Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria

Igor Mel’čuk, University of Montreal

Abstract
This paper gives a short general characterization of lexicographic definition as adopted in the Explanatory Combinatorial Dictionary (ECD). In doing so, the paper defines seven basic lexicographic concepts (lexeme, phraseme, vocable, semantic bridge, basic lexical unit of a vocable, semantic field, and lexical field). Also, six major formal principles for lexicographic definitions are formulated, discussed, and illustrated: e.g., the Univocity Principle and the Adequacy Principle; such principles underlie all definitions in the ECD. Finally, eleven heuristic criteria for definitions are introduced and discussed; these criteria are designed to aid in accounting for the free co-occurrence of a headword, in deciding about the inclusion of a given component in a definition, in distinguishing different senses of a polysemous word, and so on.

Introduction
It has often been said that an Explanatory Combinatorial Dictionary (hereafter ECD) is a formal dictionary, which, therefore, follows certain rigorous principles, principles that have been outlined in various publications: for example, in the introduction to the ECD of Modern French (Mel’čuk et al. 1984) and the ECD of Modern Russian (Mel’čuk and Žolkovskij 1984). However, in spite of the fact that these principles have been scrupulously observed in both published ECD’s, as well as in other works of ECD lexicographers, they have never been presented as a system – in a single place and according to a single canonical schema. It thus seems useful to fill this gap, by giving here a survey of certain basic lexicographic principles postulated for the ECD’s.

Two preliminary remarks are in order. First, I will examine only those lexicographic principles which specifically concern the semantic description of a lexical unit: i.e., principles for lexicographic definition. I will not consider two other types of lexicographic principles: neither general principles, such as the
strictly formal character of all descriptions, the orientation towards expression rather than towards comprehension, the rigid structure of the dictionary entry, etc; nor specific principles for the description of the syntactic behavior of lexical units (i.e., government patterns) and of restricted lexical co-occurrence (i.e., lexical functions). Instead, I will deal exclusively with the construction of definitions conforming to the general orientation of an ECD.

Second, I lay no claim to originality in regard to the principles to be discussed. Known from antiquity, the principles for the formulation of scientific definitions have been the object of general study by Essler (1970); in the lexicographic domain, they have been described by, among others, Rey-Debove (1966), Apresjan (1968, 1969a,b, 1974: 95ff.), and Mel'čuk (1974: 111; 1982: 20–21). They have been vigorously affirmed and justified in the semantic work of Wierzbicka (1972, 1980, 1985, 1987, to appear) where they have been applied to the semantic description of hundreds of lexemes (and grammemes) of English, Russian, Polish, and a number of other languages as well. My only goal in the present paper is thus a systematization of principles more or less already adopted, with a heavy emphasis on their formal character.3

In order to understand the importance of the formal principles stated below, the reader needs a general characterization of Explanatory Combinatorial Dictionaries; I will supply this in very brief outline subsequently. Furthermore, the principles under consideration must be formulated in terms of some basic lexicographic concepts: lexeme, vocable, 4 semantic field, etc. Once again, though some of these concepts have already been defined in various other works on ECD's, it seems preferable to bring their definitions together here – always with an eye toward their better systematization, especially so since in lexicography, there is no consensus on the notional content and usage of such terms. (Perhaps the first attempt to bring some order to this domain, specifically in French lexicography, is Quémada 1986.)

Finally, along with established basic principles, our team uses, in our daily lexicographic work, a number of concrete rules, recipes, or technical heuristics – i.e., rules of thumb or approximate criteria. These criteria are not yet sufficiently honed; the logical relations among them are not always clear, and even their validity can sometimes be questioned. However, they turn out to be very useful and deserve to be reviewed, although they cannot as yet be promoted to the rank of principle. (Some of these criteria are treated in more detail in Mel'čuk et al. 1988.)

The present article, then, contains the following four sections:

1. General Description of the ECD
2. Seven Basic Lexicographic Concepts
3. Six Basic Lexicographic Principles for Definitions
4. Eleven Heuristic Criteria for Definitions

---

3 The reader is referred to the works of Wierzbicka (1972, 1980, 1985, 1987, to appear) for a detailed discussion of these and other principles.

