—a property or a state of an actant, a relation between two actants of L, an event in which actants are involved, etc.;
- 2) constitute a semantic taxonomic restriction on an actant of L [‘ σ ’ is in a sense analogous to selectional restrictions in transformational grammar];
- 3) modify another semantic component, restraining its content.

For instance, in *BAKEII.1a* (‘X causes₂ that Y, which is raw¹₂ bricks₁ or pottery₂, hardens₁ by exposing Y to the action₆ of dry¹⁰ heat²_{2a} in device₁ Z’; see Subsection 5.1, p. 00), the component ‘... causes₂ that ... hardens₁ ...’ expresses a complex relation between SemAs X and Y, while the component ‘raw¹₂ bricks₁ or pottery₂’ characterizes SemA Y itself taxonomically; this characterization is necessary to block the use of the lexeme *BAKEII.1a* to name, for instance, the process of hardening a metallic alloy by heating. Note that the taxonomic

characterization of a SemA can be overt, as shown above, or covert (i.e., implicit in the decomposition of a relational component). Thus, in **BAKEII.1b** ('Y hardens₁ being baked_{II.1a} [by X] in Z'), SemA Y has no overt semantic restrictions, but since **BAKEII.1b** is defined by reference to **BAKEII.1a**, in which Y is overtly restricted to raw bricks and pottery, this restriction automatically carries over to X in **BAKEII.1b**.

Inheritance of Sem-Actants

A component 'σ' of the definition of L brings to its host all of its own SemAs, which must be explicitly accounted for in the definition. For instance, **BAKEI.2a** ('X creates₁ solid₁ food_{1b} Y from a mixture₁ ...') includes 'create₁'; the lexeme **CREATE₁** has three SemAs: who creates what from what; as a result, **BAKEI.2a** has all of these three SemAs (which are represented by variables Z, X and W). Some of the inherited SemAs may of course not be realized: they are blocked —i.e., cease to be variables. Thus, the meaning of [*to*] **COST** (as *This book cost him \$30*) presupposes a 'sell/buy' transaction, with four SemAs: *X sells Y to Z for W*. However, the seller X is not expressible with **COST**: *This book [= Y] cost him [= Z] \$30 [= W] *with John/*from John*;¹⁹ therefore, in the definition of [*to*] **COST** the corresponding participant of the transaction is represented by a generic constant, shown in small capitals:

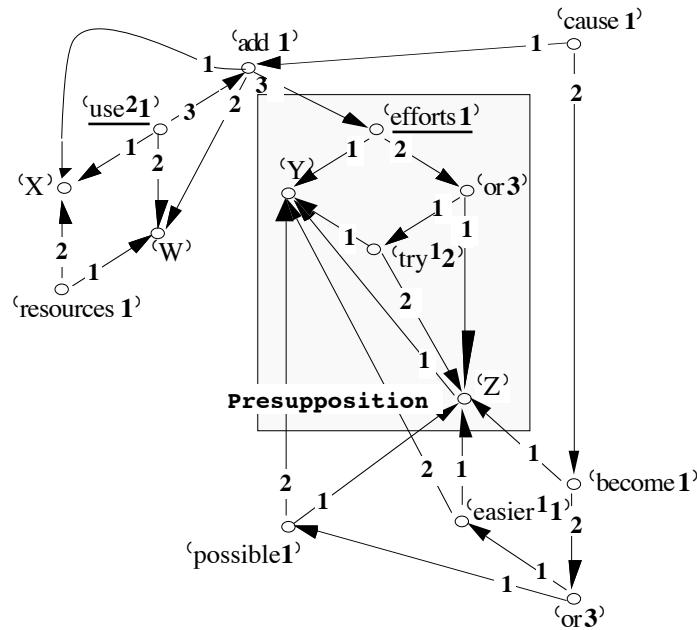
Y costs Z W ≡ 'merchandise or service₁₁ Y is paid₁ for the sum₂ W by person₁ Z
TO THE PERSON₁ from whom Z is buying₁ Y'

Now, in order to close Subsection **2.1** with a good illustration, here is yet another example of an ECD-style definition: the English verbal lexeme **HELP₁** [as in *Jack helped Mary to finish her studies with a generous gift of money*].

HELP₁

X helps Y to Z with W ≡ 'Y making efforts₁ trying₁² to do₁² or₃ doing₁² Z, ||X uses₁² X's resources₁ W by adding₁ W to Y's efforts₁ such that W causes₁ that doing₁² Z becomes₁ possible₁¹ or₃ easier₁¹ for Y'.

The corresponding Sem-Representation appears as follows:



Explanations

1. The underscoring of a semantic component (in this case, 'use²1' and 'effort¹') indicates that it constitutes the communicatively dominant node of the corresponding communicative area of the Semantic Structure.
2. As stated on p. 00, the semantic component 'cause₁' represents non-agentive causation: *The falling tree* ⟨*The bullet*⟩ **killed** *the dog*; 'cause₂' stands for the agentive causation: *John* **killed** *the dog*.
3. For simplicity's sake, the SemA 3 of 'easier' (i.e., '... than it would be otherwise') is not shown.

2.2 The Government Pattern in the ECD

The zone of syntactic cooccurrence of the headword L is subdivided in two parts: the description of passive syntactic valence of L and the description of active syntactic valence of L.

L's passive syntactic valence is L's capacity to depend syntactically on LUs of particular types; in other words, it is the set of all LUs that can, in an appropriate context, subordinate L. This set is specified in the ECD entry for L by giving L's part of speech and all its syntactic features. Since this information is very complicated, but less specific to an ECD—it is essentially anchored in L's syntax—I will not discuss it here.

By contrast, L's active syntactic valence is intimately related to the definition of L and thus constitutes an important particularity of the ECD. L's active syntactic valence is the set of particular types of LUs whose presence on some level of representation of the sentence containing L is required by L's semantic nature—i.e., by the definition of L. These LUs are L's Semantic Actants [= SemAs(L)]; the expression of L's SemAs in the text is described by L's Government Pattern.

L's Government Pattern [= GP(L)] specifies, for each of L's SemAs X, Y, Z, ..., the corresponding Deep-Syntactic Actant [= DSyntA] I, II, III, ... as follows: X \Leftrightarrow I, Y \Leftrightarrow III, etc.; for a given pair <SemA, DSyntA> all surface-syntactic and/or morphological means for expressing the DSyntA in the text are indicated. Formally, a GP is a rectangular matrix having *n* columns (designated C), numbered with Roman numerals: C_I, C_{II}, ..., with one column for each SemA, and *m* rows, numbered with Arabic numerals: I.1, I.2, ..., with one row for each syntactic-morphological means of the Deep-Synt-actant surface expression. Thus, C_{III.3} means 'Column III, row 3' and specifies the expression *with* N for the SemA Z in the GP table on next page. (For more on SemAs and SyntAs, see Mel'čuk 2004a, b.)

By convention, the obligatory character of the first DSyntA is not marked in the GP of a verb, since the DSyntA I corresponds to the subject of the finite verb, which is always obligatory in English; all other DSyntAs are taken to be optional unless otherwise indicated (with the mark '**obligatory**' in the corresponding column, see lexical entries for **BAKE** in 5.1 below).

The GP table is accompanied by numbered constraints, which specify the cooccurrence of L's different DSyntAs, the cooccurrence of surface means for expressing the DSyntAs, semantic and syntactic conditions of their use, etc. After these constraints some basic examples of possible/impossible combinations of L's actantial dependents are given.

For example, the verb **HELP1** (as in *John will help you to clean up the house*) has the following GP:

HELP1

Government Pattern

X ⇔ I	Y ⇔ II	Z ⇔ III	W ⇔ IV
1. N	1. N	1. V _{inf} 2. <i>to</i> V _{inf} 3. <i>with</i> N 4. <i>with</i> V _{ger} 5. <i>in</i> V _{ger} 6. PREP _{dir} N	1. <i>with</i> N 2. <i>by</i> V _{ger}

- 1) C_{III.1} : ‘X being directly involved in Z’ [= ‘X doing Z himself’]²⁰
 2) C_{III.2} : ‘X not being directly involved in Z’ [= ‘X not doing Z himself, but providing some resources to Y’]²¹
 3) C_{III.6} : **if** Z = ‘travel/move [something] in the direction α’,
then [III = L(‘α’) and C_{III} = C_{III.6}] **is possible**

[PREP_{dir} stands for ‘directional prepositions and adverbs’, such as *up*, *out*, *into*, *across*, *there*, ...; L(‘α’) stands for ‘L expressing the meaning ‘α’’. Constraint 3 means that, for instance, instead of *help John to climb up the stairs*, one can say *help John up the stairs*.]

- 4) C_{III.3,4} + C_{IV.1} : **undesirable**

Kathleen helped the old gentleman (to) finish his preparations ⟨with his preparations/with preparing his luggage⟩. *With her advice, Kathleen helped me in assigning the θ-roles to all arguments. Kathleen helped the boy (to) finish his studies with her generous financial assistance. She helped Jack out of his coat* ⟨up the stairs⟩ *with a hard kick in the bottom* ⟨by kicking him hard in the bottom⟩.

Undesirable: ?*Kathleen helped Arthur with his work with her advice* [by Constraint 4; correct expression: either ... *in his work with her advice* or ... *with his work by advising him*]

In the ECD the GP plays the same role as the subcategorization frame in all descendants of transformational generative grammar.

An LU may have two or more GPs, in which the same SemAs correspond to different DSyntAs. Here is an example (adapted from McCawley 1992/1993: 122-123):

PUNISH, verb

X punishes Y for Y's Z by W-ing Y = ...

Government Pattern 1

X ⇔ I	Y ⇔ II	Z ⇔ III	W ⇔ IV
1. N	1. N	1. <i>for</i> N 2. <i>for</i> V _{ger}	1. <i>with</i> N 2. <i>by</i> V _{ger}
	2. S ₁ (Z) obligatory		

The courts punish people [Y_{II.1}] for robbery/drug dealing [Z_{III.2}] with heavy prison terms [W_{IV.1}].

The courts punish robbers/drug-dealers [Y_{II.2}] by imprisoning [W_{IV.2}] them for long terms.

Government Pattern 2

X ⇔ I	Z ⇔ II	W ⇔ III
1. N	1. N 2. V _{ger} obligatory	1. <i>with</i> N 2. <i>by</i> V _{ger}

The courts punish robbery/drug dealing [Z_{II.2}] with heavy prison terms [W_{III.1}].

The verb PUNISH has the same meaning in both types of structure, i.e., it constitutes one LU that shows two different syntactic patterns in which it can be used. This peculiarity is captured by the use of two GPs in one lexical entry. The correlation between them is shown by means of the LF Conv; in this case, GP2 = Conv₁₃₄(GP1).

The same technique is used in the lexical entry for BAKEI.2a (Subsection 5.1, p. 00) in order to associate a single underlying SemR with a variety of surface syntactic realizations.

2.3 Lexical Functions in the ECD

The Lexical Cooccurrence Zone of an ECD lexical entry includes what is perhaps the best-known feature of the ECD: Lexical Functions [= LFs] (for LFs, see Žolkovskij & Mel'čuk 1966, 1967, Mel'čuk 1974: 78ff, Mel'čuk 1982, 1988a: 61ff, 1996, Mel'čuk & Zholkovsky 1984, 1988: 55-66, Frawley 1988, Alonso Ramos 1993, Polguère 2000, Steinlin *et al.* 2004). LFs describe the semantic derivations and lexical collocations of the headword L. Thus, in the entry for BAKEI.2a it is

indicated that the device people typically use to bake bread, cakes, etc. in [= S_4^{usual} = $S_{\text{instr-loc}}$] is called an *oven*, and the person whose profession is baking bread is a *baker*. If your baking is successful, you bake what you bake *to a turn* [= Ver], but if you bake a cake unsuccessfully you may *overbake* or *underbake* [= AntiVer] it; a cake baked recently is a *fresh-baked* cake; etc.

LFs allow for a thorough, systematic and homogeneous description of semantic derivations and restricted lexical cooccurrence—i.e., collocations, of any LU L. To make this statement clear, I have to begin with the corresponding definitions.

2.3.1 Semantic Derivation

Derivation is a well-know phenomenon: the lexeme L' is said to be derived from L if and only if the signified of L' includes the signified of L and the semantic difference ' L' ' – ' L ' is expressed in language \mathbf{L} by morphological means. Thus, *smoker* is derived from *smoke_v*, since '*smoker*' \supset '*smoke_v*' ('*smoker*' = 'person who smokes_v regularly') and the difference '*smoker*' – '*smoke_v*' = 'person who ... regularly' is expressed by a suffix: **-er**. The notion of semantic derivation is a generalization of the notion of derivation in the current sense.

Definition 1: Semantic derivation

The LU L' is said to be semantically derived from L in language \mathbf{L} iff the following three conditions are simultaneously satisfied:

1. The signified of L' includes that of L: ' L' ' \supset ' L '.
2. The semantic relation between L' and L is found in \mathbf{L} in several other pairs of LUs.
3. At least in some cases, the difference ' L' ' – ' L ' is expressed in \mathbf{L} by morphological means.

Thus, *blacksmith* is semantically derived from [*to*] *forge*, because '*blacksmith*' = 'person who forges regularly'. In the same vein, *terrestrial* is semantically derived from *earth*, because '*terrestrial*' = 'related to earth', cf. *ocean* ~ *ocean+ic*, *algebra* ~ *algebra+ic*, *post* ~ *post+al*, etc. All 'normal' derivations are semantic derivations

as well, but not the other way around. As I already said, an ECD takes upon itself to present all of the semantic derivations of any headword L .

2.3.2 Collocation

A *collocation* is a particular type of binary set phrase, or phraseme; more specifically, collocations are known as semi-phrasemes (see Endnote 6, p. 00), because one of the collocation's constituents is selected by the speaker freely—according to its meaning and syntactic properties—while the other one is chosen as a function of the first. Let me start with a formal definition.

Definition 2: Collocation

A phraseme $\mathbf{AB} = \langle 'S' ; /A/ \oplus /B/ ; \Sigma_{\mathbf{AB}} \rangle$ of \mathbf{L} is called a *collocation* iff it satisfies simultaneously the following three conditions:

1. The signified of \mathbf{AB} includes the signified of \mathbf{A} as its semantic pivot: 'A' is the argument of the difference ' \mathbf{AB} ' – ' \mathbf{A} ' = ' \mathbf{C} '.

[Formally: ' S ' = ' A ' \oplus ' C ' & ' C '('A').]

2. \mathbf{A} is selected by the speaker freely, i.e., independently of \mathbf{B} —for its own signified 'A'.

3. \mathbf{B} is not selected freely—it is selected for its signified 'C' restrictedly, i.e., as a function of \mathbf{A} .

The lexeme \mathbf{A} is called the *base* of the collocation \mathbf{AB} , and \mathbf{B} is its *collocate*. As far as the meaning of the collocate is concerned, two cases must be distinguished: either

— \mathbf{B} has the sense 'C' in the dictionary of \mathbf{L} (i.e., ' \mathbf{B} ' = ' \mathbf{C} '), but \mathbf{B} cannot be unrestrictedly replaced by any of its synonyms, because it is selected as a function of \mathbf{A} ; or

— \mathbf{B} does not have the sense 'C' (' \mathbf{B} ' \neq ' \mathbf{C} '), because it expresses the meaning 'C' only in combination with \mathbf{A} ; see Comment 2 below.

Comments on Definition 2

1. The meaning of a collocation \mathbf{AB} necessarily includes the meaning of one of its two lexemic components: the collocation's base \mathbf{A} is selected by the speaker strictly for its meaning. As for the other component, the collocate \mathbf{B} , its meaning

may or may not be included in the meaning of **AB**, but in any case the collocate **B** is selected restrictedly—that is, as a function of **A**, regardless of whether the use of **B** to express the meaning ‘C’ is unique (there is no lexeme ***B**‘C’ in **L**’s dictionary) or not (**B**‘C’ exists in **L**’s dictionary). In principle, one cannot replace **B** with its synonym, no matter how close this synonym is to **B** semantically.

2. The collocate **B** which has the meaning ‘C’ (so that ‘B’ = ‘C’) can be a genuine LU of **L** or not.

1) The collocate **B**‘C’ is considered a genuine LU of **L** and therefore is entered in **L**’s dictionary (= ECD) in two cases:

— The combination of **A** with **B**‘C’ is not unique, that is, the expression **B**‘C’ combines with several LUs other than **A**. This is the common case.

— The combination of **A** with **B**‘C’ is unique or quasi-unique—the expression **B**‘C’ combines only with **A** or maybe with two or three semantically close lexemes, but **B** is a monolexemic expression (= a lexeme) and **L** has no other lexemes **B**‘C’_i’ (with the same signifier and a related signified). This is the special case: unique lexemes, such as **STENTORIAN** (only with **VOICE**, **TONE** and **UTTERANCE**), **ARTESIAN** (only with **WELL**), or **PYRRHIC** (only with **VICTORY**). A unique lexeme has to be listed in **L**’s ECD.

2) The collocate **B**‘C’ is not considered a genuine LU of **L** and therefore is not entered in **L**’s dictionary in two cases as well:

— **B** is a monolexemic expression, the combination of **B**‘C’ with **A** is unique, but **B** has other senses in the dictionary of **L**. Thus, **BLACK** ‘without addition of a dairy product’ as found in *black coffee* should not have a special entry in an English ECD; this is possible since the adjective **BLACK**—with its other senses—appears in an English dictionary anyway. (But, even if **ARTESIAN** is unique in its combination with **WELL** and **STENTORIAN** is quasi-unique in its combination with **VOICE** etc., these adjectives have to be entered into the dictionary since they have no other senses, see above.) **BLACK** ‘without addition of a dairy product’ and all similar expressions are not genuine lexemes of English; they can be called pseudolexemes of English.

— **B** is a multilexemic phraseme and the combination of **B**‘C’ with **A** is also unique. Then **B** is not considered a genuine LU of **L** and is not separately stored in

its dictionary. For instance, the collocate *as a bee* in *(as) busy as a bee* does not appear as a separate entry in a dictionary of English. The same holds of *as a bug (in a rug)* in *(as) snug as a bug (in a rug)* or *as a bug's ear* in *cute as a bug's ear*, etc.: all these collocate phrases cannot be used outside of their collocation, and therefore it makes no sense to have for them a separate entry in **L**'s dictionary. On the other hand, *as a cat on a hot tin roof* can be used in several collocations: *(as) nervous as a cat on a hot tin roof*, *(as) agitated as a cat on a hot tin roof*, *(as) anxious* (*skittish, jumpy, ...*) *as a cat on a hot tin roof*; therefore, it deserves a separate entry.

3. The meaning 'C' expressed by the collocate **B** can be extremely specific—practically unique to **AB**, or, on the contrary, it can be rather general and appear in a number of different collocations.

Examples of collocations (the collocation's base is in small caps; 'C' is the meaning expressed by the collocate)

Collocations where the meaning 'C' is (quasi-)unique:

black COFFEE, French WINDOW, leap YEAR, aquiline NOSE, rancid BUTTER, artesian WELL, The HORSE neighed, ...

Collocations where the meaning 'C' is non-unique:

- *do* ⟨*make⟩ [*someone*] *a FAVOR*, *give* ⟨*deliver⟩ [*someone*] *a LOOK*, *take* ⟨*seize⟩ *a STEP*, *be* ⟨*find oneself⟩ *in DESPAIR*, *commit* ⟨*perform⟩ *a BLUNDER*, *pay* ⟨*make⟩ *a VISIT*, ... [the collocate is a light, or support, verb: ≈ 'do'];
- *strong* ⟨*powerful⟩ *COFFEE*, *heavy* ⟨*weighty⟩ *RAIN*, *BLUSH* *deeply/profusely* ⟨*profoundly⟩, *profoundly* ⟨*powerfully⟩ *AFFECT*, *as ALIKE as two peas in a pod* ⟨*as two drops of water⟩, ... [the collocate is an intensifier: ≈ 'very', 'very much', 'completely'];
- *respond (well) to a TREATMENT*, *run into an AMBUSH*, *accept an INVITATION*, *observe a RULE*, *strike a LAND MINE*, *meet a REQUIREMENT*, *heed a WARNING*, ... [the collocate is a realization verb: ≈ 'do with/for L what is expected'].

Collocations constitute the absolute majority of phrasemes in any language and represent the main challenge for any theory of phraseology (for more details, see Mel'čuk 2003). In order to describe collocations in a dictionary in a rigorous, systematic and exhaustive way, MTT proposes the apparatus of LFs.

The meaning ‘C’, which is expressed restrictedly—that is, by **B** contingent on **A**, is associated with a Lexical Function, see below. The lexeme **A**, which keeps its signified intact within the signified of the collocation and determines the expression of ‘C’ by **B**, is the argument, or keyword, of the corresponding LF; it is of course the base of the corresponding collocation.

One of the main innovations of the ECD is the emphasis on the most complete coverage possible of semantic derivations and collocations. In this subsection I concentrate on the description of collocations of an LU L, i.e., of cooccurrences of L that are not free, but whose combinability with L is determined neither by their meaning nor by their form (nor by their lexical class membership); the constraints here are purely lexical, which means that the corresponding lexemic combinations have to be stored—i.e., learned—as such. In English you *take* a step, while in Spanish you ‘give’ it (Sp. *dar un paso*) and in German you ‘make’ it (Germ. *einen Schritt machen*). In English, you *give* a lecture, while in French you ‘give’ or ‘make’ it (Fr. *donner/faire une conférence*) and in Russian, you ‘read’ it—even you don’t literally read anything (Rus. *čitat’ lekciju*). In English, a guy is *drunk completely, as a lord, as a sailor, as a skunk* or *blind-/dead-/stone-drunk*; in Russian, he is *p’jan mertveki* lit. ‘cadaver-style’/*v dosku* lit. ‘into a board’; in German, you are *besoffen wie eine Sau* ‘as a female pig’, and in French, *soûl comme un Polonais* ‘as a Pole’. Swarms of similar phrases—collocations—can be found in texts; cf. a few sentences collected in no time from a magazine (the collocations are boldfaced, with key lexeme in small caps):

- (13) a. *It’s time **to go on** a low-carb **DIET**.*
 b. *Cleaner power stations are vital to **meet** **DEMAND** for energy.*
 c. *Our **RELATIONSHIP** was **stormy** from the start.*
 d. *The legislature **is in** **SESSION**.*
 e. *The government has **kept** him **under** house **ARREST** since.*
 f. *She has **seized** **CONTROL** over U.S. foreign policy.*

This phenomenon, as is universally known, presents a serious difficulty for any text synthesizer, be it a human or a computer. At the same time, since the number of collocations is enormous (a few million), to master them is no less of a challenge for any human language learner—not only a foreigner, but a native

speaker as well. (In point of fact, to speak and write a language really well boils down to being able to use the wealth of its collocations.)

The MTT proposes to describe all semantic derivations and all collocations of each LU L in a systematic and exhaustive way: by means of LFs.

2.3.3 Lexical Function

As before, I will start with a definition.

