

**Mel'čuk, Igor (2013): *Semantics. From Meaning to Text*,  
vol. 2, Amsterdam/Philadelphia: John Benjamins Pub-  
lishing Company. pp. 259-368**

Chapter 11 Explanatory Combinatorial Dictionary	3
1 General Overview of the ECD	3
1.1 Introductory Remarks	3
1.2 The General 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 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 <b>L</b> Together and in a Similar Way	11
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 for 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	36
2.1.3 Criteria for an ECD Definition: Criteria of Type I	38
2.1.3.1 Criterion I.1: Linguistic relevance of a semantic component	39
2.1.3.2 Criterion I.2a: Cooccurrence with qualifying modifiers	41
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
2.1.4.1 ECD-style verbal definitions vs. SemRs of LUs	47
2.1.4.2 Internal structure of ECD definitions	48
2.2 The Government Pattern in the ECD	51
2.3 Lexical Functions in the ECD	54
2.3.1 Semantic Derivation	54
2.3.2 Collocation	55
2.3.3 Lexical Function	57
2.4 Illustrative Examples in the ECD	59
3 The ECD's Macrostructure: An ECD's Lexical Super-entry	60
3.1 Basic Notions for the Characterization of Lexical Super-entries	60
3.2 Criteria for the Delimitation of LUs within Vocables: Criteria of Type II	68
3.2.1 Introductory Remarks	68
3.2.2 Criterion II.1: Differentiating Lexicographic Information	70
3.2.3 Criterion II.2: Unifying Cooccurrence (= the Green-Apresjan Criterion)	74
3.2.4 A Comparison of Criteria II.1 and II.2	77
3.3 Organization of an ECD Super-entry [= Vocable]	79
3.3.1 The Vocable Synopsis	79
3.3.2 The Order of the LUs in a Vocable	79
4 Principles for Compiling the ECD	81
4.1 Formality Principle	81
4.2 Coherence Principles	83
4.2.1 Lexical Unit Internal Coherence Principle	83
4.2.2 Semantic Field Coherence Principle (= Lexical Inheritance)	88

4.3 Uniform Treatment Principles	88
4.3.1 Lexical Unit Uniform Treatment Principle	89
4.3.2 Vocabable Uniform Treatment Principle	91
4.4 Internal Exhaustivity Principle	91
4.5 Maximal Generalization Principles	92
4.5.1 Vocabable Generalization Principle	93
4.5.2 Semantic Field Generalization Principle	93
4.6 No Regularly Produced LUs Principle	94
4.6.1 Regular Derivation	94
4.6.2 Regular Polysemy	95
5 An Illustration: A Sample of an English ECD	96
5.1 Some Lexical Entries from an English ECD	96
5.2 Lexical-grammatical Problems Related to Lexical Entries for BAKE	104
Notes	
101	

## Chapter 11

# Explanatory Combinatorial Dictionary

Chapter 11 describes 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. Its description, of necessity concise and stripped down to a minimum, will be presented in five steps:

Section 1: General overview of the ECD

Section 2: The ECD's microstructure: lexical entry (= lexical unit entry)

Section 3: The ECD's macrostructure: lexical super-entry (= vocable entry)

Section 4: General principles for compiling the ECD

Section 5: Examples: several lexical entries from an English ECD

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

Since my own lexicographic practice is mostly focused on French, this chapter often uses French data; in order to simplify the presentation, the lexicographic definitions of French lexical units are formulated directly in English.

## 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 point of this chapter can be formulated as follows:

|| A formalized semantically-oriented and cooccurrence-centered lexicon is one of the two major components of a linguistic description of language **L**.

These components are a **lexicon** and a **grammar**.

- The lexicon of **L** is the set of all minimal (= elementary and quasi-elementary) lexical signs of **L**: roughly, it is the set of **L**'s lexical units [= LUs].<sup>1</sup>

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

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

- The set of all rules of **L**: semantic, syntactic, morphological, and phonological. A rule can be very general or not very general, even quite specific; but it must apply to a set of several signs, not to an individual sign. (For the notion of rule, see Part I, Ch. 1, p. 00ff.)

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 classes of **L**'s lexical signs (that is, the grammar deals with general rules).

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 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 lexical data—information related to individual LUs of **L** that are necessary for a linguistic model to go from a given SemR to all corresponding DSyntRs (and beyond), as well as vice versa.

|| In this way, the lexicon is the core element of the Semantic module of **L**.

The Meaning-Text approach puts forward 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 1960s by Alexander Zholkovsky and myself (Žolkovskij & Mel'čuk 1965, 1966, 1967). A little later, Jurij Apresjan joined us, so that many first versions of ECD's lexical entries for Russian were authored by all three of us (for instance, Apresjan *et al.* 1973). In its present form, the ECD implements many of Apresjan's ideas (see Apresjan 1969a, b, 1974, 1979b, 1980, 1986a, 1988a, b, 1990a, b, 1992a, 1995a, 2002, 2006). Three specimens of the ECD are available in printed form: Mel'čuk & Zholkovsky 1984 for Russian, as well as Mel'čuk *et al.* 1984, 1988, 1992, 1999 and Mel'čuk & Polguère 2007 for French. A simplified ECD for Spanish, or, more precisely, a dictionary of Spanish collocations—DICE (= *Diccionario de Colocaciones del Español*)—is being developed at present: Alonso Ramos 2003 and 2004b.

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 reasonable bibliography; some more references are given in the course of the exposition.

The ECD is set apart from other dictionaries by its general (= logical) and specific (= linguistic) properties. They will be presented below.

## 1.2 The General Properties of the ECD

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

- 1) theoretical orientation (**1.2.1**),
- 2) formalized character (**1.2.2**),
- 3) completeness at the level of each entry (**1.2.3**).

They 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 human language or particularities of specific languages. It is in this sense that they are called *general*, or *logical*; they are opposed to *specific*, or *linguistic*, properties of the ECD.

### 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]. 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; as stated above, the ECD constitutes an integral part of the Semantic module of an MTM.

Most current linguistic theories view a 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 useful generalization over the lexicon and, thus, secondary to it.

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

All lexicographic concepts used in an ECD (lexical unit, lexeme, phraseme, idiom, 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. Lakoff 1973: 162-164) rather than a practical conventional dictionary. More specifically, written within a clearly stated theoretical framework, an ECD purports to store all lexical knowledge shared by speakers of **L** in the form foreseen by this framework. It contains all LUs of **L**—lexemes and idioms, as prescribed by the theory—and it is designed to dovetail with the grammar ( $\approx$  syntax + morphology) of **L**.

Conventional dictionaries are 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 users. It is developed for the sake of linguistics and should satisfy the normal requirements for scientific descriptions.<sup>ii</sup>

**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 (such a dictionary is, for instance, Mel'čuk & Polguère 2007), 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; one can even say that it is a lexical database. It puts heavy emphasis on its 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 intuitively perceived semantic relatedness of two LUs is paralleled by the degree of the similarity of their entries. 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. 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, an ECD 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 Properties of the ECD

Along with the three above-mentioned general, or logical, 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 (**1.3.1**).

2) Semantic basis: the lexicographic definition of an LU L forms the logical basis of the entry for L (**1.3.2**).

3) Exhaustive coverage of the restricted lexical cooccurrence of each headword, i.e., of the LU L described in a particular lexical entry: all collocations of L are included in its entry (**1.3.3**).

4) Word list: it includes lexemes and idioms in the same order and treats them in the same way (**1.3.4**).

5) The target of a lexical entry: the headword is necessarily a monosemous LU (**1.3.5**).

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: all information about words and expressions it contains is 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 speech understanding process, since the latter requires lots of extralinguistic knowledge and common sense (which are not part of linguistic competence as such); see Part I, Ch. 3, **3.2.3**, p. 00ff). 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 of the type ‘How do you express such and such a meaning?’

For ease of reference by human users, the ECD entries—in a printed volume—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 ECD's user can expect to find:

- the term of any particular breed of dogs—BORZOI, BULLDOG, COLLIE, GERMAN SHEPHERD, GREAT DANE, GREYHOUND, POODLE, SETTER, TERRIER, etc.;
- the term for a dog of no definable breed—MONGREL, MUTT; CUR;
- 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.
- the form of the sign warning of an attack dog's presence—“BEWARE OF THE DOG”.

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 is 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 derivatives (see below, 2.3.1, p. 00ff.): for instance, for any L that denotes an event an ECD should give its action noun (DISCOVERY<sub>1</sub> for DISCOVER, FIGHT<sub>N</sub> for FIGHT<sub>V</sub>, etc.), the names of its actants (DISCOVERER [(who has discovered)], DISCOVERY<sub>2</sub> [(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.

And 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 its 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, 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 is touched upon in Ch. 16, the end of Section 6, p. 00.

The consistently synthetic orientation of the ECD does not prevent its use for analysis as well. Synthesis requires much more linguistic information than analysis: when you are trying to understand a text, 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 linguistic information an ECD contains is quite sufficient for analysis, i.e., for text understanding. The analysis requires of course huge amounts of non-linguistic information that an ECD does not—and is not supposed to—provide.

### 1.3.2 The Semantic Basis of the ECD

Because of its orientation, an ECD is a semantically based dictionary. (Hence the adjective *explanatory* in its name.) 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 of the headword and other elements in the entry.
- Third, each LU L' given in the entry for L must be associated with the description of its meaning. As indicated above, 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 preceded by its definition, so that what is given is the lexical means necessary to express a given meaning 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 its name.)

On the one hand, the ECD must faithfully reflect the properties of natural language. Restricted cooccurrence of signs—cooccurrence that is not determined semantically, i.e., by the signified, or phonologically, i.e., by the signifier—is a typical feature of natural languages; therefore, 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*, where LUs PAY and CLOSE are selected by the Speaker in a restricted way.)

On the other hand, it is the active character of the ECD that brings the restricted lexical cooccurrence to the fore. A conventional dictionary does not bother to indicate, for instance, that ILLNESS can be *grave* or *serious*, but not *\*heavy* (as it is, e.g., in Russian: *tjažĕlaja bolezn'* lit. (heavy illness)). A normal user, even not a native English speaker, when encountering the phrase *grave/serious illness* in a text, understands it without difficulty, and *\*heavy illness* will probably never be seen. 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 briefly discussed in Subsections 2.2 and 2.3 below and then developed in Chs. 13 and 14.) To the best of my knowledge, no existing conventional dictionary uses such a methodology.

### 1.3.4 The ECD Describes all LUs of L in a Similar Way

Given its formalized character, the ECD aims to cover all LUs of L in the same homogeneous manner. There are two major types of LUs (cf. Part I, Ch. 1, p. 00):

- A **lexeme**, or a minimal LU—a single word in a single well-defined sense. Lexemes constitute by far the largest part of the lexical stock of any language. Among lexemes, a particular subset is distinguished: lexemes that can be used only as an element in a compound<sub>1</sub> lexeme; they are known as **compounding elements** (or **combining forms**). In English, this type of LU is typi-

cally represented by ‘neo-classical’ LUs of the type ANTHROPO- (human<sub>N</sub>), ICHTHYO- (fish) (*ichthyophagous*, *ichthyosaur*, etc.), NEURO- (nerve), SINO- (China), or else -(O)LOGY (science), -PHILE (lover), -PHOBIA (fear). Compounding elements are typical of some languages and 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, i.e., a multilexemic LU—roughly, a set phrase whose meaning does not include the meaning of any of its components as its **semantic pivot** (see Part V, Ch. 16, 4, Def. 16.9, p. 00): κSHOOT THE BREEZE<sub>1</sub> (chat leisurely), κRIDE HERD<sub>1</sub> [*on N*] (control [N] by watching [N] closely), κCALL THE SHOTS<sub>1</sub> (being in charge, give orders), κFRIED EGGS<sub>1</sub> (dish made with the liquid contents of chicken eggs fried in a particular way), or κONCE BITTEN, TWICE SHY<sub>1</sub> (Somebody who has once suffered a misfortune is overcautious not to suffer it again), etc. (In *SMT*, an idiom is shown by raised semi-brackets κ 1. For more on idioms, see Part V, Ch. 16, 4, p. 00ff.)

Compounding elements are not so common in SAE languages and, because of this, cause no serious difficulties for a SAE 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 chapter. But idioms are a completely different story: while most conventional dictionaries fail to 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 idioms (that is, 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. And even in English, set phrases are, as a rule, described in separate dictionaries: Makkai 1975, Spears 1992, Cowie *et al.* 1993, OCD 2002.

In sharp contrast to such an attitude, an ECD includes 1) as separate full-fledged entries of the same type all lexemes and all idioms of **L** and 2) as subentries inside main entries—all pragmatemes, collocations and clichés, which number in millions. It is thus a dictionary of words

**and** (set) phrases, and the phrases numerically prevail. To make a long story short, an ECD is, in a sense, a phrasal dictionary.

The massive inclusion of idioms 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, shown in (1):

(1) French

- a.** In the expression *têtu comme un âne* lit. (stubborn as an ass) the phrase *comme un âne* is an intensifier: it means roughly (very) and is an element of the value of the lexical function [= LF; see Ch. 14]  $\text{Magn: Magn}(têtu) = \text{comme un âne}$ . 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 (stubborn).

The expression *têtu comme un âne* is a collocation of the adjective 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 (1a) no foolishness is implied:

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

In (1a) the phrase *comme un âne* (very) includes the noun ÂNE1 (he-ass, donkey) 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 useless: since the phrase occurs only with the adjective TÊTU (stubborn), being its collocata (namely, an intensifier), it is sufficient to enter it only in the lexical entry for TÊTU:

TÊTU

...

$\text{Magn}$  :  $\text{kcomme un âne1, kcomme une mule1}$

In (1b), the phrase *comme un âne* (like a fool) includes the noun ÂNE2 = (fool) and is free: 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  $OP_{ER1}$ : *allonger*, *asséner*, **neutral donner**, and  $OP_{ER2}$ : *recevoir*), etc. The two synonymous phrases *passer <sauter> du coq à l'âne* lit. (pass <jump> from the rooster to the ass) = (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 died of starvation between two identical bales of hay, being unable to choose between them) is a collocation of  $\hat{A}NE1$  (ass) = (donkey) (since its meaning includes the meaning ( $\hat{a}ne1$ ) as the semantic pivot), 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 phrasemes in the same way:<sup>iii</sup> idioms, on the one hand, and collocations as well as pragmatemes and clichés, on the other, are given different treatment.

- Idioms, which are semantically non-compositional phrasemes (the phrases of the type Fr.  $\kappa ARRIVER \langle VENIR \rangle COMME UN CHEVEU SUR LA SOUPE1$  lit. (arrive like a hair on the soup)  $\approx$  ([of a remark, etc.] be incoherent and out of place) or Fr.  $\kappa DONNER LE SEIN [à Y]1$  lit. (give the breast to Y) = (breast-feed Y)), cannot be effectively described as a function of just one of their components; they receive each a separate lexical entry—just like lexemes. A lexical entry for an idiom has (almost) the same structure and gives the same type of information as one for a lexeme.<sup>iv</sup>

- All other phrasemes, which are semantically compositional, do not receive separate entries, but are included into the entries of LUs that identify these phrasemes:

- **Pragmatemes**, that is, situationally-, or pragmatically-, bound phrasemes, of the type NO PARKING, BEST BEFORE [...] or SEE YOU LATER, are described as a function of the name of the corresponding extra-linguistic situation: NO PARKING is given (with a semantic description and a pragmatic indication “[on a street sign]”) under  $PARK_V$  [a car], and BEST BEFORE ...—under  $CAN_N$  and under PACKAGE, with the indication “[on a container of pre-packed perishable food]”. The name of the situation that binds a pragmateme is the **lexical anchor** of this pragmateme. Thus, a pragmateme is presented only in the dictionary article for its anchor, where it constitutes a subentry.

- **Collocations**, that is, the phrases of the type ACCEPT/DECLINE AN **INVITATION**, PAY A **VISIT**, KEEN **INTEREST**, etc., are best described as a function of one of their components, known as the **base** (shown here in boldface); therefore, a collocation is as well presented only in the lexical entry for its base (as a *sui generis* subentry, cf. above).

— **Clichés**, that is, compositional and semantically fully transparent expressions that are used and recognized as ready-made, global units, are also described in an ECD as non-standard LFs of their lexical anchors—LUs that identify the meanings expressed by the clichés. Thus, [*be*] *in the wrong place at the wrong moment* must be stored under WRONG; *The question is not if, but when* [ $\approx$  (This will surely happen)]—under SURE; *When in doubt, do nothing*—under DOUBT; *No rest for the wicked*—under REST; etc.

As a result of the decision to describe the phrasemes of **L** as stated above, an ECD lists, in the same alphabetical order, both lexemes and idioms. Thus, the French phrase  $\kappa$ RAT DE CAVE<sub>1</sub> lit. (cave rat)  $\approx$  (very thin long candle) = (wax taper) is an idiom having a proper definition and particular lexical cooccurrence. All conventional French dictionaries recognize this fact, but, instead of explicitly reflecting the autonomy and lexical character of the expression, they put  $\kappa$ RAT DE CAVE<sub>1</sub> 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 a practical or a logical viewpoint.

2) This treatment is never consistent. Thus, PR 2001 stores  $\kappa$ CHEMIN DE FER<sub>1</sub> lit. (iron road) = (railroad) and  $\kappa$ POMME DE TERRE<sub>1</sub> lit. (earth apple) = (potato) as single LUs and puts them, as separate entries, into general alphabetic order, while  $\kappa$ RAT DE CAVE<sub>1</sub> 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 a difficult, but completely artificial problem: to determine the ‘host’ for an idiom. Thus, should  $\kappa$ RAT DE CAVE<sub>1</sub> be stored under CAVE or under RAT? There are no obvious reasons in favor of either choice; the lexicographers of PR 2001 did not see any rationale, since they have put  $\kappa$ RAT D’HÔTEL<sub>1</sub> lit. (hotel rat) = (hotel thief) under RAT, but  $\kappa$ RAT DE BIBLIOTHÈQUE<sub>1</sub> lit. (library rat) = (bookworm), under BIBLIOTHÈQUE. However, semantically  $\kappa$ RAT D’HÔTEL<sub>1</sub> is closer to HÔTEL than  $\kappa$ RAT DE BIBLIOTHÈQUE<sub>1</sub> is to BIBLIOTHÈQUE (one can see this from the glosses); therefore, 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 κPOT DE CHAMBRE<sub>1</sub> (bedroom pot) and κVASE DE NUIT<sub>1</sub> lit. (night vase), both meaning (chamber pot), κRAT DE BIBLIOTHÈQUE<sub>1</sub> (bookworm), κRAT DE CAVE<sub>1</sub> (wax taper), κRAT D'HÔTEL<sub>1</sub> (hotel thief), κTABLE DE NUIT<sub>1</sub> (night table), etc.

— Verb phrases such as κAPPORTER DE L'EAU À [A<sub>poss-Y</sub> = mon, ton, son, ...] MOULIN<sub>1</sub> lit. (bring some water to Y's mill) ≈ (give an argument in favor of Y's viewpoint), κBRISER LA GLACE<sub>1</sub> (break the ice), κCOUCHER EN JOUE<sub>1</sub> [Y] lit. (lay [Y] in cheek) = (take aim at Y [with a rifle]), etc.

— Adjective phrases such as κCOUSU DU FIL BLANC<sub>1</sub> lit. (sewn with white thread) ≈ (sticking out a mile/as a sore thumb) or κTOUT CRACHÉ<sub>1</sub> lit. (all spit) ≈ (being the spitting image [of a person]), etc.

— Prepositional phrases such as κEN MIETTES<sub>1</sub> lit. (in crumbs) ≈ (destroyed), κPAR DESSUS LE MARCHÉ<sub>1</sub> lit. (by over the market) = (in addition, to boot), etc.

— Full clauses (proverbs and sayings) such as κAUSSITÔT DIT, AUSSITÔT FAIT<sub>1</sub> lit. (As soon said, as soon done) = (No sooner said than done), κL'ARGENT NE FAIT PAS LE BONHEUR<sub>1</sub> lit. (The money does not make the happiness) = (Money does not buy happiness), κTEL PÈRE, TEL FILS<sub>1</sub> lit. (Such father, such son) = (Like father, like son), etc.