4 The term vocable is commonly used in lexicography to denote a lexical unit that can function as a word, i.e., a unit of meaning that is distinguished from other units of meaning by its pronunciation, spelling, and sometimes its part of speech.
General Description of the ECD

The Explanatory Combinatorial Dictionary is intimately related to the linguistic Meaning-Text Theory, proposed by myself and A. Zholkovsky more than twenty years ago (Zolkovskij and Mel'čuk 1965; cf. Mel'čuk 1981). In a nutshell, the principal claim of that theory is that a natural language is a specific system of correspondences between an infinite set of meanings and an infinite set of texts:

\[
\text{Language} \\
\{\text{MEANING}_i\} \leftrightarrow \{\text{TEXT}_j\} \quad 0 < i, j < \infty
\]

A description of this correspondence by means of a complex set of formal rules is called a Meaning-Text Model (of natural language). An ECD is one of the core components of any Meaning-Text Model: it ensures the lexicalization of the initial meaning (i.e., of semantic representation), uniting bundles and configurations of semantic elements into actual lexical units and supplying the enormous bulk of syntactic and lexical co-occurrence information that accrues to the individual lexical units of the language in question.

In 1966, Zholkovsky and myself began work on an ECD of Russian; shortly thereafter, Ju. Apresjan joined us. In the following decade, some twenty researchers participated in the endeavor, with the result being a fragment of the ECD of Modern Russian, published much later (Mel'čuk and Zholkovskij 1984). After I was forced to leave the Soviet Union (for reasons that have nothing to do with linguistics in general), a permanent lexicographic team was established at the University of Montreal to develop an ECD of Modern French. The efforts of this team have been and continue to be published (Mel'čuk et al. 1984, 1988).

The ECD differs from familiar monolingual dictionaries in that it is not a commercial dictionary, which aims at a particular public, has a limited practical purpose, and exists within certain pedagogical, typographical, and financial constraints. An ECD is actually part of a scientific (= theoretical) description of language and is thus a type of THEORETICAL LEXICON (cf. Lakoff 1973: 162–64). An ECD is to a ‘normal’ dictionary what a theoretical grammar is to a language textbook (which, of course, means that a ‘normal’ dictionary ought to be sensitive to the notions and formalisms of an ECD). This fundamental fact entails the following five properties typical of an ECD:

1. An ECD is PRODUCTION-ORIENTED. It is intended to supply all the information which is conveyed by individual lexical units and which is necessary to express a given meaning in a completely idiomatic way.
2. An ECD is SEMANTICALLY BASED (hence ‘explanatory’). The definition of the entry lexeme, written in a special semantic language, serves as a basis for the description of all the paradigmatic and syntagmatic relations of this lexeme.
3. An ECD is COMBINATORIAL. It describes the syntactic and lexical collocations of the entry lexeme in the greatest possible detail.
4. An ECD is SYSTEMATIC. It places heavy emphasis on the homogeneity of the
lexicographic descriptions. It is developed by lexical fields rather than by alphabetical listing: similar lexemes have similar descriptions; all links existing between the definitions and the syntactic or lexical co-occurrence of the lexeme must be made explicit.

5. An ECD is formal. All information is presented by means of a rich lexicographic metalanguage, which ensures a rigor never before attempted in lexicography.

Consistent with these five properties, an ECD maintains a rigid standard structure for its entries, as described below. An entry of the ECD, its basic unit, corresponds to a single lexeme or phrase: i.e., one word or one set phrase taken in one separate sense. A family of dictionary entries for lexemes or phrasemes which are sufficiently close in meaning and which share the same stem(s) is subsumed under one lexical entry (see below).

An ECD entry is divided into three major zones (which, among other things, also supply more traditional types of lexicographic information not mentioned here – for example, usage labels and morphology): semantic zone, syntactic zone, and lexical co-occurrence zone. The semantic zone contains the definition of the lexeme, written in accordance with the formal principles discussed at length below. The syntactic zone contains the government pattern, by means of which the correspondence between the meaning and the deep/surface syntax of the lexeme is specified. Its format is a rectangular table, each column of which represents a semantic actant of the lexeme, with each element in the column representing one possible surface realization of the corresponding syntactic actant. The lexical co-occurrence zone contains all restricted lexical co-occurrences of the entry lexeme, represented by lexical functions, or semantic 'operators', which specify both the syntagmatic and paradigmatic relations which the entry lexeme can have with other lexemes (see, e.g., the paper by Frawley in this issue).

The upshot of the above is that the ECD is a very different sort of dictionary: formal, rigorous, and exhaustive on the level of individual lexical units. It indicates precisely not only what an entry lexeme means, but also its syntactic environment, as well as its semantic universe. To the best of my knowledge, no other dictionary can claim to do this.