Definition 3: Lexical Function

A correspondence f that associates a set $f(L)$ of lexical expressions with an LU L is called a *Lexical Function* [= LF] iff it satisfies either conditions A1-A3 or condition B:

A. f is applicable to several LUs and:

1. Semantic homogeneity of $f(L)$

For any two different LUs L_1 and L_2 , if $f(L_1)$ and $f(L_2)$ both exist, then any $L'_1 \in f(L_1)$ and $L'_2 \in f(L_2)$ bear an (almost) identical relationship to L_1 and L_2 , respectively, as far as their meaning and the DSynt-role are concerned:

$$\frac{L'_1}{L_1} \approx \frac{L'_2}{L_2}$$

2. Maximality of $f(L)$

For any two different LUs L'_1 and L'_2 , if $L'_1 \in f(L_1)$ and $L'_2 \notin f(L_2)$, then L'_2 does not stand to L_2 in the same relationship as L'_1 to L_1 :

$$\frac{L'_1}{L_1} \neq \frac{L'_2}{L_2}$$

3. Phraseological character of $f(L)$

a) At least in some cases $f(L_1) \neq f(L_2)$; and
 b) at least for some $f(L_i)$ some elements of $f(L_i)$ cannot be specified without mentioning an individual LU L_i .

B. f is applicable to only one LU L (or perhaps to a few semantically close LUs).

In $f(L)$, L , which is the argument of f , is called the keyword of f , and $f(L) = \{L_i\}$ is f 's value.

An LF that is applicable to several LUs—satisfying Conditions A1-A3—is called *normal*; an LF applicable to only one LU (or two or three semantically close LUs)—satisfying Condition B—is *degenerate*. Degenerate LFs are an extreme case of non-standard LFs (see below).

A LF is associated with a meaning or a set of meanings.

Definition 4: Standard Lexical Function

A normal LF f is called *standard* iff it satisfies the following two conditions:

1. Broadness of f 's domain

The meaning ' f ' associated with f is sufficiently general (non-specific) to be applicable to many other meanings (' f ' can even be empty²²), so that f is defined for a relatively large number of keywords.

2. Broadness of f 's range

The expressions of the meaning ' f ' are sufficiently variegated so that f has a relatively large number of elements in its possible values.

Condition 1 characterizes an LF f as a potential standard LF, and Condition 2—as an actual standard LF; it means that the set of all $f(L_i)$, for a vast variety of L_i s, is relatively rich.

Both 1) normal LFs that do not satisfy at least one of Conditions 1 and 2 of Definition 4 and 2) degenerate LFs are called *non-standard*.

Here is an example of a non-standard LF. The meaning '[a hot beverage X] with the addition of Y' has in French and in Italian at least two different expressions with the opposite structure, both being phraseologically bound: Fr. *café crème* (<*café à la crème> 'coffee with cream' vs. *café au lait* (<*café lait>); It. *caffè alla panna* 'coffee with cream' vs. *caffè latte* 'coffee with milk'. (Compare also Fr. *thé nature* 'tea without the addition of anything', but not **café nature*; *café noir*, but not **thé noir*.) Therefore, the meaning in question determines a lexical dependency which satisfies Conditions 1 and 2 of Def. 3: it is an LF. However, it fails to satisfy Conditions 1 and 2 of Def. 4, and thus it is not a standard LF: it is

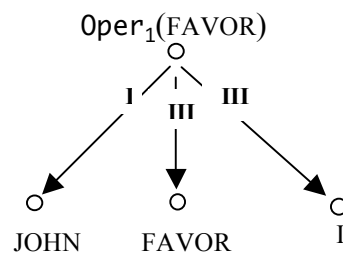
applicable just to three or four arguments and has only a couple of different expressions.

Simple standard LFs are not very numerous (about 60 in all languages examined so far); they constitute the core of the proposed description of lexical cooccurrence (for more on LFs, see the references at the beginning of Subsection 3.2, p. 00). Here I cite just two examples:

1) The LF Magn represents an intensifier:

Magn(<i>riposte</i>)	= <i>severe < crushing</i>	Magn(<i>deaf</i>)	= <i>as a post, stone-</i>
Magn(<i>cry_N</i>)	= <i>loud < deafening</i>	Magn(<i>strong</i>)	= <i>as a bull, as a horse</i>
Magn(<i>applause</i>)	= <i>loud < deafening, frenetic, frenzied, terrific</i>	Magn(<i>drunk</i>)	= <i>dead, stone-; as a lord, as a sailor, as a skunk</i>
	Magn(<i>appreciate</i>)	= <i>deeply, greatly</i>	
	Magn(<i>sleep_V</i>)	= <i>deeply, heavily, like a log</i>	
	Magn(<i>apologize</i>)	= <i>profusely</i>	

2) The LF Oper₁ represents a semantically empty (or quasi-empty) verb: a light, or support, verb such that its keyword (a predicative noun) L is its DSyntA II (i.e., its first surface object, i.e., in most cases—its Direct Object; as its DSyntA I (= as its Syntactic Subject) Oper₁ takes the DSyntA I of L; DSyntA II of L often, i.e., with some Ls, become DSyntA III of Oper₁ (not always: sometimes it DSyntA II of L). Cf. the representation of the sentence *John does me a favor* at the DSynt-level:



Oper ₁ (<i>complaint</i>)	= <i>lodge, make [ART ~]</i>	Oper ₁ (<i>talk</i>)	= <i>give [ART ~]</i>
Oper ₁ (<i>sigh</i>)	= <i>heave [ART ~]</i>	Oper ₁ (<i>despair</i>)	= <i>be [in ~]</i>
Oper ₁ (<i>flu</i>)	= <i>have [the ~]</i>	Oper ₁ (<i>attention</i>)	= <i>pay [~]</i>
Oper ₁ (<i>order</i>)	= <i>give [ART ~]</i>	Oper ₁ (<i>battle</i>)	= <i>be locked [in ~]</i>

Simple standard LFs can be combined into complex LFs and into configurations of LFs; given the limitations of space, I will not develop these two concepts here.

For some LUs, especially those that refer to complex objects having many and various uses, the number of LFs of all types—simple standard, complex (standard and non-standard), plus LF configurations—is very high. Moreover, the meaning of the LU L can be considered to have different ‘facets’ associated with different situations in which the referent of L can be used. Thus, for instance, different facets of the LU BOOK are ‘books as objects of reading/studying’, ‘books as objects of manufacturing’, ‘books as objects of selling/buying’, and ‘books as objects in library service’. Under each of such facets, an LU L can have different LFs, and the same LFs applied to L can have different values; for instance, books as object of reading are *read*, books as object of manufacturing are *printed* and *bound*, books as objects in library service are *borrowed*, *returned*, *renewed*, etc. To facilitate the presentation and retrieval of LFs in such a case, the ECD uses Thematic Groupings of LFs, corresponding to different facets of the headword L. Thus, the LF zone of the entry for the noun BLOOD must be subdivided into the following Thematic Groupings:

- **Blood as physiological liquid** (*circulates in veins, red/white cells, vessels, clot, ...*)
- **Blood as target of medical treatment** (*hematology, transfusion, blood test, leukemia, donor, blood pressure, ...*)
- **Blood as an element of injury** (*spill blood, bleed, blood oozes or spurts, stop bleeding, scab, ...*)
- **Blood as staining substance** (*bloody, blood-stained, blood-smearred, ...*)

The noun SHIP also requires Thematic Groupings in the LF zone:

- **Ship as a means of navigation** (*sails, steams, plies the waters of Y, puts into port, drops anchor, lies at anchor, weighs anchor, displacement, draught, ...*)
- **Ship as a means of transportation** (*buy a passage, go by ship, cabin, berth, ...*)

- **Ship as object of a sea disaster** (*S.O.S., capsizes, runs aground, sinks, ...*)
- **Ship as a military unit** (*navy, send to the bottom, cruiser, destroyer, corvette, ...*)

2.4 Illustrative Examples in the ECD

In an ECD, an example of the use of the headword L is a full-fledged sentence containing L and illustrating the descriptive claims made in the entry about L's meaning and use; examples are, so to speak, the linguistic flesh necessary to cover the formal skeleton of the LU that the lexicographer presents. Although formally they are not necessary, examples are of utmost importance for the ECD: they constitute the final substantiation of the lexicographic description, while helping the user to understand and to criticize it. Examples are not restricted to this special zone: examples also illustrate the Government Pattern and some (but by no means all) of the LFs. The examples in an ECD must meet the following two conditions.

First, the examples cannot be uncritically borrowed from existing texts, even from good authors. Good writers are good precisely because they stretch the capacities of word use beyond what the linguistic system permits; they experiment and play with words. But ECD illustrations must show minimal differences in meaning and cooccurrence, without being cluttered with unnecessary, albeit interesting and/or beautiful, details. Therefore all the examples must be screened by the lexicographer; in many cases they have to be doctored. In the ECD approach, textual research is an absolute must, but examples found in texts should be gone through with a fine-tooth comb by the lexicographer and adapted, when necessary, to the specific needs of each particular entry.

Second, along with positive examples (= samples of correct use), an ECD also uses negative—asterisked—ones, which are necessary to justify the restrictions the lexicographer sees fit to introduce. A linguistic constraint rules out incorrect expressions, and to illustrate such a constraint the incorrect expressions of the type barred by it must be presented. Actually, the systematic introduction of asterisked expressions as legitimate and unavoidable linguistic data has revolutionized modern linguistics—thanks to Chomsky's school of Transformational Generative Grammar in the 1960s. (Incidentally, the utility of negative examples in a dictionary was

pointed out more than 65 years ago by the Russian lexicographer L. Ščerba (1940 [1958]). Unfortunately for us, no one was listening.)

Having presented a general characterization of the lexical entries in an ECD, I will now switch to a more detailed discussion of its macrostructure (i.e. its vocables).

3 The ECD's Macrostructure: An ECD's Lexical Super-entry

The discussion of the ECD's super-entries, or vocables, will be carried out in two steps: in Subsection 3.1 three necessary basic notions are introduced; Subsection 3.2 presents formal criteria for uniting lexical entries into super-entries—or, from the opposite perspective, for distinguishing the LUs of the same vocable.

3.1 Basic Notions for the Characterization of Lexical Super-entries

The lexical entries that constitute an ECD of **L** are logically linked to each other according to two axes: in a 'horizontal' dimension, LUs of **L** are grouped into *semantic fields*, and in a 'vertical' dimension, into *vocables* (which have been already mentioned on various occasions). The 'horizontal' link is exploited by the lexicographer only when the ECD is being developed—it is not directly reflected in the ECD's organization. The 'vertical' link is likewise used at the development stage, but it is also shown in the ECD's actual structure. Both axes are semantic in nature: the LUs are grouped into semantic fields and vocables strictly on the basis of their semantic relatedness. The central notion in this respect is the *semantic bridge* between two LUs—the concept that was previously used several times, but without a definition.

Definition 5: Semantic Bridge

A *semantic bridge* between LUs L_1 and L_2 is a configuration 'σ' of semantemes shared by the lexicographic definitions of L_1 and L_2 such that it satisfies simultaneously the following two conditions:

- 1) 'σ' is sufficiently important in these definitions;
- 2) 'σ' occupies a sufficiently central position in these definitions.

Definition 5 is not precise enough—it remains unclear what exactly is required for a common semantic component to be ‘sufficiently’ important and which position is ‘sufficiently’ central in a lexicographic definition. This reflects, of course, the insufficiency of our knowledge. However, at least three points can be elaborated.

- There are some semanteme configurations that cannot constitute a semantic bridge by themselves—they are too general and therefore too common. Thus, the semanteme ‘cause₂’ is present in the meaning of MURDER_V (‘X murders Y’ includes ‘X causes₂ [that Y dies]’) and in that of CLEAN_V (‘X cleans Y’ ≈ ‘X causes₂ that Y becomes clean’); however, it is obvious that MURDER_V and CLEAN_V should not be considered as linked by a semantic bridge. Probably, it would be possible to draw up a list of very general semantemes which can never constitute a semantic bridge between two LUs by themselves, such as ‘cause_{1, 2}’, ‘act’ or ‘happen’, or else taxonomic semantic labels such as ‘state’, ‘event’, ‘period’, ‘substance’, ‘object’, ‘person’, etc., since these semantemes are semantic primitives or close enough to semantic primitives. (General semantic labels are extremely important for an ECD: they are used in the definitions in order to allow the linguist to treat LUs by their semantic class, which is encoded by the corresponding semantic label; see Polguère 2003. What I am saying here is only that very general semantic labels cannot represent semantic bridges.)

- What is important is not so much the absolute size of the would-be semantic bridge, as how big it is proportionally—i.e., how big the part it occupies is within the respective definition.

- The most central position in a definition is the generic component ‘Y’: ‘X’ ≡ ‘Y which Z’. However, a semantic bridge is not necessarily the generic component, but a semantic bridge in the position of generic component is very special, as we will immediately see.

By way of illustration, consider the vocable [to] BAKE, given below, in **5.1**. Eleven lexicographic senses of the verb BAKE—that is, eleven LUs—are distinguished; all eleven are put together, to form one vocable. The reason is that all BAKE lexemes exhibit a semantic bridge: all of them share an important semantic component—‘[cause_{1, 2} by the action of] dry heat’—while some of them share more. (Of course, ‘cause_{1, 2} ... by the action of dry heat’ need not be mentioned

explicitly in every definition: it may appear implicitly as a component of a component; thus in **BAKEI.2b**, ‘dry heat’ is implied via **BAKEI.2a**.)

On the basis of the notion of a semantic bridge, semantic field and vocable can be readily defined.

Definition 6: Semantic Field (adapted from Mel’čuk *et al.* 1995: 173-175)

A semantic field $\mathbf{F}^{\langle\sigma\rangle}$ is the set $\{L_i\}$ of LUs such that all L_i share the semantic bridge ‘ σ ’ which is 1) either the generic component in L_i ’s definition (the most common case)
or
2) is linked to the generic component by a meaning underlying an LF (the less common case).

The semantic component ‘ σ ’ is the semantic field identifier. If ‘ σ ’ is not the generic component ‘ σ^{gener} ’ of ‘L’, it can be related to this ‘ σ^{gener} ’ as:

- An actant: **PEN** is ‘an artifact designed to write with’—a Sem-Actant of **WRITE**, and therefore **PEN** belongs to $\mathbf{F}^{\langle\text{writing}\rangle}$; **SINGER** is ‘individual who sings’—equally a Sem-Actants of **SING**, so it belongs to $\mathbf{F}^{\langle\text{singing}\rangle}$; **CHOIR** is ‘set of singers’, and it also is part of $\mathbf{F}^{\langle\text{singing}\rangle}$.

- A place name: **RESTAURANT** is ‘establishment where you can eat’—an S_{loc} of **EAT**; **RESTAURANT** thus belongs to $\mathbf{F}^{\langle\text{eating}\rangle}$.

- An ‘instrument’ name: **PILLOW**, **MATTRESS**, **BED SHEET**, etc. are ‘artifacts designed to be used to sleep’— S_{instr} of **SLEEP**; the above nouns belong to $\mathbf{F}^{\langle\text{sleeping}\rangle}$.

- A ‘when’ name: **DREAMS** happen ‘when you are asleep’; so **DREAM** also belongs to $\mathbf{F}^{\langle\text{sleeping}\rangle}$, just as **SLEEPING PILLS** and **LULLABY**.

In other words, an LU belongs to $\mathbf{F}^{\langle\sigma\rangle}$ iff ‘ σ ’ is the generic component in ‘L’ or at least ‘ σ ’ is linked to the generic component of ‘L’ by a semantic relation that is more or less regular. If this relation is not prominent enough in \mathbf{L} , then L

does not belong to \mathbf{F}'_{σ} . Thus, LEG of my pants or FINGER of my glove do not belong to $\mathbf{F}'_{\text{bodyparts}}$: they only ‘cover’ the respective body part, and ‘cover’ is not a regular semantic relation in English.

As an example of a semantic field I can cite names of nationalities—[*the*] ENGLISH, CHINESE, FRENCH, GERMANS, ITALIANS, RUSSIANS, etc. all belong to $\mathbf{F}'_{\text{nationality}}$; the definition of any such name contains the semantic configuration ‘nationality₂’ as its generic component (‘nationality₁’ \approx ‘citizenship’):
 [*the*] ENGLISH: ‘NATIONALITY₂ native of England and whose mother tongue is English’;
 [*the*] CHINESE: ‘NATIONALITY₂ native of China and whose mother tongue is Chinese’; etc.

LUs belonging to the same semantic field not only have the same generic component; they also tend to have definitions featuring the same general structure, as can be immediately seen from the preceding example of nationalities.

Traditional dictionaries were, as a rule, compiled in alphabetic order of entries; ‘What letter are you at now?’—is a typical question addressed to a professional lexicographer involved in the writing of a dictionary. In sharp contrast to this, an ECD is developed by semantic fields, and it is impossible to write it in any other way. Only this technique guarantees the homogeneous description of all the LUs belonging to one semantic field.

Definition 7: Vocable

A *vocable* is the set $\{L_i\}$ of LUs such that any two LUs $L_1, L_2 \in \{L_i\}$ satisfy simultaneously the following two conditions:

- 1) L_1 and L_2 have the same signifier;
- 2) L_1 and L_2 either have a semantic bridge or are linked by a chain of semantic bridges via other LUs of the same vocable (for instance, L_1 and L' have a semantic bridge ‘ σ_1 ’, L' and L'' have a semantic bridge ‘ σ_2 ’, and L'' and L_2 have a semantic bridge ‘ σ_3 ’).

Notation

The LUs that have the same signifier, but belong to different vocables are distinguished by right superscripts: for instance, $L^1 \sim L^2 \sim L^3$, as in $\text{PEN}^1 \approx$ ‘writing

implement' vs. PEN² ≈ 'female swan' vs. PEN³ ≈ 'enclosure for animals' vs. PEN⁴ = 'penitentiary'. These LUs are homonyms of a special kind: namely, they are also homographs. Here are some more examples of homographs: nouns RENT¹ ≈ 'regular payment by a tenant' ~ RENT² ≈ 'breach, schism' ~ RENT³ slang for 'parent' or DATE¹ 'sweet fruit ...' ~ DATE² 'indication of a time moment—the name of the day, month, and year' ~ DATE³ 'romantic meeting of two people'; verbs Fr. VOLER¹ '[to] fly' ~ VOLER² '[to] steal'. Such LUs share no important common semantic components (= no semantic bridges).

Since an ECD is a dictionary that stores LUs in the written form, homographs that are not homophones should also be distinguished by superscripts: ROW¹ 'objects arranged in a line' (/rō/) vs. ROW² 'noisy quarrel' (/ra^u/).

Different LUs of the same vocable are distinguished by lexicographic numbers: Roman and Arabic numbers and small Latin letters; for instance, LI.1a ~ LI.1b ~ LII ~ LII ~ LIII.a ~ LIII.b. The numbering of LUs within a vocable is done as a function of semantic distance between two LUs. The semantic distance between LUs L₁ and L₂ is measured by two parameters considered together:

- The size of the semantic bridge (= shared semantic component) between L₁ and L₂: the bigger the semantic bridge, the closer L₁ and L₂ are.
- The regularity of the semantic distinction 'δ' between L₁ and L₂: the higher the number of lexical pairs where 'δ' appears, the closer L₁ and L₂ are.

An ECD uses thus four levels of distinctive lexicographic numbers:

- Numerical superscripts for homonymous LUs, which belong to different vocables.

The other lexicographic numbers are used to distinguish the LUs belonging to the same vocable, which is thus polysemous; such LUs feature an important enough semantic bridge.

- Roman numbers distinguish LUs of the same vocable whose semantic difference is not very regular in **L**: verbs BAKEI ('food' BAKE: bread, potatoes) ~ BAKEII ('pottery' BAKE: bricks, amphorae) ~ BAKEIII ('weather' BAKE: *We are baking here*). At the same time, Roman sense-distinguishing numerals signal well-differentiated lexeme groupings.

- Arabic numbers distinguish LUs of the same vocable whose semantic difference is relatively regular in **L**: BAKE1.1 (\approx transform edible stuff by using heat: *bake potatoes*) \sim BAKE1.2 (\approx create a product by using heat: *bake rolls*). Arabic numbers specify tighter lexical groupings.

- Lowercase Latin letters distinguish LUs of the same vocable whose semantic bridge is very important and whose semantic difference is regular in **L**: BAKE1.1a (action: *Bob baked the potatoes in 30 minutes*) \sim BAKE1.1b (process: *The potatoes baked in 30 minutes*).

Note that Roman numbers are always used in case of metaphor (but of course not exclusively; cf. BAKEIII vs. BAKEI-II), and Arabic numbers, in case of metonymy.

The use of three ranks of sense distinguishers within a vocable—rather than more or fewer—is not motivated by theory: its justification is simply that a three-fold division seems to work well in practice.

A vocable in an ECD corresponds to a polysemous entry in traditional dictionaries.

Comments on Def. 7

1. Condition 1

- The signifier of an LU L is either the signifier of the (common) radical of all of its wordforms and analytical form phrases (in the case of lexemes) or the SSynt-tree of the phrase (in the case of idioms). The signifier of the lexeme IMPROVE1.1a—in its written form—is the string of letters, extracted from *improves*, *improving*, *improved*, *has been improving*, etc.; the signifier of the idiom _SEE RED_ ‘become very angry’ is the tree SEE \circ -dir-obj \rightarrow \circ RED.

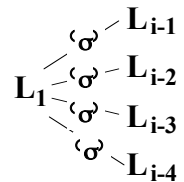
- Part-of-speech conversion is a morphological expressive means: [a] COOK is formed from [to] COOK by conversion, similarly to SMOK+ER, formed from [to] SMOKE by suffixation. Since the suffix is part of the lexeme signifier, so must be the part-of-speech conversion. Therefore, if the signifier of SMOKER is *smoker*, the signifier of [a] COOK is *cook*_N. An important corollary of this is that if SMOKER and SMOKE_V belong to different vocables, so do COOK_N and COOK_V; at the same time, COOK_N and COOK_V are by no means homonyms, since their signifiers are distinct.

2. Condition 2

Any two LUs of the same vocable need not share a semantic bridge: it is sufficient if they are linked by a chain of semantic bridges. Suppose L_1 and L_2 share a semantic bridge ‘ σ_i ’, but L_2 and L_3 share a semantic bridge ‘ σ_j ’ (without sharing ‘ σ_i ’, and ‘ $\sigma_i \neq \sigma_j$ ’). Then L_1 and L_2 are directly linked semantically, and so are L_2 and L_3 ; L_1 and L_3 are semantically linked only indirectly. That is what Condition 2 says: any two LUs of the same vocable are semantically linked—at least indirectly.

3. Two LUs L_1 and L_2 belonging to one vocable stand in the relation of *polysemy* (the example of the vocable IMPROVE at the end of Subsection 1.3, p. 00). There are two types of semantic links within any one vocable, and consequently there are two types of polysemy.

- If several LUs L_i each share one and the same semantic bridge ‘ σ ’ with the LU L_1 , we have *radial polysemy*:



For instance, let L_1 be HEAD ‘human bodypart’; then L_{i-1} is HEAD ‘animal bodypart’, L_{i-2} is HEAD \approx ‘chief’ (*head of a bank*), L_{i-3} is HEAD \approx ‘upper part’ (*head of a mushroom/of a hammer*), and L_{i-4} is HEAD \approx ‘front part’ (*head of a convoy*). All these latter HEADs are defined with a reference to the human head: by similarity.