(For lexicographic description of idioms, i.e., for idiom dictionary articles, see Ch. 16, **8**, p. 00ff.)

To ensure a better and easier access to idioms for linguistic synthesis (= speech production), the ECD systematically uses semantic cross-references. Thus, the idiom κCOUCHER EN JOUE<sub>1</sub> (take aim [with a rifle]) is semantically cross-referenced (in terms of lexical functions) under FUSIL (gun, rifle), under VISER (take aim) and under TIRER (fire<sub>V</sub>); Fr. κAPPORTER DE L'EAU À [A<sub>poss-Y</sub>] MOULIN<sub>1</sub> is referenced under ARGUMENT; κSHOOT THE BREEZE<sub>1</sub>, under CHAT; κBLACK BELT<sub>1</sub>, under KARATE; and Fr. κDONNER LE SEIN<sub>1</sub> (breast-feed), under BÉBÉ (baby), ALLAITER (nurse<sub>V</sub>), NOURRICE (nurse), and SEIN (breast). But these are only cross-references (although semantically motivated)—simple pointers to the appropriate expression; the complete lexicographic descriptions of these idioms are found in their own lexical entries.

And for analysis needs, that is, for the user who is looking up an idiom encountered in the text, the ECD has formally motivated cross-references to the idiom under all its constituents (minus, of course, structural words): κCOUCHER EN JOUE<sub>1</sub> is cross-referenced also under COUCHER and JOUE, and so forth. These, of course, are simple formal 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 semantic phrasemes of the second type—**collocations**: multilexemic expressions in which *L* retains its own meaning and syntactic-morphological properties. The number of collocations in a dictionary averages roughly a couple of dozen per lexical entry, so they actually outnumber the lexemes. (The distribution of collocations per LUs is extremely heterogeneous: some LUs have scores of collocations, while some have none.) 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<sub>N</sub>, and *lodge a complaint* under COMPLAINT<sub>N</sub>. This is done by using lexical functions [= LFs] (both standard and non-standard), a major innovation of the ECD.

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 1955 (English), Rodale (1947: English), Beinhauer (1978: Spanish), Benson *et al.* (1986: English), Ilgenfritz *et al.* (1989: French), and OCD 2002 (English); there are also specialized dictionaries of lexical cooccurrence, e.g., Deribas 1975 (Russian), Günther & Förster 1987 (Russian and German) and Bosque 2004 (Spanish). Yet such descriptions have never been carried out on a serious scale in general dictionaries; they are always presented as separate lexicographic products: dictionaries of collocations. In contrast, the ECD aims to describe all the collocations of an LU *L* within the lexical entry for *L*.

And, finally, an ECD purports to list all pragmatemes and clichés grouping them in the entries of their lexical anchors; here we talk perhaps of tens of thousands of expressions.

Summing up, I can say that the ECD corresponds fairly closely to the type of dictionary J. Becker (1975) and A. Pawley (1985) were so vigorously campaigning for.<sup>v</sup>

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

### 1.3.5 Each Article of the ECD Describes One LU, and an LU is Described by One Lexical Entry

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

The above statement is only approximately correct:

On the one hand, there are LUs such that each is described by two or more lexical entries: these are regular free compounds—for instance, Rus. SEMNADCATIËTAŽNYJ (having 17 stories) or SOROKAČETYRĚXLETNIJ (44 years old).



On the other hand, there are LUs such that two or more of them can be described by one lexical entry: these are LUs the semantic differences between which are really minor and can be specified by simple enough rules—for instance, **XXX**. This class of cases is most often described by particular **semantic constraints** to government patterns in lexical entries, see Ch. 13, p. 00.

However, these are special cases. Most often, in the standard, unmarked case, the statement “one LU—one lexical entry” holds.

Related LUs that have an identical signifier and share non-trivial semantic components in their signifieds are grouped into **vocables**; this grouping reflects polysemy (see Section 3, p. 00, where a rigorous definition of the concept is given: Definition 11.7, p. 00). Thus, the English vocable IMPROVE includes six LUs (lexemes), and each one has its own 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

IMPROVE<sup>I.1a</sup> *X improves* : (The value of the quality of X becomes higher)  
 [*The weather suddenly improved; The system will improve over time*]

IMPROVE<sup>I.1b</sup> *X improves Y* : (X causes<sup>1</sup> that Y improves<sup>I.1a</sup>)<sup>vi</sup>  
 [*The most recent changes drastically improved the system*]

IMPROVE<sup>I.2</sup> *X improves* : (The health of a sick person X improves<sup>I.1a</sup>)

[*Jim is steadily improving*]

IMPROVE<sup>I.3</sup> *X improves at Y* : (X's execution of Y improves<sup>I.1a</sup>, which is caused<sup>I</sup> by X's having practiced or practicing Y)

[*Jim is steadily improving at algebra*]

IMPROVE<sup>II</sup> *X improves Y by Z-ing* : (X causes<sup>2</sup> that the market value of a piece of real estate Y becomes higher by Z-ing in Y)

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

IMPROVE<sup>III</sup> *X improves upon Y* : (X creates a new Y' by improving<sup>I.1b</sup> Y)

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

The lexicographic (= sense-distinguishing) numbers, in accordance with existing dictionaries, capture semantic distances between LUs within a vocable: Roman numerals express the larger distances, Arabic numerals—smaller ones, and letters, the smallest distances. Semantic 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 IMPROVE<sup>I</sup> are considered to be closer to each other than to IMPROVE<sup>II</sup> and to IMPROVE<sup>III</sup>, because all of the former, but not the latter, include the semantic component (improve<sup>I.1a</sup>). IMPROVE<sup>I.1a</sup> and IMPROVE<sup>I.1b</sup> are especially close since their semantic difference—(P) ~ (cause<sup>I</sup> P)—is very regular in English. (For more on lexicographic numbers, see 3.2, p. 00ff. The semantic component (cause<sup>I</sup>) represents non-agentive causation, as in *The falling tree* <*The bullet*> **killed** *the dog*; (cause<sup>2</sup>) stands for the agentive causation: *John* **killed** *the dog* (see Part II, Ch. 5, p. 00ff).

The policy of having separate LUs as separate entries has the advantage of making it possible to indicate for each sense its morphological particularities (if any), its derivation range, its syntactic and lexical restricted cooccurrence, and so on—that is, to ensure the internal coherence of lexical entries. In conventional 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 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 fol-

lowed throughout the ECD, i.e., not only in vocables, but with respect to semantic fields as well (for the corresponding notion, see below, **3.1**, pp. 00ff).

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, a word or an idiom, taken in one particular sense (rather than a polysemous word, as in all current dictionaries).

## 2 The ECD's Microstructure: A Lexical Entry

An ECD entry, i.e., a full description of an LU L (a lexeme or an idiom), comprises three major divisions, which correspond to the three components of the linguistic sign, as it is understood in MTT (Part I, Ch. 1, Def. 1.2, p. 00). A linguistic sign **s** is a triplet:

$$\mathbf{s} = \langle (\text{Signified}) ; / \text{Signifier} / ; \Sigma(\text{syntactics}) \rangle.$$

Since an LU L is presented in the ECD as a linguistic sign (recall the convention formulated in Note 1, p. 00), each component of this triplet is described in a 'zone' of the ECD lexical entry for L. Each zone, in its turn, is subdivided into sub-zones (indicated below 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 **lexicographic definition** of L, which fully specifies L's meaning; and L's **connotations** (meanings that language **L** associates with L, but that are not part of its definition; see Ch. 15, p. 00ff).

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 this connection, a detailed study Apresjan 1990b), as well as L's **spelling**.

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 cooccurrence sub-zone

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 cooccurrence sub-zone

This sub-zone includes three main divisions. 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 cooccurrence sub-zone

Here are presented all lexical links between L and other LUs of **L**—in terms of LFs. **Lexical Functions** cover both semantic derivations and collocations of L with individual LUs or very small and irregular groups of LUs, as well as pragmatemes and clichés semantically associated with L.

The stylistic cooccurrence sub-zone

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

The pragmatic cooccurrence sub-zone

**Pragmatic Clues** pinpoint the real-life situations in which L 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) <sup>#</sup>*Parking forbidden* or <sup>#</sup>*Prohibition to park*, while in French exactly the opposite is true: the appropriate formula is *Défense de stationner* lit. (Prohibition to park), rather than <sup>#</sup>*Aucun stationnement* (No parking) (the symbol “#” indicates pragmatic inappropriateness). The expression NO PARKING, which is a pragmateme, is given as a value of a non-standard LF in the lexical entry for the verb PARK<sub>V</sub> and must be supplied with the indication “[on a street sign]”; the expression BEST BEFORE ..., given in the entry for the noun CAN<sub>N</sub> [tinplate container], has the indication “[on a container of pre-packed food]”; etc.

Pragmatic clues accompany not only pragmatemes, for which they are definitorial, but also various lexemes, collocations and idioms, for instance: *Roger!* (I understood you) [in a radio communication], *Push/Pull* (to open, push/pull) [on a door in a public building], *Hold the line!* (I

ask you not to hang up for some time) [**in a telephone conversation**], *At ease!* (I permit you light movement given your right foot does not leave the ground) [**a military command**], etc.

4) To the above three major zones, the ECD adds a fourth one, 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 four of the sub-zones mentioned above will be considered here:

- from the Semantic Zone, the **Definition** (to the exclusion of connotations, which are treated a special chapter: Ch. 15, p. 00ff);
- from the Cooccurrence Zone, the **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. In this chapter only the lexicographic definition is treated in some detail; the government pattern and lexical functions are sketched very briefly, since they are described more fully in separate chapters: Chs. 13 and 14.

## 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 element 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 Sem-Comm-areas (Rheme ~ Theme, Given ~ New, Foregrounded ~ Backgrounded ~ Neutral, etc.), as should be in the framework of Meaning-Text semantics (cf. Part II, Ch. 4, p. 00ff). An ECD definition is written ‘verbally’—in a special semantic metalanguage, which is basically **L** itself, but submitted to particular constraints, that is, this metalanguage is a subset of **L**. (In point of fact, this is a very large subset of **L**: the majority of **L**'s LUs are

admitted into it.) The verbal definition of L and its standard SemR stand of course in one-to-one correspondence.

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

$$\textit{definiendum} \equiv (\textit{definiens}),$$

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); “ $\equiv$ ” is the symbol of equivalence. The ECD definition will be discussed in four sub-subsections:

- Substantive requirements for ECD definitions: **2.2.1**
- Rules for the well-formedness of ECD definitions: **2.2.2**
- Criteria of linguistic truthfulness of ECD definitions: **2.2.3**
- General characteristics of ECD definitions: **2.2.4**

### 2.1.1 Substantive Requirements for 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.

1) L’s denotational potential is the class **D** of extralinguistic entities or facts to which L (or, more precisely, the meaning (L) of L) can be applied, i.e., which (L) denotes, when L is stored in the lexicon. 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 a specific element of class **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. Therefore:

An ECD lexicographic definition must by all means avoid including information about the real world (i.e., encyclopedic information), beyond what is strictly necessary to fully characterize 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 scien-

tific concept “CAT” aims at distinguishing cats from other mammals, that is, at identifying a cat; the lexicographic definition of the English lexeme CAT, on the other hand, should be concerned only with the use of the lexeme as applied to the prior somehow identified animal (cf. the relevant remarks in McCawley 1986). In other words, an ECD definition does not reflect the concepts speakers have of the corresponding object or event, but tries to encapsulate the subconscious speakers’ knowledge of the LU defined 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 never 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: 100-103; cf. 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 LUs.<sup>vii</sup>

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, Def. 11.5, p. 00). From this viewpoint, the definition of L must allow for the following two things:

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

*X escapes from Y by way of Z to W*  $\alpha$  (|[X being in place Y<sup>1</sup> where Y<sup>2</sup> is, such that (something related to) Y<sup>2</sup> threatens X and it is possible that X will not be able to move away from Y<sup>1</sup> before the threat by Y<sup>2</sup> to X is realized,]|

X succeeds in intentionally moving away from Y<sup>1</sup> via Z to place W, which causes<sup>1</sup> that the threat by Y<sup>2</sup> is not realized)

#### Comments

1. Y<sup>1</sup> and Y<sup>2</sup> represent a **split variable** Y: namely, Y = Y<sup>1</sup>/Y<sup>2</sup>. The technique of split variables allows us to cover by the same definition such different expressions as *escape from Cambodia/from the prison camp* [= Y<sup>1</sup>] and *escape from one’s captors/from the execution* [= Y<sup>2</sup>]. 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., they do not correspond to two different SemA-slots. Cf. Ch. 12, 3.4.2.2, p. 00.

2. The symbols “[ ... ]” enclose a presupposition; see below, 2.1.4, p. 00. Presuppositions have been characterized in Part II, Ch. 6, 3.7, p. 00ff.

3. (Cause1) stands for non-agentive causation ( $\approx$  (be the cause of)), Part II, Ch.5, 2, p. 00ff.

This definition allows one to replace ESCAPE1.1c with FLEE in some contexts (*Dith Pran escaped Cambodia*  $\approx$  *Dith Pran fled Cambodia*), but not in others (*Dith Pran barely escaped*  $\langle$  \*fled  $\rangle$  *Cambodia*); this is ensured by the 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 1992a on a program for a Russian dictionary of synonyms, in which the lexicographic data underlie definitions that satisfy the above requirement—that is, these definitions guarantee the correct selection of L in paraphrasing; for the dictionary itself, see Apresjan 2004a.)

- 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”, not about the lexeme SNOW), but because English has expressions like SNOWY (pure white), [*white*]  $\kappa$ AS SNOW<sub>1</sub>, SNOW-[*white*], and SNOW<sub>3</sub> ([*white*] cocaine powder), which are perceived as semantically related to SNOW; the definition of SNOW must make this speakers' perception ex-

plicit. 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, **2.1.3.1**, p. 00.

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

- The definition of L must ensure the proper combinability of L with L's non-restricted lexical partners, that is, with all LUs that can/cannot cooccur with L according to semantic considerations only. Thus, GRAFT<sup>2</sup> [corruption] cannot be defined as (PRACTICE of obtaining money or advantage by the dishonest use of influence or power...) (LDOCE), 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 GRAFT<sup>2</sup>, which is better in this respect, could be (obtaining money or advantage ...), since the morphosyntactic behavior of gerunds is closer to that of GRAFT<sup>2</sup> 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 contribute to the description of its restricted lexical cooccurrence. Such cooccurrence is covered by syntagmatic lexical functions of L—not by the definition; but the definition should contain components that can accept meanings expressed by L's syntagmatic LFs. Thus, since English has the expression *pass an exam* [*pass* =  $\text{Real}_2(\text{exam})$ ,  $\text{Real}_2$  being 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, **2.1.3.2**, p. 00.)

Thus, a lexicographic definition in an ECD 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., predicting L’s semantically controlled lexical combinability, both free and restricted).

To make an ECD definition conform to these three requirements, two types of lexicographic instructions are needed: instructions that control formal correctness (= well-formedness) of the definition and instructions that take care of its linguistic truth. A definition that is not well-formed, cannot be seriously challenged; but being well-formed is not sufficient for a lexicographic definition: it must also truthfully represent the particular LU L. The well-formedness of a definition is covered by ECD-definition rules (2.1.2); as for the linguistic truth of ECD definitions, they must satisfy several lexicographic criteria (2.1.3).

### 2.1.2 Well-Formedness Rules for ECD Definitions

Five rules are proposed to control the formal correctness of lexicographic definitions: Rule 1 concerns the form of the definiendum, Rules 2 through 4, the form of the definiens, and Rule 5, the equivalence “*definiens*  $\equiv$  (definiendum)” itself.

#### 2.1.2.1 Rule 1: Propositional Form Rule

Let the LU L, which is to be defined, have a signified that is a semantic predicate or a quasi-predicate (see Part II, Ch. 4, 2.2.2, p. 00ff):  $(L(X_1 ; \dots ; X_n))$ . The meanings that correspond to the arguments  $X_1, \dots, X_n$  of  $(L)$  are known as **semantic actants** [= SemAs] of L (for SemAs, see Ch. 12, 3.4, p. 00ff).

L’s definiendum is 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).<sup>viii</sup>

Unlike 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., an LU whose meaning is a predicate or a quasi-predicate), the definiendum is the name of L supplied with variables that stand for L’s SemAs. These variables are symbolized by uppercase Latin letters X, Y, Z, etc. Thus, in order to define REPROACH<sub>V</sub> (as in *The Senate leader reproaches him for recent remarks*), the ECD actually defines the expression *X reproaches Y for Z*; for CHANGE<sub>V</sub> (in *Life has changed into an endless succession of shocks*), the definiendum is *X changes into Y*; and for SKIRMISH<sub>N</sub>, the ECD has to define *skirmish between X and Y over Z*.

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 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 full LUs  $L_1, L_2, \dots, L_n$  such that 1)  $(L) = (L_1) \oplus (L_2) \oplus \dots \oplus (L_n)$  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 uniting of two meanings (see Part I, Ch. 1, p. 00).

#### Exceptions

Two types of LUs do not undergo semantic decomposition:

- 1) semantic primitives, like (not), (feel) or (set);
- 2) absolute synonyms, of which one is semantically decomposed and all the others are simply referred to it: COUGAR (puma) and  $\kappa$ MOUNTAIN LION<sub>1</sub> (puma).

In what follows, these exceptions are ignored.

#### Comments

**1. ‘Two or more full LUs.’** Since the lexicographic definition of  $L$  must be a decomposition of the meaning  $(L)$ , the definiens must include no less than two full LUs. Grammatical, i.e., structural LUs, are not counted, as are not counted ‘light,’ or support, verbs. Thus, the verb ATTACK<sub>V</sub> cannot be defined as  $*(\text{carry out an attack})$ , because CARRY OUT is here a support verb, which has no lexical meaning:  $(\text{attack}_N) = (\text{attack}_V)$ ; therefore, (carry out an attack) is not a legitimate definiens.

**2. ‘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 in a technical sense: ( $X$  is included in the definition of  $Y$ , while  $Y$  cannot be included in the definition of  $X$ ). Therefore, the semantic relation “be semantically simpler than,” as understood here, does not apply to every pair of meanings: it is applicable only to a pair of meanings such that one can 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 a rather obvious example. Which is simpler:  $(\text{manII}) (= (\text{adult male human}))$  or  $(\text{woman}) (= (\text{adult female human}))$ ? Following A. Wierzbicka (1972: 44ff), we say that the meaning  $(\text{woman})$  can—and must—be defined without mentioning  $(\text{manII})$ , exclusively by a woman's capacity of giving birth; but  $(\text{manII})$  is impossible to define without mentioning  $(\text{woman})$ , since  $(\text{manII})$  has to be defined by the capacity of causing<sup>2</sup> that a woman gives birth.<sup>ix</sup> As a result, the meaning of the lexeme  $\text{MAN}^1$  (adult male person) includes the meaning of the lexeme  $\text{WOMAN}$  (adult female person), but not vice versa;  $\text{WOMAN}$  is thus semantically simpler than  $\text{MAN}^1$ .

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  $\text{ASTRONOME}$  /astronOm/ (astronomer) and  $\text{ASTRONOMIE}$  /astronOm+i/ (astronomy).<sup>x</sup> One can say that  $(\text{astronome})$  is  $\approx$  (person who does astronomy), defining  $\text{ASTRONOME}$  via  $\text{ASTRONOMIE}$ . Inversely, it is also possible to say that  $(\text{astronomie})$  is  $\approx$  (science done by astronomers), and then  $\text{ASTRONOMIE}$  is defined via  $\text{ASTRONOME}$ . Both statements are factually true. However, from a lexicographic viewpoint, only the first is acceptable as an ECD definition; here is why.