A central reason for the success of a project like an ECD is that it is based on clear principles. It is to these principles that I now turn, providing, first, the definitions of seven concepts needed in the discussion.

**Seven Basic Lexicographic Concepts**

The unit of lexicographic description in the ECD is the *lexical unit*, which can be either a *lexeme* or a *phrase*.

**Definition 1: Lexeme**

A *lexeme* is a word taken in one, well-specified sense and supplied with all the information specifying its behavior when it is used in this sense.
A lexeme is printed in large capitals: e.g., RESPECT, SNOW, STOMACH, etc.

In a text, lexemes come together in phrases. A phrase \( L_1 + L_2 + \ldots + L_n \) is free if and only if all its semantic and syntactic properties are completely determined by the respective properties of its constituent lexemes (and by the general rules of syntax); otherwise, it is bound. (Clearly, free phrases are of no interest for the dictionary.)

Definition 2: **Phraseme**
A phraseme is a bound phrase taken in one, well-specified sense and supplied with all the information specifying its behavior when it is used in this sense.

A phraseme is also printed in large capitals and surrounded by raised half-brackets: e.g., "ON FOOT", "AT ONCE", "SHOOT THE BREEZE", etc.

For each single lexical unit, the ECD has a corresponding dictionary entry, and each entry in the ECD treats only a single lexical unit. Every lexical unit is thus the ***headword*** of the corresponding dictionary entry.

The phrase well-specified sense (of a lexical unit) refers to the signified of the lexical unit, which is represented in the ECD by an expression written in a special language, derived from the object language: in an English ECD, this special language is derived from English; in a French ECD, it is derived from French; and so on. This expression, which must satisfy several formal conditions, is called the **lexicographic definition**.

Now, if the definition (= the signified) "S_1" of a lexical unit has a non-trivial component in common with the definition (= the signified) "S_n" of another lexical unit, then we will say that the two units are **semantically linked** directly or that their signifieds are linked directly. If "S_1" is directly linked to "S_2", "S_2" to "S_3", ..., and "S_n-1" to "S_n", then "S_1" and "S_n" are simply linked (not directly, but via several intermediate direct links).

**Definition 3: Vocabile**
A vocabile is the set of all lexical units such that (i) their signifiers are identical and (ii) the signifieds of any two units are linked (directly or indirectly).

The semantic link between two lexical units is a **semantic bridge**.

**Definition 4: Semantic Bridge**
A semantic bridge between lexical units \( L_1 \) and \( L_2 \) is the non-trivial component common to their definitions, which formally expresses the semantic link felt between the two.

As an example, I will illustrate the concept of semantic bridge by showing the common components in the lexemes belonging to the English vocabile SPEED.

1. **Speed Y [of P] of X=** (a) Magnitude characterizing the displacement P of X by the value Y, which is the distance that X travels in a unit of time, or (b) this value [the speed of the car].
2a. **Speed Y [of P] of X=** (a) Magnitude characterizing the rotation P of X
by the value Y, which is the number of turns in a unit of time, or (β) this value 
\[ \text{the speed of the engine}. \]

(The semantic bridge with SPEED 1.1 is ‘Magnitude characterizing ... P of X by the value Y ... in a unit of time, or this value’.)

1.2b. Speed Y [of P] of X = (α) Magnitude characterizing X’s treatment P of information by the value Y, which is the number of units that X treats in a unit of time, or (β) this value [the speed of the laser printer].

(The semantic bridge with SPEED 1.1 is ‘Magnitude characterizing ... P ... of X by the value Y ... in a unit of time, or this value’.)

1.3. Speed [of X] = Great speed 1.1, 2a, 2b of X [Paul loves speed (of a car, of an engine, of a typewriter, etc.)].

(The semantic bridge with SPEED 1.1, 2a, 2b is self-evident: by inclusion.)

II. Nth speed of Y [of X] = the Nth step in the gear system of a bicycle Y designed to allow the rider X of Y to change it as a function of desirable speed 1.1 of Y [the tenth speed of the bicycle].

(The semantic bridge with SPEED 1.1 is self-evident: by inclusion.)

III. Speed of P = The fact that an action or process P is accomplished quickly [the speed of the decision].

(The semantic bridge with SPEED 1.3 is indirect, and appears in the process of semantic decomposition of ‘great speed 1.1, 2a, 2b’, which contains ‘small time interval between two consecutive events’.)

Definition 5: Basic Lexical Unit of a Vocale
The basic lexical unit of a vocable is the lexical unit which has a semantic bridge [= is semantically linked directly] with the majority of the other lexical units of the vocable.