- If the LU L_{i-1} shares the semantic bridge ‘ σ_i ’ with the LU L_1 , L_{i-2} shares the semantic bridge ‘ σ_j ’ with L_{i-1} , and L_{i-3} shares the semantic bridge ‘ σ_k ’ with L_{i-2} , we have *chain polysemy*:

$$L_1 \text{---} \sigma_i \text{---} L_{i-1} \text{---} \sigma_j \text{---} L_{i-2} \text{---} \sigma_k \text{---} L_{i-3}$$

An English example is as follows: L_1 = BODY ‘human body [as opposed to the soul]’; L_2 = BODY ‘group of humans’ (*large bodies of unemployed*) has a semantic bridge with L_1 , and this is ‘human’; but L_3 = BODY ‘organization’ (*governing body*) has a semantic bridge with L_2 (= ‘group’); L_4 = BODY ‘main part of the human body’ [\approx ‘torso’] again has a semantic bridge with L_1 (= ‘human body’), but L_5 = BODY ‘main part’ (*the body of a plant/a text*)—only with L_4 (= ‘main part’). Another example (suggested by E. Marshmann) is L_1 = BUG1.1 ‘insect’, L_2 = BUG1.2 ‘virus—

as if it were a *bugI.1*’, and $L_3 = \text{BUGII}$ ‘error in a computer program—as if it were a *bugI.2*’.

In actual practice, both types of polysemy are often found inside the same vocable, as shown in the above example with *BODY*.

From a formal viewpoint, a vocable is a lexical *super-entry*—a set of individual dictionary entries brought together because of their semantic and phonological relatedness. The use of vocables in an ECD achieves three goals:

— It allows for important generalizations. Thus, the part of speech and many, if not all, morphological characteristics accrue to all LUs of a vocable and must be extracted from individual entries in order to be associated directly with the name of the vocable. In a similar way, some values of LFs can be shared by all LUs of the vocable, and they also can be ‘raised’ to the vocable, to avoid tedious repetition (see below, p. 00).

— It reflects the intuition of the speakers, who perceive different LUs of the same vocable as belonging to one polysemous ‘word.’

— It allows for a greater compactness and better surveyability and thus presents obvious advantages for the user, who can better grasp the commonalities and differences between LUs.

3.2 Criteria for the Delimitation of LUs within Vocables:

Criteria of Type II

3.2.1 Introductory Remarks

Delimiting (or separating) LUs within a vocable—i.e., distinguishing the lexicographic senses of a polysemous word or of a polysemous set phrase—is one of the most difficult tasks that a linguist encounters in dictionary-writing; the decisions he makes in this respect entail serious and long-reaching consequences. The differentiation of word senses is in fact one of the central problems not only of lexicology and lexicography, but also of theoretical semantics:

How should one distinguish, on the one hand, between ambiguity and generality (= vagueness) of meaning of a lexical item, and, on the other hand, in case of ambiguity, between homonymy and polysemy?

(For special studies of the question, as well as a rich bibliography, see Dean 1988 and Tuggy 1993; in my view, the most important contribution on the topic is Wierzbicka 1996: 258ff.)

The task of sense discrimination cannot be considered here in all its ramifications; I will simply put more logical order into what has been known for quite a long time, but not systematically used.

The problem of the unity of a lexical item L arises in a case when we perceive that 1) L has different uses that refer to two (or more) different entities/facts of the real world, but, at the same time, 2) these different uses involve common semantic configurations. For instance, the verb [*to*] PAINT and the noun AUNT are problematic. Indeed, in (14), [*to*] PAINT appears in either of the two possible senses:

(14) *Alain painted the ceiling of the hall.*

a. *painted* = ‘has covered [the ceiling] with paint’:

Alain carried out a renovation or refurbishing.

b. *painted* = ‘has covered [the ceiling] with artistic images’:

Alain created a work of art.

[I ignore a third possible sense: Alain painted an artistic image of the hall ceiling.]

In both uses the same semantic component is present: ‘apply the paint’.

In (15), AUNT can be understood even in three ways:

(15) *This is my aunt Joan.*

a. *aunt* = ‘a sister of the mother’;

b. *aunt* = ‘a sister of the father’;

c. *aunt* = ‘the wife of an uncle’.

Again, in all three uses the same semantic component appears: ‘sister or wife of a brother of a parent’.

At first glance, examples (14) and (15) are parallel: in both the uses of a lexical item—[*to*] PAINT in (14) and AUNT in (15)—correspond to different extralinguistic realities, two in the case of PAINT, three in the case of AUNT; at the same time, these uses show semantic bridges. Does this, however, mean that in (14) we have two LUs PAINT and in (15) three LUs AUNT? The answer is that there is no parallelism, and these examples represent two different cases. [*To*] PAINT in (14) is ambiguous and corresponds to two LUs: PAINT₁ ‘cover with paint’ and PAINT₂ ‘cover with artistic images’; but AUNT in (15) is not ambiguous: it represents one LU with a disjunctive definition (AUNT = ‘a sister of the mother, or a sister of the

father, or the wife of an uncle’). To help distinguish lexical items of type [to] PAINT from those of type AUNT we introduce Criteria of Type II, or criteria for distinguishing lexicographic senses of a polysemous lexical item. They supply a formal frame to help the linguist make a decision concerning LUs: given a ‘suspect’ lexical item L, how many actual LUs does it cover?

A ‘suspect’ lexical item is a lexical item L ‘... σ_1/σ_2 ...’ such that 1) L has in its meaning two mutually exclusive semantic components ‘ σ_1 ’ and ‘ σ_2 ’ that correspond to two different referents, but 2) along with ‘ σ_1 ’ and ‘ σ_2 ’, L carries enough semantic material to make it look like a single unit. Therefore, intuitively, L is a candidate for the status of LU, but it may also turn out to ‘hide’ two different LUs. The problem resides exactly in the semantic components ‘ σ_1 ’ and ‘ σ_2 ’; namely, we want to know which of the following alternatives is true:

a) Either ‘ σ_1 ’ and ‘ σ_2 ’ are inside one lexicographic definition, related by the disjunction ‘or3’; the unity of L is upheld, so that L is one single LU with a logical disjunction in its meaning, i.e., with a disjunctive definition:

$$L \text{ ‘... } \sigma_1 \text{ or3 } \sigma_2 \text{ ...’};$$

b) Or ‘ σ_1 ’ and ‘ σ_2 ’ belong to two lexicographic definitions; L should be split in two LUs, so that we have

$$L_1 \text{ ‘... } \sigma_1 \text{ ...’ and } L_2 \text{ ‘... } \sigma_2 \text{ ...’}.$$

To make the correct choice between a) and b), Criteria II must be applied to any suspect lexical item L.

3.2.2 Criterion II.1: Differentiating Lexicographic Information

The hypothetical lexical entry for the suspect item L is, formally speaking, a set of lexicographic information units: $\mathbf{I} = \{\mathbf{i}_n\}$. In an ideal case, any $\mathbf{i}_i \in \mathbf{I}$ is valid for L independently of the distinction ‘ σ_1 ’ vs. ‘ σ_2 ’; but in a less than ideal case, some \mathbf{i}_i are true only if L is taken to mean ‘ σ_1 ’, but not ‘ σ_2 ’, or vice versa. Then \mathbf{I} can be partitioned in two different subsets, $\mathbf{I}_1(\text{‘}\sigma_1\text{’})$ and $\mathbf{I}_2(\text{‘}\sigma_2\text{’})$, the first being the lexicographic information related to the meaning ‘ σ_1 ’, and the second—the lexicographic information related to the meaning ‘ σ_2 ’. Lexicographic information that constitutes a subset of the information in a hypothetical lexical entry such that it is related to a particular meaning within the hypothetical definition rather than to

the whole of the definition is called differentiating lexicographic information.

The presence of differentiating lexicographic information is a strong indication that the suspect L must be split into two LUs. The strength of this indication depends on at least three factors:

- The relation between the subsets $\mathbf{I}_1(\sigma_1)$ and $\mathbf{I}_2(\sigma_2)$: the inclusion of one in another gives the weakest indication, their intersection adds to its strength, and if they are disjoint, the indication is the strongest.
- The number of information units in the difference between the subsets. If, for instance, $\mathbf{I}_1(\sigma_1)$ and $\mathbf{I}_2(\sigma_2)$ differ by only one element—say, the noun L ‘... σ_1 ...’ has only the singular, while the noun L ‘... σ_2 ...’ has both numbers, the indication is very weak.
- The type of information units in the difference between the subsets. If the differentiating information units are local—of the same type (say, they are all in declension or all in the GP), the indication is weaker; if they are not local—some in declension, some in lexical cooccurrence, etc.), the indication is stronger.

Unfortunately, I do not yet know how to compute the overall strength of the indication supplied by differentiating lexicographic information on the basis of the above parameters. As a result, I am forced to simplify the picture and work just with two values of the said indication: the presence/absence of differentiating lexicographic information. The only allowance is as follows:

If $\mathbf{I}_1(\sigma_1)$ and $\mathbf{I}_2(\sigma_2)$ differ just by one element and this element is marginal, their difference is ignored.

The formal differences in L’s behavior that depend on the ‘ σ_1 ’ vs. ‘ σ_2 ’ choice can be found in four domains:

- 1) in morphological properties (e.g., different inflection patterns for different uses of L);
- 2) in the Government Pattern (different means for the expression of actants with different uses of L);
- 3) in the Lexical Functions, and more specifically
 - 3a) in semantic derivations (different derivations possible for different L’s uses);
 - 3b) in collocates (different collocates for different L’s uses).

These domains should be carefully checked when applying Criterion II.1, which can now be formulated:

If the semantic difference between two uses of L is correlated with two subsets $I_1('σ_1')$ and $I_2('σ_2')$ of differentiating lexicographic information which show more than one formal difference,
then L should be split in two LUs $L_1('σ_1')$ and $L_2('σ_2')$.

The idea behind Criterion II.1 is obvious: if the semantic difference observed between two uses of the suspect item L is paralleled by several formal differences in L's behavior, this is a decisive argument in favor of splitting L in two LUs. If, however, this difference is not correlated at all to a formal difference, Criterion II.1 gives no recommendation and, as the default case, we can keep L as a single unit. If the semantic difference is paralleled by only one formal difference, the latter can be accommodated within one single LU—by, so to speak, an amendment to L's lexical entry; Criterion II.1 again does not say anything.

NB : In practice, one can be even more lenient and allow more than one formal difference to be disregarded, especially if these are local. Everything depends on good intuitive judgements.

Examples

The semantic difference in L is not correlated with any formal difference in L's behavior

This is the case of AUNT: taken in any of its three possible uses, this noun has the same morphology, syntax and cooccurrence; absolutely nothing in its behavior points to the choice of the referent. Criterion II.1 thus gives a negative result: it does not prevent us from upholding the unity of the LU AUNT. (Criterion II.2 confirms this decision: see (22)b, p. 00.)

The semantic difference in L is related to just one formal difference in L's behavior

(16) a. *He **wiped** his hands **with** a handkerchief* [rubbing the handkerchief against his hands].

vs.

b. *He **wiped** his hands **on** his pants* [rubbing his hands against the pants].

The semantic difference consists in what is rubbed against what; it is linked to just one formal difference—different prepositions used to express the corresponding actants (there is no other formal difference in the behavior of WIPE). Therefore, we are allowed to try a unified disjunctive definition:

X wipes Z with/on W = 'Person₁ X removes₂ liquid²₁ or₂ dirt₁ from Z by rubbing₁ W against Z or₂ Z against W'.

The GP of [to] WIPE indicates which preposition corresponds to which movement; otherwise, there are no other differences between the two uses of WIPE. (Criterion II.2 confirms the above decision: *He wiped his hands with a new towel and his feet on the mat.*)

The semantic difference in L is related to more than one formal differences in L's behavior

Consider the French verb VENDRE '[to] sell' used in the sense of '[to] prostitute oneself':

- (17) a. *Elle vendait ses faveurs* ⟨*ses charmes, son corps*⟩ *aux matelots ivres* ⟨*aux touristes étrangers, au premier venu*⟩
 lit. 'She was selling her favors ⟨her charms, her body⟩ to drunken sailors ⟨to foreign tourists, to anybody⟩'.

Now, is this the same verb VENDRE as that found in (17)b, or a different lexeme?

- b. *Ils vendaient des voitures d'occasion* ⟨*les services touristiques, les souscriptions*⟩ *aux gens du quartier*
 'They were selling second-hand cars ⟨tourist services, subscriptions⟩ to the people from the neighborhood'.

The verb VENDRE, as illustrated by (17)b, can be defined as follows:

- (18) a. $X \text{ vend } Y \text{ à } Z \text{ pour } W \equiv$ 'X gives to Z the right 1) of permanent possession of an entity Y^1 or 2) of obtaining service Y^2 —in exchange for money W'

VENDRE in (17)a shows a semantic difference with respect to this definition: the service Y^2 is not any service, but just 'having sex'; this semantic difference is correlated to a formal one, namely—for VENDRE in (17)a the actant Y must be lexically expressed by $A_{\text{poss}=X}$ FAVEURS, CHARMES, or CORPS.

Had this been the only formal difference in the behavior of the 'prostitution' VENDRE with respect to the 'normal' VENDRE, it still could have been covered by the definition (18)a, with the above indication added to its GP. But there are other formal differences:

- *la vente de ces voitures, de services touristiques* vs. **la vente de ses faveurs, de son corps*
- *la vendeuse de ces voitures, des services touristiques* vs. **la vendeuse de ses faveurs, de son corps*
- *Elle lui a vendu ses voitures.* vs. **Elle lui a vendu ses faveurs.*

- *Ses voitures ont été vendues.* vs. **Ses faveurs ont été vendues.*

Unlike ‘normal’ VENDRE, the ‘prostitution’ VENDRE denotes exclusively an activity, but not an action/an event; that is why it has neither the *passé composé* nor a passive. All that leads us, in conformity with Criterion II.1, to split VENDRE in (at least) two senses: one described by the definition (18)a and the other, by (18)b:

- b.** *X vend Y à Z pour W* ≡ ‘X gives to Z the right of obtaining from X sexual service Y in exchange for money W’

(Criterion II.2 buttresses this decision: **Elle vendait de vieux bouquins et ses faveurs* lit. ‘She was selling old books and her favors’.)

Still another example, dealing with differences in derivation, can be useful. Take Fr. ÉLEVER ‘educate, breed, cultivate’; its two uses are illustrated in (19):

- (19) **a.** *Toute sa vie, Jeanne a élevé des enfants* ‘Her whole life, Jeanne educated children’:

Jeanne is a teacher in a daycare institution; *élever* ≈ ‘educate’.

- b.** *Toute sa vie, Jeanne a élevé des cochons* ‘Her whole life, Jeanne bred pigs’:

Jeanne is a pig farmer; *élever* ≈ ‘breed’.

In (19)b the verb ÉLEVER has an action noun ÉLEVAGE (*Jeanne s’occupe d’élevage de cochons* lit. ‘J. does pig breeding’) and an agent noun ÉLEVEUR/ÉLEVEUSE: *Jeanne est éleveuse de cochons* ‘J. is a pig breeder’. But for (19)a, these derivatives are impossible: **Jeanne s’occupe d’élevage d’enfants* lit. ‘J. does child breeding’; **Jeanne est éleveuse d’enfants* lit. ‘J. is [a] child breeder’); here suppletive lexical derivatives can be used: INSTITUTEUR ‘teacher’, _JARDINIÈRE D’ENFANTS_ ‘daycare worker’, ÉDUCATION, FORMATION, etc.

Criterion II.1 allows for the distinction of still another sense of ÉLEVER:

- c.** *Toute sa vie, Jeanne a élevé du vin* lit. ‘Her whole life, Jeanne grew wine’:

Jeanne is a winegrower.

Even if the agent noun ÉLEVEUR/ÉLEVEUSE *de vin* is possible with this sense, there is no *ÉLEVAGE *de vin*. (Criterion II.2 also confirms the splitting of ÉLEVER: **Toute sa vie, Jeanne a élevé des chiens et des enfants*, **Toute sa vie, Jeanne a élevé du vin et des cochons*.)

As far as the example with [to] PAINT is concerned (see (14)), Criterion II.1 recommends the split:

- PAINT₁, but not PAINT₂: has an objectal-copredicative complement (*paint the room green/a bright color*); has an S₀ PAINTING₁; has an S_{res} PAINTWORK; has a non-standard LF REPAINT; etc.

- PAINT₂, but not PAINT₁: has a SemA representing the images, as in *sarcophagus painted with Homeric scenes*; has an S₁ ARTIST and an S₂ PAINTING₃, PICTURE; has for its SemA 4 such expressions as *in oils, in water colors*; has ‘alternations’ of the following types: 1) *paint Mr. Polgùere ~ paint the portrait of Mr. Polgùere* or 2) *paint an old church* = ‘represent an old church on a picture’ or ‘cover the walls of an old church with paintings’; has a semantic derivation HAND-PAINTED; etc.

3.2.3 Criterion II.2: Unifying Cooccurrence (= the Green-Apresjan Criterion)²³

A powerful means of testing the unity of a suspect lexical item L ‘... σ_1/σ_2 ...’ is tentative coordination of clause elements linked to its two hypothetical senses:

L—synt—[L’ and L’’].

If the result is a normal sentence, this constitutes a strong recommendation in favor of a single L; on the contrary, nothing prevents us from splitting L if the result is a zeugma (a pun, a wordplay of a particular kind, as in *She took [= L] a lover [= L’] and a huge risk [= L’’]* or *Physician: a person on whom we set [= L] our hopes [= L’] when ill and our dogs [= L’’] when well* [A. Bierce]). More formally:

A sentence that contains a suspect lexical item L ‘... σ_1/σ_2 ...’ is said to manifest unifying cooccurrence of L if and only if L is syntactically linked:

1) either simultaneously to two coordinated LUs L’ and L’’ such that L’ is semantically related to ‘ σ_1 ’ and L’’ to ‘ σ_2 ’,

2) or to an LU L’ that is semantically simultaneously related to ‘ σ_1 ’ and ‘ σ_2 ’ (which must be clear from the context).

Now Criterion II.2 can be formulated:

If it is possible to construct a sentence that manifests unifying cooccurrence for L, so that we have

$$L'(\sigma_1) \text{---} L' \dots \sigma_1 / \sigma_2 \dots \text{---} L''(\sigma_2)$$

or

$$L'(\sigma_1, \sigma_2) \text{---} L' \dots \sigma_1 / \sigma_2 \dots ,$$

then L should not be split; as a result, we have one LU L with disjunction in its definition:

$$L' \dots \sigma_1 \text{ or } \sigma_2 \dots '$$

Examples

First let me illustrate the two cases.

Case 1: the LU [*to*] COOL ‘become cooler or cool’ [*The gas cooled a bit but still was very hot; The cake should cool completely*]. The sentence manifesting unifying cooccurrence for [*to*] COOL is, for instance, (20):

(20) *The gas cooled first by only a few degrees [= L'], and then completely [= L'']*.

Case 2: the French LU BELLE-MÈRE ‘mother of the spouse or the wife of the father, who has replaced the deceased/divorced mother’ = ‘mother-in-law or step-mother’; it can be found in a sentence with unifying cooccurrence for BELLE-MÈRE:

(21) *Mes [= L'] deux belles-mères s'entendaient parfaitement*
lit. ‘My two mother-in-law/step-mother got along perfectly’.

Now I will provide more examples to better show the working of Criterion II.2:

(22) a. (i) BOMBARD ‘drop bombs [= ‘ σ_1 ’] or3 hurl heavy artillery shells [= ‘ σ_2 ’]—one verb or two?’

The following sentence is perfectly OK:

(ii) *In October 1944, Allied planes [= ‘L’] and three British cruisers [= ‘L’] bombarded the dykes in Walcheren, Zeeland, causing considerable flooding.*

Therefore, BOMBARD is described as one lexeme—with a disjunction in its definition.

- b. (i) For AUNT ‘sister of the mother or the father or the wife of an uncle’ Criterion II.2 also requires one definition with disjunction:
- (ii) *All my [= L'] aunts were there—the older sister of my mother, the three sisters of my father, and the pretty wife of Uncle Jim.*
- c. (i) What about the French verb FLAMBER ‘[to] blaze’ in the sentences *Son gosier flambait* lit. ‘His throat was blazing [= ‘burning’]’, *Son visage flambait* lit. ‘His face was blazing [= ‘flaming’]’, and *Ses yeux flambaient* lit. ‘His eyes were blazing’?

Criterion II.2 does not give a precise answer for sense 1 (‘feeling of burning’) with respect to two other senses:

- (ii) **Son gosier et son visage <ses yeux> flambaient.*

This means, however, that we are not forced to unite these two meanings under the same definition. Now, for senses 2 and 3 (‘have an abnormally red coloring’ and ‘have an abnormal shine’) a sentence with unifying cooccurrence is possible:

- (iii) *Son visage et ses yeux flambaient de fièvre*

lit. ‘His face and eyes were blazing with fever’.

Consequently, these two senses should be united under one lexeme with disjunction in the definition (see FLAMBERVII.b in Mel’čuk *et al.* 1988 [= DEC-2], contrasting with FLAMBERV).

The Criterion of Unifying Cooccurrence is known also in a different form (Lakoff & Ross 1976).

‘Do so’ Criterion

If it is possible to construct a sentence of the form *X V-es L, and so does X* in which the verb *V* involves ‘ σ_1 ’ and *does so* involves ‘ σ_2 ’,

then L should not be split; as a result, we have one LU L with disjunction in its definition:

L ‘... σ_1 or σ_2 ...’

It is convenient to have at one’s disposal this version of Criterion II.2, because in some cases the application of its ‘classical’ version can be blocked by unfavorable grammatical conditions. (In languages that, unlike English, do not have the DO SO construction, one can use a similar construction with a lexeme meaning ‘as well, just as’.) The existence of two versions of Unifying Cooccurrence

Criterion gives the lexicographer more flexibility. In all cases analyzed until now, if both versions of Criterion II.2 apply, they give the same results:

(21') *The gas cooled first only a few degrees [= L'], and then **did so completely** [= L'']*.

(22') **a'**. *In October 1944, Allied planes [= 'L'''] bombarded the dykes in Walcheren, Zeeland, and **so did** three British cruisers [= 'L'''], causing considerable flooding.*

(22') **c'**. (iii) *Son visage flambait de fièvre **tout comme** ses yeux*

lit. 'His face was blazing with fever, just as his eyes were'.

3.2.4 A Comparison of Criteria II.1 and II.2

Criterion II.1 is based on the structure and the contents of the hypothetical lexical entry for L, i.e., on lexicographic information that this entry supplies for L; this is an internal lexicographic criterion. It is aimed at **differentiating** two lexicographic senses; when its premise is satisfied, it gives the linguist a strong incentive to split the suspect L in two LUs.

Criterion II.2, on the contrary, is based on L's behavior in sentences; it is an external lexicographic criterion. It is aimed at **uniting** two semanteme configurations inside one disjunctive definition; when its premise is satisfied, it gives the linguist a strong incentive not to split the suspect L.