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

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

$$(\text{astronomie}) = (\text{science done by people who study celestial bodies}).$$

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

$$(\text{astronomie}) = (\text{science of celestial bodies}).$$

In this way, we come back to the first statement:  $(\text{astronome})$  must be defined as (person who does astronomy), but not the other way around.<sup>xi</sup>

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.

**3. Semantic decomposition.** Rule 2 means that a definition of (L) must of necessity be (L)’s decomposition; the methodology of semantic decomposition, launched in the 1960s by A. Wierzbicka, is central to MTT. 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 (retort, react) (*Elle riposta en éclatant d’un rire fou* (She reacted [to this] by bursting into hysterical laughter)) simply by *répondre* (answer), *répliquer* (rejoin) or *réagir* (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 *répondI.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, for instance, the following cases:

(2) French

**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 ).

[The reference is to the word *Merde !* (Shit!), uttered by Marshal Cambronne, the commander of the Old Guards in the battle of Waterloo the 18<sup>th</sup> of June 1815, as an answer to the British proposal to surrender.]

**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 (*répondreI.2*), so that the latter is a poorer synonym of the former; therefore, using (*répondreI.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 riposte )]. 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 ).

- In an ECD, vicious circles in the system of lexicographic definitions, which are one of the current plagues of practically all existing dictionaries, are successfully avoided.<sup>xii</sup>

**4. 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 started and vigorously developed by A. Wierzbicka; the reader is referred to her work: Wierzbicka 1972, 1980, 1987, 1991 and 1996, as well as Goddard & Wierzbicka 1994.

For the time being, the Meaning-Text approach does not have a more or less definitive list of semantic primitives, or **semes**. I believe that they must appear as a result of thorough lexicographic research: successive decompositions of lexical meanings of a language inevitably lead to semes. But an educated guess as to what could be good candidates for seme status in English can be made already now; for instance, the following meanings are, in all probability, semantic primitives:

( something )	( set ) [in mathematical sense]
( more.than )	( space )
( no/not )	( time )
( and )	( say )
( or )	( this.speech.act )

But, however interesting and important, the question of semantic primitives cannot be seriously discussed here. I will limit myself to two remarks.

- It is not the case that semantic primitives of **L** are not decomposable at all: they cannot be decomposed 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 “ $\neg$ ” (= NO/NOT) is defined easily:

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

This is a good definition, yet it cannot appear in a dictionary of English. Other examples of the same type of lexicographically inadmissible definitions include WATER  $\equiv$  (H<sub>2</sub>O), LIGHT  $\equiv$  (electromagnetic waves of frequency  $\phi$ ), or CAT  $\equiv$  (*felis felis*) (the meanings of these LUs are not semantic primitives: they should and can be properly defined, i.e., decomposed, in an ECD-style dictionary). Such definitions characterize the thing denoted by the LU in question, 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 MTT, we should speak of the semantic primitives of English, Chinese, Swahili, Totonac, etc. This does not preclude the sets of semantic primitives for different languages being (or, rather, almost being) in one-to-one correspondence; 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 (= each one is carrying different meanings) nor 2) synonymous expressions (= several ones are carrying the same meaning).

#### Comments

**1. Non-ambiguity of defining elements.** The first constraint means that each elementary 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 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; yet for a non-native it can create problems, and from a scientific viewpoint it is unacceptable. To illustrate the problem of ambiguous expressions in definitions, consider the definition of the French lexeme HAUTEUR<sub>I.1</sub> (height [e.g., of a tower]) as presented in PR 2001:

HAUTEUR<sub>I.1</sub> α (dimension dans le sens vertical, de la base au sommet)  
 [(dimension, in the vertical direction, from the base to the top)]<sup>xiii</sup>

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 under analysis has to be rewritten as follows:

HAUTEUR<sub>I.1</sub> ≡ (dimension<sub>I.2</sub> dans le sens<sup>2</sup> vertical, de la base<sub>I.1</sub> au sommet<sub>I</sub>)

Lexicographic distinctive numbers here are borrowed from the same dictionary, which has them, but does not use them in its own definitions (for a 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 chapter will use the disambiguated elements, with lexicographic numbers borrowed from *LDOCE Online*. (Structural words, used in the definitions for readability, are of course not disambiguated. Since the sense differentiation in *LDOCE* is not always satisfactory, the disambiguation shown here is, on many occasions, quite approximate—which is not too serious a drawback 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 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 servicing 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*, *servicing* [*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 lexicographic number it remains difficult to process.) *Allowing*, *servicing* 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) 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 will be reformulated as follows:

WATCH : ( device<sup>1</sup> designed<sup>2</sup> for showing<sup>5</sup> the time<sup>1</sup><sub>2</sub> ... )

HAMMER : ( tool<sup>1</sup> designed<sup>2</sup> for striking<sup>1</sup><sub>2</sub> ... )

KNIFE : ( instrument<sup>1</sup> designed<sup>2</sup> for cutting<sup>1</sup><sub>2</sub> ... )

SPOON : ( utensil designed<sup>2</sup> for carrying<sup>1</sup><sub>1</sub> food<sup>1</sup> to the mouth<sup>1</sup><sub>1</sub> ... )

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 Sem-configuration  $(L_1) \oplus (L_2) \oplus \dots \oplus (L_n)$  such that it is semantically equivalent to the meaning of a LU  $L'$  that exists in  $\mathbf{L}$ , so that

$$(L_1) \oplus (L_2) \oplus \dots \oplus (L_n) = (L'),$$

then  $(L')$ , and not the above Sem-configuration, must appear in the definition.

The semanteme  $(L')$  is the maximal block with respect to the Sem-configuration  $(L_1) \oplus (L_2) \oplus \dots \oplus (L_n)$ .

In other words, Rule 4 (first formulated in Apresjan 1969a: 14, 1969b: 421; see also 1974: 95) requires obligatory semantic reduction within a lexicographic definition; the semantic decomposition of a meaning 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.

In contrast to Rules 1 - 3, Rule 4 is not logical—in the sense that a good definition that follows this rule is equivalent to a good one that does not. However, it is important methodologically: Rule 4 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 one 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  $\mathbf{L}$  directly in terms of semantic primitives must first have at his disposal a well-established set of semantic primitives of  $\mathbf{L}$ , 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  $\mathbf{L}$ .

- A definition written only in terms of semantic primitives is very long and complex, which makes it unwieldy: not only the dictionary user, but the lexicographer himself will find it difficult

to work with. Worse still, the speaker's intuition balks at the evaluation of such definitions. 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 shortcomings that the notion of the maximal block is introduced.

#### 2.1.2.5 Rule 5: Mutual Substitutability Rule

A lexicographic definition of L in an ECD 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 metalinguistic ones)—*salva significatione* (i.e., stylistic elegance or even normal lexical cooccurrence may be violated).

The substitutability of L and its definition required here is not the substitutability *salva veritate*, i.e., with the preservation of the truth-value, which is expected in many philosophical approaches to synonymy and hence to the theory of defining. The substitutability that Rule 5 requires is the substitutability *salva significatione*, i.e., with the preservation of the same meaning. Once again, we see to what extent 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*  $\equiv$  (|[information Y being hitherto hidden<sup>1</sup> from Z by someone and X believing that Z would like to know Y,]|  
X directly<sup>1</sup> causes<sup>2</sup> that Y becomes<sup>1</sup> known<sup>3</sup> by Z)  
(cf. Wierzbicka 1987: 308-309)

**b.** *Y becomes W*  $\equiv$  (Y begins<sup>1</sup> to be<sup>2</sup> W)

and

**c.** *Y is [= BE<sup>2</sup>1] known by Z*  $\equiv$  (information Y is<sup>2</sup> in Z's mind<sup>1</sup>)

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

**d.** *X reveals Y to Z*  $\equiv$  (|[information Y being hitherto hidden<sup>1</sup> from Z by someone and X believing that Z would like to know Y,]|  
X directly<sup>1</sup> causes<sup>2</sup> that Y begins<sup>1</sup> to be<sup>2</sup> in Z's mind<sup>1</sup>).

Indeed, *John revealed to all his colleagues that he had traveled to Coruña* does mean (John directly<sup>1</sup> caused<sup>2</sup> [by saying, writing, or showing something] that the information «John has traveled to Coruña», which was hitherto hidden<sup>1</sup> from all John's colleagues, began<sup>1</sup> to be in the mind<sup>1</sup> of all John's 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 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 Wierzbicka's effort, the requirement of 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 a particular natural language. On the other hand, where this rule is classified is not that important—provided it is observed.

The rule of mutual substitutability of the definiendum and the definiens guarantees, among other things, the correct constitution of the definiens: it must contain only necessary semantic components, and the set of all its semantic components must be sufficient for substitutability. The presence of a non-necessary component ('L is overdefined') leads to (L) being not

substitutable by its decomposition in certain contexts; and the absence of a necessary component ('L is underdefined') entails the decomposition of (L) being substitutable not only for (L), but for some other semantemes different from (L).

### 2.1.3 Criteria for Linguistic Truthfulness of 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; in order to be sufficient it must also be:

- explicitly linked with all semantically related definitions in the dictionary;
- 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 a definition 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 Criteria of Type I, or Criteria I, since later I will introduce another type of lexicographic criteria, Criteria II—for distinguishing LUs within a vocable (see 3.2 below, p. 00).

Let me start with the criterion 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 denotation 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 real-world entities that are not **Ps**. 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: 2.1.1, p. 00.) Criterion I.1 helps to give the answer.

### 2.1.3.1 Criterion I.1: Linguistic relevance of a semantic component

Let there be  $(\sigma)$ , a semanteme or a configuration of semantemes, which is a candidate for the inclusion in the definition of L;  $(\sigma)$  reflects a non-constitutive, but typical property of the referent of L, so that one feels tempted to have  $(\sigma)$  as a component in the definition of L. However, the necessity of  $(\sigma)$  in the definition of L is not immediately seen— $(\sigma)$  is, so to speak, a dubious component.

|| A dubious semantic component  $(\sigma)$  must be included in the definition of L iff language **L** has at least one other LU  $L'$  that is formally linked to L and has  $(\sigma)$  in its meaning.

The existence of the LU  $L'$  with the indicated characteristics demonstrates the linguistic relevance of  $(\sigma)$  in the definition of L.

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

- Polysemy:  $L'$  is another LU of the same vocable to which L belongs.
- Derivation:  $L'$  is an LU derived from L.
- Phraseology:  $L'$  is a phraseme that contains L.

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

#### Examples

##### Polysemy

(4) L = CLOUDI (accumulation<sup>2</sup> of grayish white<sup>1</sup> substance<sup>1</sup> ... that partially hides<sup>1</sup> the sky);

$(\sigma) = (... \text{ that partially hides}^1 \text{ the sky});$

$L' = \text{CLOUDIII} [on N_Y] (\text{fact X ... that (partially) spoils}^1 \text{ the positive}^1 \text{ character}^3 \text{ of the fact Y [kas if}^1 \text{ X were a cloudI that partially hides}^1 \text{ the sky]})$  (as in *This sad news was the only cloud on the otherwise excellent vacation* or *Chandra casts a cloud on the anti-matter theory*).

The existence of CLOUD<sup>III</sup> shows the linguistic relevance of the component ( $\sigma$ ) in the definition of CLOUD<sup>I</sup>. The semantic link between CLOUD<sup>III</sup> and CLOUD<sup>I</sup> is obvious to an English speaker—it is a comparison with CLOUD<sup>I</sup> [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. 11.3 below, p. 00), two lexemes of the same vocable should explicitly manifest their semantic bridge. As a result, we have to include the component ( $\sigma$ ) in the definition of CLOUD<sup>I</sup>, which allows us to have the component ( $\kappa$ as if<sup>1</sup><sub>1</sub> X were a cloud<sup>I</sup> that (partially) hides<sup>1</sup><sub>2</sub> the sky ... ) in the definition of CLOUD<sup>III</sup>, and the semantic link—a semantic bridge—is ensured.

### Derivation

#### (5) Russian

L = KIRPIČI (brick<sup>1</sup><sub>1</sub>) = (artifact designed to be used as a building material—block made of baked<sup>II.1a</sup> clay, of a medium size, heavy, (of a dark brownish red color) and having the form of rectangular parallelepiped);

( $\sigma$ ) = (of a dark brownish red color);

L' = KIRPIČNYJII [*cvet* (color)] (dark brownish red color—that of a kirpičI) (*kirpičnyj rumjanec* (dark brownish red face color)).

The component (of a dark brownish red color) in the definition of KIRPIČI ensures a semantic bridge with the derived adjective KIRPIČNYJII; it indicates the basis of an obvious comparison. (The semantically equivalent English noun BRICK<sup>1</sup><sub>1</sub> does not have a corresponding component: *brick* is not used to denote a color.) Note that this component is weak (which is shown by parentheses): it can be suppressed by the context: it is quite normal to say, for instance, *seryj kirpič* (grey brick<sup>1</sup><sub>1</sub>).

The presence of the component (having the form of a rectangular parallelepiped) is justified by the existence in Russian of KIRPIČII (manufactured quantity of solid substance X having the form of a rectangular parallelepiped—as if it were a brick<sup>1</sup><sub>1</sub>) and KIRPIČIII (‘No Entry’ traffic sign, which is a red rectangle having the form of a long parallelepiped face and reminding one of a brick<sup>1</sup><sub>1</sub>). Here, as we see, the polysemy considerations are involved.

The component (heavy) is based on the existence of the noun KIRPIČIIV, a *plurale tantum*, meaning (something heavy): — *Tam u tebjā čto, kirpiči, čto li?* (What do you have there, some bricks or what?); — *Èta sumka kirpičami nabita* (This bag is filled with bricks); cf. also the idiom *s duši kirpič [svalilsja]* lit. (from the soul a brick [fell]) = (X felt relief).

Phraseology

(6) L = SNOW<sub>I</sub> (white<sup>1</sup><sub>I</sub> cold<sup>1</sup><sub>I</sub> substance<sub>I</sub>...);

(σ) = (white<sup>1</sup><sub>I</sub>);

L' = [WHITE]<sub>KAS</sub> SNOW<sub>I</sub>, SNOW-[WHITE]

Let it be added that the presence of the component (white<sup>1</sup><sub>I</sub>) in the lexicographic definition of SNOW<sub>I</sub> is confirmed by the polysemy and the derivation tests:

- English has SNOW<sub>II</sub> (cocaine in powder form—WHITE as SNOW<sub>I</sub>).
- English has SNOW<sub>YII</sub> (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<sup>1</sup><sub>I</sub>). Thus, our approach is strictly lexicological, not at all encyclopedic.

**2.** The semantic bridge between L and L' can be based on a connotation of L (see Ch. 15) rather than on a component in the definition of L. Thus, STONE<sub>N</sub> has a connotation (cruel indifference), which is justified by the existence of the adjective STONY, as in a *stony heart/stare*: (insensitive—as if it were of stone).

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, that is, L's cooccurrence with qualifying modifiers, quantifying modifiers and negation. (For a special discussion of the links between semantic components of L's ECD-type definition and the cooccurrence of L, see Kahane 2003b.)

**2.1.3.2 Criterion I.2a: Cooccurrence with qualifying modifiers**

|| The definition of L must explicitly reflect L's cooccurrence with qualifying modifiers: it must include a semantic component (σ) capable of 'accepting' the meaning of the given modifier M, i.e., technically, of being the argument of the corresponding predicate (M).

The modifier M can be free or lexically restricted, i.e., expressed by an LF.

**Examples**Adjectival modifier

The noun APPLAUSE readily accepts adjectival modifiers of the type  $\text{Magn}/\text{AntiMagn}$ : *deafening* <*frenetic, frenzied, thunderous*> or *scattered* <*subdued, thin*>, which express intensification/attenuation; therefore, the definition of APPLAUSE must include a semantic component ( $\sigma$ ) that allows for this kind of qualification. Here is a tentative definition (( $\sigma$ ) here and below is shown in small caps):

*X's applause to Y for Z*  $\equiv$  (Repeated clapping<sup>1</sup><sub>1</sub> by X as a sign<sup>1</sup><sub>2</sub> of approval<sub>2</sub> by X of Y's Z, THE FORCE<sup>1</sup><sub>4</sub> AND RATE<sup>1</sup><sub>1</sub> of the clapping<sup>1</sup><sub>1</sub> being<sup>2</sup><sub>4</sub> proportional to the degree<sub>3</sub> of X's approval<sub>2</sub>).

Adverbial modifier

Fr. BATTRE<sub>II</sub> (beat, defeat) (as in *Jean a battu Pierre au tennis* (J. beat P. at tennis)): the verb *battre* can take adverbial intensifiers of the type  $\text{Magn}$ , 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 BATTRE<sub>II</sub> cannot use as the central component (avoir le dessus) = (have the upper hand), defining BATTRE<sub>II</sub> as (avoir le dessus sur ...) = (have the upper hand over ...) (as does, e.g., PR 2001), because  $\kappa\text{AVOIR LE DESSUS}_1$  is an idiom —i.e., one LU, so that it does not represent the decomposition of the meaning (battre<sub>II</sub>). (Moreover, the meaning *avoir le dessus* is not easily intensifiable: *avoir le dessus* <sup>?</sup>*complètement* <<sup>?</sup>*totalement*> (have completely <totally> upper hand).) Here is what can be proposed as a better definition:

*X bat*<sub>II</sub> *Y dans Z pour W* E (|[X and Y being<sup>2</sup><sub>3</sub> opposed<sub>1</sub> in struggle<sub>1</sub> Z over W,] X causes<sub>2</sub> such damage<sup>1</sup><sub>1</sub> to Y that Y is<sup>2</sup><sub>3</sub> UNABLE to continue<sub>1</sub> Z, as a result of which Y does not obtain<sub>1</sub> W).<sup>xiv</sup>

The choice of the above examples is not fortuitous: the modifiers shown are restricted lexical cooccurents of L—elements of the values of the LF  $\text{Magn}$ . 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 of L and L's restricted lexical cooccurrence (= L's LFs) is one of guiding principles in the developing and presenting ECD entries.

Note that there can be more difficult cases where it is not immediately obvious what component of the definition is targeted by a given modifier. Thus, in the French collocation *célibataire endurci*  $\approx$  (confirmed bachelor), the adjective ENDURCI is a Magn of BACHELOR; a legitimate question arises about the component that ‘accepts’ this modifier. But ENDURCI means here (who wants to remain bachelor) and bears on the central component of (bachelor)—on (man). See the discussion of this case below, in 4.2.1, p. 00, example (22b).

### 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 *\*des ails* <*\*des aulx*> ([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 means only (different sorts of ...) rather than (several units of ...). Moreover, unlike the three other nouns, AIL does not combine with numerals:

(7) French

a. *Apporte-moi un* <trois> *oignon* <s>/*une* <trois> *carotte* <s>/*un* <trois> *chou* <x> !

(Bring me an <three> onion <s>/a <three> carrot <s>/a <three> cabbage <s> !)

but

b. *\*Apporte-moi un ail* <trois ails/aulx> ! lit. (Bring me a <three> garlic <s> !)

The only way to say (7b) correctly is to use the ‘counter’ TÊTE (head): *Apporte-moi une tête d’ail/ trois têtes d’ail* (Bring me a head/three heads of garlic).

Examples in (7) show that the four nouns cannot be defined in the same way. AIL (garlic) is (assaisonnement<sub>1</sub> qui ...) = (seasoning that ...), so that AIL is a substance, not a unit; Fr. ASSAIS-SONNEMENT<sub>1</sub> (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, the semanteme (seasoning) will impose on AIL its own morphosyntactic behavior. (The noun AIL has another sense: (domestic plant whose bulbs produce ail<sub>1</sub>), see below.) OIGNON

(onion) , CAROTTE (carrot) and CHOU (cabbage) in (7) can be tentatively defined as (unité de légume qui ...) = (unit<sub>1</sub> of vegetable<sup>1</sup> that ...).