For example, in the set of lexemes of the vocable SPEED, the basic lexeme is SPEED 1.1 since all the other lexemes (SPEED 1.2a, 1.2b, 1.3, etc.) are semantically linked to SPEED 1.1 directly or indirectly. In other words, the definition of SPEED 1.1 serves as the basis for the definitions of all the other lexemes of SPEED. According to ECD conventions, the dictionary entry corresponding to the basic lexical unit of a vocable is placed FIRST: before all the other dictionary entries under that vocable.

Definition 6: Semantic Field
A semantic field is the set of all lexical units that share an explicitly distinguished non-trivial semantic component.

For example, the semantic field ‘parts of the human body’ includes the lexemes HEAD 1.1a, FINGER 1a, STOMACH 1, BACK 1.1a, etc. The non-trivial component common to all these lexemes is ‘human body part’.

Definition 7: Lexical Field
A lexical field is the set of all vocables whose basic lexical units belong to the same semantic field.

For example, the lexical field ‘parts of the human body’ contains full vocables HEAD, FINGER, STOMACH, BACK, etc. That is, this lexical field includes
not only lexemes which designate human body parts, but also those which do not, while belonging to the same vocable: *head* of the table, *finger* of fate, *back* of the house, etc.

**Six Basic Lexicographic Principles for Definitions**

Heretofore, this paper has been concerned with fundamental lexicographic concepts; the remainder of the paper will treat the construction and justification of definitions themselves. Whereas this discussion will address the major properties of ECD definitions, the following four important features of ECD definitions will not be specifically studied:

1. The definiendum is, generally speaking, a **propositional form** including the lexical unit in question with variables representing its semantic actants (rather than just a lexical unit); for instance, for *fill*, the definiendum is as follows: *X fills Y with Z*. The same variables appear equally in the definiens.

2. The definiens must render explicit, in the most general way possible, the **semantic invariants** found in the definiendum (rather than simply enumerating, one by one, the semantic features of the definiendum, no matter how relevant these features are).7

3. The definiens must avoid all overly **idiomatic expressions** in favor of more abstract and, therefore, more general formulations.

4. The definiens does not necessarily have to conform completely to the **stylistic norms** of the object language although it must not violate these norms needlessly. Note, e.g., the use of the expression *X causes that Y ...* in the ECD; note also that numerous deviations from standard usage might be due to the elimination of overly idiomatic expressions.

In what follows, I will assume the reader's familiarity with this general type of definition and I will present only the six aforementioned principles bearing on the formulation of definitions in the ECD. These principles can be divided into four groups, according to their application to the following:

A. The defining language
B. A particular definition
C. The system of definitions within a vocable
D. The system of vocables within a lexical field

**A. The defining language**

**Univocity Principle.** The defining language must not contain any ambiguous or synonymous terms.

Adherence to this principle means at least two things. First, each term used in the definitions of an ECD must have one, and only one, sense. Polysemous expressions that have to appear in a definition must thus be previously
disambiguated with the help of distinctive numbers specifying the intended sense and introduced elsewhere in the dictionary (see Mel'čuk et al. 1984: 5, 24). This practice is unknown in other dictionaries: none, to my knowledge, utilizes the numbers of the senses within its definitions. Thus, the Petit Robert (1977: 801), for example, defines FONCTIONNAIRE 'civil servant' as 'person qui remplit une fonction publique' = 'person who fulfills a public function', while PERSONNE has, according to the same dictionary, seven different senses, REMPLIR (headword for 'remplit') six, FONCTION 'nine, and PUBLIC (headword for 'publique') nine also, making the quoted definition theoretically 3,402 ways ambiguous ($7 \times 6 \times 9 \times 9$).

However, it should also be noted that complete disambiguation of the components of the definiens is, at present, but a program. Actually, at the moment, we are able to number only those senses (= lexical units) that have already been described in the ECD. Consequently, we are forced to use, in the definitions, non-disambiguated lexical units, sometimes supplied with provisional distinctive numbers, borrowed from some existing dictionary. This violation of the Univocity Principle represents only a temporary compromise, pending completion of our lexical description.