Both criteria are valid only in the positive sense; if their premise is not satisfied, they remain, strictly speaking, silent. (However, perhaps we should consider the negative result of a Criterion II application as at least a mild indication to the contrary?) The criteria must concur: if Criterion II.1 is positive, Criterion II.2 must be negative, and vice versa; in other words, they should not contradict each other. But what happens if they are in conflict? Criterion II.2 seems stronger, since it is more objective: it checks the actual behavior of L in the text, while Criterion II.1, being system-specific, concerns the coherence and elegance of the internal organization of the lexicographic description. However, given the purely functional nature of Meaning-Text modeling in general, systemic considerations play a crucial role in this approach. Therefore, until enough factual data is available, it is not possible to pass a general judgment on the comparative power of the two criteria. Yet it is useful to consider a particular case of their conflict: two LUs, BAKEL1 and

BAKEI.2, which are distinguished in spite of the recommendation of Criterion II.2, see 3.2.4, p. 00.

To conclude the discussion of LU-distinguishing criteria, let us consider a presumed conflict of Criteria II.1 and II.2 and demonstrate its successful resolution. The French verb PRENDRE ‘take’ in *prendre un médicament* ‘take a medication’ and *prendre une bière* lit. ‘take a beer’ (*prendre* is here an element of the value of the LF Real₁): one lexeme or two? In the first case, an action noun exists (*la prise d’un médicament*), but not in the second (**la prise de bière*). However, this is the only formal divergence between the two uses, and under Criterion II.1 we are allowed to disregard it—i.e., to treat it in the entry for a single lexeme as an exception. Criterion II.2 strongly recommends the unity of PRENDRE in these contexts, since it is possible to say *Tu prends trop de bière et de médicaments en même temps* lit. ‘You take too much beer and medication at the same time’ or *J’ai pris une aspirine et une bière* lit. ‘I took an aspirin and a beer’. As the final result, we have one PRENDRE in both cases, with a special constraint: $S_0 = \text{PRISE} \mid \mathbf{Y \text{ is medication}}$.

However, the conflict resolution does not always come so easily. Take the distinction of ‘potato’ BAKEI.1 vs. ‘bread’ BAKEI.2, proposed in the ECD sample in Subsection 5.1: it can in fact be questioned. English dictionaries do not draw this distinction, and Criterion II.2 recommends a unified description:

(23) a. *John bakes bread and cakes as well as potatoes, apples, ham and fish like a wizard.*

or

b. *For dinner, I’ll bake some potatoes and a fruitcake.*

Yet I believe that the distinction is valid and should be maintained, following instead Criterion I.1. There are significant differences in the GPs and in the two sets of LFs for both lexemes; if BAKEI.1 and BAKEI.2 were to be united under a disjunctive definition, the resulting lexical entry would be clearly separated in two disjoint parts: one for the ‘potato’ BAKEI.1, and another for the ‘bread’ BAKEI.2. (Conventional dictionaries escape from this problem simply because they do not supply all necessary lexicographic information.) Thus, Criterion I.1 turns out to be stronger.

My guess is that some restrictions must be imposed on Criterion II.2, although for the time being I am not sure which ones. Perhaps this could be a particular type of polysemy that is allowed to violate Criterion II.2? In this case, the polysemy ‘transformation T of something’ ~ ‘creation of a product by

transformation T of something' would allow for the coordination of two different lexicographic senses without producing a zeugma. Let it be stressed that the distinction between Change-of-State Verbs (like the 'potato' BAKEI.2) and Creation Verbs (like the 'bread' BAKEI.2) is in general very typical of English; cf., e.g., Miller *et al.* 1988: 202-206.²⁴

In any event, it is easier to lump together than to distinguish; should we prove wrong on the point of the sense distinction in question, there will be fewer problems in merging the two lexemes. Until some convincing argument one way or the other comes to light I proceed on the assumption that Criterion II.1 can be given priority.

3.3 Organization of an ECD Super-entry [= Vocable]

In an ECD, a superentry, or a vocable, is a structured collection of entries each of which deals with an LU. To briefly characterize the structure of an ECD superentry, the following two points have to be made: the vocable synopsis and the ordering of LUs within a vocable.

3.3.1 The Vocable Synopsis

A presentation of an ECD vocable begins with an introductory *synopsis*: a table of contents of the vocable, so to speak. The synopsis lists all the LUs of the vocable in question—in the order in which they are arrayed, identifying each one by a truncated version of the definition and by an example. (The truncated definition is supposed to be understandable even if incomplete.) This helps the user not only to find the lexeme he needs easier, but also to form a compact picture of the vocable as a whole. (Such synopses are not unknown in traditional lexicography; for instance, they are employed systematically in *Dictionnaire du français contemporain* and sporadically in LDoCE 1978—e.g., under MAKE.)

The vocable synopsis has, strictly speaking, no logical value, but it is very useful for the user and for the ECD lexicographer himself. (This is one of the pedagogical concessions made by the ECD.)

3.3.2 The Order of the LUs in a Vocable

The order of entries in a vocable is determined by the lexicographic numbers assigned to LUs, that is, word senses, constituting the vocable. As is the habit in all

conventional dictionaries, the senses in an ECD are ordered in such a way as to reflect their semantic proximity. To illustrate this, we will consider the ordering of LUs within the vocable *BAKE*, given in 5.1, p. 00ff; to put it differently, we will explain the adopted numbering of particular *BAKE* lexemes.

1) *BAKE*_I, *II*, *III* are ordered in this way since in contemporary English, *BAKE* is primarily a verb of cooking.

— This belief is based not on frequency but on psychological salience: a decontextualized sentence *They are baking* is likely to be understood as referring to cooks/foodstuffs rather than to bricks/pottery or sweltering sunbathers.

— ‘Pottery’ *bake* is closer to ‘cooking’ *bake* than is ‘sunbathing’ *bake*: in fact, ‘pottery’ *bake* shares with ‘cooking’ *bake* the semantic component ‘enclosed space’, which ‘sunbathing’ *bake* does not. Even more importantly, ‘pottery’ *bake* represents a specialization of ‘cooking’ *bake*, while ‘sunbathing’ *bake* is of course a metaphor.

— Having made ‘cooking’ *bake* *BAKE*_I, we are forced into describing ‘pottery’ *bake* as *BAKE*_{II}, and ‘sunbathing’ *bake*, as *BAKE*_{III}. The decision to put ‘cooking’ *bake* as *BAKE*_I has the additional advantage that it makes more perspicuous the parallelism between *BAKE* and other ‘cooking’ verbs (such as *ROAST*, *FRY*, *BOIL* or *STEW*)—all of which must have their ‘cooking’ lexemes first for the same reason.

2) Within *BAKE*_I, ‘potato’ *BAKE* [= *BAKE*_{I.1}] is ordered before ‘bread’ *BAKE* [= *BAKE*_{I.2}]. Traditionally, their difference is correctly described as that between an affected-object verb (Causing-Change-of-State Verb) and an effected-object verb (Creation Verb), cf. Atkins *et al.* 1988: 87, Miller *et al.* 1988: 202ff. Lexicographically, this means that in *BAKE*_{I.1} the food actant Y has the same name it had before it was baked: e.g., a baked potato is called a potato still; whereas in *BAKE*_{I.2} the food actant Y has the name it obtains only after baking: *bread* rather than **baked dough*, *cake* rather than **baked batter*. The affected-object *BAKE*_{I.1} is semantically poorer than the effected-object *BAKE*_{I.2} because the former describes causing a change of state of the same thing while the latter describes causing a change of one thing into another. (That is why *BAKE*_{I.2} has an extra actant with respect to *BAKE*_{I.1}—namely, the actant that corresponds to the created thing.)

3) Within *BAKE*_{I.1}, *I.2* and *II.1*, the transitive *BAKE*_{I.1a}, *I.2a* and *II.1a* are ordered before the intransitive *BAKE*_{I.1b}, *I.2b* and *II.1b*, because the three transitive **a**-*BAKE*

are semantically simpler than their intransitive counterparts: ‘bakeI.1b’ includes the component ‘being bakedI.1a’, and so on. (BAKEI.1 and BAKEI.2 are not completely symmetrical in this respect, see 5.2, Item 1, p. 00.)

4) ‘Oven/kiln’ BAKE [= BAKEI.1c, I.2c and II.1c] are placed after the intransitive BAKEI.1b, I.2b and II.1b, because they are semantically even more remote from BAKEI.1a, I.2a and II.1a: the instrument of baking appears as their syntactic subject, and they describe properties of this instrument rather than actions/events. They constitute the ‘instrumental’ senses of the action transitive verbs BAKEI.1a, I.2a and II.1a (see the discussion in 4.6.2, p. 00; the fact that this phenomenon is not fully predictable grammatically justifies our isolating these three BAKE as separate lexemes).

As a result of the ordering of LUs of a vocable, we obtain the *basic* LU of the vocable: the LU that comes first because all the other LUs of this vocable are, in a sense, ‘derived’ from it—that is, they refer to it, one way or another. In the BAKE vocable, the basic LU is BAKEI.1a (the ‘potato’ transitive BAKE).

4 Principles for Compiling the ECD

All dictionaries try to be logical and consistent in the organization of their entries, in the selection of their lexical stock, in their presentation, etc. However, the ECD is probably the first dictionary that raises the logical rigor and consistency of its articles to the status of an absolute law. An ECD lexicographer is obliged to stick to the policy of zero tolerance towards violations of this law, even if the consequence is additional complexity of the description. The requirement of logical rigor and consistency can be expressed more concretely in the form of the following nine principles underlying the work of compiling an ECD:

- The Formality Principle (4.1)
- Two Coherence Principles (4.2)
- Two Uniform Treatment Principles (4.3)
- The Internal Exhaustivity Principle (4.4)
- Two Maximal Generalization Principles (4.5)
- No Regularly Produced LUs Principles (4.6)

The topics raised by these principles have already been touched upon, one way or another, but it seems useful to review them in a systematic way.

4.1 The Formality Principle

I have already said that anything stated within the framework of an ECD is formal. This actually means that lexicographic descriptions in an ECD have the following two characteristics:

- All lexicographic statements in an ECD must be
- 1) written in a pre-established metalanguage,
 - and
 - 2) completely explicit.

Lexicographic metalanguages

An ECD-style lexicographic description is carried out in a pre-established formal metalanguage—or, to be more precise, in several specialized metalanguages (dealing with semantics, syntax, lexical cooccurrence, etc.). Of course, all existing dictionaries use some kind of lexicographic metalanguage, but as a rule, this metalanguage is limited to morphology (declension and conjugation types), as well as usage labels; where this is called for, a traditional dictionary also formalizes the presentation of the pronunciation (phonetic transcription). With a few exceptions,²⁵ the meaning, the syntactic behavior and, especially, the restricted lexical cooccurrence of the head word are not described by means of a precise and sufficiently rich metalanguage. In sharp contrast, the ECD puts the emphasis on formal metalanguages sufficiently expressive to cover the semantic description, i.e., the definition, of the head word *L*, its syntactic active valence description, i.e., *L*'s Government Pattern, and the description of *L*'s semantic derivations and restricted lexical cooccurrence (Lexical Functions). All specialized ECD metalanguages now in use are specified by strict formation rules; taken together, they allow the lexicographer to describe all observed lexicographic phenomena.

Complete explicitness

An ECD-style lexicographic description is fully explicit: nothing is left to the user's intuition. Thus, a French ECD cannot define the noun *MAGAZINE* as 'usually illustrated periodical publication', as does PR 2001: this definition does not distinguish magazines, on the one hand, from illustrated newspapers and journals, on the other hand. A magazine is different from an illustrated newspaper in that its pages are smaller and attached together (unlike those of a newspaper); it is different from a review/journal in that it is designed to entertain (cf. **magazine*

mathématique ‘mathematical magazine’ vs. *revue mathématique* ‘mathematical review’). These two defining features of magazines—book format and entertaining character—must be made explicit in an ECD definition:

Fr. MAGAZINE ≡ ‘(illustrated)²⁶ periodical designed to entertain and having book format with soft cover’.

Moreover, a French ECD cannot illustrate the article for MAGAZINE by a single example *magazine féminin* lit. ‘feminine magazine’, because this example is not explicit enough:

- In principle, the expression *magazine féminin* could mean ‘magazine dedicated to women’s issues’ (cf. *magazine sportif* ≡ ‘... dedicated to sports issues’) or ‘magazine addressed to women’ (cf. *magazine pour enfants* ≡ ‘... addressed to children’); in actual practice it means only ‘magazine addressed to women’ = ‘women’s magazine’, and an ECD must state this fact explicitly.

- An ECD must also somehow indicate that one cannot say by analogy **magazine enfantin* lit. ‘infantile magazine’ = ‘magazine addressed to children’: the only correct expression is *magazine pour enfants* lit. ‘magazine for children’, while for the meaning ‘women’s magazine’ both expressions are possible: *magazine féminin* and *magazine pour femmes*.

- An ECD must also indicate that we have [*magazine*] *politique/sportif* pour dire ‘dedicated to politics/sports issues’.

Considering these examples leads the linguist to the following description:

Fr. MAGAZINE
magazine au sujet de Y pour Z ≡ ‘(illustrated) periodical designed to entertain the public Z, (dedicated to subject Y)²⁷ and having book format with soft cover’.

We can observe here the necessity of variables in the definition, since MAGAZINE turns out to be a quasi-predicate: although it denotes a physical object, its signified presupposes semantic arguments, or Sem-Actants. Once these are introduced, we have to specify the possible ways of expressing them:

X = *politique, sportif, de cinéma* (<**cinématographique*>), *de théâtre* (<**théâtral*>), *humoristique, de mots croisés, ...*

Y = *féminin/pour femmes, pour enfants, pour les jeunes, ...*

Of course, the statements about the expression of the Sem-Actants must be presented in a special formal metalanguage—in a GP, see 2.2, p. 00ff.

The reader thus can see to what extent the requirement of complete explicitness is productive: it pushes the linguist to find and present substantial amounts of information that otherwise may escape his attention.

4.2 Coherence Principles

The ECD aims at complete coherence in two aspects:

- between the elements of a lexical entry—that is, coherence inside an LU’s description;
- between semantically related lexical entries, that is, coherence inside a semantic field.

Consequently, the lexicographer must follow two coherence principles.

4.2.1 Lexical Unit Internal Coherence Principle

Logical rigor at the level of the microstructure of the dictionary (= within a particular lexical entry) entails the following principle:

In an ECD lexical entry, the semantic, syntactic and cooccurrence descriptions of the head LU L should be in complete agreement.

By ‘agreement’ I understand here mutual correspondences, explicitly indicated, between semantic components in the definition of L, its actantial syntactic dependents (that is, L’s Deep-Syntactic actants) and its semantic derivations and restricted lexical cooccurrences. The problem of correspondences between the meaning of L and its syntactic actantial pattern has been actively explored in linguistics, where it is known as *linking* (see, e.g., Levin & Rappaport 1995 and 2005). However, lexicography has not incorporated many valuable ideas and findings of these studies. The situation is even worse as far as correspondences between the meaning of L and its restricted lexical cooccurrence are concerned.

Therefore, to clarify the idea of internal ‘agreement’ in an ECD entry, let us consider an example of links between L’s meaning and its lexical cooccurrences. The French noun CÉLIBATAIRE_N (masculine gender) can be tentatively defined as follows:

CÉLIBATAIRE_{N(masc)} ≡ ‘man₁ who is²₃ not and has never₁ been²₃ married₁’ = ‘bachelor’.

The definition is quite OK, except that it does not account for two restricted lexical cooccurents: the adjectives VIEUX ‘old’ and ENDURCI ‘hardened’.

The first problem with these adjectives is that they apply to men only, although the noun CÉLIBATAIRE_N used in the feminine gender can refer to a woman.²⁸

(24) a. *Cette célibataire de 36 ans dirige trois ateliers*
 ‘This single woman of 36 manages three workshops’.

b.

Pierre est un $\left\{ \begin{array}{l} \text{vieux célibataire} \\ \text{célibataire endurci} \end{array} \right\}$ ~ **Marie est une* $\left\{ \begin{array}{l} \text{vieille célibataire} \\ \text{célibataire endurcie} \end{array} \right\}$

‘Pierre is $\left\{ \begin{array}{l} \text{an old} \\ \text{a confirmed} \end{array} \right\}$ bachelor’. ~ *‘Marie is $\left\{ \begin{array}{l} \text{an old} \\ \text{a confirmed} \end{array} \right\}$ bachelor’.

The solution here is straightforward: CÉLIBATAIRE_{N(fem)} is a different lexeme, and both of the collocates above must be mentioned in the entry for CÉLIBATAIRE_{N(masc)} only. (For a discussion of the masculine ~ feminine noun pairs in French as different lexemes, see Mel’čuk 2000.)

The second problem is that these adjectives are perceived as intensifiers: *vieux célibataire*, *célibataire endurci* ≈ ‘très célibataire’ lit. ‘very bachelor’. But the above definition of CÉLIBATAIRE_{N(masc)} does not have a component ready to accept the intensification. Thus, *vieux* intensifies the period of time during which X has remained *célibataire*, but there is no time component in the definition. It is even worse for *endurci*: what is characterized by ‘very’ when you say *célibataire endurci*?

The LU Coherence Principle does not allow us to brush aside this problem—something must be done about it. Logically, two solutions are possible.

- Solution I: we introduce still another lexeme, CÉLIBATAIRE_{N(masc)}², whose definition is adapted to the two adjectives. Namely, it could be as follows:

‘man₁ who is²₃ not and has never₁ been²₃ married₁ for a considerable period of time and who wants¹₁ to be²₃ not married₁’.

The formation of a feminine counterpart would not be allowed for CÉLIBATAIRE_{N(masc)}². And the adjectives VIEUX and ENDURCI will be compatible with CÉLIBATAIRE_{N(masc)}² only; the component ‘for a considerable period of time’

will accept the intensification by *vieux*, and the component ‘want’ will accept that by *endurci*. However, this solution is flawed, since the noun CÉLIBATAIRE_{N(masc)} itself—that is, when it is taken alone, does not carry the above meaning. Sentences (25) constitute a proof of this:

- (25) a. ?*Pierre ne l’épousera jamais, c’est un célibataire*
 ‘Pierre will never marry her, he is a bachelor’.
- b. *Pierre a été le célibataire le plus convoité, mais pour très peu de temps — il a épousé Marie presque tout de suite*
 ‘Pierre was the most coveted bachelor, but for a very short time—he married Mary almost immediately’.
- c. *Vrai, c’est encore un célibataire, mais il est à la recherche d’une épouse*
 ‘True, he is still a bachelor, but he is looking for a wife’.

Note also that CÉLIBATAIRE_{N(masc)} is used as a technical term to describe the family status of a man.

We are forced to conclude that CÉLIBATAIRE_{N(masc)} does not contain the semantic components ‘for a considerable period of time’ and ‘... who wants to remain unmarried’. The adjectives VIEUX and ENDURCI cannot thus be simple intensifiers of CÉLIBATAIRE_{N(masc)}, bearing semantically on some components within its definition, because the latter does not have components to be intensified. As a result, Solution II must be preferred.

• Solution II: we keep one lexeme CÉLIBATAIRE_{N(masc)} and describe the two adjectives under discussion in its lexical entry as follows:

Magn^{time}_['has been'] : vieux | antepos

who wants to remain C., Magn : endurci | postpos

Explanations

1. The superscript ^{time} with the name of the LF Magn indicates the semantic nuance: that is time that is intensified; the subscript _['has been'] identifies the semantic component in the definition of CÉLIBATAIRE_{N(masc)} that accepts the intensification.
2. The vertical bar | separates the LU presented as an element of the value of the LF from the conditions of its use.
3. The abbreviations ‘antepos’/‘postpos’ specify the obligatory anteposition/postposition of the adjective: **célibataire vieux*, **endurci célibataire*.

For VIEUX, we can use the standard LF Magn supplied with a superscript and a subscript. The superscript ^{time} indicates explicitly that what is intensified is duration; and the subscript [_{has been}] indicates the duration of what is meant. The result is that an ECD explicitly specifies that this collocation denotes a man who is not married and has never been, for a long time.

For ENDURCI, a different technique is used: a non-standard LF, which has in its description the necessary additional meaning (‘wanting to remain unmarried’).

Interestingly, the things are different with the Russian equivalents of *célibataire endurci*: *ubežděnnij* lit. ‘convinced’, *zakorenelyj* lit. ‘well-rooted’ *xolostjak* ‘bachelor’. What is crucial, the noun XOLOSTJAK cannot be used in Russian to describe the family status in a neutral way, that is, e.g., in an official document: for this, you have to use the adjective XOLOST ‘single’ or NEŽENAT ‘unmarried’. Therefore, its definition is different from that of CÉLIBATAIRE_{N(masc)}:

XOLOSTJAK = ‘man₁ who is²³ not and has never₁ been²³ married₁ (and who is²³ accustomed₂ to and wants¹ to be²³ not married₁)’

With such a definition, ZAKORENELYJ and UBEŽDĚNNYJ are obvious intensifiers: the first intensifies the component ‘accustomed [to]’, and the second, the component ‘want’. Both adjectives are elements of the value of the standard LF Magn, but for each of them the targeted component of the definition must be indicated:

Magn_[‘accustomed’] : zakorenelyj [‘very accustomed to being unmarried’]

Magn_[‘want’] : ubežděnnij [‘seriously wanting to continue unmarried’]

Another adjective that often combines with XOLOSTJAK is STARYJ ‘old’; it is also an element of Magn, but it intensifies the period of time during which X has not been married, which can be presented in the same way as in French:

Magn_[^{time}_{has been}] : staryj [‘having been unmarried for a long time’]

These examples show to what extent the restricted lexical cooccurrence of L must be ‘dovetailed’ with its definition.

Other LUs may require still other treatment. Returning to French, let us consider the collocation [*un*] *grand blessé* ‘[a] seriously injured person’. The meaning the adjective GRAND lit. ‘big’ has in this expression should not be described the same way we proposed for VIEUX and ENDURCI with CÉLIBATAIRE_N.

The expression *grand blessé* means ‘injured living person [= *blessé*_N] whose wound [= *blessure*] does much harm to his health’; here the semantic contribution of the adjective GRAND is just ‘much’ = ‘very’, the rest—‘[injury] does harm to his health’—being part of the definition of the noun BLESSURE ‘wound’, which is part of the definition of BLESSÉ_N:

BLESSÉ_N ≡ ‘living¹₁ person₁ who has²_a one¹₁ or³ more¹₁ injuries₁ [= *blessure(s)*].’

[The component ‘living’ is necessary here since a cadaver with even the worst injuries cannot be called *un blessé*.]

The adjective GRAND is here a real intensifier, so that the current definition of BLESSURE ‘wound’—‘visible lesion inflicted on the tissues of a living being by an external agent’ [adapted from PR 2001]—has to be modified. The Principle of Internal Coherence requires us to include in the definition a component capable of being intensified: ‘... and which harms the health of the being’. The resulting definition of BLESSURE ‘wound’ then reads as follows:

BLESSURE

blessure de X par Y ≡ ‘visible¹₁ lesion₁ which has been inflicted on the tissues₁ of the living¹₁ being¹₃ X by an external₁ agent₃ Y and which harms²₁ the health₂ of X’.