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

(8) **a.** *manger de l'oignon* <des oignons>/<\*de la carotte> *des carottes/du chou* <des choux>

lit. (eat of the onion <of.the onions>/<of the carrot>/of.the carrots/of.the cabbage <of.the cabbages> )

**b.** *aimer bien l'oignon* <<sup>?</sup>les oignons>/<<sup>?</sup>la carotte> *les carottes/le chou* <<sup>?</sup>les choux>

lit. ([to] like the onion <the onions>/<the carrot> the carrots/the cabbage <the cabbages> )

**c.** *L'oignon* <Les oignons>/*La carotte* <Les carottes>/*Le chou* <Les choux>/*L'ail* <\*Les ails/aulx> *pousse*<nt> *bien dans cette région*

lit. (The onion <The onions>/The carrot <The carrots>/The cabbage <The cabbages>/The 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 (7) - (8):

1. (Unit<sub>1</sub> of vegetable<sup>1</sup> U that ... [size, form, color, consistency, taste, ...]) [= (7)]
2. (Edible substance<sub>1</sub> of U ... ) [= (8a-b); this sense is also valid for AIL, since (seasoning) = (edible substance<sub>1</sub> that ...)]
3. ((Class<sub>3</sub> of) plant(s)<sup>2</sup> U that produce(s)<sub>2</sub> U ... ) [= (8c)]

These definition 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 substance<sub>1</sub> 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* (see (8b)).

Criterion I.2b not only facilitates the differentiation of the lexemes of a vocable, but also helps choose the central, i.e., generic, component of a lexeme's definition. Thus, in a Type 1 sense ((unit<sub>1</sub> of vegetable<sup>1</sup> U that ...)), the generic component (unit<sub>1</sub>) 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 WIDOW<sub>1</sub> (characterization of the family status of a woman) is formulated like this (adapted from *LDOCE Online*):

(9) a. *X is a widow* ≡ (X is a woman<sub>1</sub> who has lost<sub>6</sub> her husband<sub>1</sub> and has not remarried<sub>1</sub>).

NB: In *John's widow later married my brother* we have a different LU—WIDOW<sub>2</sub>: *X is the widow of Y*.

With the definition in (9a), the sentence *Zhu is not a widow* would mean (Zhu is not a woman<sub>1</sub> who ...). But in reality this sentence negates the facts that 1) Zhu lost her husband and 2) has not remarried, but affirms that Zhu is an adult woman: according to the meaning of (9a), Zhu cannot be a male or a little girl. To reflect this property in the definition of WIDOW<sub>1</sub>, the component (woman<sub>1</sub>) must have a status different from that of the component (has lost<sub>6</sub> her husband<sub>1</sub> and has not remarried<sub>1</sub>): only the latter can be negated, when we negate the whole meaning (widow). More specifically, (woman) is a semantic taxonomic restriction on a semantic actant (its semantic type), which functions as a presupposition, while (has lost<sub>6</sub> her husband<sub>1</sub> and has not remarried<sub>1</sub>) 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 woman<sub>1</sub>, has lost<sub>6</sub> her husband<sub>1</sub> and has not remarried<sub>1</sub>).

The presuppositions can also be indicated in an equivalent way: by the symbols “[ ... ]”, put around presuppositions:

(|[X being a woman<sub>1</sub>,] X has lost<sub>6</sub> her husband<sub>1</sub> and<sub>1</sub> has not remarried<sub>1</sub>).

In this book the latter notation is used.

In the reformulated form of (9b), the definition ensures the correct description of the combination with negation:

c. *X is not a widow*  $\equiv$  ( $[[X$  being a woman<sup>1</sup>,] $||$  X has not lost<sup>6</sup> her husband<sup>1</sup> or<sup>3</sup> has re-married<sup>1</sup> ).<sup>xv</sup>

**Example 2**

The verb PERMIT (= ( permit<sup>1</sup> ), i.e.,  $\approx$  ( say yes )) represents a more complex case.

(10) *Leo<sub>X</sub> did not permit me<sub>Z</sub> to go<sub>Y</sub> to France.*

Firstly, X *permits* to do Y only to Z who wants to do Y, and X must know about this wish; sentence (10), 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 PERMIT.

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 (10): 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 to do Y to Z*  $\equiv$  ( $[[$ Knowing<sup>1</sup> that Z wants<sup>1</sup> to do<sup>2</sup> Y, which Z is<sup>2</sup> not supposed<sup>1</sup> to do<sup>2</sup> against X's will<sup>3</sup>,] $||$  X communicates<sup>1</sup> to Z that, according to some reasons<sup>2</sup>, Z's doing<sup>2</sup> Y is<sup>2</sup> not against X's will<sup>3</sup> ) (cf. Wierzbicka 1987: 108-111)<sup>xvi</sup>

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 to do Y to Z*, according to the above definition, is not a negation of *X permitted to do Y to Z*:

*X permitted to do Y to Z*  $\approx$  ( X communicated<sup>1</sup> to Z that Z's doing<sup>2</sup> Y is<sup>2</sup> not against X's will<sup>3</sup> )

vs.

*X did not permit to do Y to Z*  $\approx$  ( X did not communicate<sup>1</sup> to Z that Z's doing<sup>2</sup> Y is<sup>2</sup> not against X's will<sup>3</sup> )

In point of fact,

*X did not permit to do Y to Z*  $\approx$  ( X communicated<sup>1</sup> to Z that Z's doing<sup>2</sup> Y is<sup>2</sup> against X's will<sup>3</sup> ), 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 ( be<sup>2</sup> not against X's will<sup>3</sup> ), giving ( **not** be<sup>2</sup> not against ) = ( be<sup>2</sup> against ). But 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' syntactic negation, but an antonym of PERMIT<sub>v</sub>—a separate LU of English. This has to be stated explicitly in the lexical entry for PERMIT:

PERMIT<sub>v</sub><sup>1</sup>

...

Anti<sub>['not contrary']</sub> : *do not permit*

[Anti—‘antonym’—is a lexical function, see 2.3 below.]

### 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 of L vs. a SemR of L (2.1.4.1); 2) internal structure of ECD definitions (2.1.4.2).

#### 2.1.4.1 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 subareas (i.e., an indication of the division of the meaning into subnetworks marked ‘Rheme ~ Theme’, ‘Given ~ New’, ‘Focalized ~ Non-Focalized’, ‘Presupposed ~ Asserted’, etc.); however, actual ECD definitions are sentences or phrases in a natural language—they are verbal and linear. Thus, in an English ECD definitions must be 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 accept compromises.

- The second reason is more profound: in an English ECD, a formulation of a definition in an English-based semantic metalanguage is much more readily accessible to the linguistic intuition of speakers, including the lexicographer himself (this point was drawn to my attention by A. Polguère; 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.

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 representa-

tions for the meaning of an LU: a network SemR and a verbal definition, plus an algorithm for transforming one into another. However, for the time being, this is still a dream.

#### 2.1.4.2 Internal structure of ECD definitions

The structuring of an ECD definition involves, in the first place, at least the following four facets of the definition: 1) communicative status of its components, 2) logical status of its components, 3) different structural roles played by its components, and 4) inheritance of semantic actants.

##### Communicative 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 (between the symbols “[ ... ]|”). A presupposition within 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. Values of Sem-Communicative oppositions and related notions have been presented in Part II, Ch. 6, p. 00ff.

**NB:** On the distinction of different logical-communicative layers in ECD definitions, see, in particular, Apresjan 1980: 49ff. E. Padučeva proposes ‘formatted definitions’ as the main tool in lexicographic work; see Padučeva 2002 and 2004: 525ff, where four types of lexicographic parameters for structuring 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 each of the actants (‘liquid’ ~ ‘person’ ~ ...). For a general discussion of the problem, see Iordanskaja & Mel’čuk 1990.

##### Logical status of a component in an ECD definition

In a different vein, a semantic components of a lexicographic definition can be a default components: it is present in the given meaning, if nothing in the context contradicts it, but 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 (Zaliznjak 1987: 138ff). Typically, Russian masculine names of profession or nationalities, when in the singular, refer to a male, but in the plural cover both men and women. Thus, consider the following case:

## (11) Russian

a. POËT<sub>masc</sub> (person (of masculine sex)<sub>pl</sub> who writes poetry):

*Ja očen' cenju ètogo poèta<sub>sg</sub>* (I appreciate this [male] poet very much).

vs.

*Ja očen' cenju ètix poètov<sub>pl</sub>* (I appreciate these poets [men and/or women] very much).

b. ISPANEC<sub>masc</sub> (person (of masculine sex)<sub>pl</sub> who belongs to the nationality<sup>2</sup> native of Spain and whose mother tongue is Spanish):

*Ja perepisyvajus' s odnim ispancem<sub>sg</sub>* (I correspond with a [male] Spaniard).

vs.

*Ja perepisyvajus' s neskol'kimi ispancami<sub>pl</sub>*

(I correspond with several Spaniards [men and/or women]).

Here, (of masculine sex) is a weak component; it disappears in the plural, which is indicated by the subscript “<sub>pl</sub>” to the parentheses enclosing it. For more on weak components in lexicographic definitions, see Note 22, p. 00. The derived feminine nouns POËTESSA<sub>fem</sub> (female poet) and ISPANKA<sub>fem</sub> (female Spaniard) refer exclusively to women.

### Different structural roles played by the components in an ECD definition

Each semantic component ( $\sigma$ ) 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 ( $\sigma$ ) 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.;
- 3) constitute a semantic taxonomic restriction on an actant of L;
- 4) modify another semantic component, restraining its content.

For instance, in BAKEII.1a ((X causes<sup>2</sup> that Y, which is raw<sup>1</sup><sub>2</sub> bricks<sup>1</sup> or<sup>3</sup> pottery<sup>2</sup>, hardens<sup>1</sup>—by exposing Y to the action<sup>6</sup> of dry<sup>10</sup> heat<sup>2a</sup> in device<sup>1</sup> Z); see Subsection 5.1, p. 00), the component (... causes<sup>2</sup> that ... hardens<sup>1</sup> ...) expresses a complex relation between SemAs X and Y, while the component (raw<sup>1</sup><sub>2</sub> bricks<sup>1</sup> or pottery<sup>2</sup>) characterizes SemA Y taxonomically; this characterization is necessary to block the use of the lexeme BAKEII.1a to name, for instance, the process of hardening a liquid substance by heating it and making it boil. Taxonomic characterization of a SemA can be overt, as shown above, or covert, i.e., implicit in the decomposition of a

component. Thus, in **BAKEII.1b** (( Y hardens<sub>1</sub> being.baked<sub>II.1a</sub> [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

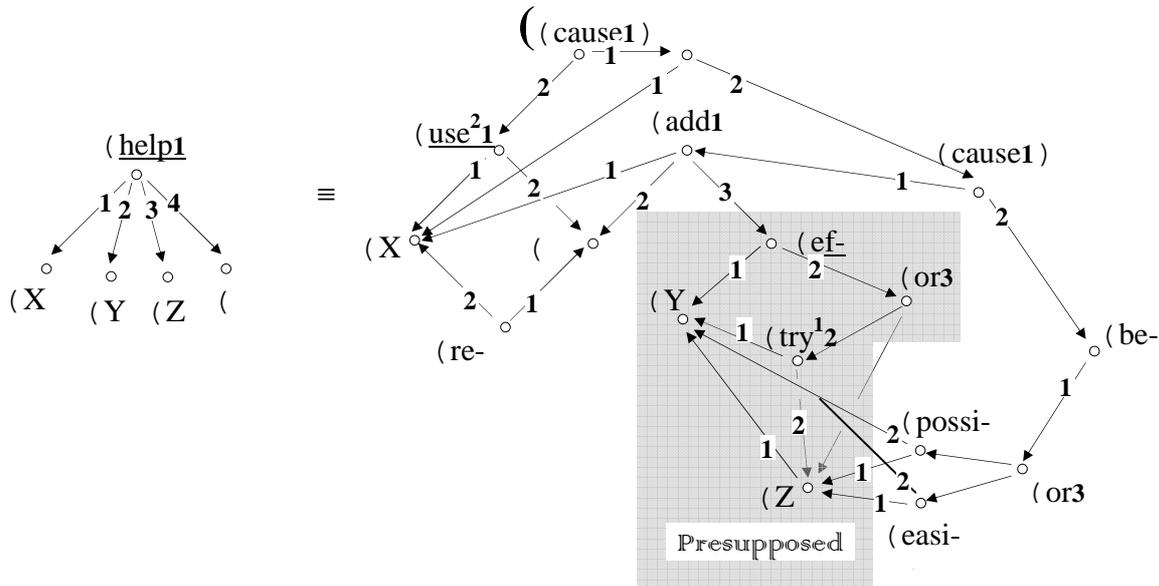
A component ( $\sigma$ ) in the definition of L brings to (L) all of its own SemAs, which must be explicitly accounted for in the definition. For instance, **BAKEI.2a** (( X creates<sub>1</sub> solid<sub>1</sub> food<sub>1b</sub> Y from a mixture<sub>1</sub> ... )) includes (create<sub>1</sub>); the 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 X, Y 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 the verb **COST<sub>V</sub>** (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<sub>Y</sub> cost him<sub>Z</sub> \$30<sub>W</sub> \*with John/\*from John;*<sup>xvii</sup> in the definition of **COST<sub>V</sub>** the corresponding participant is represented by a generic constant, shown in small capitals: *Y costs Z W*  $\equiv$  (merchandise or service<sub>11</sub> Y is paid<sub>1</sub> for the sum<sub>2</sub> W by person<sub>1</sub> Z TO THE PERSON<sub>1</sub> from whom Z is buying<sub>1</sub> 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**.

*X helps<sub>1</sub> Y to Z with W*  $\equiv$  [( Y making efforts<sub>1</sub> trying<sub>12</sub> to do<sub>21</sub> or<sub>3</sub> doing<sub>21</sub> Z,)] X uses<sub>21</sub> X’s resources<sub>1</sub> W by adding<sub>1</sub> W to Y’s efforts<sub>1</sub>, this causing<sub>1</sub> that doing<sub>1</sub> Z becomes<sub>1</sub> possible<sub>1</sub> or<sub>3</sub> easier<sub>1</sub> for Y (, caused<sub>1</sub> by X’s wanting this).

The **weak**, or optional, component ((, caused<sub>1</sub> by X’s wanting this)) is necessary to account for the possible use of the verb **HELP1** to refer to involuntary help: one can say, for instance, *Benazir unknowingly helped Musharraf gain support within the army*. On the other hand, in the sentence *Whole Foods staffers wholeheartedly helped a little girl in need*, the verb **HELP1** refers of course to a voluntary action of helping. The component under discussion can be suppressed by an explicit contradicting context (like *unwillingly* or *unknowingly*); but without such a context, a sentence with **HELP1** is understood as implying that X’s using W is aimed at facilitating Z.

The corresponding Sem-Representation appears as follows:



This is of course a specific lexical Lex-Sem-rule—of the type that were considered in Part IV, Ch. 10, 2.1.2, p. 00ff.

**Comments**

1. The underscoring of a semantic component in a SemS (in this case, (use<sup>2</sup><sub>1</sub>) and (effort<sub>1</sub>)) indicates that it constitutes the Comm-Dom-node of the corresponding Sem-Comm-area.
2. The semantic component (cause<sub>1</sub>) represents non-agentive causation, as in *The falling tree* <*The bullet*> **killed** the dog; (cause<sub>2</sub>) 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 classes of 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—it will not be discussed here. In order to give the reader a vague

idea about passive syntactic valence of an LU, let me mention only three relevant syntactic features:

- The capacity of an adjective to be an adnominal modifier or a copular attributive. Such an adjective as ILL or WET [in the sense of (wearing wet cloths)] cannot be an adnominal modifier (*\*an ill child, \*a wet soldier*), while such an adjective as ANCIENT or FORMER cannot be a copular attributive (*\*This language is ancient, \*This minister is former*).
- The capacity of an LU to be anteposed or postposed in spite of the common property of the corresponding part of speech: for instance, the adposition NOTWITHSTANDING can be anteposed or postposed to the noun it governs, and the adposition AGO can be only postposed to it; or the conjunctions  $\kappa$ SO AS TO<sub>1</sub> or  $\kappa$ SO THAT<sub>1</sub> (together with their subordinate clause) can only follow the main clause.
- Being a particle, that is, depending only on the immediately following wordform and excluded from any coordination.

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**. What is offered here is just a short sketch; for details on the government pattern, see Ch. 13, p. 00ff.

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:  $X \Leftrightarrow \mathbf{I}$ ,  $Y \Leftrightarrow \mathbf{III}$ , etc. For a given pair  $\langle \text{SemA(L)}, \text{DSyntA(L)} \rangle$ , the GP(L) specifies:

- all surface-syntactic dependents of L, actantial and non-actantial, that can realize in the text the DSyntA under consideration;
- all surface-syntactic and/or morphological means for expressing the DSyntA in the text.

Formally, a GP is a rectangular matrix having  $n$  columns (designated C), numbered with Roman numerals: C<sub>I</sub>, C<sub>II</sub>, ..., 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<sub>III.3</sub> stands for "Column III, row 3" and specifies the SSynt-expression "**–oblique-objectival→with N**" (see the GP table for HELP<sub>1</sub> on next page), which means that the SemA

Z of HELP1 can be realized in the text by an Oblique Object having the form of a WITH-phrase. (For more on SemAs and SyntAs, see Ch. 12, p. 00ff.)

As a rule, the form of a SSynt-Dependent is correlated with its SSynt-role: a prepositional phrase is an Oblique Object, a ‘bare’ noun phrase is the Subject (if it corresponds to DSyntA I) or the Direct Object (if it corresponds to DSyntA II), an adjectival phrase is a Modifier, etc. This allows us in most cases not to indicate explicitly the SSynt-role of the Dependent presented in a GP. The corresponding indication is supplied only when we have to deal with an ambiguity.

For each DSyntA its optional/obligatory character has to be stated. The default case is optionality, which is not marked. The obligatory character of a DSyntA is indicated with the mark “obligatory” in the corresponding column (see lexical entries for BAKE in 5.1 below). By convention, the obligatory character of DSyntA I is not marked in the GP for the languages such as English, French, German, Russian, etc., since with a finite verb’s DSyntA I corresponds to the Subject, which is always obligatory in these languages.

The GP table is accompanied by numbered constraints, which specify the cooccurrence of L’s different DSyntAs among themselves, and that 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, verb

**Government Pattern**

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

- 1) C<sub>III.1</sub> : (X being directly involved in Z) [= (X doing Z himself)]<sup>xviii</sup>
- 2) C<sub>III.2</sub> : (X not being directly involved in Z) [= (X not doing Z himself, but providing some resources to Y)]<sup>xix</sup>
- 3) C<sub>III.6</sub> : **if** Z = (travel/move [something] in the direction α),  
**then** [III = L((α)) **and** C<sub>III</sub> = C<sub>III.6</sub>] is possible

[PREP<sub>dir</sub> stands for ‘directional prepositions and adverbs,’ such as *up, out, into, across, there, ...*; L(( $\alpha$ )) stands for ‘L expressing the meaning ( $\alpha$ ).’ 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<sub>III.3,4</sub> + C<sub>IV.1</sub> : 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  $\theta$ -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; the corresponding cases will be presented in Ch. 13, p. 00.

## 2.3 Lexical Functions in the ECD

The Lexical Relations Zone of an ECD entry includes lexical functions [= LFs]; a detailed discussion of LFs is presented in Ch. 14, so that here a brief characterization will suffice.

For a headword L, LFs describe what are known as institutionalized lexical relations of L: L’s semantic derivations and L’s restricted lexical cooccurrence, i.e., L’s collocations. Thus, in the entry for **BAKE1.2a** (p. 00) it is indicated that the device people typically use to bake bread, cakes, etc. in [= S<sub>4</sub><sup>usual</sup> = S<sub>instr-loc</sub>] is called an *oven*, and the person whose profession is baking bread [= professional-S<sub>1D</sub>] is a *baker*; **OVEN** and **BAKER** are semantic derivations of **BAKE1.2a**. If your baking is successful, you bake what you bake *to a turn* [= **VER**], and a cake baked recently [= A<sub>2recently.Perf</sub>] is *fresh-baked*; **BAKE TO A TURN** and **FRESH-BAKED** are collocations of **BAKE1.2a**. Both types of expressions are specified in a lexical entry of an ECD by LFs.