Second, each distinct semantic component must always be represented by the same linguistic expression. For example, in an ECD, the emission of a cry by a being and the production of a noise by an object cannot be designated in two different ways, as can be done, for instance, in a current dictionary: HOWL 'to emit a ...' vs. CRACK 'to make a ...' (see Webster's New Collegiate Dictionary). The definitions of an ECD follow the same formula in all cases – in choosing a defining expression which is a maximum block (see below) and, even more importantly, the least idiomatic expression possible, in this case 'produce':

\[
\begin{align*}
\text{MEOW} & \quad \text{BARK} \\
\text{HOWL} & \quad \text{YELP} \\
\text{MOO} & \\
& \quad \text{'to produce a cry...'} \\
\text{CRACK} & \quad \text{SNAP} \\
\text{CRUNCH} & \quad \text{BONK} \\
\text{BANG} & \\
& \quad \text{'to produce a noise...'}
\end{align*}
\]

Similarly, for all artifacts, an ECD says 'intended for...', instead of the several variants found in other dictionaries ('serving...,' 'having the goal of...', 'designed for...,' 'for...', 'that one uses for...', and so forth).
B. A particular definition

Adequacy Principle. In the definition of a lexical unit L, each component must be necessary, and the set of all components must be sufficient, for the definition to identify L uniquely in all imaginable uses.

In other words, following the old saying, we need to define the whole definiendum and nothing but the definiendum.

(a) Necessity.

The terms necessary and unnecessary are used here in their mathematical sense; unnecessary thus means ‘should not be present’ rather than simply ‘superfluous’ (as in common usage). Thus, in our view, an unnecessary semantic component is not simply redundant: it can be harmful. For example, it can falsely narrow the usage of a defined unit in comparison to its observed usage. The Petit Robert (1977: 152) indicates for BAIE ‘bay’ that its ‘opening is narrow’, a fact which contradicts the normal usage of the term: note the French sentence La baie avait une entrée très large ‘The bay had a very wide opening’.

Even worse, an unnecessary component may be not only too restrictive, but also contradictory, thereby destroying the definition: a well known example is French TABOURET ‘stool’ (see Rey-Debove 1966: 77), defined traditionally as ‘chaise sans dossier’= ‘chair without a back’, while CHAISE ‘chair’ is defined as ‘siège à dossier...’= ‘seat with a back’. TABOURET is then ‘siège à dossier sans dossier...’= ‘seat with a back without a back’.

In the best case, an unnecessary component does not change the intension of the definition, but affects it nevertheless, making it too cumbersome and cluttered. For example, the Petit Robert (1977: 1827) has the following definition of French SOLDAT ‘soldier’: ‘un homme qui sert dans une armée, en temps de paix ou en temps de guerre, comme mercenaire ou engagé volontaire...; ou, de nos jours, en vertu d’une obligation civique ou professionnelle...’= ‘a man who serves in an army, in peacetime or in wartime, as a mercenary or enlisted volunteer...; or, in modern times, by virtue of a civil or professional obligation...’ The entire portion of the definition to the right of the vertical bar is superfluous and is proscribed by the Adequacy Principle. Indeed, it exhausts all the logical possibilities and adds nothing to the first part: there is no other time except peacetime or wartime. (The presence of superfluous components in current lexicographic definitions, as is the case in the example of soldat, is often due to psychological considerations or to confusion between semantic (i.e., linguistic) information and encyclopedic (i.e., extralinguistic) information – see below, Criterion 4, for further discussion.) The rejection of unnecessary components derives from considerations of economy and represents a particular case of the general scientific principle known as Occam’s razor.

It must be stressed that the concept of necessary semantic component is much less evident than it might seem at first. Of course all distinctive components are necessary: they contribute to uniquely identifying the signified under analysis...
and distinguish it from all contiguous signifieds. But a non-distinctive component may also be necessary, though for other reasons. On the one hand, it may be necessary in order to make explicit the possible co-occurrence of the lexeme in question with other lexemes in a sentence, to describe the semantic interaction of the lexeme with its syntactic partners, etc. (i.e., syntagmatic reasons). On the other hand, it may be necessary in order to make explicit perceived semantic links between the lexeme in question and its lexical 'relatives' (i.e., paradigmatic reasons). In any case, the term necessary, such as it is applied to components of ECD definitions, requires special study.

(b) Sufficiency.

The absence of a necessary component extends the usage of a definiendum beyond its observable usage. Thus, take the definition of French COMBLER I.1a ‘fill up’ (X comble Y = ‘Corps’ I.1a, b X solide rempli† I.1a (α) entièrement un espace Y limité tridimensionnel...’ = X fills Y = ‘Solid body X fills entirely a limited three-dimensional space Y...’); if the component entièrement ‘entirely’ is removed, the definition ceases to be sufficient since incorrect usages are not prevented, such as *La boue comblait la fosse aux trois quarts, lit. ‘The mud filled (up) three-quarters of the hole’ (while La boue comblait la fosse ‘The mud filled (up) the hole’ is a normal sentence).