[A cadaver can have many *blessures*, yet they must have been inflicted on a living being.]

The intensifiable component ‘[to] harm²₁’ allows us to account, in a natural and systematic way, for the collocates of BLESSURE, such as [*blessure*] *grave* ‘grave’ ⟨*sérieuse* ‘serious’, *mortelle* ‘mortal’, *légère* lit. ‘light’ = ‘minor’, ...). The same component, inherited by BLESSÉ (adjective and noun), accounts for its collocates: [*blessé*] *grave* ‘grave’ ⟨**lourd* ‘heavy’⟩ vs. [*blessé*] *léger* ‘light’; cf. *grièvement* ‘seriously’ *blessé* ⟨**malade* ‘ill’⟩ vs. *gravement* ‘gravely’ *malade*, etc.

In the entry for BLESSÉ_N, the adjective GRAND appears as an element of the value of the standard LF Magn:

Magn[‘injuries’] : grand | antepos

Summing up: The Internal Coherence Principle requires that within a lexical entry all elements be ‘well tuned’ to each other; it does not allow us to have collocates that do not perfectly fit the definition. (See a discussion of this problem in Iordanskaja & Polguère 2005.)

As we have seen, simultaneous and comparative processing of semantic, syntactic and lexical-cooccurrence data, driven by the quest for coherence, gives interesting results concerning all three types of lexicographic information. Sometimes, the lexicographer has to change the starting definition—such is the case for the noun BLESSURE in our example. In other cases, he chooses to change the description of particular cooccurents by particular LFs; we have done so for VIEUX and ENDURCI (in the entry for CÉLIBATAIRE_{N(masc)}), and for UBEŽDĚNNYJ and ZAKORENELYJ (in the entry for XOLOSTJAK).

Of course an ECD lexicographer has to deal with many completely different and more complex cases—for instance, the correspondence between the definition of L and the inventory of L's syntactic actants. Thus, one necessarily *writes* in a language; therefore, the semanteme 'language' must appear in the definition of WRITE. Therefore, the denomination of a language that modifies the verb WRITE in a sentence must be considered to be the expression of L's semantic actant (and, consequently, syntactic actant). Now, the genuine actants feature a rather idiomatic, constrained, irregular behavior, but the phrase of the type *in English* with WRITE is absolutely regular and is used freely (*She submitted her thesis in French; He gives his classes in Hebrew; Leo made his declaration of love in excellent Catalan*). The question arises as to whether IN ENGLISH, IN FRENCH, etc. are really expressions of a SemA of WRITE. According to the definition of Semantic Actant (Mel'čuk 2004a: 38), they are. However, in order to be certain, we have to develop numerous lexical entries—for WRITE and dozens of related verbs (EXPOUND, PRESENT, DESCRIBE, ...), as well as for the phrases of the form *in X* (where X is a language name). For the time being, it is better to remain faithful to our principles—that is, to give absolute priority to logic and to the available definitions; as a consequence, we will consider that the language name is a Sem-actant of the verb WRITE and its semantic relatives.

4.2.2 Semantic Field Coherence Principle (= Lexical Inheritance)

The communicatively dominant node of the definition of the LU L is the semanteme of another LU L'; as a rule, L 'inherits' not only the semantic properties of L', but also its syntactic and lexical-cooccurrence properties (at least to some

extent). This phenomenon, known as ‘lexical inheritance’ (Mel’čuk & Wanner 1996) is captured by the following principle:

|| In an ECD, LU L’s entry should be in complete agreement with the entry for L’ that expresses the communicatively dominant node in The definition of L.

This principle forces the lexicographer, first of all, to systematically check that all the Sem- and DSynt-Actants of L’ are inherited by L. This does not of course mean that L must necessarily feature all the Sem-/DSynt-Actants of L’ separately: some of them can coincide with L’s own Sem-/DSynt-Actants, or become saturated (i.e., turned into generic constants). Then the restricted lexical cooccurrence of L should be systematically compared to that of L’: much of it can be also inherited. Although the inheritance of the values of LFs seems to be not very consistent and regular (cf. again the results of Mel’čuk & Wanner 1996), an ECD lexicographer is supposed to go through all semantic derivations and collocations of L’ to harmonize them with what L has.

4.3 Uniform Treatment Principles

Uniform treatment of lexical material is required in an ECD in two respects:

- for all LUs within the same semantic field; and
- for all vocables within the same lexical field.

Let us consider them in turn.

4.3.1 Lexical Unit Uniform Treatment Principle

In an ECD, descriptions of semantically related LUs must be carried out in the same—or at least in a parallel—way. For instance, the entries for nouns MAGAZINE, REVIEW, JOURNAL and NEWSPAPER must be in complete agreement as to the content and organization of lexicographic data supplied, while any manifestation of disagreement has to be explained and justified. (Otherwise, a disagreement reflects a mistake.) This requirement is expressed as follows:

|| All LUs belonging to the same semantic field must be described together and similarly to the extent that the language L allows.

In spite of its obvious character this principle is not properly observed in existing dictionaries.²⁹ Consider, as an illustration, the description of nationality names in PR 2001.

— [Un] ALLEMAND ‘[a] German’ has no special entry; however, the entry for the adjective ALLEMAND contains a division marked enigmatically as follows: \diamond **n**. *Les Allemands* [note the plural!—IM]; no definition is proposed, nor is the indication of the feminine given.

— [Un] CHINOIS ‘[a] Chinese’ is defined [sense II.1] as ‘person living in or being native of China’, which is obviously false (there are many foreigners living in China); note the singular here. The feminine, on the contrary, is presented: *Une Chinoise*.

— [Un] ESPAGNOL ‘[a] Spaniard’ also has a special entry [sense 2 under the adjective ESPAGNOL], but without any information in it; it is also in the singular; the feminine is indicated in a strange way: *Un, une Espagnole*, which is, strictly speaking, incorrect (**un Espagnole*).

— [Un] FRANÇAIS ‘[a] Frenchman’ is defined [sense 2] in the singular as ‘person of French nationality’, which is again inaccurate: *un Français* should be defined as ‘person of French nationality OF MASCULINE SEX’ (a woman cannot be called *un Français*). But the indication of the feminine is this time correct: *Un Français, une Française*.

— [Un] RUSSE ‘[a] Russian’ has an entry, which is empty: it contains neither definition nor indication of the feminine.

There is no point in continuing with such examples: it seems that PR 2001 does not have two nationality names with fully parallel descriptions. But in an ECD such treatment is unthinkable: all nationality names should be described in an identical way. More specifically, they must be described in the plural³⁰ with the reference to the corresponding geographical place and the mother tongue, for instance:

[les] ALLEMANDS :	‘nationality2 native to Germany	whose mother tongue is German’
[les] CHINOIS :	‘nationality2 native to China	whose mother tongue is Chinese’
[les] ESPAGNOLS :	‘nationality2 native to Spain	whose mother tongue is Spanish’
[les] FRANÇAIS :	‘nationality2 native to France	whose mother tongue is French’
[les] RUSSES :	‘nationality2 native to Russia	whose mother tongue is Russian’

For each nationality name, the masculine and the feminine forms should be indicated in the same explicit way, for instance: *un Allemand, une Allemande; un Chinois, une Chinoise*; etc.

This is no more than a rough sketch: we have yet to define the terms ‘nationality2’, ‘native [of N]’ and ‘mother tongue’.³¹ In addition, our definition schema does not foresee non-prototypical cases: for instance, how to describe a Chinese born in France to Chinese parents, but who does not speak a word of Chinese? Or a Jew born in Russia? The idea is simply to show how one can ensure the uniformity of the lexicographic descriptions of related LUs in an ECD.

NB: The requirement of uniformity of LU lexicographic descriptions does not mean that ALL related LUs should be described in an absolutely identical way. This is of course impossible: thus, the noun [*les*] SUISSES cannot be defined as *‘nationality2 native of Switzerland...’; the good definition in this case is [*un*] SUISSE = ‘male person native to [or: citizen of?] Switzerland’. Similarly, the noun [*les*] BERBÈRES has to be defined as ‘nationality2 whose mother tongue is Berber’ (without specifying the country of origin), etc. The Lexical Unit Uniform Treatment Principle imposes uniformity only where it is possible and based in linguistic reality.

It is clear that only compiling a dictionary by semantic fields can guarantee the validity of the LU Uniform Treatment Principle. Even if the final product—a printed ECD—uses an alphabetical arrangement of entries, this is no more than a concession to the convenience of consultation. An ECD is a dictionary based on semantic fields, and the concept of the semantic field entails the obligation, on the part of an ECD lexicographer, to determine, for each semantic field $\mathbf{F}_{\sigma}^{\text{sem}}$ processed, the generalized schema, or standard format, for lexicographic description of the LUs belonging to $\mathbf{F}_{\sigma}^{\text{sem}}$. For nationalities, considered above, such a schema for the definition could resemble the following:

les $L_{(\text{nationality}2)}\text{-s}$ = ‘nationality2 native of the country ... and whose mother tongue is ...’

Using general schemas is of course not limited to the lexicographic definition: generalized schemas are valid for all parts of a lexical entry (in particular, for the Government Pattern and Lexical Functions). Such generalized schemas belong to the lexicographic metalanguages we have been discussing above.

4.3.2 The Vocabal Uniform Treatment Principle

The most homogeneous treatment possible is necessary not only for all the LUs belonging to the same semantic field. The ECD requires generalized schemas for lexicographic description even at a higher level: namely, for a uniform treatment of vocables belonging to the same lexical field. Let me first define the notion of lexical field and then introduce the Vocabal Uniform Treatment Principle.

Definition 8: Lexical Field

A lexical field \mathbf{F}^{lex} is the set $\{\tilde{L}_i\}$ of vocables such that their basic LUs L_i^1 belong to the same semantic field.

Now the principle itself can be readily formulated:

Vocable Uniform Treatment Principle

Two vocables belonging to the same lexical field must be presented, all else being equal, according to the same schema: the related LUs of either vocable should be described in a parallel fashion:

- (i) their definitions must be formulated as similarly as possible;
- (ii) they must appear in the same order within each vocable;
- (iii) the semantic distances between them must be represented as similarly as possible (i.e., by the same or almost the same means).

As a good example, I can mention the general schema for lexicographic description of French vocables in the lexical field «BODYPARTS» (Jumarie & Iordanskaja 1988).

4.4 The Internal Exhaustivity Principle

Like all big conventional dictionaries, the ECD also strives to describe the lexical stock of \mathbf{L} as fully as possible—that is, to describe all the LUs of \mathbf{L} known at this particular moment. This goal can be qualified as ‘external’ exhaustivity. However, given the complexity of its lexical entries, an ECD cannot compete in this respect with normal dictionaries.³² The ECD’s main struggle is along a different axis: that of ‘internal’ exhaustivity, which concerns the description of an LU rather than of \mathbf{L} ’s whole lexical stock.

In an ECD, the lexical entry for L must contain all lexicographic data concerning L that are necessary for two goals:

- 1) to utilize L correctly in any possible context
- and
- 2) to find any other LU L' which is semantically linked to L .

Thus, the entry for the French LU (= quasi-idiom) `_MOYEN DE TRANSPORT_` ‘transportation means’ must contain:

- the names of all existing means of transportation (on earth: VOITURE ‘car’, CAMION ‘truck’, AUTOBUS ‘bus’, AUTOCAR ‘tourist bus’, TROLLEYBUS ‘trolley bus’, TRAMWAY ‘streetcar’, MÉTRO ‘subway’, ...; on water: BATEAU ‘ship’, RADEAU ‘raft’, ...; in the air: AVION ‘airplane’, HÉLICOPTÈRE ‘chopper’, FUSÉE ‘rocket’, ...);
- numerous LUs such as CARGAISON ‘load, cargo’, _TITRE DE TRANSPORT_ ≈ ‘ticket’, BILLET ‘ticket’, VOYAGEUR ‘passenger’, CORRESPONDANCE ‘transfer_N’, RÉSEAU ‘network’, DESSERVIR ‘serve’, TRANSPORTER ‘transport_V’, VOYAGER ‘travel_V, ride_V’, COMPOSTER ‘punch_V, stamp_V [the ticket]’, ... and even RESQUILLER ‘fare-jump’.

As another example, let us consider animals. For VACHE ‘cow’ an ECD gives MEUGLER ‘[to] moo’ and MEUH ! ‘moo!’; for CANARD ‘duck’, it offers CANCANER ‘[to] quack’ and COIN-COIN ! ‘quack!’. In the same vein, under each LU the ECD has to list all relevant interjections: for DÉGOÛT ‘disgust’—POUAH ! ‘yuck!’ or BEURK ! ‘yuck!’; for PLAISIR ‘pleasure’—AH ! ‘mmm!’ or MIAM ! ‘yum-yum!’; for DOULEUR ‘pain’—AÏE ! ‘ow!’, AÏE-AÏE-AÏE ! ‘ow!’, OUILLE ! ‘ouch!’, etc.

Moreover, for interjections the ECD must specify the exact prosody, since Fr. AH ! of pleasure and AH ! of amazement are not pronounced in the same way. Even if this means very long and very complex entries, the ECD has to supply for an LU L, all the LUs {L_i} that the speaker might need one day.

4.5 Maximal Generalization Principles

As is typical of all scientific work, the ECD places high value on capturing generalizations. In practice, this means that recurrent lexicographic information should be ‘factored out’ as much as possible and specified only once. In a printed version of the dictionary this must of course be done with great caution, in order not to render the consultation of the dictionary too cumbersome. This general requirement can be concretized in the form of the following two principles.

4.5.1 The Vocabular Generalization Principle

Lexicographic information valid for all LUs of a vocable should be extracted from individual LU entries and transferred directly to the vocable name (thus, it will be stated only once, for the whole of the vocable).

This principle is more or less regularly observed in conventional dictionaries: it concerns, as a rule, such data as the pronunciation/spelling, the part of speech and the inflectional type of the LUs within the vocable; I need not dwell on it.

In parallel to the above principle, an ECD uses another compacting technique: if some LFs of an LU L_2 have the same values as the LFs of another LU L_1 in the same vocable, a cross-reference is used, instead of repeating the values several times. This is done in the following way:

L_2

...

LFs $f_1, f_2, \dots : \uparrow L_1$

(That is, “For the values of the LFs listed here go to the entry for L_1 .”)

4.5.2 Semantic Field Generalization Principle

Lexicographic information valid for all LUs of a semantic field F^{sm} should be extracted from individual LU entries belonging to F^{sm} and transferred directly to the entry for the LU that is the name of F^{sm} .

This principle, which reflects Lexical Inheritance, concerns mostly LFs (and, to a lesser extent, GPs). Thus, the name of any particular illness Y normally collocates with the following expressions:

be stricken with Y, battle Y, get over Y, recover from Y, come down with Y, succumb to Y, etc.

And of course it allows all the semantic derivations characteristic of an illness in general, such as MEDICINE, DOCTOR, NURSE, HOSPITAL, MEDICATION, PAIN, BED, etc. All these collocates and semantic derivatives should not be repeated under PNEUMONIA, TYPHOID, CHOLERA, FLU, AIDS, SYPHYLIS, etc. They can be given only once, under ILLNESS; in the entry for a particular illness name, only the corresponding pointer is inserted, for instance:

PNEUMONIA

...

IncepOper₁, AntiReal₁, try.to.Liqu₁Func₀, ...: \uparrow ILLNESS

Two complications can be encountered in this connection:

— The specific illness name L has some derivatives or collocates that ILLNESS does not have. In this case, the corresponding expressions are simply given in L 's

entry; when using the entry, they are added to the elements provided by the ILLNESS entry.

— The specific illness name L does not have some derivatives or collocates that ILLNESS and (some) other LUs of the semantic field under description have. Then in L's entry these elements must be explicitly indicated as inadmissible.

(I will not discuss here some more specific cases; the techniques foreseen by semantic field generalization are presented in detail in Mel'čuk & Wanner 1996.)

4.6 The No Regularly Produced LUs Principle

If an LU L' of **L** is related to another LU L in a completely regular way, that is, the lexical entry for L' can be computed by general rules from L's lexical entry,
then L' should not be explicitly entered in an ECD.

I will comment on two types of such cases: regular derivation and regular polysemy.

4.6.1 Regular Derivation

Suppose that **L** possesses a derivateme 'δ' that, when applied to an LU L, always changes the lexical entry for L in the same way. Then L', derived by the application of 'δ' to L, should not appear in the dictionary with its own lexical entry; it is enough to indicate in L's entry that for L this derivation is possible. As an example, consider names of female Xs in French: *étudiant* 'student' ~ *étudiant+e* 'female student', *baron* 'baron' ~ *baronn+e* 'female baron', *directeur* 'director' ~ *directric+e* 'female director', *cuisinier* '[a] cook' ~ *cuisinière+e* '[a] female cook', [un] *Belge* '[a] Belgian' ~ [une] *Belge* '[a] female Belgian', etc. This derivation is semantically absolutely regular: except for the component 'of the feminine sex', the lexical entry for the derived noun has no differences with respect to the entry for the starting noun. However, it is not completely regular formally: some nouns do not have a female counterpart, which would be semantically quite plausible—such as *docteur* 'doctor', *agent* 'agent', *auteur* 'author' or *écrivain* 'writer'. Therefore, a French ECD must indicate for a noun L referring to a person that it has a 'female' derivation L' and the form thereof; this can be done by means of a non-standard LF

Fem (= ‘female counterpart of’). However, there is no need to put L’ in the dictionary with its own lexical entry. Thus, we have:

DIRECTEUR _(masc) ‘director’	RUSSE _{N(masc)} ‘[a] Russian’	PAYSAN _{N(masc)} ‘peasant’
Fem : <i>directrice</i>	Fem : <i>Russe</i>	Fem : <i>paysanne</i>

The feminine nouns DIRECTRICE, RUSSE_{N(fem)} and PAYSANNE_{N(fem)} need not be entered in an ECD.

However, this approach does not work all the time. While for many human nouns the corresponding ‘feminine’ lexemes are quite regular in French (as far as their meaning and their lexical cooccurrence are concerned), for many others they are not: thus, as we have seen above, CÉLIBATAIRE_{N(fem)} ‘single woman’ does not have the same LFs as CÉLIBATAIRE_{N(masc)} ‘single man’ and therefore needs its own lexical entry. The lesson to be extracted from this is that a lexicographer must proceed with great caution and consider each case individually; jumping at sweeping generalizations is easy but can be harmful.

4.6.2 Regular Polysemy

L can have many pairs of LUs $L_{1-i} \sim L_{1-j}$, $L_{2-i} \sim L_{2-j}$, ..., $L_{m-i} \sim L_{m-j}$, such that the semantic difference between **j**-member and **i**-member is the same in all pairs: ‘ L_j ’ – ‘ L_i ’ = ‘ σ ’; this is a case of regular polysemy. If the lexical entry for L_j has no other differences with respect to L_i but the semantic component ‘ σ ’ (and other changes that automatically accompany it), then, as is the case with regular derivation, L_j need not be entered in the dictionary: it is sufficient to indicate its possibility under L_i . This technique could be applied to the lexical entries of BAKE that present its ‘instrumental’ senses (see below).

Many transitive action verbs that semantically presuppose an instrument have a special sense which accepts the name of this instrument as its syntactic subject:

John cuts the cardboard with a knife. ~ *The knife **cuts** the cardboard.*

John writes his cards with this pen. ~ *This pen **writes** well.*

John solved the equation on his computer. ~ *The computer **solved** the equation.*

This sense can be called ‘instrumental.’ The question rises: Should an ECD store the instrumental sense for each verb that has it—as a separate lexical entry?

First, not all action verbs have instrumental senses: *John nails boxes with a hammer.* ~ **The hammer **nails** boxes* (<*nails well*>); or *John killed the wolf with a gun.* ~ **The gun **killed** the wolf* (<*kills easily*>). And second, some instrumental senses describe a property of the instrument, but cannot refer to an action (*This pen writes well.* ~ **This pen wrote two letters.*), while some others do both (*This knife cuts poorly.* ~ *The knife cut through the cardboard in 3 minutes.*). These facts show that the entry for the basic verb should have the indication on whether the instrumental sense exists and, if so, of what type; in most cases, a separate entry for an instrumental sense is not needed. I will not go into a semantic analysis and classification of instrumental senses here, limiting myself to the following remark. It is possible to avoid actually storing the LUs **BAKEI.1c**, **BAKEI.2c** and **BAKEII.1c** in an English ECD; it is sufficient to indicate their existence and their type in the corresponding lexical entries:

BAKEI.1a	BAKEI.2a	BAKEII.1a
...
Has an instrumental sense (property and action)	Has an instrumental sense (property and action)	Has an instrumental sense (property and action)
:	:	:
BAKEI.1c	BAKEI.2c	BAKEII.1c

With such indications in place, **BAKEI.1c**, **BAKEI.2c** and **BAKEII.1c** can be absent from the dictionary.

5 An Illustration: A Sample of an English ECD

In this section, first, several model lexical entries are presented as they are meant to appear in an ECD of English (5.1); then linguistic comments are offered on some phenomena that are intimately related to the lexicon, but must nevertheless be considered as part of the grammar and therefore are not reflected in an ECD (5.2).

5.1 Some Lexical Entries from an English ECD

BAKE, verb, regular conjugation

'Food' BAKE

- I.1a.** X cooks₁ solid Y ... in device Z¹ ... [*John baked the potatoes in the oven*].
- 1b.** Solid Y cooks₂ ... in device Z¹ ... [*The potatoes baked in 20 minutes*].
- 1c.** Device Z is used [by X]³³ in baking_{I.1a} Y [*This microwave bakes potatoes in 10 minutes*].
- 2a.** X creates solid food Y from W in device Z ... [*John baked good rolls from corn flour*].
- 2b.** Y is baked_{I.2a} [by X] from W in device Z [*The rolls baked quickly in the new oven*].
- 2c.** Device Z is used [by X] in baking_{I.2a} Y [*This oven bakes good bread*].

'Hardening' BAKE

II.1a. X causes that Y, which are raw bricks or pottery, become hard in device Z ... [*John baked the bricks*].

1b. Y is baked **II.1a** [by X] in device Z [*The bricks are baking now*].

1c. Device Z is used [by X] in baking **II.1a** Y [*This kiln bakes excellent tiles*].

2. Substance X becomes hard ... [*The mud on the shore baked under the July sun*].

'Heat' BAKE

III. People X¹ in location X² feel affected by heat ... [*Boston was baking in a heat-wave*].

'Food' BAKE**'Potato' BAKE****I.1a.****Definition**

X bakes Y in Z : X cooks₁ solid¹₁ Y by submitting Y to the indirect₁ action₆ of dry¹⁰ heat²₂ in an device₁ Z¹ or in contact₁ with source₁ Z² of heat²₂.

Government Pattern

X ⇔ I	Y ⇔ II	Z ⇔ III
1. N	1. N obligatory	1. Prep _{loc} N

Jane baked the apples (in the oven/over the coals).