### 2.3.1 Semantic Derivation

Derivation is a well-known linguistic phenomenon: lexeme **B** is said to be **derived** from lexeme **A**, or to be **A**’s **derivative**, iff the signified of **B** includes the signified of **A** and the semantic dif-

ference  $(\delta) = (B) - (A)$  is expressed in language  $\mathbf{L}$  by morphological means. Thus, *smoker* is derived from *smoke<sub>v</sub>*, since  $(\text{smoker}) \supset (\text{smoke}_v)$  ( $(\text{smoker}) = (\text{person who } \mathbf{smokes} \text{ regularly})$ ) and the difference  $(\text{smoker}) - (\text{smoke}_{1v}) = (\text{person who } \dots \text{ regularly})$  is expressed by a suffix: **-er**. The notion of **semantic derivation** is a generalization of derivation in the current sense.

### Definition 11.1: Semantic derivative of the LU $\mathbf{A}$

The LU  $\mathbf{B}$  is a **semantic derivative** of the LU  $\mathbf{A}$  in language  $\mathbf{L}$  iff it satisfies simultaneously Conditions 1 - 3:

1. The signified of  $\mathbf{B}$  includes that of  $\mathbf{A}$  (i.e.,  $\mathbf{B}$  is lexicographically defined in terms of  $\mathbf{A}$ ):  
 $(B \supset (A))$ .
2. The semantic difference  $(\delta)$  between  $(B)$  and  $(A)$  [ $(\delta) = (B) - (A)$ ] either (i) is regular, i.e., appears between members of many lexical pairs in  $\mathbf{L}$  [that is,  $(B_i) - (A_i) = (\delta)$ , where  $i$  is sufficiently large], or (ii) is not regular, but then the element  $(\delta')$  that makes it irregular is negligible ( $\approx$  neutralizable in some contexts), so that it can be ignored and the difference  $(\delta)$  becomes regular.
3. In some cases, the semantic difference  $(\delta)$  is expressed in  $\mathbf{L}$  by morphological means.

Thus, *baker* is semantically (and morphologically) derived from the verb *bake<sub>I.2a</sub>*, because  $(\text{baker}) = (\text{person who } \mathbf{bakes}_{I.2a} \text{ professionally})$ ; *blacksmith* is semantically (but not morphologically) derived from the verb *forge<sub>v</sub>*, because  $(\text{blacksmith}) = (\text{person who } \mathbf{forges} \text{ professionally})$ . In the same vein, *terrestrial* is semantically derived from *earth*, because  $(\text{terrestrial}) = (\text{related to earth})$ , 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  $\mathbf{L}$ . (For more on semantic derivation, see Ch. 14, **1.2**, p. 00).

### 2.3.2 Collocation

A **collocation** is a particular type of set phrase, or **phraseme**; more specifically, one of a collocation’s components 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 defini-

tion (it is repeated, with detailed comments, in Ch. 16, 4.2: Definition 16.10, p. 00; it is followed by a typology of collocations).

### Definition 11. 2: Collocation

A semantic phraseme  $\mathbf{AB} = \langle (S) ; /A/ \oplus /B/ ; \Sigma_{\mathbf{AB}} \rangle$  of  $\mathbf{L}$  is called a **collocation** of the LU  $\mathbf{A}$  iff it satisfies simultaneously Conditions 1-3 :

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:  $(S) = (\mathbf{A}) \oplus (\mathbf{C})$ , such that  $(\mathbf{C})((\mathbf{A}))$ .]
2.  $\mathbf{A}$  is selected by the Speaker unrestrictedly, i.e., independently of  $\mathbf{B}$ —for its own signified  $(A)$ .
3.  $\mathbf{B}$  is not selected freely—it is selected restrictedly: as a function of  $\mathbf{A}$ .

The lexeme  $\mathbf{A}$  is called the **base** of the collocation  $\mathbf{AB}$ , and  $\mathbf{B}$  is  $\mathbf{A}$ 's **collocate**.

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

- *do*  $\langle *make \rangle$  [*someone*] a FAVOR, *give*  $\langle *deliver \rangle$  [*someone*] a LOOK, *take*  $\langle *seize \rangle$  a STEP, *be*  $\langle *find \text{ oneself} \rangle$  in DESPAIR, *commit*  $\langle *perform \rangle$  a BLUNDER, ... [the collocate is a light, or support, verb:  $\approx$  (do); the meaning  $(C)$  is (near-)empty];
- *strong*  $\langle *powerful \rangle$  COFFEE, *heavy*  $\langle *weighty \rangle$  RAIN, *BLUSH* deeply/profusely  $\langle *profoundly \rangle$ , *profoundly*  $\langle *powerfully \rangle$  AFFECT, *as ALIKE as two peas in a pod*  $\langle *as \text{ two drops of water} \rangle$ , ... [the collocate is an intensifier; the meaning  $(C) \approx$  (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; the meaning  $(C) \approx$  (do with/for L what is expected)].

The meaning  $(C)$ , which is expressed in a collocation restrictedly—it expressed by  $\mathbf{B}$  contingent on  $\mathbf{A}$ —is associated with a lexical function, see below. The lexeme  $\mathbf{A}$ , which keeps its signified intact within the signified of the collocation and determines the expression of  $(C)$  by  $\mathbf{B}$ , that is, the base of collocation, is the argument, or **keyword**, of the corresponding LF.

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

### 2.3.3 Lexical Function

Again, I will start with a definition.

#### Definition 11.3: Lexical Function

A correspondence  $\mathbf{f}$  that associates a set  $\mathbf{f}(L)$  of lexical units with an LU  $L$  is called a **lexical function** [= LF] iff it satisfies either conditions **I.1-3** or condition **II**:

**I.**  $\mathbf{f}$  is applicable to several LUs and:

1. Semantic homogeneity of  $\mathbf{f}(L)$

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

$$\frac{L'_1 \in \mathbf{f}(L_1)}{L_1} \approx \frac{L'_2 \in \mathbf{f}(L_2)}{L_2}$$

2. Maximality of  $\mathbf{f}(L)$

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

$$\frac{L'_1 \in \mathbf{f}(L_1)}{L_1} \neq \frac{L'_2 \notin \mathbf{f}(L_2)}{L_2}$$

3. Phraseological character of  $\mathbf{f}(L)$

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

**II.**  $\mathbf{f}$  is applicable to only one LU  $L$  (or perhaps to a few semantically close LUs).

Being applicable to only one LU means, in itself, a restricted cooccurrence and thus phraseological character.

$L$ , which is the argument of  $\mathbf{f}$ , is called the **keyword** of  $\mathbf{f}$ , and  $\mathbf{f}(L) = \{L'_i\}$  is  $\mathbf{f}$ '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, which will be systematically presented in Ch. 14, 2.4, p.00ff).

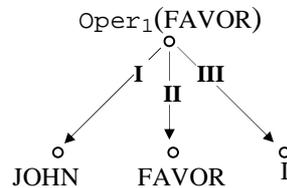
The definition of LF is taken up in Ch. 14 (p. 00), whether major types of LFs are introduced and discussed. Here I will limit myself to just two examples.

**Examples of LFs**

1) The LF *Magn* represents an intensifier:

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

2) The LF *Oper<sub>1</sub>* 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), *Oper<sub>1</sub>* takes the DSyntA I of L; DSyntA II of L often, i.e., with some Ls, becomes DSyntA III of *Oper<sub>1</sub>* (sometimes it remains DSyntA II of L). Cf. the representation of the sentence *John does me a favor* at the DSynt-level:



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

For some LUs, especially those that refer to complex objects having many and various uses, the number of LFs of all types can be very high. Moreover, the meaning of the LU L can be considered under different ‘facets,’ such that in each one, the same LFs can have different values. To facilitate the presentation and retrieval of LFs in such a case, the ECD uses thematic groupings of LFs. 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, artery, clot, ...*)
- Blood as target of medical treatment (*hematology, hemophilia, transfusion, blood test, leukemia, donor, blood pressure, blood coagulation, ...*)
- 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, pool of blood, ...*)

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 (*SOS, mayday, capsizes, runs aground, sinks, life boat, life raft, life vest, ...*)
- Ship as a military unit (*navy, naval battle, send to the bottom, cruiser, destroyer, corvette, submarine, ...*)

## 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, tex-

tual research is an absolute must, but examples found in texts should be gone through with a fine-tooth comb 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 Generative-Transformational Grammar school in 1960s. (Incidentally, the utility of negative examples in a dictionary was emphasized more than 65 years ago by the Russian lexicographer L. Ščerba (1940). Unfortunately, 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:

- Three basic notions, necessary for the discussion of DEC’s super-entries: semantic bridge, semantic field, and vocable (3.1).
- Formal criteria for distinguishing the LUs within a vocable (3.2).

#### 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 on the basis of their semantic relatedness (to be grouped into one vocable, the LUs must also feature a common signifier: the stem of all their wordforms). 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 11.4: Semantic Bridge**

A **semantic bridge** between LUs  $L_1$  and  $L_2$  is a Sem-configuration ( $\sigma+$ ) that is shared by the lexicographic definitions of  $L_1$  and  $L_2$  and satisfies simultaneously Conditions 1 - 2:

1. ( $\sigma+$ ) constitutes a sufficiently important part of these definitions.
2. ( $\sigma+$ ) occupies a sufficiently central position in these definitions.

Definition 11.4 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 Sem-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<sub>V</sub> ((X murders Y) includes (X causes2 [that Y dies])) and in that of CLEAN<sub>V</sub> ((X cleans Y)  $\approx$  (X causes2 that Y becomes clean)); however, it is obvious that MURDER<sub>V</sub> and CLEAN<sub>V</sub> 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 by themselves a semantic bridge between two LUs: such as (cause1, 2), (act1, 2), (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 several 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 is the part it occupies within the respective definition.

- The most central position in a definition is the **generic component** (Y): (X)  $\equiv$  (Y which Z). Although a semantic bridge is not necessarily the generic component, but if a semantic bridge is the generic component it plays a very special role, as we will immediately see.

A special type of semantic bridge should be mentioned—(as if it were ...) component, used to represent a metaphorical semantic link between two lexemes of the same vocable. Thus, PIG2  $\equiv$  (very untidy and unpleasant person—as if he were a pig1) (PIG1 being (farm animal that

...). Without the (as if ...) component there would be no semantic bridge between PIG1 and PIG2, and the two lexemes would not be put into the same vocable. Three further examples:

BOIL<sub>V5</sub> ≡ (be in a very agitated state—as if X were boiling<sub>1</sub>) (*He was boiling with rage*)

TON2 ≡ (very big quantity—as if X were a ton<sub>1</sub>) (*He brought with him a ton of books*)

COLD5 ≡ (lacking normal human feelings and unfriendly—as if X were cold<sub>1</sub>) (*He was as cold as ice when he spoke*)

By way of a more detailed illustration, consider the vocable [to] BAKE, given below in 5.1. Eleven lexicographic senses of the verb BAKE, i.e., eleven LUs, are distinguished; all are put together, to form one vocable. The reason is that BAKE lexemes exhibit a semantic bridge: all of them share an important semantic component—(cause<sub>1, 2</sub> ... by the action of dry heat), while some of them share more. (Of course, the component (cause<sub>1, 2</sub> ... 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.)

Now semantic field and vocable can be readily defined.

### Definition 11.5: Semantic field

A semantic field  $\mathbf{F}_{(\sigma)}$  is the set  $\{L_i\}$  of all LUs that share the semantic bridge  $(\sigma+)$  that satisfies one of Conditions 1- 2:

1. either  $(\sigma+)$  is the generic component in  $L_i$ 's definition [the most common case];
- 2) or  $(\sigma+)$  is linked to the generic component by a Sem-component underlying an LF

[the

less common case].

The semantic component  $(\sigma+)$  is the semantic field identifier. If  $(\sigma+)$  is not the generic component  $(\sigma^{\text{gener}})$  of  $(L)$ , it can be related to this  $(\sigma^{\text{gener}})$  as, for example:

- An actant or a characteristic of an actant: PEN (an artifact designed to write with) is a Sem-actant of WRITE, and therefore PEN belongs to  $\mathbf{F}_{(\text{writing})}$ ; SINGER (individual who sings) is equally a Sem-actant of SING, so it belongs to  $\mathbf{F}_{(\text{singing})}$ ; CHOIR (set of singers) is also part of  $\mathbf{F}_{(\text{singing})}$ .

- A place name: RESTAURANT (establishment where people serve food to other people) is an  $S_{\text{loc}}$  of EAT; RESTAURANT thus belongs to  $\mathbf{F}_{(\text{eating})}$ .

- An 'instrument' name: PILLOW, MATTRESS, BED SHEET, etc.  $\approx$  (artifacts designed to be used to sleep) are  $S_{\text{instr}}$  of SLEEP; the above nouns belong to  $\mathbf{F}_{(\text{sleeping})}$ .

- A ‘when’ name: DREAMS happen (when you are asleep); so DREAM also belongs to  $\mathbf{F}_{(\text{sleeping})}$ .
- A ‘to-be-used-for’ name: SLEEPING PILLS and LULLABY.

In other words, an LU belongs to  $\mathbf{F}_{(\sigma)}$  iff  $(\sigma)$  is the generic component in  $(L)$  or at least  $(\sigma)$  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  $\mathbf{F}_{(\sigma)}$ . Thus, LEG of my pants or FINGER of my glove do not belong to  $\mathbf{F}_{(\text{body parts})}$ : they only (cover) the respective body part, and (cover) is not a regular semantic relation in English.

As an example of a semantic field whose LU share the generic component of their definitions one can cite names of nationalities. The nouns ENGLISH, CHINESE, FRENCH, GERMANS, ITALIANS, RUSSIANS, etc. all belong to  $\mathbf{F}_{(\text{ethnic group})}$ ; the definition of any such name contains the semanteme (nationality2) as its generic component ( $(\text{nationality1}) \approx (\text{citizenship})$ ):

[*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 Sem-field because they have the same generic component also tend to have definitions with the same general structure, as can be immediately seen from the preceding example of nationalities.

The semantic field of sleeping  $\mathbf{F}_{(\text{sleeping})}$  is different: it includes many LUs whose generic components are not identical, for instance (the generic component is in small caps, and the semantic bridge, which is here the semantic field identifier, is in boldface):

BED1 (PIECE OF FURNITURE designed for X to **sleep** in)

SNORE1 (X, who is **sleeping**, PRODUCES NOISE with X’s throat)

SLEEPY ([person X] FEELS NEED to **sleep**)

κWAKE UP1 (X CEASES to **sleep**)

κALARM CLOCK1 (CLOCK designed to make an alert sound at a set moment in order to cause that X ceases to **sleep**)

The definitions of such members of the same Sem-field do not of course have the same structure of definitions, but still they might show certain parallelisms.

Traditional dictionaries are, as a rule, compiled in alphabetic order of entries; ‘What letter are you at now?’ is a typical question 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 impos-

sible to write it in any other way. Only this technique guarantees the homogeneous description of all the LUs belonging to one semantic field.

### Definition 11.6: Vocable

A **vocable** is the set  $\{L_i\}$  of LUs such that any two LUs  $L_1, L_2 \in \{L_i\}$  satisfy simultaneously Conditions 1- 2:

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  $(\sigma_1)$ ,  $L'$  and  $L''$  have a semantic bridge  $(\sigma_2)$ , and  $L''$  and  $L_2$  have a semantic bridge  $(\sigma_3)$ ].

### 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  $PEN^1 \approx$  (writing implement) vs.  $PEN^2 \approx$  (female swan) vs.  $PEN^3 \approx$  (enclosure for animals) vs.  $PEN^4 =$  (penitentiary). These LUs are homonyms of a special kind: namely, they are also **homographs**. Here are some more examples of homographs: nouns  $RENT^1 \approx$  (regular payment by a tenant)  $\sim$   $RENT^2 \approx$  (breach, schism)  $\sim$   $RENT^3$  slang for (parent); or  $DATE^1$  (sweet fruit ...)  $\sim$   $DATE^2$  (indication of a time moment—the name of the day, month, and year)  $\sim$   $DATE^3$  (romantic meeting of two people); verbs Fr.  $VOLER^1$  ([to] fly)  $\sim$   $VOLER^2$  ([to] steal). Such LUs share no important common semantic components (= they have no semantic bridges).

Since an ECD is a dictionary that stores LUs in the written form, homographs that are not homophones should be also distinguished by superscripts:  $ROW^1$  (objects arranged in a line) (/rō/) vs.  $ROW^2$  (noisy quarrel) (/ra<sup>u</sup>/).

Different LUs of the same vocable are distinguished by lexicographic numbers: Roman and Arabic numbers and small Latin letters; for instance,  $LI.1a \sim LI.1b \sim LII \sim LII \sim LIII.a \sim LIII.b$ . The numbering of LUs within a vocable is done as a function of semantic distance between two LUs. The semantic distance between LUs  $L_1$  and  $L_2$  is measured by two parameters considered together:

—The size of the semantic bridge (= shared semantic component) between  $L_1$  and  $L_2$ : the bigger the semantic bridge, the closer  $L_1$  and  $L_2$  are.

—The regularity of the semantic distinction ( $\delta$ ) between  $L_1$  and  $L_2$ : the higher the number of lexical pairs where ( $\delta$ ) appears, the closer  $L_1$  and  $L_2$  are.

An ECD uses thus four levels of 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 a 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 in the city*). 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*) ~ **BAKE1.2** ( $\approx$  create a product by using heat: *bake rolls*); cf. *boil potatoes* ~ *boil some stew*, *fry potatoes* ~ *fry pancakes*, *cook meat* ~ *cook soup*, etc. Arabic numbers specify tighter lexical groupings.

- Lowercase Latin letters distinguish LUs of the same vocable whose semantic bridge is really important and whose semantic difference is regular in **L**: **BAKE1.1a** (action: *Bob baked the potatoes in 30 minutes*) ~ **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 in practice.

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

#### Comments on Def. 11.6

##### 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 common string of letters, extracted from *improves*, *improving*, *improved*, *has been improving*, etc.; the signifier of the idiom **kSEE RED1** (become very angry) is the tree **SEE-dir-obj**→**RED**.

- Part-of-speech conversion is a morphological expressive means: [*a*] **COOK** is formed from [*to*] **COOK** by conversion, just as **SMOK+ER** is formed from [*to*] **SMOKE** by suffixation.<sup>xx</sup> 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<sub>N</sub>*. An important corollary of this

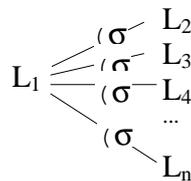
is that if  $\text{SMOKER}$  and  $\text{SMOKE}_V$  belong to different vocables, so do  $\text{COOK}_N$  and  $\text{COOK}_V$ ; at the same time, the lexemes  $\text{COOK}_N$  and  $\text{COOK}_V$  are by no means homonyms, since their signifiers are distinct.

**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  $(\sigma_1)$ , but  $L_2$  and  $L_3$  share a semantic bridge  $(\sigma_j)$  (without sharing  $(\sigma_1)$ , and  $(\sigma_1) \neq (\sigma_j)$ ). 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  $\text{IMPROVE}$  at the end of Subsection **1.3**, p. 00). There are two types of semantic links within a vocable, and consequently two types of polysemy.

- If several LUs  $L_i$  each share one and the same semantic bridge  $(\sigma_+)$  with the LU  $L_1$ , we have **radial polysemy**:



For instance, let  $L_1$  be  $\text{HEAD}^1_{1a}$  (human body part); then  $L_2$  is  $\text{HEAD}^1_{1b}$  (animal body part),  $L_3$  is  $\text{HEAD}^1_4 \approx$  (person in charge) (*head of a bank*),  $L_4$  is  $\text{HEAD}^1_{10} \approx$  (upper part) (*head of a mushroom/of a hammer*), and  $L_5$  is  $\text{HEAD}^1_5 \approx$  (front part) (*head of a convoy*). All these latter HEADS are defined with a reference to the human head: by similarity of function or position; the semantic bridge  $(\sigma_+)$  is ([human]  $\text{head}^1_1$ ).