The operational corollary of the Adequacy Principle is the absolute substitutability of the definiens and definiendum: they must be mutually substitutable salva significatone in all imaginable contexts:

1. In linguistic contexts: i.e., we require the commutability of whatever lexical unit with its definition in whatever sentence of the language.
2. In extralinguistic contexts: i.e., we require the commutability of whatever lexical unit with its definition to designate whatever state of affairs (given a state of affairs and a lexical unit L to designate it, the definition of L designates it as well, and designates no other state of affairs, and vice versa).
3. In metalinguistic contexts: i.e., we require the commutability of whatever lexical unit with its definition within all lexicographic definitions (in whatever definiens, any term can be replaced by its own definition – and vice versa – without affecting the validity of the initial definition).

Mutual substitutability of the definiens and the definiendum is of primary importance for the present lexicographic approach; we always use substitutability to justify the presence or absence of a semantic component. It is likewise the principal technique in Wierzbicka’s (1972, 1980, 1985, 1987, to appear) semantic research.

**Decomposition Principle.** The definition of a lexical unit L must contain only terms that are semantically simpler than L.

Two important comments are in order here. First, simpler must not be construed in a psychological sense: as ‘easier to understand’, ‘more evident
from the viewpoint of linguistic intuition', or 'more common'. By \( L_1 \) is semantically simpler than \( L_2 \) we mean only that \( L_2 \) can be defined in terms of \( L_1 \), but not the inverse. (It can be demonstrated that the 'inversibility' of definitions does not exist.) I emphasize that the definiens may contain terms that are much less current and much less known to readers than the definiendum; such is the expression 'if (to) cause that ...' (in the sense of 'bring about that') used in the definitions of most current transitive verbs. The only valid consideration in the choice of the components of the definiens is the direction of possible decompositions.

Second, \( L_1 \) does not mean 'simple': except for semantic primitives, every component of a definiens is semantically complex and must be decomposed in turn (in the corresponding dictionary entry).

According to the Decomposition Principle, the meaning of the lexical unit defined is always decomposed, in its definition, into simpler meanings. Applied methodically, such decompositions guarantee the absence of vicious circles in the system of definitions and eventually lead to semantic primitives.8

A definition in an ECD may deviate from the Decomposition Principle in only the following two cases:

1. If lexical units \( L_1 \) and \( L_2 \) are perfect synonyms, then one may be defined by the other, which, itself, must have a regular definition (i.e., a decomposition). More precisely, the definition, say, of \( L_1 \) is replaced by the reference to the dictionary entry of \( L_2 \). For example, in French, 'DOIGTS DE PIED', lit. 'foot fingers', means the same as ORTEILS 'toes', and therefore the phrase 'DOIGTS DE PIED' is defined as follows: 'DOIGTS DE PIED' = 'orteils'. The direction of the reference, from \( L_1 \) to \( L_2 \) or vice versa, is a function of convenience (e.g., from less current to more current) and has no scientific relevance.

2. If a lexical unit \( L_1 \) is a syntactic derivative of another lexical unit \( L_2 \), then the definition of \( L_1 \) is replaced by the symbol of the corresponding lexical function with the lexical unit \( L_2 \) as the argument (see Melčuk et al. 1984: 20). For example, DESTRUCTION is defined as \( S_0(DESTROY) \), PAYMENT as \( S_1(PAY) \), etc. (We use symbols of the type \( S_0 \) where existing dictionaries use expressions such as state of or action of, phrases which represent an unfortunate confusion between purely semantic elements of the definition and elements of the grammatical metalanguage.)

Interestingly, the reductionist character of all definitions (and, in particular, of lexicographic definitions) was observed twenty-five hundred years ago by Aristotle, who required that a definition always reduce what is conceptually 'posterior' to what is conceptually 'anterior.'

Maximum Block Principle. If a definition contains a free phrase which is composed of lexical units \( L_1 + L_2 + \ldots + L_n \) and is semantically equivalent to a lexical unit \( L \), then this phrase must be replaced by \( L \).
The lexical unit L is called the maximum block with respect to the definition in question.
This principle is justified by three considerations:

1. It ensures gradual decomposition into ‘immediate semantic constituents’, without bypassing a constituent; such a procedure contributes to making explicit, in a consistent manner, all the hierarchical semantic relations among lexical units.
2. It makes definitions which would otherwise be excessively long and complex easier to read and manageable.
3. It facilitates intuitive judgments of substitutions when such substitutions are carried out to verify definitions.