Lexical Functions

Syn : make baked Y S_{instr} : baking tray/sheet; baking tin,
Syn_c : cook₁ cake tin; foil
Syn_n : roast₁ Ver : to a turn
Gener : cook₁ too much, AntiVer : //overbake
Conv₂₃ : bake_{1.1b} not enough, AntiVer: //underbake
S₃^{usual} = S_{instr-loc} : oven

Examples

—Keep quiet, Jane is baking the potatoes! —How about some baked potatoes, my darling? In a microwave, you can bake Sturmers better ⟨softer⟩ than Golden Delicious. ~ In a microwave, Sturmers bake [= passive-potential] better ⟨softer⟩ than Golden Delicious. Ruritians adore baking zucchini in hot ashes and eating them stuffed with eggs. Wrap the fish in foil and bake (for twenty minutes) (in a hot oven).

Linguistic comments

1) The component 'indirect [action of dry heat]' in the definition is necessary because of such contrasts as *BAKE potatoes in their jackets* ⟨*chicken in tin foil*⟩ vs. *ROAST potatoes* [must be peeled first and/or cut into pieces] ⟨*chicken* [must be directly exposed to the heat, for at least part of the cooking period]⟩; cf. Lehrer 1974: 34, 182.

- 2) The component ‘submit’ is an abbreviation for ‘cause₂ to undergo’; the corresponding sense is absent from LDoCE 1978.
- 3) Here and below, Prep_{loc} stands for a locative preposition (which is chosen by the speaker according to its meaning): *in, on, over, under, ...*
- 4) Instead of *X bakes_{1.1a} Y*, some speakers spontaneously use the construction *X makes baked Y*, as in *For dessert, I’ll make some baked apples*; the preference depends on Y. This fact is taken into account by putting MAKE BAKED Y as a synonym of BAKE_{1.1a}.

I.1b.

Definition

Y bakes in Z : solid Y cooks₂ _as a result_²₆ of being baked_{1.1a} [by X] or₃ of undergoing the indirect₁ action₆ of dry₁₀ heat²₂ in device₁ Z¹ or in contact₁ with source₁ Z² of heat²₂.

Government Pattern

Y ⇔ I	Z ⇔ II
1. N	1. Prep _{loc} N

The apples baked (in the oven ⟨on hot bricks⟩).

Lexical Functions

Syn_C : cook₂ S₂^{usual} = S_{instr-loc} : oven Caus : bake_{1.1a}
 Syn_∅ : roast₂ S_{instr} : baking tray/sheet;
 Gener : cook₂ baking tin, cake tin; foil

Examples

—Keep quiet, the potatoes are baking! Sturmers bake better ⟨softer⟩ than Golden Delicious.

Linguistic comments

BAKE_{1.1b} marginally admits situations of non-agentive baking. As a result, BAKE_{1.1b} and BAKE_{1.2b} are not quite parallel: *The potatoes baked* does not necessarily presuppose that someone baked them on purpose (they could have baked ‘themselves’ inadvertently—for instance, in a house fire); but *The bread baked* presupposes a conscious creator. To account for the bizarre event of self-baking potatoes, the definition of BAKE_{1.1b} contains the disjunctive component ‘... or undergoing the indirect₁ action₆ of dry₁₀ heat²₂’.

I.1c.**Definition**

Z bakes Y : Device₁ Z used²₁ [by X] in baking_{1.1a} Y causes₁ Y to bake_{1.1b}.

Government Pattern

Z ⇔ I	Y ⇔ II
1. N	1. N

This oven baked the potatoes in 15 minutes. This microwave bakes quite well.

Lexical Functions

Ver : to a turn | M₂ ≠ Λ

'Bread' BAKE**I.2a.****Definition**

X bakes Y from W in Z : 'X creates₁ solid₁ food_{1b} Y from a mixture₁ W¹ of W², which is flour¹ made¹₁ of grain₁ W³ or other powder-like foodstuff, with liquid²₁—by baking_{1.1a} W¹ in Z'.

Government Pattern 1

X ⇔ I	Y ⇔ II	W ⇔ III	Z ⇔ IV
1. N	1. N	1. <i>from</i> N 2. <i>out of</i> N 3. <i>with</i> N	1. Prep _{loc} N

1) C_{III.1,2} without C_{II} : undesirable

2) C_{III.3} : N ≠ W³

Bob bakes good bread (from (= out of, with) imported flour) (in his new oven).

Bob bakes on Fridays.

Undesirable : ?*Bob bakes from (out of) imported flour* [by Constraint 1; correct expression: *Bob bakes with imported flour*].

Impossible : **Bob bakes good bread with imported wheat* [by Constraint 2; correct expression: *Bob bakes good bread from (out of) imported wheat*].

Government Pattern 2

X ⇔ I	Z ⇔ II	Y ⇔ III	W ⇔ IV
1. N	1. N obligatory	1. <i>into</i> N obligatory	1. Prep _{loc} N

1) C_{II} : N ≠ W³

Leo baked imported flour into good bread (in his new oven).

Impossible : **Leo baked imported wheat into good bread* [by Constraint 1]

Lexical Functions

Syn_c : make

Syn_n : cook₁

shop which carries

mostly baked goods

or pastry ≈ S_{loc} : bakery; pastry

professional-S _{1▷}	: baker; confectioner	shop, pâtisserie
S ₂ ^{usual}	: baked goods	Ver : to a turn
S _{2▷} ^{usual}	: pie; pastry, charlotte; meatloaf	too much, AntiVer : //overbake
S ₃ ^{usual}	: dough, batter; flour	not enough, AntiVer : //underbake
S ₄ ^{usual} = S _{instr-loc}	: oven; Dutch oven	A ₂ recentlyPerf : //fresh-baked

Examples

—Keep quiet, Jane is baking the cake! In an old-fashioned oven, you can bake your buns to a turn, provided you don't overbake them. Ruritians adore fresh-baked bread. Alain bakes very well: for instance, he is an excellent baker of scones. Transfer the batter into a cake pan and bake for 20 minutes. Robert baked Caroline a beautiful birthday cake and promised to bake another for her friend next month. —From now on, you will bake in this new oven. Jamie offered me a fire cake (a mess of flour and water baked on a hot stone) [G. Vidal]. This nan [local bread] was baked directly over dried camel dung.

I.2b.**Definition**

Y bakes in Z : Y becomes₁ ready₁ to be eaten₁ _as a result_₂ of being baked_{1.2a}
[by X from W] in Z.

Government Pattern

Y ⇔ I	Z ⇔ II
1. N	1. Prep _{loc} N

The bread baked (in his new oven <over the fire>) for 20 minutes.

Lexical Functions

Ver : to a turn

Examples

—Your rolls baked to a turn in about twenty minutes.

I.2c.**Definition**

Z bakes Y : Device Z used [by X] in baking_{1.2a} Y causes₁ Y to bake_{1.2b}.

Government Pattern

Z ⇔ I	Y ⇔ II
1. N	1. N

This oven bakes rolls in 30 minutes.

Lexical Functions

Ver : to a turn

‘Hardening’ BAKE**II.1a.****Definition**

X bakes Y in Z : X causes² that Y, which is raw bricks or pottery, hardens¹ by exposing Y to the action⁶ of dry¹⁰ heat^{2a} in device¹ Z.

Government Pattern

X ↔ I	Y ↔ II	Z ↔ III
1. N	1. N obligatory	1. <i>in</i> N

Greeks baked their amphorae (in primitive kilns).

Lexical Functions

Syn : fire_v
 S_2^{usual} : bricks; pottery
 $S_3^{usual} = S_{instr-loc}$: kiln
 Magn : hard

Examples

They first dry their bricks and then bake them hard.

II.1b.**Definition**

Y bakes in Z : Y hardens¹ being baked^{II.1a} [by X] in Z.

Government Pattern

Y ↔ I	Z ↔ II
1. N	1. <i>in</i> N

This amphora should bake (in the kiln) for 20 minutes.

Lexical Functions

Syn : fire
 S_1^{usual} : bricks; pottery
 $S_2^{usual} = S_{instr-loc}$: kiln
 Magn : hard < stone-hard

Examples

Such bricks can bake in a regular oven.

II.1c.**Definition**

Z bakes Y : Device¹ Z used²¹ [by X] in baking^{II.1a} Y causes¹ Y to bake^{II.1b}.

Government Pattern

Z ⇔ I	Y ⇔ II
1. N	1. N

This kiln bakes good bricks.

Examples

Leander's kiln bakes tiles in huge batches.

II.2.**Definition**

X bakes in Y : Substance₁ X hardens₁ by the action₁ of dry¹⁰ heat²₂ caused₁ by phenomenon₁ Y.

Government Pattern

X ⇔ I	Y ⇔ II
1. N	1. Adv ₂ Caus N

The mud baked (in the sun).

Examples

The earth floors baked stone-hard when the raging fire destroyed this modest dwelling.

'Heat' BAKE**III.**

X bakes in Y : Humans² X¹ in location₁ X² feel¹₂ intensely affected²₁ by dry¹⁰ heat²₂ caused₁ by the sun¹_{2a} or artificial₁ rays²₁ Y.

Government Pattern

X ⇔ I	Y ⇔ II
1. N	1. Adv ₂ Caus N

Jane was baking (in the sun (under the rays of the tropical sun)).

Lexical Functions

Syn : fry X = X², A₂ : baking₁
 Syn₀ : roast; burn; cook_{II}, swelter as if the skin
 shriveled, Magn + AntiVer : **Infml** to a frazzle

Examples

Quebec was baking in a heat wave. Sylvain liked baking on the sunbed. —You'll bake to a frazzle if you stay naked three minutes more! —Let me out, I'm baking in this stuffy room. Then I started treatment at Ospedale Maggiore for bending the knees, baking in a box of mirrors with violet rays, massage and bath [E. Hemingway].

Linguistic comments

1) The impossibility of **Jane was baking with fever* [correct expression: *Jane was **burning** with fever*] is foreseen by the constraints on Y; the correct expression is supplied by the entry for FEVER, since *burn* = [Magn + Oper₁](*fever*).

2) Consider the expression *bake in the sun* ‘be exposed to direct strong sun’ as seen in sentences (26):

(26) a. *For years, teens have spent hours baking in the sun in pursuit of the perfect golden tan.*

b. *You can almost smell the garbage baking in the sun.*

This expression is a collocation of SUN¹2a—namely, MagnInvol_v21(*sun*¹2a)—and therefore it should not be covered by the lexical entry for BAKEIII.

BAKED ALASKA, idiom, nominal phrase, countable

baked Alaska : dessert consisting of ice cream topped² with meringue slightly browned² by bakingI.2a it.

BAKED BEANS, idiom, nominal phrase, pl, uncountable

baked beans : dish [= food] consisting of beans boiled and then bakedI.1b (in a thick sauce, usually tomato-based).

Examples

Baked beans were served, and then some dessert. Baked beans typically come in cans and may be eaten hot or cold.

BAKING, adjective

1. Location U is such that people bakeIII in U [*It is baking in here*].

2. So [hot] that people bakeIII [*baking hot*].

1. predicative use only.

Definition

It is baking in U : It is so hot_n^o in location₁ U that humans² bakeIII in U.

Government Pattern

U ↔ I
1. Loc _{in} N

It was baking (in the stuffy room ⟨here, in here⟩).

Lexical Functions

Syn_c : roasting

Syn_n : boiling; steaming

2. modifying use only; adverb-modifying

Baking [P] : [P] so that people bake_{III} | Magn(hot_n^o).

Examples

The room became baking hot.

Linguistic comments

The relevant sense of HOT—'high temperature of surrounding air'—is not present in LDoCE 1978; that is why I write it as HOT_n^o. The necessity of having this lexeme as a separate LU is shown by the impossibility of **baking hot water* or **baking hot fire*.

SUN-BAKED, adjective

sun-baked [X] : [_Body part_ X of a human²] perceived as _dried out_2 or3
burnt₁₃ by the sun¹_{2a} [as if X were mud₁ baked_{II.2} in the sun¹_{2a}].

Example

sun-baked lips <cheeks, hands>; a sun-baked forehead <face>

5.2 Lexico-grammatical Problems Related to Lexical Entries for BAKE

The ECD aims to achieve maximal completeness of lexicographic information at the level of individual LUs (cf. 4.4). With such a policy in place, some checks are necessary in order to avoid including too much data in the dictionary entries—information that is either encyclopedic rather than linguistic or grammatical rather than lexical. I already mentioned the criterion of linguistic relevance (used to screen components of lexicographic definitions); now I will consider, as a modest illustration, five problems involving grammatical properties of LUs, in order to see whether they (or some of them) can be factored out and transferred to the grammar.

1. Decausatives vs. Passives vs. Passive-Potentials

The second member of the pairs BAKE_{I.1a} ~ BAKE_{I.1b}, BAKE_{I.2a} ~ BAKE_{I.2b} and BAKE_{II.1a} ~ BAKE_{II.1b}, or the **b**-lexeme, differs from the **a**-lexeme in the communicative organization of its meaning. In the **a**-lexeme the central, i.e., communicatively dominant, component is 'cause₂', but in the **b**-lexeme this component is present, so to speak, only on the periphery: here, the central component is the designation of a change of state, to which the causation component is attached as a modifier. Thus, the central component in the meaning of

BAKE1.1a is ‘cook1’ \approx ‘cause2 to become ready for eating’, while the central component in the meaning of BAKE1.1b is ‘cook2’ \approx ‘become ready for eating’; ‘cause2’ is relegated here to a modifier status: ‘[cook2] being baked1.1a’ (‘bake1.1a’ contains ‘cause2’). To put it differently, BAKE1.1a refers to an action or an activity, while BAKE1.1b designates a process, resulting from this action/ activity. The same relationship holds in two other verb pairs in our illustration. We will call the **b**-member the *decausative* (of the **a**-member).

NB: The term *decausative* is not ideal, since it is used in many different (as a rule, much broader) senses; cf., for instance, Padučeva 2001. But in the present context this is not very relevant, so I can afford using it strictly in the sense specified above—after this warning. Note that another current term for this types of verb is *inchoatives* (cf., for instance, Levin 1993: 4-5).

The decausatives call for the following three remarks.

- Decausatives are separate lexemes with respect to their transitive counterparts (cf. Atkins *et al.* 1988), contrary to passives (as in *be baked*): passives are inflectional forms of the same lexeme. All active transitive verbs—with a few exceptions—can be passivized by a very general rule, so that passives should not be described in the dictionary (although they can, of course, be illustrated there—like any other inflectional form). On the other hand, decausatives are possible with a great many verbs, but by no means with all: cf. *The pie baked* \langle *prepared \rangle *in twenty minutes*, *The houses sold* \langle *built \rangle *in twenty months*, *The bridge wore out* \langle *destroyed \rangle *in a few years*, *The books shipped* \langle *sent \rangle *this morning*, etc. The constraints on decausatives are too capricious to allow the formulation of a general grammatical rule. Unless such a rule is available, we have to regard decausatives, unlike passives, as separate dictionary entries (= separate lexemes).

- However, even if they are not 100% productive, decausatives are productive enough, in the sense that English has many pairs of the type ‘[transitive] verb V’ \sim ‘decausative of V,’ and in all of them the decausative changes the meaning of the starting V in the same way:

if	‘V’	=	‘cause in the way α that Y enters into the state S’,
then	‘decausative(V)’	=	‘Y enters in the state S as a result of causation in the way α ’.

This warrants our marking the ‘V’~ ‘decausative of V’ contrast with **a** vs. **b** lexicographic distinguishers: this is a typical case of regular polysemy (in the sense of Apresjan 1973 and 1992: 213ff).

- Decausatives contrast with passive-potentials³⁴ of the type *These judges bribe easily* or *This blouse washes well*. The passive-potential of a verb V (V-Π→Y) means ‘Y has such properties that it is Z-y to V Y’. Because of such a meaning, passive-potentials cannot be used in ‘completive’ or ‘punctual’ statements—they tend to be generic. What is important for us here is that they are sufficiently productive—in any event, more so than the decausatives: **This manual read in two days* [decausative], but *This manual read like a novel* [passive-potential]; **This carpet wore in a year* [decausative], but *These carpets wore too quickly* [passive-potential]; etc. Perhaps passive-potentials can be considered as inflectional forms of the verb (like ‘normal’ passives) formed by conversion; in this case they should not constitute separate lexical entries (they can of course be illustrated in the lexical entries for the basic active verbs, along with normal passives). If, however, they are not regular enough, that is, too lexically constrained, then they represent derivation—so that a passive-potential constitutes a separate lexical unit. But even in this case, it need not have its own real lexical entry in the dictionary, since all its cooccurrence properties remain those of the basic (active) form. It is sufficient to indicate in the basic lexical entry the possibility of its formation.

2. Object-Shuffling

Examples like *Bob bakes good bread out of even the cheapest flour* and *Bob bakes even the cheapest flour into good bread* represent purely syntactic modifications of a single lexeme, having the same semantic content and the same derivations and collocations in both patterns; in the lexical entry, we show this overtly by assigning both usages to one lexeme [= BAKE1.2a] and supplying this lexeme with two GPs. Thus, in an ECD, Object-Shuffling is treated in a way that is intermediate between the treatment of passives/passive-potentials and that of decausatives. Verbs of the BAKE1.2a type have to be lexically marked, since not all creation verbs admit Object-Shuffling: e.g., CREATE itself does not [**God created dust into man*]. They need not, however, be split into two different lexemes since in both patterns they have the same meaning and the same restricted cooccurrence; they can be described by the same definition.

NB: In this an Object-Shuffling verb is different from a decausative. A decausative changes the semantic content of the basic verb—it introduces a new central component in V’s meaning and thereby modifies the taxonomic class of V; as a consequence, it cannot be described simply by different syntactic patterns associated with the same definition.

Object-Shuffling verbs are but a special subset of what can be called Actant-Shuffling verbs, which also include such types of verbs as LOAD or SWARM (cf. Fillmore 1968: 68, 1977). Many of the latter cannot be assigned to a single lexeme without gross injustice to the semantic differences between their syntactic variants: *load bricks on the wagon* does not presuppose that the whole wagon will be filled, while *load the wagon with bricks* does; *teach arithmetic to children* does not presuppose that the children learn it, but *teach children arithmetic* does. (For a different treatment of Object-Shuffling in the case of BAKE, see Atkins *et al.* 1988; for a general discussion of Actant-Shuffling, or ‘diathesis alternation,’ see Atkins *et al.* 1986, Katz & Levin 1988, Levin 1993, and Levin & Rappaport 1995, 2005, where further relevant references can be found.)

3. Object-Deletion

Atkins *et al.* 1988 point out that in cookery books Object-Deletion with BAKE (and similar verbs) is common, as in *Bake for thirty minutes*. However interesting this phenomenon, it should not be reflected in the dictionary (except in the examples). This kind of Object-Deletion is a feature of instructional literature in general, and it is essentially limited to imperatives or prescriptive impersonals (*And then you bake for 30 minutes*) therein; therefore, it is not the responsibility of the dictionary, but that of the grammar: more specifically, of its stylistic component. Under appropriate conditions, the corresponding grammatical rule overrides the explicit mark ‘obligatory’ in the GP. Thus, Object-Deletion is indeed the deletion, in specifiable circumstances, of an otherwise obligatory object, rather than the non-expression of an optional object, as in BAKE1.2a (*Bob bakes on Fridays*).

4. Benefactives

The problem of so-called benefactive complements, or benefactives, exemplified by *Bob baked Mary a cake* (= ‘Bob baked a cake for Mary’), is again a grammatical one. As stated in Atkins *et al.* 1988, all creation verbs (with the exception of certain Latinate verbs having the stress on a non-initial syllable, such as CREATE itself, CONSTRUCT, etc.—but not, e.g., PREPARE) admit the benefactive (‘Dative’) construction, which means ‘... with the purpose that the beneficiary Z

directly uses the created object'. Therefore, the description of English benefactives is likewise the responsibility of the grammar, not the dictionary; see, for instance, Mazurkiewich & White 1984, where a grammatical rule for the Dative Beneficiary construction is postulated; a list of verbs which admit/do not admit a benefactive is given in Levin 1993: 48-49. Two remarks seem in order at this juncture.

- We subsume under 'creation verbs' also all verbs of, so to speak, 'negative creation,' i.e. destruction, such as BREAK [*her a glass*]. Although we are unlikely to hear *Dick broke her a glass* very often, the sentence is perfect—if, for some reason, she actually needed a broken glass and Dick obligingly broke her one.

- A verb of creation may appear with a different GP such that it ceases to be a verb of creation. In this usage, it rejects a benefactive (Levin 1993: 3):

(27) a. *Martha carved a toy out of wood for the baby.* =

b. *Martha carved **the baby** a toy out of wood.*

c. *Martha carved some wood into a toy for the baby.*

d. **Martha carved **the baby** some wood into a toy.*

In (27)d) CARVE is not used as a verb of creation, and a benefactive with it is impossible.

5. Subject/Object-Oriented Resultative Complements

Resultative complements, as in *We baked the pot **hard*** or *The pot baked **hard** in the kiln*, are possible with Creation Verbs or Change-of-Physical-State Verbs (as in *I knocked it **flat***). They must be treated as syntax- or semantics-driven phenomena, i.e., they must be described by general rules rather than by particular lexical entries; they can be shown in the Illustrative Zone of a lexical entry, and then only when they are considered of particular interest.

To sum up: The treatment of the above five problems depends on whether one regards them as lexically conditioned or syntactically/semantically conditioned. If lexically conditioned (like Decausatives or Object-Shuffling), a phenomenon should be accounted for in the dictionary—either via the definition, as with Decausatives (which leads to two different lexemes), or via the GP, as with some cases of Object-Shuffling (which leads to two different GPs within one lexeme). If syntactically/semantically conditioned (like benefactives, Object-Deletion or subject/object-oriented resultative complements), a phenomenon should be accounted for in

the grammar and has no place in the dictionary, except—for pedagogical purposes—in the examples.

Acknowledgments

I would like to express my deepest gratitude to all those with whom I have been working on the ECD for about 40 years, but this is of course impossible—it would result in too long a list. Therefore, I will limit myself to the following three groups of names:

- A. Zholkovsky, my coworker and coauthor for 10 years from the very beginning, whose numerous ideas and proposals have been embodied in the ECD which we see today, and Ju. Apresjan, with whom I have had so many lexicographic conversations and discussions that now I have trouble to distinguish his contribution from the rest.

- The first and the last reader of this text, L. Iordanskaja; she shared with me all the years of intensive lexicographic work.

- M. Alonso Ramos, D. Beck, P. Hanks, E. Marshman, J. Milićević, and A. Polguère, who read the subsequent version of the present text and whose remarks have helped me to improve it drastically.

Igor Mel'čuk

Observatoire de linguistique Sens-Texte [OLST]

Département de linguistique et de traduction, Université de Montréal

POB 6128 Centre-ville Montreal H3C3J7 Canada

e-mail: igor.melcuk@umontreal.ca

web-site: <http://www.olst.umontreal.ca/melcuk>

Notes

¹ (1, p. 00) This paper draws heavily on several previously published studies, especially Ilson & Mel'čuk 1989, Mel'čuk 1995 and Mel'čuk *et al.* 1995.