- If the LU  $L_2$  shares the semantic bridge  $(\sigma_{+i})$  with the LU  $L_1$ ,  $L_3$  shares the semantic bridge  $(\sigma_{+j})$  with  $L_2$ , and  $L_4$  shares the semantic bridge  $(\sigma_{+k})$  with  $L_3$ , etc., we have **chain polysemy**:

$$L_1 - (\sigma_{+i}) - L_2 - (\sigma_{+j}) - L_3 - (\sigma_{+k}) - L_4 - \dots - (\sigma_{+1}) - L_n$$

Thus, with the English noun  $\text{BODY}$  we have three ‘parallel’ polysemy chains:

•  $L_1 = \text{BODY1a}$  (human body [as opposed to the soul]) —  $L_2 = \text{BODY3a}$  (group of humans) (*large bodies of unemployed*; the semantic bridge with  $L_1$ : (human)) —  $L_3 = \text{BODY3b}$  (organization) (*governing body*; the semantic bridge with  $L_2$ : (group of humans)).

•  $L_1 = \text{BODY1a}$  (human body) —  $L_4 = \text{BODY1b}$  (main part of the human body) ( $\approx$  (torso); the semantic bridge with  $L_1$ : (human body)) —  $L_5 = \text{BODY4b}$  (main part) (*the body of a plant/a text*; the semantic bridge with  $L_4$ : (main part)).

•  $L_1 = \text{BODY1a}$  (human body) —  $L_6 = \text{BODY2}$  (dead person) (the semantic bridge with  $L_1$ : (human body))

Another example of chain polysemy:  $L_1 = \text{BUGI.1}$  (insect),  $L_2 = \text{BUGI.2}$  (virus—as if it were a bugI.1), and  $L_3 = \text{BUGII}$  (error in a computer program—as if the error 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.

Now it is easy to formulate the essential difference between polysemy and homonymy. Let there be LUs  $L_1$  and  $L_2$  whose signifiers coincide:  $/L_1/ = /L_2/$ , but the signified do not:  $(L_1) \neq (L_2)$ . Then:

|| Iff  $(L_1)$  and  $(L_2)$  are linked by a (chain of) semantic bridge(s),  $L_1$  and  $L_2$  stand in a relation of **polysemy**; otherwise, they stand in a relation of **homonymy**.

As it must be clear now,  $L_1$  and  $L_2$  that stand in a relation of polysemy belong to the same vocable;  $L_1$  and  $L_2$  that stand in a relation of homonymy belong to two different vocables.

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 extracted and ‘raised’ to the vocable, to avoid tedious repetition.

—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 for 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 when we perceive that L has different uses that 1) denote two (or more) different classes of entities/facts of the real world, but, at the same time, 2) involve common configurations of semantemes. For instance, the verb PAINT<sub>V</sub> and the noun AUNT are problematic. In (12), PAINT<sub>V</sub> appears in either of the two possible senses:

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

**a.** *painted* ≡ (has covered [the ceiling] with paint) :

Alain carried out a renovation or refurbishing of a room.

**b.** *painted* ≡ (has covered [the ceiling] with artistic images) :

Alain created a work of art on the ceiling.

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

In both uses the same semantic component is present: (apply the paint).

In (13), AUNT can be understood even in three ways:

(13) *This is my aunt Joan.*

**a.** *aunt* ≡ (a sister of the mother)

**b.** *aunt* ≡ (a sister of the father)

c. *aunt*  $\equiv$  ( the wife of an uncle )

In all three uses the same semantic component appears: ( sister or wife of a brother of a parent ).

At first glance, examples (12) and (13) are parallel: in both the uses of a lexical item—*PAINT<sub>V</sub>* in (12) and *AUNT* in (13)—correspond to two, respectively, three different extralinguistic realities, while these uses show semantic bridges. Does this mean that in (12) we have two LUs *PAINT* and in (13) three LUs *AUNT*? The answer is that there is no parallelism, and these examples represent two different cases. *PAINT<sub>V</sub>* in (12) is ambiguous and corresponds to two LUs: *PAINT<sub>V1</sub>* and *PAINT<sub>V2</sub>*, that is, it manifests polysemy. But *AUNT* in (13) is not ambiguous, but vague: it represents one lexeme with a disjunctive definition (*AUNT*  $\equiv$  ( a sister of the mother, or a sister of the father, or the wife of an uncle )). To solve such cases, that is, to distinguish lexical items of type [*to*] *PAINT* from those of type *AUNT*, Criteria of Type II, or Criteria II—for distinguishing lexicographic senses of a polysemous lexical item—are introduced. 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\** (...  $\sigma_1/\sigma_2$  ...) such that 1) *L\** has in its meaning two mutually exclusive semantic components ( $\sigma_1$ ) and ( $\sigma_2$ ) that correspond to two different real world entities/facts, but 2) along with ( $\sigma_1$ ) and ( $\sigma_2$ ), *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 ( $\sigma_1$ ) and ( $\sigma_2$ ); namely, we want to know which of the following alternatives is true:

a) Either ( $\sigma_1$ ) and ( $\sigma_2$ ) are inside one lexicographic definition, related by the semanteme (or3). The unity of *L\** is upheld, so that *L\** is one single LU *L* with a logical disjunction in its meaning, i.e., with a disjunctive definition: *L* (...  $\sigma_1$  or3  $\sigma_2$  ...).

b) Or ( $\sigma_1$ ) and ( $\sigma_2$ ) belong to two lexicographic definitions; *L\** should be split in two LUs, so that we have *L*<sub>1</sub> (...  $\sigma_1$  ...) and *L*<sub>2</sub> (...  $\sigma_2$  ...).

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  $\mathbf{I}$  for the suspect lexical item  $L^*$  is, formally speaking, a set of lexicographic information units  $\mathbf{i}_n: \mathbf{I} = \{\mathbf{i}_n\}$ . In an ideal case, any  $\mathbf{i}_i \in \mathbf{I}$  is valid for any use of  $L^*$ , independently of the distinction  $(\sigma_1)$  vs.  $(\sigma_2)$ ; but in a less than ideal case, some  $\mathbf{i}_i$  are true only if  $L^*$  is taken to mean  $(\sigma_1)$ , but not  $(\sigma_2)$ , or vice versa. Then the set  $\mathbf{I}$  can be partitioned in two different subsets,  $\mathbf{I}_1((\sigma_1))$  and  $\mathbf{I}_2((\sigma_2))$ , the first being the lexicographic information related to the meaning  $(\sigma_1)$ , and the second—the lexicographic information related to the meaning  $(\sigma_2)$ . 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 item  $L^*$  must be split into two LUs. The strength of this indication depends on at least three factors:

- The relation between the subsets  $\mathbf{I}((\sigma_1))$  and  $\mathbf{I}((\sigma_2))$ : the inclusion of one into 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,  $\mathbf{I}((\sigma_1))$  and  $\mathbf{I}((\sigma_2))$  differ by only one element—say, the noun  $L^* (\dots \sigma_1 \dots)$  has only the singular, while the noun  $L^* (\dots \sigma_2 \dots)$  has both numbers, the indication is 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 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  $\mathbf{I}((\sigma_1))$  and  $\mathbf{I}((\sigma_2))$  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  $(\sigma_1)$  vs.  $(\sigma_2)$  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 checked when applying Criterion II.1, which can now be formulated:

### Criterion II.1

**If** the semantic difference between two uses of  $L^*$  is correlated with subsets  $\mathbf{I}_1((\sigma_1))$  and  $\mathbf{I}_2((\sigma_2))$  of differentiating lexicographic information which show more than one formal difference,  
**then**  $L^*$  should be split in two LUs— $L_1((\sigma_1))$  and  $L_2((\sigma_2))$ .

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  $L$ . 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 differences 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 lexical cooccurrence; absolutely nothing in its behavior points to the choice of the denotation. Criterion II.1 thus gives a negative result: it does not prevent us from upholding the unity of the lexeme AUNT. (Criterion II.2, introduced below, confirms this decision: see (20b), p. 00.)

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

(14) **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:

$X \text{ wipes } Z \text{ with/on } W = (\text{Person } X \text{ removes liquid or dirt from } Z \text{ by rubbing } W \text{ against } Z \text{ or } Z \text{ against } W).$

The GP of WIPE<sub>V</sub> indicates which preposition corresponds to which movement; otherwise, there are no other differences between the two uses of WIPE<sub>V</sub>. (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 (sell) used in the sense of (prostitute oneself):

(15) **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 (15b), 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 (15b), can be defined as follows:

(16) **a.**  $X \text{ vend } Y \text{ à } Z \text{ pour } W \alpha$  (X gives to Z the right of permanent possession of an entity Y<sup>1</sup> or of obtaining service Y<sup>2</sup>—in exchange for money W)

VENDRE in (15a) shows a semantic difference with respect to this definition: the service Y<sup>2</sup> is not any service, but just (having sex); this semantic difference is correlated to a formal one, namely—for VENDRE in (15a) the actant Y must be lexically expressed by A<sub>poss-X</sub> FAVEURS ⟨CHARMES, 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 (16a), 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 splitting VENDRE in (at least) two senses: one described by the definition (16a) and the other, by (16b):

- b.** *X vend Y à Z pour W* α ( 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 (17):

- (17) **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 (17b) the verb ÉLEVER has an action noun ÉLEVAGE (*Jeanne s’occupe d’élevage de cochons* lit. ( Jeanne does pig breeding )) and an agent noun ÉLEVEUR/ÉLEVEUSE: *Jeanne est éleveuse de cochons* ( Jeanne is a pig breeder ). But for (17a), these derivatives are impossible: \**Jeanne s’occupe d’élevage d’enfants* lit. ( Jeanne does child breeding ); \**Jeanne est éleveuse d’enfants* lit. ( Jeanne is [a] child breeder ); here suppletive lexical derivatives can be used: INSTITUTEUR ( teacher ), κJARDINIÈRE D’ENFANTS<sub>1</sub> ( 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 (12), Criterion II.1 recommends the split:

- *PAINT1*, but not *PAINT2*, has an objectal-copredicative complement (*paint the room **green/a bright color***); it has an  $S_0$  *PAINTING1* and an  $S_{RES}$  *PAINTWORK*; it has a non-standard LF *REPAINT*; etc.

- *PAINT2*, but not *PAINT1*, has a SemA representing the images, as in *sarcophagus painted with Homeric scenes*; it has an  $S_1$  *ARTIST* and two  $S_2$  *PAINTING3* and *PICTURE*; it has for its SemA **4** such expressions as *in oils, in water colors*; it shows ‘alternations’ of the following types: 1) *paint Mr. Polguère ~ paint the portrait of Mr. Polguère* or 2) *paint an old church* = (represent an old church on a picture) or (cover the walls of an old church with paintings); it has a semantic derivation *HAND-PAINTED*; etc.

### 3.2.3 Criterion II.2: Unifying Cooccurrence (= the Green-Appresjan Criterion)<sup>xxi</sup>

A powerful means of testing the unity of a suspect lexical item  $L^*$  (...  $\sigma_1/\sigma_2$  ...) is, roughly speaking, coordination of two clause elements that depend on  $L$  and such that each of them is semantically linked to a different one of its two semantic components that are being tested. If the result is a normal sentence, this constitutes a strong recommendation in favor of a single LU  $L$ ; on the contrary, nothing prevents us from splitting  $L$  if the result is a *zeugma* (a pun, i.e., a word-play of a particular kind, as in *She **took** a lover and a huge risk* or in *Physician: a person on whom we **set** our hopes when ill and our dogs when well* [A. Bierce]).

More precisely, one has to distinguish two cases that involve two different constructions, as we will immediately see in the formulation of the criterion.

#### Criterion II.2

Let there be a suspect lexical item  $L^*$  (...  $\sigma_1/\sigma_2$  ...).

**If** it is possible to construct a sentence that manifests unifying cooccurrence for  $L^*$ , so that we have one of the following two cases:

1) either  $L^* - \text{synt} - [L' \text{ and } L'']$ , such that  $(L') - \text{synt} - (\sigma_1) \ \& \ (L'') - \text{synt} - (\sigma_2)$ ;

or  
 2)  $L'((\sigma_1), (\sigma_2))$ –**synt**– $L^*(\dots \sigma_1/\sigma_2 \dots)$ ,  
**then**  $L^*$  should not be split; as a result, we have one genuine LU  $L$  with disjunction in its definition:  $L(\dots \sigma_1 \text{ or } \sigma_2 \dots)$ .

**Examples**

Case 1: the lexeme  $\text{COOL}_V$  (become cooler or cool) [*The gas cooled a bit but still was very hot; The cake should cool completely*]. A sentence manifesting unifying cooccurrence for  $\text{COOL}_V$  is, for instance, (18):

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

Case 2: The French lexeme BELLE-MÈRE means (mother of the spouse or the wife of the father, who has replaced the deceased or divorced mother) = (mother-in-law or step-mother); it can be found in a sentence with unifying cooccurrence for BELLE-MÈRE:

(19) *Mes [=  $L'$ ] deux belles-mères s'entendaient parfaitement*  
 lit. (My two mothers-in-law got along perfectly).

Here are a few more examples to better show the working of Criterion II.2.

(20) a. (i) BOMBARD (drop bombs [=  $(\sigma_1)$ ] or<sub>3</sub> hurl heavy artillery shells [=  $(\sigma_2)$ ):  
 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, 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 (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—as the DO SO Test (Lakoff & Ross 1976):

### 'Do so' Test

Let there be a suspect lexical item  $L^*$  (...  $\sigma_1/\sigma_2$  ...);  $N$  and  $N'$  are any semantically convenient nouns and  $V$ , a convenient verb.

**If** it is possible to construct a grammatically correct sentence of the form

$$N V\text{-es } L^*, \text{ and so does } N'$$

in which the verb  $V$  semantically involves ( $\sigma_1$ ) and the phrase *so does* involves ( $\sigma_2$ ),

**then**  $L^*$  should not be split; as a result, we have one LU  $L$  with disjunction in its definition:  $L$  (...  $\sigma_1$  or  $\sigma_2$  ...).

In a language that, unlike English, does not have the DO SO construction, one can use a similar construction with a lexeme meaning (as well) or (just as).

It is convenient to have at one's disposal Criterion II.2 in this form because in some cases its application in the 'classical' form can be blocked by unfavorable grammatical conditions. In all cases analyzed until now, if both versions of Criterion II.2 apply, they give the same results:

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

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

(21') **c'**. (iii) Fr. *Son visage flambait de fièvre **tout comme** ses yeux*  
lit. (His face was blazing with fever, just as his eyes were).

We will see the DO SO Test again in the next chapter: in a slightly different version, it is used there for differentiating between actants and circumstantials (Ch. 12, 5, p. 00).

### 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 a suspect lexical item  $L^*$ , 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 lexical item  $L^*$  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 Sem-configurations inside one disjunctive definition; when its premise is satisfied, it gives the linguist a strong incentive not to split the suspect lexical item  $L^*$ .

Both criteria are valid only in 'one direction:' if their premise satisfied. If it 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 (and requires the splitting of  $L^*$  into two LUs), Criterion II.2 must be negative (that is, not require to keep the unity of  $L^*$ ), and vice versa. In other words, Criterion II.1 and Criterion II.2 should not contradict each other. But what happens if they do? Criterion II.2 could seem 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 lexemes, *BAKEI.1* and *BAKEI.2*, which are distinguished in spite of the recommendation of Criterion II.2, see immediately below.

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  $LF_{\text{Real}_1}$ ): 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:

$$S_0 = \text{PRISE} \mid Y \text{ is medication.}$$

However, the conflict resolution does not always come so easily. Take the distinction of 'potato' BAKE1.1 vs. 'bread' BAKE1.2, proposed in the ECD sample in 5.1: it can in fact be questioned. English dictionaries do not draw this distinction, and Criterion II.2 recommends a unified description:

(21) **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 BAKE1.1 and BAKE1.2 were to be united under a disjunctive definition, the resulting lexical entry would be clearly separated in two disjoint parts: one for the 'potato' BAKE1.1, and another for the 'bread' BAKE1.2. (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' BAKE1.2) and Creation Verbs (like the 'bread' BAKE1.2) is in general very typical of English; cf., e.g., Miller *et al.* 1988: 202-206.<sup>xxii</sup>

In any event, it is easier to lump together than to distinguish; should we be proven 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 I.1 can be given priority.

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

In an ECD, a super-entry, or a *vocable*, is a structured collection of entries each of which deals with an LU. To briefly characterize the structure of an ECD super-entry, 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 systematically employed 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, to word senses that constitute 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, I will explain the adopted numbering of particular BAKE lexemes.

1) BAKEI, II, III are ordered in this way for the following reasons.

— In contemporary English, BAKE is primarily a verb of cooking. This belief is based not simply on frequency but on psycholinguistic 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.

— Having made ‘cooking’ *bake* BAKEI, we are forced into describing ‘pottery’ *bake* as BAKEII, and ‘sunbathing’ *bake*, as BAKEIII. The decision to put ‘cooking’ *bake* as BAKEI 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 BAKEI, ‘potato’ BAKE [= BAKEI.1] is ordered before ‘bread’ BAKE [= BAKEI.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 BAKEI.1 the food actant Y has the same name it had before it was baked: e.g., a baked potato is still called a potato; whereas in BAKEI.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 BAKEI.1 is semantically poorer than the effected-object BAKEI.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 BAKEI.2 has an extra actant with respect to BAKEI.1, namely the actant that corresponds to the created thing.)

3) Within BAKEI.1, I.2 and II.1, the transitive BAKEI.1a, I.2a and II.1a are ordered before the intransitive BAKEI.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 up 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:

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

The topics described 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 Formality Principle

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 statement in an ECD must be
1) written in a preestablished metalanguage
and
2) completely explicit.

#### Lexicographic metalanguages

An ECD-style lexicographic description is carried out in a preestablished 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 excep-

tions,<sup>xxiii</sup> the meaning, the syntactic behavior and especially the restricted lexical cooccurrence of the headword 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 is insufficient, it 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

*magazine au sujet de Y pour Z*  $\alpha$  ((illustrated)<sup>xxiv</sup> periodical designed to entertain the public Z, (dedicated to subject Y) and having large book format with soft cover).

We can observe here the necessity of variables in the definition, since MAGAZINE is 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:

$L((Y)) = \textit{politique, sportif, de cinéma} <*\textit{cinématographique}>, \textit{de théâtre} <*\textit{théâtral}>, \textit{humoristique, de mots croisés}, \dots$

$L((Z)) = \textit{féminin/pour femmes, pour enfants, pour les jeunes}, \dots$

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' are understood 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: they are not studied in linguistics and completely ignored in conventional dictionaries.

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<sub>N</sub> (masculine gender) can be tentatively defined as follows:

CÉLIBATAIRE<sub>N(masc)</sub> (bachelor)  $\alpha$  (man<sup>1</sup> who is<sup>2,3</sup> not and has never<sup>1</sup> been<sup>2,3</sup> married<sup>1</sup>).

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

The first problem with these adjectives is that they apply to men only, although the noun CÉLIBATAIRE<sub>N</sub> used in the feminine gender,<sup>xxv</sup> can refer to a woman:

(22) **a.** *Cette célibataire de 36 ans dirige trois ateliers*

( This single woman of 36 manages three workshops ).

**b.** *Pierre est un*  $\left\{ \begin{array}{l} \text{vieux célibataire} \\ \text{célibataire endurci} \end{array} \right\}$ .  $\sim$  \**Marie est une*  $\left\{ \begin{array}{l} \text{vieux célibataire} \\ \text{célibataire endurci} \end{array} \right\}$   
 ( Pierre is  $\left\{ \begin{array}{l} \text{old bachelor} \\ \text{confirmed bachelor} \end{array} \right\}$  ).  $\sim$  \* ( Marie is  $\left\{ \begin{array}{l} \text{old bachelor} \\ \text{confirmed bachelor} \end{array} \right\}$  ).

The solution here is straightforward: CÉLIBATAIRE<sub>N(fem)</sub> is a different lexeme, and both of the collocates above must be mentioned in the entry for CÉLIBATAIRE<sub>N(masc)</sub> 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*  $\approx$  ( very bachelor ). But the above definition of CÉLIBATAIRE<sub>N(masc)</sub> 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 this problem aside; something must be done about it. Logically, two solutions are possible.

- Solution I: we introduce still another lexeme, CÉLIBATAIRE<sub>N(masc)</sub><sup>2</sup>, whose definition is adapted to the two adjectives. Namely, it could be as follows:

( man<sup>I</sup> who is<sup>2</sup><sub>3</sub> not and has never<sup>1</sup> been<sup>2</sup><sub>3</sub> married<sup>1</sup> for a considerable period of time and who wants to be<sup>2</sup><sub>3</sub> not married<sup>1</sup> )

The formation of a feminine counterpart would not be allowed for CÉLIBATAIRE<sub>N(masc)</sub><sup>2</sup>. The adjectives VIEUX and ENDURCI will be compatible with CÉLIBATAIRE<sub>N(masc)</sub><sup>2</sup> 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<sub>N(masc)</sub> itself, that is, when it is taken alone, does not carry the above meaning. Sentences (23) constitute a proof of this:

(23) **a.** <sup>?</sup>*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<sub>N(masc)</sub> is used as a technical term to describe the family status of a man.

We are forced to conclude that CÉLIBATAIRE<sub>N(masc)</sub> 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<sub>N(masc)</sub> that bear 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<sub>N(masc)</sub> and describe the two adjectives under discussion in its lexical entry as follows:

Magn<sup>time</sup><sub>['has been']</sub> : vieux | antepos

who wants to remain c., Magn<sub>['want']</sub> : endurci | postpos

#### Comments

1. The superscript “<sup>time</sup>,” with the name of the LF Magn indicates the semantic nuance: it is the time of being a bachelor that is intensified; the subscript “[‘has been’]” identifies the semantic component in the definition of CÉLIBATAIRE<sub>N(masc)</sub> that accepts the intensification. For superscripts, which specify semantic subtypes of LFs, see Ch. 14, 6.1.2, p. 00.

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 use the standard LF Magn supplied with a superscript and a subscript. The superscript “<sup>time</sup>” 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 mixed LF—with a non-standard part that indicates the additional meaning (( wanting to remain unmarried)). Under this description, the component ( wanting to remain bachelor ) is part of the meaning of the adjective ENDURCI.

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 French *CÉLIBATAIRE<sub>N(masc)</sub>*:

Rus. *XOLOSTJAK* (bachelor): (man<sup>1</sup> who is<sup>2,3</sup> not and has never<sup>1</sup> been<sup>2,3</sup> married<sup>1</sup> (and who is<sup>2,3</sup> accustomed<sup>2</sup> to and wants<sup>1</sup> to continue<sup>4</sup> being<sup>2,3</sup> unmarried<sup>1</sup>))

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*<sub>['accustomed']</sub> : *zakorenelyj* [(very accustomed to being unmarried)]

*Magn*<sub>['want']</sub> : *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*<sub>['has been']</sub><sup>time</sup> : *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 of the adjective *GRAND* lit. (big) in this expression should not be described the same way we proposed for *VIEUX* and *ENDURCI* with *CÉLIBATAIRE<sub>N</sub>*. The expression *grand blessé* means (living injured person [= *blessé<sub>N</sub>*] whose wound [= *blessure*] harms<sup>2,1</sup> his health very much); the semantic contribution of the adjective *GRAND* is just (much) = (very), the rest—([injury] harms<sup>2,1</sup> his health)—being part of the definition of the noun *BLESSURE* (injury, wound), which is part of the definition of *BLESSÉ<sub>N</sub>*:

*BLESSÉ<sub>N</sub>* α (living<sup>1</sup> person<sup>1</sup> who has<sup>2,2a</sup> one<sup>1</sup> or<sup>3</sup> more<sup>1</sup> injuries<sup>1</sup> [= *blessure(s)*]).

(The component (living) is necessary 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 (injury)—(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 this definition a component capable of being intensified: (... and which harms<sup>2</sup><sub>1</sub> the health of the being). The resulting definition of BLESSURE (injury) then reads as follows:<sup>xxvi</sup>

BLESSURE

*blessure de X par Y* α (visible<sup>1</sup> lesion<sup>1</sup> which has been inflicted on the tissues<sup>1</sup> of the living<sup>1</sup> being<sup>1</sup><sub>3</sub> X by an external<sup>1</sup> agent<sup>3</sup> Y and which harms<sup>2</sup><sub>1</sub> the health<sup>2</sup> of X).

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

The intensifiable component ([*to*] harm<sup>2</sup><sub>1</sub>) 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), ... This component, inherited by BLESSÉ (noun or adjective), accounts for the collocates of the latter: *grand blessé*<sub>N</sub>, déjà mentionné, ainsi que *blessé*<sub>N</sub> *grave* (grave) vs. *blessé*<sub>N</sub> *léger* (light), *grièvement* (seriously)/*gravement* lit. (gravely) *blessé*<sub>A</sub> vs. *légèrement* (lightly) *blessé*<sub>A</sub>, etc.

In the entry for BLESSÉ<sub>N</sub>, the adjective GRAND appears as an element of the value of the standard LF<sub>Magn</sub>, with the indication that it intensifies the component ([*to*] harm<sup>2</sup><sub>1</sub>):

Magn<sub>[‘harm’]</sub> : grand | antepos

To sum 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 cooccurrents that do not perfectly fit the definition. (A discussion of this problem is found 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 French noun BLESSURE in our example. In other cases, he chooses to change the description of particular cooccurrents by particular LFs; we have done so for VIEUX and ENDURCI (in the entry for CÉLIBATAIRE<sub>N(masc)</sub>), 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, consider the verb WRITE. One writes necessarily in a language; therefore, the semanteme (language) must appear in the definition of WRITE. The denomination of a

language that modifies the verb WRITE in a sentence must then be considered the expression of one of L's semantic actants (and, consequently, of a syntactic actant). Now, the genuine actants feature a rather idiomatic, constrained behavior, but the phrase of the type *in English* with WRITE is absolutely regular and is used with no constraint (*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 (Ch. 12, 3.4.1, p. 00ff), they are: they are syntactic expressions of a SemA. However, in order to be certain, we have to develop numerous lexical entries—for dozens of verbs related to WRITE (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 the logic and to the available definitions; 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 (= Lexical Inheritance) Principle

The communicatively dominant node of the lexicographic definition of an 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 to systematically check that all 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 blocked (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 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 in a similar way—to the extent that language **L** allows.

In spite of its obvious character this principle is not properly observed in existing dictionaries.<sup>xxvii</sup> Consider, as an illustration, the description of nationality names in *Petit Robert* [= 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 plural<sup>xxviii</sup> 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 ).<sup>xxix</sup> 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.

**N:** 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 or maybe specifying just the geographical area—in this case, Northern Africa ), 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 ethnic groups, considered above, such a schema for the definition could resemble the following:

Fr. *les*  $L_{(\text{nationality}2)}\text{-s} \equiv$  ( 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 schemas belong to the lexicographic metalanguages we have been discussing above.

### 4.3.2 Vocabable 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 general schemas for lexicographic description even at a higher level: namely, for a uniform treatment of vocabables belonging to the same lexical field.

#### Definition 11.7: Lexical field

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

Now the principle itself can be readily formulated:

#### Vocabable Uniform Treatment Principle

Two vocabables belonging to the same lexical field must be presented, everything else being equal, according to the same schema: the related LUs of either vocabable 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 vocabable;
- (iii) the semantic distances between them must be represented as similarly as possible (i.e., by the same or almost the same means).

A good example is the general schema for lexicographic description of French vocabables in the lexical field «BODY PARTS» (Arbatchewski-Jumarie & Iordanskaja 1988).

### 4.4 Internal Exhaustivity Principle

Like all 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.<sup>xxx</sup> 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 idiom  $\kappa$ MOYEN DE TRANSPORT<sub>1</sub> (transportation means) must contain:

- The names of all existing means of transportation:

— on earth: VOITURE (car), CAMION (truck), AUTOBUS (bus), AUTOCAR (tourist bus), TROLLEY-BUS (trolley bus), TRAMWAY (streetcar), MÉTRO (subway), ...;

— on water: BATEAU (ship), NAVIRE (ship), BARGE (barge), RADEAU (raft), ...;

— in the air: AVION (airplane), HÉLICOPTÈRE (chopper), FUSÉE (rocket), ...

- numerous LUs such as CARGAISON (load, cargo),  $\kappa$ TITRE DE TRANSPORT<sub>1</sub>  $\approx$  (ticket), BILLET (ticket), VOYAGEUR (passenger), CORRESPONDANCE (transfer<sub>N</sub>), RÉSEAU (network), DESSERVIR (serve), TRANSPORTER (transport<sub>V</sub>), VOYAGER (travel<sub>V</sub>, ride<sub>V</sub>), COMPOSTER (punch<sub>V</sub>, stamp<sub>V</sub> [the ticket]), ... and even RESQUILLER (fare-jump).

Another example could be the names of 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! (ow!), OUILLE! (ouch!), etc.

Moreover, for interjections the ECD must specify the exact prosody, since, for instance, 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 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  $F^{sem}$  should be extracted from individual LU entries belonging to  $F^{sem}$  and transferred directly to the entry for the LU that is the name of  $F^{sem}$ .

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

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

And of course it admits 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

...

IncepOper<sub>1</sub>, AntiReal<sub>1</sub>, try.to.Liqu<sub>1</sub>Func<sub>0</sub>, ...: ↑ILLNESS

Two complications can be encountered in this connection:

— A 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.

— A 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.

Some more specific cases will not be discussed here; the techniques foreseen by semantic field generalization are presented in detail in Mel'čuk & Wanner 1996.

#### 4.6 No Regularly Produced LUs in the Lexicon Principle

**If** an LU L' of **L** is related to another LU L in a completely regular way, i.e., 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 as a separate lexical entry (it must be specified in the lexical entry for L).

Such a situation can present itself in three types of cases: regular compounding, regular derivation, and regular polysemy. I will comment on the two last types. (Potential lexical units, that is, 100% regularly compounded<sub>1</sub> and derived<sub>1</sub> LUs, were discussed in Ch. 7, 2.1.2, p. 00 and Ch. 10, 2.1.2, p. 00.)

##### 4.6.1 Regular Derivation

Suppose that **L** possesses a derivateme ( $\delta$ ) 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 ( $\delta$ ) 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<sub>1</sub> is possible. (Derivation<sub>1</sub> is opposed to derivation<sub>2</sub>: the latter covers fossilized derivatives, which are derived only historically.) 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 is semantically plausible—such as *docteur* (doctor), *agent* (agent), *auteur* (author) or *écrivain* (writer). Therefore, a French ECD indicates for a human noun  $L$  that it has a female derivation  $L'$  and the form thereof; this can be done by means of a non-standard LF  $F_{em}$  (= ‘female counter-part of’). However, there is no need to put  $L'$  in the dictionary with its own full lexical entry. Thus, we have:

DIRECTEUR <sub>(masc)</sub> (director)	RUSSE <sub>N(masc)</sub> ([a] Russian)	PAYSAN <sub>N(masc)</sub> (peasant)
$F_{em}$ : <i>directrice</i>	$F_{em}$ : <i>Russe</i>	$F_{em}$ : <i>paysanne</i>

The feminine nouns DIRECTRICE, RUSSE<sub>N(fem)</sub> and PAYSANNE<sub>N(fem)</sub> 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<sub>N(fem)</sub> (single woman) does not have the same LFs as CÉLIBATAIRE<sub>N(masc)</sub> (single man<sup>11</sup>) and therefore needs its own lexical entry. The lesson to be drawn 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:  $(L_j) - (L_i) = (\sigma)$ ; this is **regular polysemy**. If the lexical entry for  $L_j$  has no other differences with respect to  $L_i$  but the semantic component  $(\sigma)$  (and other differences that automatically accompany it), then, as is the case with regular derivation,  $L_j$  need not be stored in the dictionary as a separate entry: 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 English transitive action verbs that semantically presuppose an instrument have a special sense that accepts the name of this instrument as its syntactic subject:

<i>John cuts the cardboard with a knife.</i>	~	<i>This knife <b>cuts</b> the cardboard.</i>
<i>John writes his cards with this pen.</i>	~	<i>This pen <b>writes</b> well.</i>
<i>John solved the equation on his computer.</i>	~	<i>The computer <b>solved</b> the equation.</i>

This sense can be called ‘instrumental.’ The question is, 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 this gun.* ~ \**This 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 yes, 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, limiting myself to the following remark:

It is possible to avoid actually storing such LUs as **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:

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

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<sup>xxxi</sup>

In this section, first, several 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’ BAKEI

- I.1a.** X cooks<sub>1</sub> solid Y ... in device Z<sup>1</sup> ... [*John baked the potatoes in the oven*].
- 1b.** Solid Y cooks<sub>2</sub> ... in device Z<sup>1</sup> ... [*The potatoes baked in 20 minutes*].
- 1c.** Device Z is used [by X]<sup>xxxii</sup> 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*].



2) The component (submit) is an abbreviation for (cause<sub>2</sub> to undergo) (as in ... *submit it to the action of strong heat in a crucible*); the corresponding sense is absent from *LDOCE Online*.

3) Here and below, Prep<sub>loc</sub> stands for a locative preposition (which is chosen by the speaker according to its meaning): *in, on, over, under, ...*

4) Instead of *X bakes<sub>I.1a</sub> Y*, some speakers use spontaneously 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<sub>I.1a</sub>.

### I.1b.

#### Definition

*Y bakes in Z* : solid<sup>1</sup> Y cooks<sub>2</sub> as a result<sup>2</sup><sub>6</sub> of being baked<sub>I.1a</sub> [by X] or<sub>3</sub> of undergoing the indirect<sub>1</sub> action<sub>6</sub> of dry<sub>10</sub> heat<sup>2</sup><sub>2</sub> in device<sub>1</sub> Z<sup>1</sup> or in contact<sub>1</sub> with source<sub>1</sub> Z<sup>2</sup> of heat<sup>2</sup><sub>2</sub>.

#### Government Pattern

Y ⇔ I	Z ⇔ II
1. N	1. Prep <sub>loc</sub> N

*The apples baked (in the oven <on hot bricks>).*

#### Lexical Functions

Syn<sub>C</sub> : cook<sub>2</sub>      S<sub>2</sub><sup>usual</sup> = S<sub>instr-loc</sub> : oven      Caus : bake<sub>I.1a</sub>  
 Syn<sub>∩</sub> : roast<sub>2</sub>      S<sub>instr</sub> : baking tray/sheet;  
 Gener : cook<sub>2</sub>      baking tin, cake tin; foil

#### Examples

—Keep quiet, the potatoes are already baking! | Macintosh bake better <softer> than Golden Delicious.

#### Linguistic comments

BAKE<sub>I.1b</sub> marginally admits situations of non-agentive baking. As a result, BAKE<sub>I.1b</sub> and BAKE<sub>I.2b</sub> 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<sub>I.1b</sub> contains the disjunctive component (... or<sub>3</sub> of undergoing the indirect<sub>1</sub> action<sub>6</sub> of dry<sub>10</sub> heat<sup>2</sup><sub>2</sub>).

### I.1c.

#### Definition

*Z bakes Y* : Device<sub>1</sub> Z used<sup>2</sup><sub>1</sub> [by X] in baking<sub>I.1a</sub> Y causes<sub>1</sub> Y to bake<sub>I.1b</sub>.

**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 | M<sub>2</sub> ≠ Λ

[“M<sub>2</sub> ≠ Λ” means ‘DSyntA II of BAKE is not empty’, i.e., it is obligatory: \**My oven bakes to a turn* vs. *My oven bakes rolls to a turn*]

**‘Bread’ BAKE I.2**

**I.2a.**

**Definition**

*X bakes Y from W in Z* : X creates<sup>1</sup> solid<sup>1</sup> food<sup>1b</sup> Y from mixture<sup>1</sup> Z<sup>1</sup> of Z<sup>2</sup>, which is flour<sup>1</sup> made<sup>1</sup> of grain<sup>1</sup> Z<sup>3</sup> or other powder-like foodstuff, with liquid<sup>2</sup><sup>1</sup>—by baking<sup>I.1a</sup> Z<sup>1</sup> in W.

**Government Pattern 1**

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

1) C<sub>III.1,2</sub> without C<sub>II</sub> : **undesirable**

2) C<sub>III.3</sub> : N ≠ Z<sup>3</sup>

*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<sup>xxxiii</sup>**

X ⇔ I	Z ⇔ II	Y ⇔ III	W ⇔ IV
1. N	1. N <b>obligatory</b>	1. <i>into</i> N <b>obligatory</b>	1. Prep <sub>loc</sub> N

1) C<sub>II</sub> : N ≠ Z<sup>3</sup>

*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 <sub>c</sub>	: make	shop which carries
Syn <sub>o</sub>	: cook <sup>1</sup>	mostly baked goods
professional-S <sub>1D</sub>	: baker; confectioner	or pastry ≈ S <sub>1oc</sub> : bakery; pastry shop,

$S_2^{\text{usual}}$	: baked goods		pâtisserie
$S_{2\supset}^{\text{usual}}$	: pie; pastry, charlotte; meat-loaf	Ver	: to a turn
$S_3^{\text{usual}}$	: dough, batter; flour	too much, AntiVer	: //overbake
$S_4^{\text{usual}} = S_{\text{instr-loc}}$	: oven; Dutch oven	not enough, AntiVer	: //underbake
		$A_2\text{recentlyPerf}$	: //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 becomes<sup>1</sup> ready<sup>1</sup> to.be.eaten<sup>1</sup> as a result<sup>2</sup> of being baked<sup>1.2a</sup> [by X from W] in Z.

**Government Pattern**

$Y \Leftrightarrow \mathbf{I}$	$Z \Leftrightarrow \mathbf{II}$
1. N	1. Prep <sub>loc</sub> 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<sup>1.2a</sup> Y causes<sup>1</sup> Y to bake<sup>1.2b</sup>.

**Government Pattern**

$Z \Leftrightarrow \mathbf{I}$	$Y \Leftrightarrow \mathbf{II}$
1. N	1. N

*This oven bakes rolls in 30 minutes.*

**Lexical Functions**

Ver : to a turn



**II.2.****Definition**

*X bakes in Y* : Substance<sup>1</sup> X hardens<sup>1</sup> by the action<sup>1</sup> of dry<sup>10</sup> heat<sup>2</sup> caused<sup>1</sup> by phenomenon<sup>1</sup> Y.

**Government Pattern**

X ↔ I	Y ↔ II
1. N	1. Adv <sub>2</sub> 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<sub>III</sub>****III.a.****Definition**

*X bakes in Y* : Humans<sup>2</sup> X<sup>1</sup> in location<sup>1</sup> X<sup>2</sup> feel<sup>1</sup><sub>2</sub> intensely affected<sup>2</sup><sub>1</sub> by dry<sup>10</sup> heat<sup>2</sup> caused<sup>1</sup> by the sun<sup>1</sup><sub>2a</sub> or artificial<sup>1</sup> rays<sup>2</sup><sub>1</sub> Y.

**Government Pattern**

X ↔ I	Y ↔ II
1. N	1. Adv <sub>2</sub> Caus N

*Jane was baking (in the sun (under the rays of the tropical sun)).*

**Lexical Functions**

Syn : fry

X = X<sup>2</sup>, A<sub>2</sub>

: baking<sub>1</sub>

Syn<sub>∩</sub> : roast; burn; cook<sub>II</sub>, swelter

as if X's 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 + Oper<sub>1</sub>](*fever*), where [Magn + Oper<sub>1</sub>] is a configuration of LFs, see Ch. 14, 4.2, p. 00.