² (1.1, p. 00) **Lexical Units and Linguistic Signs** (cf. Subsection 1.3.4, p. 000)
Strictly speaking, an LU L of **L** is not a sign of **L**: it is the set of signs {s_i} (of **L**) of a particular type. Namely:

- if L is a lexeme (such as STUDENT, SING or HIGH), then $\{s_i\}$ are inflectional forms of L—wordforms and analytical form phrases;
- if L is a compounding element (such as SINO- or SOCIO-), then $\{s_i\}$ are its morphological variants;
- if L is an idiom (such as _SHOOT THE BREEZE_), then $\{s_i\}$ are also inflectional forms of L—set phrases that manifest this idiom.

From the signs s_i which are elements of L, a common minimal sign \bar{s} can be extracted. If L is a lexeme, this \bar{s} is the shared stem of all its wordforms and analytical form phrases; if L is a compounding element, the sign \bar{s} is the shared part of all its variants; and if L is an idiom, the sign \bar{s} is the shared configuration of stems (= a SSynt-tree) of all its set phrases. The sign \bar{s} represents the LU L, standing to \bar{s} in one-to-one correspondence. To say that a lexicon of **L** is a list of lexical signs of **L** is to use an obvious abbreviation.

³ (1.2, p. 00) **Theoretical lexicology vs. practical lexicography**

It is of course impossible to delve here into the heated debate opposing theoretical lexicology, which is a part of linguistics, and practical lexicography, which boils down to dictionary making. Their goals, methodologies, means and, most importantly, constraints are so different that to many it is difficult to see how one can benefit from the other; a serious discussion of corresponding matters would require a book in itself. Nonetheless, in the actual fact, they can benefit greatly from each other, and as a theoretical lexicologist with keen interest for dictionaries, I will allow myself to formulate the following two remarks about the stormy relationship of the two disciplines.

- Many shortcomings and outright mistakes in conventional dictionaries could be avoided if lexicographers followed some of the guidelines offered by theoretical lexicology. More coherence and better semantic analysis are quite compatible with regular commercial requirements. (For instance, an important improvement would be developing the dictionary by semantic fields rather than by alphabet. This would immediately produce a better consistency with no additional costs.) These directives are generally ignored because of centuries-long traditions.

• B. Atkins (1992/1993: 8-9), while defending the AHD 1981 against McCawley, who reproached it that the nouns BEAN and RICE were not marked as ‘count’ vs. ‘mass’, says: “Does the AHD, designed principally for native speakers of American English, need to hold such information?” The implied answer is ‘Of course not, since the users intuitively know all this.’ Well then, they also know, in the same sense, what BEAN and RICE mean and how these nouns are used; do native speakers need to open the AHD to look them up? And here I come to a completely heretical statement:

|| I do not know what a native speaker needs a monolingual dictionary for. Personally, I have never in my life used a monolingual dictionary of Russian as a speaker—only as a linguist, and then just in order to find faults in it. Of course, native speakers, including myself, need to look up rare words and expressions—dialectal, obsolete, or technical—for their meaning, pronunciation, or grammatical characteristics; and they do. Thus, an article published by Associated Press on the Web 10 December 2005—<http://news.yahoo.com/s/ap/20051210>—indicates that the ten top words looked up by Americans in the electronic Merriam-Webster dictionary are the following ones: 1. *integrity*, 2. *refugee*, 3. *contempt* [of the court], 4. *filibuster*, 5. *insipid*, 6. *tsunami*, 7. *pandemic*, 8. *conclave*, 9. *levee*, 10. *inept*, all of them highly technical terms. But this does not justify compiling huge monolingual dictionaries where the absolute majority of entries are such words as I, FOR, TABLE, WHITE or WALK. More than that, I never got a reasonable answer to the above question from a lexicographer...

On the other hand, I believe that a scientifically oriented monolingual dictionary can be an extremely powerful tool in teaching logic, semantic analysis and formal reasoning to general public. High school students are not supposed to be afraid of trigonometry, chemical formulas and computers; why should they be of analytical lexicographic definitions, syntactic patterns and collocation descriptions? The creation of really scientific, reasonably formalized monolingual dictionaries should go hand in hand with new ways of teaching the mother tongue in schools—with the emphasis on formal description of the lexicon and the grammar rather than simply on spelling. This, however, is no more than a free flight of my fancy...

⁴ (1.3, p. 00) **Communicatively dominant position**

The semantic component ‘ σ ’ is said to be in *communicatively dominant position* within the meaning ‘S’ if and only if ‘ σ ’ constitutes a minimal paraphrase of ‘S’: replacing ‘S’ with ‘ σ ’ brings about a loss of information, but not its distortion.

The communicatively dominant semantic component ‘ σ ’ within the meaning ‘S’ is shown by underlining: ‘S’ = ‘... σ ...’.

⁵ (1.3, p. 00) For Lexical Function, see below, Subsection 2.3, p. 00ff.

⁶ (1.3, p. 00) Phrasemes (\approx set phrases) are divided into three major classes:

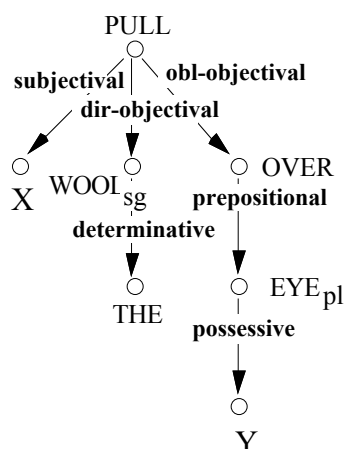
- full phrasemes, or idioms;
- semi-phrasemes, or collocations; and
- quasi-phrasemes, or quasi-idioms.

For details, see Mel’čuk 1995b and 2003.

⁷ (1.3, Item 1, p. 00) **Lexical entries for idioms**

Strictly speaking, the dictionary entry for an idiom is formally different from that of a lexeme—with respect to the latter, it contains some additional fields, which are related to its linguistic nature, namely being syntactically a phrase, and not a word.

First of all, the dictionary entry for an idiom must contain its Surface-Syntactic Structure; thus, the idiom *_PULL THE WOOL OVER Y’S EYES_*, *X pulls the wool over Y’s eyes* \equiv ‘X deceives Y by feigning₁ good₂ intentions₂ *_with the purpose_* to hide X’s true₁ bad₂ intentions₂ from Y and thereby *_gain an end_*’ receives the following Surface-Syntactic tree:



Second, an idiom must be accompanied by indications about the possible syntactic transformations and linearization of its SSynt-tree (when these operations cannot be carried out according to available general rules). Thus, in this particular case, it has to be stated that the idiom admits a passive-like transformation: *They had the wool pulled over their eyes more than once by this sleazy young adventurer*. Compare this with the proverbial idiom *_KICK THE BUCKET_*, where the passivization is impossible: **The bucket was kicked by Jim*. The expression *[to] pull strings*, which has a normal passive (*Strings ⟨A couple of strings⟩ were pulled by my powerful uncle*) represent a different case: this is a collocation of the noun *STRINGS*II ≈ ‘hidden influence or control’ (a *plurale tantum*).

⁸ (1.3, p. 00) A similar approach to developing a computational phrasal dictionary is proposed in the interesting study Zernik & Dyer 1987; Jackendoff 1995 makes a strong case for a phrasal dictionary as well.

⁹ (1.3, p. 00) Note the use, in the definition, of an ungrammatical expression *causes1 that ...* In our semantic metalanguage, i.e., in semantic decompositions used as definitions in an ECD, we admit a few clumsy and even ungrammatical expressions, insofar as this is unavoidable for the sake of semantic precision or clarity.

¹⁰ (2.1.1, p. 00) What has been said so far should by no means be construed as opposition to inclusion of encyclopedic information in lexical entries. My objections are leveled only at confusing the semantic and the encyclopedic information in lexicographic definitions. I do not mind putting into the entry for L

as much encyclopedic information about the denoted object or event as might be judged useful to better characterize the usage of L—under the condition, however, that encyclopedic information is explicitly marked as such and is kept in a special zone, strictly apart from purely linguistic information.

¹¹ (2.1.2.1, Rule 1, p. 00) A *quasi-predicate* is a meaning whose referent is an entity (substance, person, object, etc.) that is normally involved in a situation from which it borrows its Semantic Actants. Typical examples of quasi-predicates include names of functions (*professor of mathematics*), of actors (*winner of the prize*), of artifacts, in particular—instruments (*his saw*), of institutions (*Women's Hospital*), etc.

¹² (2.1.2.1, Rule 1, p. 00) **Propositional form**

1. The name *propositional form* is due to the fact that, in the prototypical case—when L is a predicate—the expression of the form $X Ls Y$ represents a logical proposition: it is sufficient to fill in the variables, and this expression becomes a proposition. However, in the case of a quasi-predicate the result is not a proposition: the expression *minister of [country] Y for [domain] Z* does not give rise to a proposition. Therefore, our terminology is a bit sloppy. Nevertheless, no real harm is done, since we can think of the expression *X is a minister of Y for Z*, which does underlie a proposition.

2. Let me emphasize that the presentation of the proposed propositional form has no scientific value in the following sense: it does not participate in any formal manipulation or discussion. Its vocation is purely pedagogical: it provides a minimal logical framework for the subsequent definition of the lexeme as well as a minimal pattern for typical syntactic constructions in which the lexeme would actually be used. In fact, the propositional form is a simplified and user-oriented presentation of L's government pattern (see below), which helps to relate it in a more perspicuous way to the definition. What is scientifically important is the specification of L's SemAs; therefore, one could write the propositional form for L as $L(X; Y)$.

¹³ (2.1.2.2, p. 00) To show this in a clearer way, here are the corresponding definitions from Wierzbicka 1972:

‘woman’ = ‘human being that could be someone’s mother’
 ‘man^{II}’ = ‘human being that could cause another human being to be
 someone’s mother’
 ‘mother of Y’ = ‘human being inside whose body there was once something that
 was becoming Y’s body’

¹⁴ (2.1.2.2, p. 00) Here is a simple demonstration of why vicious circles in the dictionaries are so bad. Suppose one defines A as follows: (i) $A = B + C$; B, in its turn, is defined as (ii) $B = D + E$, and D as (iii) $D = A + F$. By substituting D in (ii) by $A + F$ (in virtue of (iii)), we get (iv) $B = A + F + E$; finally, by substituting B in (i) by $A + F + E$, we obtain (v) $A = A + F + E + C$. This is an absurdity: A is declared to be equal to itself plus a lot of other things.

¹⁵ (2.1.2.3, Rule 3, Comment 1, p. 00) This definition is factually not satisfactory: for example, a table or a car has a height, but it does not have a ‘base’ or a ‘summit’. However, this detail is not relevant here, since we are dealing exclusively with the formal aspect of the definitions.

¹⁶ (2.1.3.2, Criterion I.2a, p. 00) It is interesting to compare BATTRE^{II} with VAINCRE ‘win’:

$X \text{ vainc } Y \text{ dans } Z \text{ [pour } \alpha \text{]} \cong \text{‘} X \text{ and } Y \text{ being confronted in struggle1 } Z \text{ over } \alpha, \parallel X \text{ gets2a } \alpha \text{’}$

VAINCRE is about X obtaining the intended result in the struggle against the adversary Y, while BATTRE^{II} is about X putting Y out of the struggle. The standard intensifier of BATTRE^{II} is *à plate couture*, and it characterizes the state of Y; VAINCRE does not have a really idiomatic intensifier, but the noun VICTOIRE ‘victory’ has some (although not very idiomatic ones): *grande* ‘great’, *large* ‘big’ | *antepos*, *complète* ‘complete’; they all express the degree to which X obtains α or the importance that obtaining α has for X.

¹⁷ (2.1.3.4, Criterion I.2c, (11)a), p. 00) The improved definition satisfies one of DeMorgan’s logical rules, relevant in this case: $\neg(A \wedge B) = \neg A \vee \neg B$ (“the negation of a conjunction is equal to the disjunction of negations”).

¹⁸ (2.1.3.4, p. 00) PERMIT2 \approx '[to] enable', as in *His illness did not permit us to continue the journey.*

¹⁹ (2.1.4, p.00) In reality, the situation with [to] COST is more complex. In *This book costs \$30 at Chapters* the phrase *at Chapters* may be considered as an expression of the Payee: it can be not a simple locative Circumstantial, isofunctional with *in Boston* or *on a plane*. To cover such uses as well the verb [to] COST will have the corresponding SemA-slot, supplied with necessary semantic restrictions.

²⁰ (2.2, p. 00) C_{III.1} is possible only with the active form of HELP_V, but this fact should not be stated here—i.e., in the GP of this particular lexeme, since a 'bare' infinitive cooccurs with no verbal lexeme in the passive (*We saw him cross* <*to cross> *the street.* ~ *He was seen to cross* <*seen cross> *the street.*), the only exception being the passive of the verb LET: *He was let go* <*let to go>. This is a general rule of English syntax.

²¹ (2.2, p. 00) Constraints 1 and 2 became obsolete in Contemporary American English. Sentences in which [to] HELP takes a bare infinitive even if X does not directly participate in Z are quite current: *The advocate helped him obtain compensation; These voters helped him win in 2001; She helped me save some money;* etc. Native speakers find them perfectly grammatical. I keep Constraints 1 and 2 as an illustration of an interesting theoretical possibility.

²² (2.3.3, Def. 4, p. 00) As an example of empty LFs we can cite light, or support, verbs of the type Oper_i, see below, p. 00.

²³ (3.2.3, p. 00) Criterion II.2 is a rephrasing of Apresjan's criterion, postulated for a logical disjunction of components inside a lexicographic definition (Apresjan 1974: 85). The prototype of this criterion was proposed by G. Green (1969), which allows us to call it the *Green-Apresjan Criterion*.

²⁴ (3.2.4, p. 00) **Sense superposition, or admissible zeugmas**

Such usages as that reflected in (23)—that is, constructions involving one lexical occurrence expressing two lexemes at once—are known in other cases as well:

(i) a. *Her tender, courageous **heart** was thumping in her chest*

[HEART as the organ of feelings and HEART as a physiological organ].

b. *They were **told** that John was absent and to leave immediately*

[TELL as ‘communicate’ and TELL as ‘instruct, order’].

c. *They **asked** whether John was absent and for permission to leave immediately* [ASK as ‘question’ and ASK as ‘request’].

d. *I believe in the **purity** of these snows and their souls*

[PURITY in a literal (‘no dirt’) and a metaphorical (‘elevated feelings’) sense].

This is what is known as ‘superposition of senses’; see Percova 1988, where some Russian examples are collected and three types of regular polysemy are indicated under which such superposition is possible: 1) ‘a human organ’ ~ ‘its function’, 2) ‘a place’ ~ ‘people who are in this place’ and 3) ‘information’ ~ ‘carrier of this information’. Therefore, an allowance for the fact that Criterion II.2 is not absolute (sometimes it admits coordination of the dependents of two different lexicographic senses of the same word) should not seem so exotic.

Interestingly, a parallel, although different, phenomenon exists in the domain of morphology: one wordform can manifest the superposition of two different forms of a lexeme. Here are a couple of examples:

(ii) **a.** Ger. *Ich habe gegessen **was** übrig war*

lit. ‘I have eaten what was remaining’

[**was** is simultaneously the form of the accusative, governed by *habe gegessen* ‘[I] have eaten’ and the form of the nominative, governed by *war übrig* ‘was remaining’].

b. Pol. *Kogo on lubi a Jerzy nienawidzi?* lit. ‘Whom he likes and Jerzy hates?’

[**kogo** is simultaneously the form of the accusative, governed by *lubi* ‘likes’, and the form of the genitive, governed by *nienawidzi* ‘hates’].

For a detailed analysis of the latter case, see Dalrymple & Kaplan 2000.

²⁵ (4.1, p. 00) As an example of such exceptions in the domain of definition writing, LDoCE 1978 can be cited. It utilizes, in its definitions, about 2 000 words—the only ones admitted in the definitions of all the LUs in the dictionary. A rigorous usage of such a defining metalanguage, as R. Quirk says in his ‘Preface,’ has “in many cases resulted in a fresh and revealing semantic analysis.” As for syntactic

combinatorics, the same dictionary uses a special coding, developed and introduced more than half a century ago by A.S. Hornby (*Oxford Advanced Learner's Dictionary of Current English*, 1948; I am referring to the seventh printing from 1977). Each code identifies a syntactic construction in which L can participate—among a few dozen constructions listed and exemplified at the beginning of the dictionary. For instance, the verb PLUNGE is marked ‘X9’, which means that this verb is used with a Direct Object and requires a directional complement: [to] PLUNGE *something somewhere*.

Another dictionary of the same company, the *Longman Language Activator* (1993), has made serious advances in the formalization of lexicographic metalanguages. Not only does it use a controlled defining vocabulary of 1,052 lexemes, but it also proposes a well-developed formalization of the description of syntactic and lexical cooccurrence of its LUs, as well as that of semantic and lexical fields. The *Activator* is a good model of a dictionary that closely corresponds to our ideas about what a dictionary for general public should be: it combines, in a judicious way, the high level of formalization and rigor, on the one hand, and quite a pedagogical presentation, on the other. *Activator* is ‘living’ proof that a logically organized and formalized dictionary can be made accessible to an average user—and be a commercial success.

²⁶ (4.1, p. 00) The semanteme ‘illustrated’ is a weak component in this definition: there can be a *magazine* that does not have illustrations, such as *magazine humoristique* ‘humoristic magazine’ or *magazine de mots croisés* ‘crossword magazine’, etc.

²⁷ (4.1, p. 00) The expression in parentheses is again is a weak component: a magazine can deal with several different subjects.

²⁸ (4.2, p. 00) CELIBATAIRE_{N(masc)} and CÉLIBATAIRE_{N(fem)} are of course two different LUs of French.

²⁹ (4.3.1, p. 00) The reason for this, however ridiculous, is that, currently, dictionaries are compiled in alphabetical order, and not by semantic fields, as an ECD necessarily is.

³⁰ (4.3.1, p. 00) By their nature, national characteristics can be attributed only to collectivities: one can define only ‘the Chinese’; ‘[a] Chinese’ has to be defined as an element of the collectivity ‘the Chinese’. (Cf. the correspondence ‘[the] English’ ~ ‘[an] Englishman’, ‘[the] French’ ~ ‘[a] Frenchman’, ‘[the] British’ ~ ‘[a] Briton’, ‘[the] Spanish’ ~ ‘[a] Spaniard’.) Other cases in which the LU being defined must be in the plural include the names of organs and devices that consist of several entities, most often, two: TEETH, CLAWS, EYES, LEGS, SHOES, SKIS, etc. Nouns of this type are, in a sense, *pluralia tantum*. Thus, EYES (not an EYE!) is ‘organ of vision that consists of two openings ...’. The corresponding morphological singular means ‘an element of the set in question’; this singular can be pluralized in its turn, which produces the meaning ‘several elements of the set in question’. The plural form of such nouns is thus ambiguous: SHOES means either ‘a pair of shoes’ [a *plurale tantum*], or ‘several separate shoes’ [a genuine plural]. (It can also mean ‘several pairs of shoes— as in *She was buying only most expensive shoes*, but with no possibility of quantifying: *five shoes* means only ‘five separate shoes’; with a quantifier one needs to use PAIR: *five pairs of shoes*.)

Two remarks seem in order here.

1. The semantic correlation between the singular and the plural forms of the LUs such as [*les*] FRANÇAIS ~ [*un*] FRANÇAIS or [*les*] GANTS ~ [*un*] GANT is very regular and productive, and an ECD should contain a rule explicitly stating the corresponding generalization. This type of rule belongs to what Apresjan aptly called, more than 30 years ago, the *grammar of the dictionary*. However, notwithstanding the importance of this concept for the ECD, it is impossible to dwell on it here.

2. In some languages the semantic correlation of the same type that we see between EYES and EYE in English is explicitly shown by morphological means. Thus, in Hungarian, the meaning ‘eyes [of a person]’, i.e., a pair of eyes, is expressed by a singular noun SZEM /sem/; its plural, *szemek*, denotes several pairs of eyes, as in *All eyes turned to her*. One ‘eye’ is called in Hungarian FÉLSZEM lit. ‘half of SZEM’, and its plural, *félszem+ek*, refers to several single eyes, as in *He had five right eyes ready for transplanting*.

³¹ (4.3, p. 00) Here is a tentative definition for the LU NATIONALITY2:

*nationality*₂ = ‘set₂ of individual₁ that share₁ linguistic₁, cultural, social₁, physical₁ and/or psychological₁ characteristics, which is caused₁ by common₁ origins₂’.

³² (4.4, p. 00) Of course there is nothing in the nature of the ECD that prevents it from achieving any degree of external exhaustivity as well. The only problem is organizational (or, if you wish, financial).

³³ (5.1, p. 00) When a variable corresponding to a Sem-Actant appears in square brackets in a definition, it indicates that this actant cannot be expressed syntactically; it loses its communicative importance and is ‘suppressed.’

³⁴ (5.2, p. 00) Passive-potentials are often called *middles* (for instance, in Levin 1993) or *medio-passives*. They are possible mostly, but not only, for verbs that denote a change of state, and semantically presuppose an Agent. On English middles (in the above sense), see Fagan 1988 and 1992, Akema & Schoorlemmer 1995, Rosta 1995, and Yoshimura & Taylor 2004.

References

- Ackema, Peter, & Maaike Schoorlemmer (1995) Middles and Nonmovement. *Linguistic Inquiry*, 26: 2, 173-197.
- AHD (1981) *The American Heritage Dictionary of English Language*. Boston etc.: Houghton Mifflin Co.
- Alonso Ramos, Margartita (1993) *Las Funciones Léxicas en el modelo lexicográfico de I. Mel’čuk*. [PhD thesis, Universidad Nacional de educación a distancia, Madrid.]
- Alonso Ramos, Margartita (2003) Hacia un diccionario de colocaciones del español y su codificación. In: Fernández Montraveta, A., Martí Antonin, A. & Vázquez García, G., eds., *Lexicografía computacional y semántica*, Barcelona: Universidad de Barcelona, 11-34.
- Alonso Ramos, Margartita (2004) Elaboración del Diccionario de colocaciones del español y sus aplicaciones. In: Bataner, P. & DeCesaris, J., eds., *De lexicografía: Actes del I Symposium internacional de lexicografía*, Barcelona: IULA, 149-162.

- Altman, Joel, & Alain Polguère (2003) La BDéf : base de définitions dérivée du Dictionnaire explicatif et combinatoire. In: *Actes de la première Conférence Internationale sur la Théorie Sens-Texte*, Paris: École Normale Supérieure, 43-54.
- Apresjan, Jurij (1969a) Tolkovanie leksičeskix značenij kak problema teoretičeskoj semantiki [Definition of Lexical Meanings as a Problem of Theoretical Semantics]. *Izvestija Akademii Nauk SSSR, Serija lit. i jazyka*, 28: 1, 11-23.
- Apresjan, Jurij (1969b) O jazyke dlja opisanija značenij slov [On a Language for the Description of Lexical Meanings]. *Izvestija Akademii Nauk SSSR, Serija lit. i jazyka*, 28: 5, 415-428.
- Apresjan, Jurij (1973) Regular Polysemy. *Linguistics*, 142: 5-32.
- Apresjan, Jurij (1974) *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka* [Lexical Semantics. Synonymic Means of the Language]. Moscow: Nauka. [See a corrected and extended translation: Apresjan 1992a.]
- Apresjan, Jurij (1980) *Tipy informacii dlja poverxnostno-semantičeskogo komponenta modeli Smysl ⇔ Tekst* [Types of Information for the Surface-Semantic Component of the Meaning-Text Model]. Wien: Wiener Slawistischer Almanach.
- Apresjan, Jurij (1988a) Morfologičeskaja informacija dlja tolkovogo slovarja [Morphological Information in a Monolingual Dictionary]. In: Ju. Karaulov, red., *Slovarnye kategorii*, Moskva: Nauka, 31-59.
- Apresjan, Jurij (1988b) Tipy kommunikativnoj informacii dlja tolkovogo slovarja [Types of Communicative Information for a Monolingual Dictionary]. In: *Jazyk: sistema i funkcionirovanie*, Moskva: Nauka, 10-22.
- Apresjan, Jurij (1990a) Tipy leksikografičeskoj informacii ob označajuščem leksemy [Types of Lexicographic Information on a Lexeme's Signifier]. In: *Tipologija i grammatika*, Moskva: Nauka, 91-108.
- Apresjan, Jurij (1990b) Formal'naja model' jazyka i predstavlenie leksikografičeskix znanij [A Formal Model of Language and Representation of Lexicographic Knowledge]. *Voprosy jazykoznanija*, n° 6, 91-108.
- Apresjan, Jurij (1992a) *Lexical Semantics: User's Guide to Contemporary Russian Vocabulary*. Ann Arbor, MI: Karoma].

- Apresjan, Jurij (1992b) O novom slovare sinonimov ruskogo jazyka [On a New Dictionary of Synonyms of Russian]. *Izvestija RAN, Otdelenie Lit-ry i Jazyka*, 51: 1, 18-39.
- Apresjan, Jurij (1995) *Izbrannye trudy. Tom II. Integral'noe opisanie jazyka i sistemnaja leksikografija* [Selected Writings. Vol II. An Integral Linguistic Description and Systemic Lexicography]. Moskva: Škola «Jazyki ruskoj kul'tury».
- Apresjan, Jurij (2002) Leksikografičeskaja koncepcija «Novogo Bol'sogo Anglo-Ruskogo slovarja» [Lexicographic Philosophy of the *New Big English-Russian Dictionary*]. In: Ju. Apresjan & È. Mednikova, *Novyj Bol'soj anglo-russkoj slovar'*, Moskva: Russkij jazyk, 6-17.
- Apresjan, Jurij (2004) *Novyj ob'jasnitel'nyj slovar' sinonimov ruskogo jazyka* [New Explanatory Dictionary of Russian Synonyms]. Moskva—Vena: Jazyki slavjanskoj kul'tury—Wiener Slawistischer Almanach.
- Arbatchewsky-Jumarie, Nadia & Lidija Iordanskaja (1988) Le champ lexical 'Parties du corps': description sémantique des lexèmes et structure des vocables. In: Mel'čuk *et al.* 1988: 63-72.
- Atkins, Beryl T.S. (1992/1993) Theoretical Lexicography and Its Relation to Dictionary Making. In: W. Frawley, ed., *Forum on the Theory and Practice of Lexicography* [= *Dictionaries*, 14], 4-43.
- Atkins, Beryl, Judith Kegl, & Beth Levin (1986) *Explicit and Implicit Information in Dictionaries*. [= *Lexicon Project Working Papers*, 12]. Cambridge, MA: M.I.T.
- Atkins, Beryl, Judith Kegl, & Beth Levin (1988) Anatomy of a Verb Entry: From Linguistic Theory to Lexicographic Practice. *International Journal of Lexicography*, 1: 2, 84-127.
- Becker, Joseph D. (1975) The Phrasal Lexicon. In: R. Schank & B.L. Nash-Webber, eds, *Proceedings of Interdisciplinary Workshop on Theoretical Issues in Natural Language Processing*, 70-73.
- Beinhauer, Werner (1978) *Stilistisch-phraseologisches Wörterbuch spanisch-deutsch*. München: Max Hueber.

- Benson, Morton, Evelyn Benson, & Robert Ilson (1986) *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam/Philadelphia: John Benjamins.
- Binon, Jean, Serge Verlinde, Jan Van Dyck, & Ann Bertels (2000) *DAFA : Dictionnaire d'apprentissage du français des affaires*. Paris : Didier.
- Bosque Muñoz, Ignacio (2004) *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM.
- Cowie, Antony, Ronald Mackin & Isabel McCaig (1993) *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.
- Dalrymple, Mary & Ronald Kaplan (2000) Feature Indeterminacy and Feature Resolution. *Language*, 76: 4, 759-798.
- Deane, Paul (1988) Polysemy and Cognition. *Lingua*, 75, 325-361.
- Deribas, V. (1975/1983) *Ustojčivye glagol'no-imennye slovosočetanija ruskogo jazyka* ['Verb + Noun' Set Phrases in Russian]. Moskva: Russkij jazyk.
- Fagan, Sarah (1988) The English Middle. *Linguistic Inquiry*, 19: 2, 181-203.
- Fagan, Sarah (1992) *The Syntax and Semantics of Middle Constructions*. Cambridge: Cambridge University Press.
- Fillmore, Charles (1968) The Case for Case. In: E. Bach & R.T. Harms, eds., *Universals in Linguistic Theory*, New York etc.: Holt, Rinehart and Winston, 1-89.
- Fillmore, Charles (1977) The Case for Case Reopened. In: P. Cole & J.M. Sadock, eds., *Grammatical Relations [= Syntax and Semantics 8]*, New York etc.: Academic Press, 59-81.
- Frawley, William (1988) New Form of Specialized Dictionaries. *International Journal of Lexicography*, 1: 3, 189-213.
- Goddard, Cliff & Anna Wierzbicka (1994) *Semantic and Lexical Universals. Theory and Empirical Findings*. Amsterdam/Philadelphia: John Benjamins.
- Green, Georgia (1969) On the Notion 'Related Lexical Entry'. *Papers from the Fifth Regional Meeting, Chicago Linguistic Society*, 76-88.
- Günther, Erika, & Waldtraut Förster (1987) *Wörterbuch verbaler Wendungen. Deutsch-Russisch. Eine Sammlung verbal-nominaler Fügungen*. Leipzig: Enzyklopädie.

- Hanks, Patrick (1992/1993) Lexicography: Theory and Practice. In: W. Frawley, ed., *Forum on the Theory and Practice of Lexicography* [= *Dictionaries*, 14], 97-112.
- Ilgenfritz, Peter, Nicole Stephan-Gabinel & Gertraud Schneider (1989) *Langenscheidts Kontextwörterbuch Französisch-Deutsch*. Berlin etc.: Langenscheidt.
- Ilsou, Robert & Igor Mel'čuk (1989) English BAKE revisited (BAKE-ing an ECD). *International Journal of Lexicography*, 8: 4, 325-345.
- Iordanskaja, Lidija & Igor Mel'čuk (1984) Connotation en sémantique et lexicographie. In: Mel'čuk *et al.*, *Dictionnaire explicatif et combinatoire du français contemporain : Recherches lexico-sémantiques I*. Montréal: Presses de l'Université de Montréal, 33-40.
- Iordanskaja, Lidija & Igor Mel'čuk (1990) Semantics of Two Emotion Verbs in Russian: *BOJAT'SJA* '[to] be afraid' and *NADEJAT'SJA* '[to] hope'. *Australian Journal of Linguistics*, 10: 2, 307-357. [See also in: I. Mel'čuk, *The Russian Language in the Meaning-Text Perspective*, 1995, Moskva—Wien, Škola «Jazyki russkoj kul'tury» —Wiener Slawistischer Almanach, 81-124.]
- Iordanskaja, Lidija & Igor Mel'čuk (2006) Connotation. In: *The Slavic Languages: An International Handbook of their History, their Structure and their Investigation*, eds. T. Berger, K. Gutschmidt, S. Kempgen & P. Kosta, Berlin New York: Walter de Gruyter.
- Iordanskaja, Lidija & Alain Polguère (2005) Hooking Up Syntagmatic Lexical Functions to Lexicographic Definitions. In: Ju. Apresjan & L. Iomdin, eds., *East-West Encounter: Second International Conference on the Meaning ⇔ Text Theory*, Moscow: Jazyki slavjanskoj kul'tury, 176-186.
- Jackendoff, Ray (1995) The Boundaries of the Lexicon. In: M. Everaert, E.-J. van der Linden, A. Schenk, R. Schreuder, eds., *Idioms. Structural & Psychological Perspectives*, Hillsdale, N.J.—Hove, UK: Lawrence Erlbaum Associates, 133-165.

- Katz, Boris & Beth Levin (1988) *Exploiting Lexical Regularities in Designing Natural Language Systems*. [= *Lexicon Project Working Papers*, 22]. Cambridge, MA: M.I.T.
- Lakoff, George & John Ross (1976). Why you can't *do so* into the sink. In *Syntax and Semantics, vol. 7: Notes from the Linguistic Underground*, J. McCawley (ed.), New York: Academic Press, 101-111.
- Lakoff, Robin (1973) Lexicography and Generative Grammar. II: Context and Collocation in the Dictionary. In: R. McDavid & A. Duckert, eds, *Lexicography in English* [= *Annals of the New York Academy of Sciences*, 211], New York, N.Y.: New York Academy of Sciences, 154-164.
- LDoCE (1978) *Longman Dictionary of Contemporary English*. Harlow—London: Longman.
- Lehrer, Adrienne (1974) *Semantic Fields and Lexical Structure*. Amsterdam — London: North-Holland.
- Levin, Beth (1993) *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, Beth, & Malka Rappaport (1995) *Unaccusativity. At the Syntax—Lexical Semantics Interface*. Cambridge, MA—London: The MIT Press.
- Levin, Beth, & Malka Rappaport (2005) *Argument Realization*. Oxford: Oxford University Press.
- McCawley, James (1986) What Linguists Might Contribute to Dictionary Making If They Could Get Their Act Together. In: V. Raskin & P. Bjarkman, eds., *The Real-World Linguist*, Norwood, N.J.: Ablex, 3-18.
- McCawley, James (1992/1993) How to Achieve Lexicographic Virtue through Selective & Judicious Sinning. In: W. Frawley, ed., *Forum on the Theory and Practice of Lexicography* [= *Dictionaries*, 14], 120-129.
- Makkai, Adam (1975) *A Dictionary of American Idioms*. Woodbury, N.Y.: Barron's Educational Series.
- Mazurkewich, Irene, & Lydia White (1984) The Acquisition of the Dative Alternation: Unlearning Overgeneralizations. *Cognition*, 16, 261-283.
- Mel'čuk, Igor (1973) Towards a Linguistic "Meaning ⇔ Text" Model. In: F. Kiefer, ed., *Trends in Soviet Theoretical Linguistics*, Dordrecht: Reidel, 33-57.

- Mel'čuk, Igor' (1974) *Opyt teorii modelej Smysl ⇔ Tekst*. Moskva: Nauka.
[Reprinted: 1999, Moscow: Škola «Jazyki russkoj kul'tury».]
- Mel'čuk, Igor (1981) Meaning-Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, 10: 27-62.
- Mel'čuk, Igor (1982) Lexical Functions in Lexicographic Description. *Proceedings of the VIIIth Annual Meeting of the Berkeley Linguistics Society*, 427-444.
- Mel'čuk, Igor (1988a) *Dependency Syntax: Theory and Practice*. Albany, N.Y.: State University of New York Press.
- Mel'čuk, Igor (1988b) Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria. *International Journal of Lexicography*, 1: 3, 165-189.
- Mel'čuk, Igor (1989) Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory. *Quaderni di semantica*, 10: 1, 65-102.
- Mel'čuk, Igor (1993) *Cours de morphologie générale. Vol. 1. Introduction et Première Partie: Le mot*. Montréal/Paris: Les Presses de l'Université de Montréal/CNRS Éditions.
- Mel'čuk, Igor (1995a) The Future of the Lexicon in Linguistic Description and the Explanatory Combinatorial Dictionary. In: I.-H. Lee, ed., *Linguistics in the Morning Calm 3 (Selected Papers from SICOL-1992)*, Seoul, 1995, 181-270.
- Mel'čuk, Igor (1995b) Phrasemes in Language and Phraseology in Linguistics. In: M. Everaert, E.-J. van der Linden, A. Schenk, R. Schreuder, eds., *Idioms. Structural and Psychological Perspectives*, Hillsdale, N.J.—Hove, UK: Lawrence Erlbaum Associates, 167-232.
- Mel'čuk, Igor (1996) Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In: L. Wanner, ed., *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: John Benjamins, 37-102.
- Mel'čuk, Igor (1997) *Vers une linguistique Sens-Texte* [Leçon inaugurale]. Paris: Collège de France.

- Mel'čuk, Igor (2000) Un FOU/une FOLLE : un lexème ou deux ? In: *Lexique, syntaxe et sémantique. Mélanges offertes à Gaston Gross à l'occasion de son soixantième anniversaire* [BULAG, numéro hors série], 95-106.
- Mel'čuk, Igor (2003) Collocations dans le dictionnaire. In: Th. Szende, éd., *Les écarts culturels dans les Dictionnaires bilingues*, Paris: Honoré Champion, 19-64.
- Mel'čuk, Igor (2004a) Actants in Semantics and Syntax I: Actants in Semantics. *Linguistics*, 42: 1, 1-66.
- Mel'čuk, Igor (2004b) Actants in Semantics and Syntax II: Actants in Syntax. *Linguistics*, 42: 2, 247-291.
- Mel'čuk, Igor (2006) *Aspects of the Theory of Morphology*. Berlin: Mouton de Gruyter.
- Mel'čuk, Igor, André Clas & Alain Polguère (1995) *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Mel'čuk, Igor, & Alain Polguère (1987) A Formal Lexicon in the Meaning-Text Theory (Or How to Do Lexica with Words). *Computational Linguistics*, 13: 3-4, 261-275.
- Mel'čuk, Igor, & Alain Polguère (to appear) *Lexique Actif du Français*. Louvain-la-Neuve: Duculot.
- Mel'čuk, Igor, & Leo Wanner (1996) Lexical Functions & Lexical Inheritance for Emotion Lexemes in German. In: L. Wanner, ed., *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: John Benjamins, 209-278.
- Mel'čuk, Igor, & Alexander Zholkovsky (1984) *Explanatory Combinatorial Dictionary of Modern Russian. Semantico-syntactic Studies of Russian Vocabulary*, Vienna: Wiener Slawistischer Almanach.
- Mel'čuk, Igor, & Alexander Zholkovsky (1988) The Explanatory Combinatorial Dictionary. In: M. Evens, ed., *Relational Models of the Lexicon*, Cambridge etc.: Cambridge University Press, 41-74.

- Mel'čuk, Igor, *et al.* (1984, 1988, 1992, 1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*. Montréal : Les Presses de l'Université de Montréal [= DEC-1/2/3/4].
- Miller, George, Christine Fellbaum, Judith Kegl, & Katherine Miller (1988) WORDNET: An Electronic Lexical Reference System Based on Theories of Lexical Memory. *Revue Québécoise de Linguistique*, 17: 2, 181-216.
- OCD (2002) *Oxford Collocations Dictionary*. Oxford: Oxford University Press.
- Padučeva, Elena (2001) Kauzativnyj glagol i dekauzativ v russkom jazyke [The Causative Verb and the Decausative in Russian]. *Russkij jazyk v naučnom osveščanii*, No. 1, 52-79.
- Padučeva, Elena (2002) O parametrah leksičeskogo značenija glagola: taksonomičeskij klass učastnika [On the Parameters of a Verb's Lexical Meaning: The Taxonomic Class of a Participant]. *Russkij jazyk v naučnom osveščanii*, No. 1 (3), 87-111.
- Padučeva, Elena (2004) *Dinamičeskie modeli v semantike leksiki* [Dynamic Models in Lexical Semantics]. Moskva: Jazyki slavjanskoj kul'tury.
- Pawley, Andrew (1985) On Speech Formulas and Linguistic Competence. *Linguas modernas*, 12: 84-104.
- Percova, Natal'ja (1988) *Formalizacija tolkovanija slova* [Formalization of Lexicographic Definitions]. Moskva: Izd-vo MGU.
- Polguère, Alain (1992) Remarque sur les réseaux sémantiques Sens ↔ Texte. In: A. Clas, réd., *Le mot, les mots, les bons mots*, Montréal : Les Presses de l'Université de Montréal, 109-148.
- Polguère, Alain (2000) Towards a Theoretically-Motivated General Public Dictionary of Semantic Derivations and Collocations for French. In: U. Heid, S. Evert, E. Lehmann & C. Rohrer, eds., *Proceedings of 9th Euralex International Congress EURALEX 2000*, Stuttgart, 517-527.
- Polguère, Alain (2003) Étiquetage sémantique des lexies dans la base de donnée DiCo. *Traitement Automatiques des langues*, 44: 2, 39-68.
- PR (2001) *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Paris : Dictionnaires Le Robert.

- Prčić, Tvrko (2005) Prefixes vs. Initial Combining Forms in English: A Lexicographic Perspective. *International Journal of Lexicography*, 18: 3, 313-334.
- Reum, Albrecht (1953) *Le petit dictionnaire de style à l'usage des Allemands*. Leipzig: VEB Bibliographisches Institut.
- Reum, Albrecht (1955) *A Dictionary of English Style*. Leverkusen: Gottschalksche Verlagsbuchhandlung. [Third edition: 1961, München: Max Hueber.]
- Rey, Alain (1990) Definitional Semantics: Its Evolution in French Lexicography. In: J. Tomaszczyk & B. Lewandowska-Tomaszczyk, eds., *Meaning and Lexicography*, Amsterdam/Philadelphia: John Benjamins, 43-55.
- Rodale, Jerome Irving (1947) *The Word Finder*. Emmaus, PA: Rodale Books.
- Rosta, Andrew (1995) 'How Does this Sentence Interpret?' The Semantics of English Mediopassives. In: B. Aarts & Ch. F. Meyer, eds., *The Verb in Contemporary English*, Cambridge: Cambridge University Press, 123-144.
- Spears, Richard A. (1992) *Common American Phrases in Everyday Contexts*. München: Orbis.
- Steinlin, Jacques, Sylvain Kahane, Alain Polguère, Adil El Ghali (2004) De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux. In: G. Williams & S. Vessier, eds., *Proceedings of the 11th EURALEX International Congress EURALEX 2004*, Lorient: Université de Bretagne, 177-186.
- Ščerba, Lev (1940 [1958]) Opyt obščej teorii leksikografii [Outline of a general theory of lexicography]. In: L.V. Ščerba, *Izbrannye raboty po jazykoznaniju i fonetike*, Leningrad: Izd-vo LGU, 54-91.
- Tuggy, David (1993) Ambiguity, Polysemy, and Vagueness. *Cognitive Linguistics*, 4: 3, 273-290.
- Wanner, Leo (2003) Definitions of Lexical Meanings: Some Reflections on Purpose and Structure. In: *Actes de la première Conférence Internationale sur la Théorie Sens-Texte*, Paris: École Normale Supérieure, 55-65.
- Wierzbicka, Anna (1972) *Semantic Primitives*. Frankfurt am Main: Athenäum.
- Wierzbicka, Anna (1980) *Lingua Mentalis: The Semantics of Natural Language*. Sydney etc: Academic Press.
- Wierzbicka, Anna (1985) *Lexicography and Conceptual Analysis*. Ann Arbor, MI: Karoma.

- Wierzbicka, Anna (1986a) The Semantics of 'Internal Dative' in English. *Quaderni di semantica*, 7: 1, 121-135.
- Wierzbicka, Anna (1986b) The Semantics of the Internal Dative—A Rejoinder. *Quaderni di semantica*, 7: 1, 155-165.
- Wierzbicka, Anna (1987) *English Speech Act Verbs. A Semantic Dictionary*. Sydney etc.: Academic Press.
- Wierzbicka, Anna (1991) Semantic Complexity: Conceptual Primitives and the Principle of Substitutability. *Theoretical Linguistics*, 17: 1-2-3, 75-97.
- Wierzbicka, Anna (1996) *Semantics. Primes and Universals*. Oxford—New York: Oxford University Press.
- Yoshimura, Kimihiro, & John R. Taylor (2004) What Makes a Good Middle? The Role of Qualia in the Interpretation and Acceptability of Middle Expressions in English. *English Language and Linguistics*, 8: 2, 293-321.
- Zaliznjak, Anna (1987) O tipax vzaimodejstvija semantičeskix priznakov [Types of Interaction of Semantic Features]. In: *Èksperimental'nye metody v psixolingvistike*, Moskva: Akademija Nauk SSSR, 136-196.
- Zernik, Uri, & Michael Dyer (1987) The Self-Extending Phrasal Lexicon. *Computational Linguistics*, 13: 3-4, 308-327.
- Žolkovskij, Aleksandr, & Igor Mel'čuk (1965) O vozmožnom metode i instrumentax semantičeskogo sinteza [On a Possible Method and Tools for Semantic Synthesis], *Naučno-texničeskaja informacija*, n° 5, 23-28.
- Žolkovskij, Aleksandr, & Igor Mel'čuk (1966) O sisteme semantičeskogo sinteza. I. Stroenie slovarja [On a System for Semantic Synthesis. I. Structure of the Dictionary], *Naučno-texničeskaja informacija*, n° 11, 48-55.
- Žolkovskij, Aleksandr, & Igor Mel'čuk (1967) O semantičeskom sinteze [On Semantic Synthesis]. *Problemy kibernetiki*, 19: 177-238. [Translated into French as: Zholkovskij & Mel'chuk 1970.]

Index of Terms

- assertion (= assertive part of a lexicographic definition)
- benefactive (complement)

collocate
collocation
communicatively dominant node
communicatively dominant position
compounding element
cooccurrence zone (of a lexical entry)
decausative
differentiating lexicographic information
Government Pattern
headword (of a lexical entry)
homonyms
idiom
keyword (= argument) (of a Lexical Function)
lexeme
lexical entry
lexical super-entry (= vocable)
lexical field
Lexical Function
 complex ~
 configuration of ~s
 degenerate ~
 non-standard ~
 standard ~
 thematic grouping of ~s
lexical unit
lexicographic number
linguistic relevance (of a semantic component)
linking (of Syntactic Actants to Semantic Actants)
maximal block (in a lexicographic definition)
passive-potential
phonological/graphematic zone (of a lexical entry)
phraseme

polysemy

chain ~

radial ~

presupposition (in a lexicographic definition)

propositional form (of a lexicographic definition)

pseudo-lexeme

quasi-idiom

quasi-predicate

restricted lexical cooccurrence

semanteme

Semantic Actant

semantic bridge

semantic decomposition

semantic derivation

semantic field

semantic pivot (of a signified)

semantic primitive

semantic taxonomic restriction (on a Semantic Actant)

semantic zone (of a lexical entry)

semantically simpler

sign, linguistic

signified (of a linguistic sign)

signifier (of a linguistic sign)

split variable (in a lexicographic definition)

synopsis (of a vocable)

syntactic valence, active

syntactic valence, passive

syntactics (of a linguistic sign)

unifying cooccurrence (of a lexical unit)

vocable