2) Consider the expression *bake in the sun* (be exposed to strong sun) as seen in sentences (24):

(24) a. *For years, teens have spent hours **baking in the sun** in pursuit of the perfect tan.*

**b. You can almost smell the garbage *baking in the sun*.**

This expression is a collocation of SUN<sup>1</sup>2a—namely,  $\text{MagnInvolV}_{21}(\text{SUN}^1 2a) - \text{II} \rightarrow \text{SUN}^1 2a$ —and therefore it should not be covered by the lexical entry for BAKE<sup>III</sup>. The same is true of the following sentence:

*They set it [= the geranium] out and let the hot sun bake it [F. O'Connor];*

BAKE here is a collocation of SUN (InvolV) and should be described in the entry for SUN.

**III.b****Definition**

*Y bakes X* : Intense heat Y causes<sup>1</sup> that people X bake<sup>III.a</sup> in Y.

**Government Pattern**

Y ⇔ I	X ⇔ II
1. N	1. N

*The heat wave was baking Boston.*

**Lexical Functions**

Syn : fry

Syn<sub>∩</sub> : roast; burn; cook<sup>II</sup>

**BAKED ALASKA**, idiom, nominal phrase, countable

*baked Alaska* : dessert consisting of ice cream topped<sup>2</sup> with meringue slightly<sup>1</sup> browned<sup>2</sup> by baking<sup>I.2a</sup> it.

**BAKED BEANS**, idiom, nominal phrase, pl, uncountable

*baked beans* : dish [= food] consisting of beans boiled and then baked<sup>I.1b</sup> (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 bake<sup>III</sup> in U [*It is baking in here*].

2. So [hot] that people bake<sup>III</sup> [*baking hot*].

1. predicative use only.

**Definition**

*It is baking in U* : It is so hot<sup>n°</sup> in location<sup>1</sup> U that humans<sup>2</sup> bake<sup>III</sup> in U.

**Government Pattern**

U ⇔ I
1. LOC <sub>in</sub> N

*It was baking (in the stuffy room ⟨here, in here⟩).*

**Lexical Functions**

Syn<sub>C</sub> : roasting                      Syn<sub>∩</sub> : boiling; steaming

2. modifying use only; adverb-modifying

*Baking* [P] : [P] so that people bake<sub>III</sub> | Magn(hotn°).

**Lexical Functions**

Syn : roasting, scorching

**Examples**

The room became baking hot.

**Linguistic comments**

The relevant sense of HOT—(high temperature of surrounding air)—is not present in *LDOCE Online*; that is why I write it as HOTn°. The necessity of having this sense as a separate LU is shown by the impossibility of *\*baking hot water* or *\*baking hot fire*; cf. also *roasting ⟨scorching⟩ hot*, possible only about weather.

**SUN-BAKED**, adjective

*sun-baked* [X] : [Body part X of a human<sup>2</sup>] perceived as dried out<sup>2</sup> or<sup>3</sup> burnt<sup>13</sup> by the sun<sup>12a</sup>  
—as if X were mud<sup>1</sup> baked<sub>II.2</sub> in the sun<sup>12a</sup>.

**Example**

sun-baked lips ⟨cheeks, hands⟩ | a sun-baked forehead ⟨face⟩

**5.2 Lexical-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 BAKEI.1a ~ BAKEI.1b, BAKEI.2a ~ BAKEI.2b and BAKEII.1a ~ BAKEII.1b, or the **b**-lexeme, differs from the **a**-lexeme in the communicative organization of its meaning. In the **a**-lexeme, the communicatively dominant (= generic) component is (cause2), but in the **b**-lexeme, this component is present, so to speak, only on the periphery: here, the generic component is the designation of a change of state, to which the causation component is attached as a modifier. Thus, the generic component in the meaning of BAKEI.1a is (cook1) ≈ (cause2 to become ready for eating), while the generic component in the meaning of BAKEI.1b is (cook2) ≈ (become ready for eating); (cause2) is relegated here to a modifier status: ([cook2] being bakedI.1a) ((bakeI.1a) contains (cause2)). To put it differently, BAKEI.1a expresses an action or an activity, while BAKEI.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. Another current term for this types of verb is *inchoative* (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 general rule, so that passives should not be described in the dictionary (although they can, of course, be illustrated there in the examples—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* ⟨\*prepared⟩ *in twenty minutes*, *The houses sold* ⟨\*built⟩ *in twenty months*, *The bridge wore out* ⟨\*destroyed⟩ *in a few years*, *The books shipped* ⟨\*sent⟩ *this morning*, etc. The constraints on decausative formation are too capricious to allow for 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’ ~ ‘decausative of V,’ and in all of them the decausative changes the meaning of the starting V in the same way:

**if** (V) = (cause2 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<sub>2</sub> in the way  $\alpha$ ).

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 1974 [1992: 213ff]).

- Decausatives contrast with passive-potentials<sup>xxxiv</sup> of the type *These judges bribe easily* or *This blouse washes well*. The passive-potential of a verb V (V-II→Y) means (Y has such properties that it is Z-y to do V to Y). Because of this, passive-potentials cannot be used in ‘completive’ or ‘punctual’ statements—they tend to be generic. What is important 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 judge bribed many times* [decausative], but *This judge bribed with no effort* [passive-potential]; etc. Passive-potentials are inflectional forms of the verb (like ‘normal’ passives) and should not constitute separate lexical entries. (They can of course be illustrated in the lexical entry for the basic transitive verbs, along with normal passives.)

## 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. (What is different is the DSynt-Comm-Structures of both sentences; here we allow ourselves to ignore this fact.) The lexical entry of the corresponding verb shows 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. A verb of the BAKE1.2a type has to be lexically marked, since not all creation verbs admit Object-Shuffling: e.g., CREATE itself does not (\**God created dust into man*). It need not, however, be split into two different lexemes since in both patterns it has the same meaning and the same restricted lexical cooccurrence; both patterns 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 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 BAKE12a (*Bob bakes on Fridays*).

### 4. Benefactives

The problem of so-called Beneficiary 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 Z uses Y created by X). 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 are 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 government pattern such that it ceases to be a verb of creation. In this usage, it rejects a benefactive (Levin 1993: 3):

(25) **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 (25d) CARVE is not used as a verb of creation, and a benefactive with it is impossible.

### 5. Subject/Object-Oriented Resultative Complements

Resultative complements are possible with Creation Verbs (as in *We baked the pot **hard*** or *The pot baked **hard** in the kiln*, 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 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.

## Notes

<sup>i</sup> (1.1, p. 00) As was stated in Part I, Ch. 1, p. 00, an LU L is not a linguistic sign in the strict sense: it is a set of linguistic signs. However, this set can be always represented by one genuine sign: the stem for a lexeme or a configuration of syntactically linked stems (= a syntactic tree) for an idiom. Therefore, as a convenient abbreviation, I allow myself to speak of LUs as signs.

<sup>ii</sup> (1.2, p. 00) **Theoretical Lexicology vs. Practical Lexicography**

It is impossible to delve here into the heated debate opposing theoretical lexicology, which is a part of linguistics, and practical lexicography, which boils down to commercial 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 fact, they can greatly fertilize 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 possibilities are generally ignored simply because of centuries-long traditions.

**NB:** Compiling a dictionary by semantic and/or lexical fields does not of course mean that it has to be presented in print by these fields. For a standard user of a printed dictionary, the best disposition remains the alphabetical order, which is used in ECD. Note, however, that for electronic dictionaries of the future the problem of ordering does not arise.

- B. Atkins (1992/1993: 8-9), while defending the *American Heritage Dictionary* [= 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 traditional 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. R. Moon (2002) tells a very instructive story: in a survey of needs of poor British families, 53% of respondents named a dictionary as a necessity—above cars, VCRs or even a newspaper! But there is no clear answer to the question what does a family need a complete monolingual dictionary for... More than that, I never got a reasonable answer to the above question from a lexicographer. (To check the spelling or pronunciation or to find out the meaning of rare, archaic or technical LU a smaller and cheaper specialized dictionary would be more than sufficient.)

On the other hand, I believe that a scientifically oriented monolingual dictionary can be a powerful tool in teaching logic, semantic analysis and formal reasoning to ordinary people. 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 scientific, reasonably formalized monolingual dictionaries should go hand in hand with new ways of teaching the mother tongue in schools—with the emphasis on the 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...

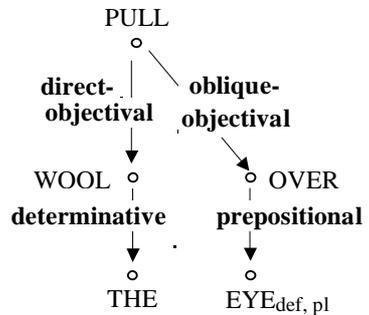
<sup>iii</sup> (1.3.4, p. 00) Phrasemes ( $\approx$  set phrases) are divided into four major classes:

- pragmatic phrasemes,                      or 1) pragmatemes
- semantic phrasemes:
  - compositional phrasemes,              or 2) collocations and 3) clichés
  - non-compositional phrasemes,      or 4) idioms

For details about phrasemes, see Ch. 16, p. 00ff.

<sup>iv</sup> (1.3.4, Item 1, p. 00) **Lexical Entries for Idioms**

Strictly speaking, the dictionary entry for an idiom is formally different from that of a lexeme: it contains some additional fields, which are related to its linguistic nature, namely—being syntactically a phrase, not a word. The dictionary entry for an idiom must contain its Surface-Syntactic Structure; thus, the idiom κPULL THE WOOL OVER Y’S EYES<sub>1</sub>, *X pulls the wool over Y’s eyes* ≡ (X deceives Y by feigning<sub>1</sub> good<sub>2</sub> intentions<sub>2</sub> with the purpose to hide X’s true<sub>1</sub> bad<sub>2</sub> intentions<sub>2</sub> from Y and thereby gain an end) receives the following Surface-Syntactic tree:



**Figure 11-1:** The SSyntS of the idiom κPULL THE WOOL OVER Y’s EYES<sub>1</sub> in its lexical entry in an English ECD

In addition, the government pattern of the idiom must contain indications as to how the expression of DSynt-actant II [ ↔ ( Y ) ] is joined to the body of the phrase:

**Government Pattern**

Y ↔ II	
1. A <sub>poss</sub> (N)	←determ–EYE <sub>pl</sub>
2. N’s	←possess–EYE <sub>pl</sub>
3. of N	←nom-compl–EYE <sub>pl</sub> –determ→THE <b>obligatory</b>

*pull the wool over **her** eyes* ⟨over *Mary’s* eyes, over the eyes *of* all these *people*⟩

For more on the description of idioms in an ECD, see Ch. 16, 7 and 8, p. 00ff.

Second, an idiom must be accompanied by indications about the possible syntactic transformations and linearization of its SSynt-tree (when these operations cannot or should not be carried out according to available general rules). Thus, in this particular case, it has to be stated that the idiom in question 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 κKICK THE BUCKET<sub>1</sub>, where the passivization is impossible: *\*The bucket was kicked by Jim.*)

This can be specified in the lexical entry of the idiom  $\text{K PULL THE WOOL OVER } Y\text{'S EYES}_1$  by the indication of the lexical function  $\text{CONV}_{21}$ :

$\text{CONV}_{21}$  : *Y has the wool pulled over Y's eyes by X*

Lexical entries for idioms are discussed in more detail in Section 5 of this chapter and in Ch. 16, 7, p. 00ff.

The expression [to] *pull strings*, which has a normal passive—*Strings*  $\langle$ A couple of strings $\rangle$  *were pulled by my powerful uncle*—represent a different case: this is a collocation of the noun  $\text{STRINGS}_{II} \approx$  (hidden influence or control) (a *plurale tantum*). This case and similar cases are discussed in Ch. 16, 4.1, p. 00.

<sup>v</sup> (1.3.4, 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.

<sup>vi</sup> (1.3.5, p. 00) On the ‘ungrammatical’ metalinguistic expression (cause<sub>1</sub> that ...), see Part II, Ch. 5, Note 17, p. 00.

<sup>vii</sup> (2.1.1, p. 00) What has been said should by no means be construed as the opposition to inclusion of encyclopedic information into 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.

<sup>viii</sup> (2.1.2.1, Rule 1, p. 00) **Propositional Form (= the Definiendum in the ECD)**

1. The name *propositional form* is due to the fact that, in the prototypical case—that is, when L is a genuine 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* (*Britain's Defense minister*) 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. The propositional form has no scientific value in the following sense: it does not participate in any formal reasoning. Its vocation is purely pedagogical: it provides a minimal logical framework for the subsequent definition of the LU L as well as a minimal pattern for typical syntactic constructions in which L would actually be used. The propositional form is a simplified 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  $L(X; Y; \dots)$ .

<sup>ix</sup> (2.1.2.2, p. 00) To show this in a clearer way, here are the corresponding definitions from Wierzbicka 1972:

(mother of Y) = (human being inside whose body there was once something that was becoming Y's body)

(woman) = (human being that could be someone's mother)

(man<sup>1</sup>) = (human being that could cause<sup>2</sup> another human being to be someone's mother)

<sup>x</sup> (2.1.2.2, Comment 2, p. 00) The phonemic transcription of the signifiers of the signs considered is needed because the French orthography does not correctly represent the relationships between the signifiers involved.

<sup>xi</sup> (2.1.2.2, Comment 2, p. 00) '**Reverse Derivation**'

An additional problem is created here by the fact that the morphological relation between French lexemes ASTRONOME and ASTRONOMIE is inverse with respect to their semantic relation: *astronome*  $\subset$  *astronom+ie* (i.e., 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: we have a case of what is known as **reverse derivation** (Mel'čuk 2006: 529-532). In English the situation with the equivalents of ASTRONOME and ASTRONOMIE is different: ASTRONOM+ER and ASTRONOM+Y are formally of equal complexity (morphological intersection instead of inclusion), so there is no immediate conflict with semantic inclusion.

<sup>xii</sup> (2.1.2.2, p. 00) **Vicious Circles in the System of Lexicographic Definitions**

Here is a simple demonstration of why vicious circles in a dictionary are so bad. Suppose that A is defined 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. Such a situation obtains because of the vicious circle in (iii): D, which is used in the definition of A, has A in its own definition.

<sup>xiii</sup> (2.1.2.3, Rule 3, Comment 1, p. 00) This definition is not satisfactory also from the viewpoint of linguistic facts: 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.

<sup>xiv</sup> (2.1.3.2, Criterion I.2a, p. 00) It is interesting to compare BATTREII with VAINCRE (win):

*X vainc Y dans Z [pour α] E* (|[X and Y being<sup>2</sup>3 opposed<sup>1</sup> in struggle<sup>1</sup> Z over α,] X gets<sup>2a</sup> α)  
 VAINCRE is about X obtaining the intended result in the struggle against the adversary Y, while BATTREII is about X putting Y out of the struggle. The standard intensifier of BATTREII, as we have said, is *à plate couture*, and it characterizes the state of Y. The verb VAINCRE does not have a really idiomatic intensifier, but the noun VICTOIRE 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.

<sup>xv</sup> (2.1.3.4, Criterion I.2c, (9c), p. 00) The improved definition satisfies one of DeMorgan’s logical laws, 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”).

<sup>xvi</sup> (2.1.3.4, p. 00) As indicated by Wierzbicka, the component (according<sub>to1</sub> some reasons<sub>2</sub>) opposes PERMIT<sub>V</sub> to ALLOW: *Although visits were not permitted, the chief surgeon allowed them to enter.*

<sup>xvii</sup> (2.1.4, p.00) In reality, the situation with the verb COST<sub>V</sub> is more complex. In *This book costs \$30 at Chapters* the phrase [at] *Chapters* may be considered as an expression of the Payee: it is not a simple locative Circumstantial, isofunctional with *in Boston* or *on a plane*. To cover such uses as well the verb COST will have the corresponding SemA-slot, supplied with necessary semantic restrictions.

<sup>xviii</sup> (2.2, p. 00) C<sub>III.1</sub> is possible only with the active form of HELP<sub>V</sub>, but this need not be stated in the GP of this particular lexeme, since a ‘bare’ infinitive never cooccurs with a verbal lexeme in the passive (*We saw him cross* ⟨\**to cross*⟩ *the street*. ~ *He was seen to cross* ⟨\**was 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.

<sup>xix</sup> (2.2, p. 00) Constraint 2 became obsolete in American English. Sentences in which HELP<sub>V</sub> 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 Constraint 2 as an illustration of an interesting theoretical possibility, which was a reality not so long ago.

<sup>xx</sup> (3.1, comments on Def. 11.6, p. 00) Conversion is a morphological process where the expressive means is a change of a feature in the syntactics of the sign to which conversion is applied. Thus, the English verb [*to*] BOMB ( [*to*] submit to the action of bombs ) is derived from the noun [*a*] BOMB by the following conversion (the relevant change is boldfaced):

$$\mathbf{C}^{\text{submit}} = \langle (X \text{ submits } Y \text{ to the action of } \dots); \mathbf{N} \Rightarrow \mathbf{V}; \Sigma = \mathbf{N}_{(\xi)}, \dots \rangle$$

$\mathbf{N}_{(\xi)}$  indicates the class of nouns that accept this conversion; they are specified by the appropriate lexical function in their lexical entry. For instance, BOMB<sub>N</sub> has in its entry:

Labreal<sub>12</sub> :  $\mathbf{C}^{\text{submit}}$  (i.e., [*to*] BOMB)

Similarly for OIL<sub>N</sub> ~ [*to*] OIL, SALT<sub>N</sub> ~ [*to*] SALT, SAW<sub>N</sub> ~ [*to*] SAW, MACHINE GUN<sub>N</sub> ~ [*to*] MACHINE-GUN, BUS<sub>N</sub> ~ [*to*] BUS, etc. (For more on conversion, see Mel’čuk 2006: 288ff, especially p. 304.)

<sup>xxi</sup> (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*.

<sup>xxii</sup> (3.2.4, p. 00) **Sense Superposition, or Admissible Zeugmas**

Such usages as that reflected in (21, 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, the fact that Criterion II.2 is not absolute (sometimes it allows for coordination of the dependents of two different lexicographic senses, ‘illegitimately’ combined under one signifier) 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 inflectional 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.

<sup>xxiii</sup> (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 DirO 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 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.

<sup>xxiv</sup> (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. The semanteme (dedicated to subject Y) is also is a weak component: a magazine can deal with several different subjects. See 2.1.4.2, p. 00.

<sup>xxv</sup> (4.2.1, p. 00) CELIBATAIRE<sub>N(masc)</sub> and CÉLIBATAIRE<sub>N(fem)</sub> are two different LUs of French.

<sup>xxvi</sup> (4.2.1, p. 00) See interesting considerations about the lexicographic description of BLESSURE in Kahane 2003b.

<sup>xxvii</sup> (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.

<sup>xxviii</sup> (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/a Brit), (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, very 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 bearing on a pair of shoes, one needs to use the lexeme 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 Fr. [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, *szem+ek*, denotes several pairs of eyes, as in *All eyes turned to her*. One (eye) is called in Hungarian FÉLSZEM lit. (half-SZEM), and its plural, *félszem+ek*, refers to several single eyes, as in *He had five right eyes ready for transplanting*.

<sup>xxix</sup> (4.3.1, p. 00) Here is a tentative definition for the LU NATIONALITY<sub>2</sub>:

*nationality*<sub>2</sub> ≡ (set<sub>2</sub> of individual<sub>1</sub> that share<sub>1</sub> linguistic<sub>1</sub>, cultural, social<sub>1</sub>, physical<sub>1</sub> and/or psychological<sub>1</sub> characteristics, which is caused<sub>1</sub> by common<sub>1</sub> origins<sub>2</sub>).

<sup>xxx</sup> (4.4, p. 00) 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).

<sup>xxxi</sup> (5, p. 00) This section uses the data and analyses from the paper Ilson & Mel'čuk 1989.

<sup>xxxii</sup> (**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.

<sup>xxxiii</sup> (**5.1**, **BAKEI.2a**, p. 00) On the ordering of alternative GPs in one dictionary article, see Ch. 13, **3**, p. 00.

<sup>xxxiv</sup> (**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.