For example, let us consider the definition of PROMISE (in Peter promised this trip to Mary) as ‘X causes explicitly that Z know that X is certain that there will be an event (related to) Y ...’ This is substantially adequate, but formally incorrect: it violates the Maximum Block Principle since, in current English, ‘cause explicitly that someone know that’ is also INFORM; therefore, the definition of PROMISE ought to read ‘X informs Z that X is certain that ...’

The Decomposition Principle establishes the lower limit for the size of the definition: not less than two semantic components (except, of course, in the case of synonyms and syntactic derivatives mentioned above). The Maximum Block Principle establishes the higher limit: as few semantic components as possible from the viewpoint of the stock of defining elements (i.e., lexical units) available in the language.

C. The system of definitions within a vocable

Semantic Bridge Principle. The definitions of any two lexical units of the same vocable must be explicitly linked: whether by a semantic bridge or by a sequence of semantic bridges.

It should be noted that semantic links between two lexical units can be established not only via a component common to their definitions, but also via a connotation (which is not part of the definition, but which is included in the semantic zone of the dictionary entry: see Mel’čuk et al. 1984: 33ff.).

The very definition of a vocable predicts semantic links among the lexical units constituting it. What this principle adds is that such links must be formally expressed in the definitions: by semantic bridges. Thus, if the intuition of the lexicographer places the units L₁ and L₂ in the same vocable, the Semantic Bridge Principle guarantees the explicit expression of this intuition by a semantic bridge in the definitions of L₁ and L₂. For example, in an ECD of Modern English, one cannot have two different lexemes for PIG, PIG I and PIG II, defined as: I – ‘domestic animal...’ and II – ‘slovenly person...’ The Semantic Bridge Principle requires that PIG II be defined as ‘slovenly person...’, as if he were a pig I’ and that the lexical connotation ‘slovenly’ be introduced into the semantic zone of the entry for PIG I.
It is important to emphasize that a semantic bridge may be contained in a definition indirectly and appear only in later stages of semantic decomposition. An example of this is found above in the case of SPEED I.3 and SPEED III. Similarly, in Mel'čuk et al. (1984: 31), the French lexeme TÊTE III ‘head’ (as in the head of the parry) is linked to TÊTE I.1a ‘head’ = ‘the top part of the body...’ via the component ‘À LA TÊTE’ ‘at the head’, which, in turn, contains TÊTE I.1a: [Y] étant à la tête de X = [Y] dirigeant ou gérant X [as if Y were the head of X and exercised control over X]’.

In order to have only directly ‘visible’ semantic bridges, our definitions should always be formulated in terms of basic semantic elements: i.e., in terms of semantic primitives. In the ECD, as I have indicated, we adopt the opposite strategy: the definitions are virtually never written in terms of basic elements; we prefer gradual semantic decomposition. As a result, semantic bridges that are visible only indirectly are unavoidable, making the task of reading the dictionary more difficult; in fact, in certain cases, the user may be forced to consult more than one definition in his search for a semantic bridge. But this does not affect the essence of the Semantic Bridge Principle: the authors of an ECD must be sure, in putting two lexemes in the same vocable, that the necessary semantic bridges exist, and they must know at what stage of decomposition these bridges will be visible.

D. The system of vocables within a lexical field

Uniformity Principle. Two vocables belonging to the same lexical field must be presented, everything else being equal, according to the same schema: (i) the related lexical units of these vocables must appear in the same order within each vocable; (ii) the semantic distances between the related lexical units must be represented identically.

This principle is explicated more fully in Mel'čuk et al. (1984: 31–32). One can see, however, the uniformity of presentation in the following table, taken from Mel'čuk et al. (1984: 32), which shows the parallel semantic structures for the vocables TÊTE ‘head’ and COEUR ‘heart’ in the ECD of Modern French:

<table>
<thead>
<tr>
<th>TÊTE</th>
<th>COEUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. partie du corps I.1a d'une personne</td>
<td>1a. organe d'une personne</td>
</tr>
<tr>
<td>1b. partie du corps I.1b d'un animal</td>
<td>1b. organe d'un animal</td>
</tr>
<tr>
<td>2. produit alimentaire</td>
<td>2. produit alimentaire</td>
</tr>
<tr>
<td>TÊTE</td>
<td>COEUR</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>3a. visage [partie antérieure de la tête 1.1a]</td>
<td>3. partie de la poitrine sous laquelle se trouve le cœur 1.1a</td>
</tr>
<tr>
<td>4. organe de la raison</td>
<td>4a. organe des sentiments</td>
</tr>
<tr>
<td></td>
<td>4b. organe de l’intuition</td>
</tr>
<tr>
<td>5a. propriété de la personnalité liée à la raison</td>
<td>5a. propriété de la personnalité liée aux sentiments</td>
</tr>
<tr>
<td>5b. personne possédant cette propriété</td>
<td>5b. personne possédant cette propriété</td>
</tr>
<tr>
<td>II 1a. partie supérieure d’un objet</td>
<td>2a. partie centrale d’un espace topographique</td>
</tr>
<tr>
<td>1b. partie antérieure d’une suite d’objets</td>
<td>2b. partie centrale d’une plante</td>
</tr>
<tr>
<td>3. partie autonome</td>
<td>1a. partie principale</td>
</tr>
<tr>
<td></td>
<td>1b. élément principal</td>
</tr>
<tr>
<td>4. unité de légumes ayant la forme sphéroïdale</td>
<td>3. objet ayant la forme du cœur 1.1a</td>
</tr>
<tr>
<td></td>
<td>4. couleur 2 des cartes à jouer dont les points ont la forme du cœur 1.1a</td>
</tr>
</tbody>
</table>

(From Mel’čuk et al. 1984: 32, numbering therefrom).

**Eleven Heuristic Criteria for Definitions**

As I have already said, I will briefly review some heuristic criteria that our team presently uses, even though we do not fully understand their logical nature, and sometimes we are not even certain of their validity. I will divide them into three groups, according to their application to the following:

A. The structure of a particular definition  
B. Distinguishing lexical units within a vocable  
C. Ordering lexical units within a vocable
A. Criteria for the structure of a particular definition

**Criterion 1.** The definition of L must explicitly reflect the free co-occurrence of L with quantifiers, especially with numerals, if L is a noun, or with the expression \( n \) times, if L is a verb.

This technique contributes to the choice of the central component of a definition. At the same time, co-occurrence with quantifiers facilitates the distinction of lexemes within a vocable (cf. Mel'čuk 1979). We can cite, here, the case of French PLUIE I ‘rain’ vs. NEIGE I.a ‘snow’ (analyzed in more detail in Mel'čuk et al. 1988). Each of these two lexemes designates a substance (water and white flakes, respectively) and the falling of this substance, since one can say *La pluie fouettait ... ‘The rain beat ...’* [the substance] and *La neige couvrait ... ‘The snow covered ...’* [the substance] and *La pluie-La neige a commencé à trois heures ‘The rain/The snow began at three o’clock’* [the falling]. However, *trois pluies*, lit. ‘three rains’, is possible, but *trois neiges* ‘three snows’ is not. This forces us to define PLUIE I as ‘chute d’eau ... ou cette eau qui tombe’ = ‘falling of water ... or this water that falls’ since, in French, CHUTE ‘falling’ is countable, but NEIGE I.a as ‘substance ... ou la chute de cette substance’ = ‘substance ... or the falling of this substance’ since SUBSTANCE (in the intended meaning) is not countable in the same sense in which CHUTE is countable (*trois substances* ‘three substances’ means ‘three different substances’ or ‘three types of substances’ rather than triple repetition of the same).

**Criterion 2.** The definition of L must explicitly reflect the free co-occurrence of L with all types of modifiers – adjectivals, if L is a noun, and adverbials, if L is a verb or adjective.

Of all modifiers, intensifiers would seem to be the most revealing with respect to the selection of semantic components. For example, if, in English, one says *wild applause* with *wild* functioning as an operator displacing a measurable value towards the high end of the corresponding scale, then the definition of *applause* must include at least one component capable of ‘accepting’ this operator. In fact, this component for *applause* is ‘strike’ (‘the striking of the hands against each other ...’), which, in turn, includes ‘contact having a certain force and certain frequency’: operators like *wild* displace the values for force and frequency toward the high end of the scale (or toward the low end: *polite applause*).

**Criterion 3.** The definition of L must explicitly reflect the way in which L combines with negation.

The part of the meaning of L that is never affected by negation constitutes either a *PRESUPPOSITION*, which must be expressed by a participial phrase outside the scope of the negation, or a *SEMANTIC RESTRICTION*, which must be expressed by a relative clause modifying the corresponding variable. Thus, ALLOW, as in *X allows Z to do Y*, implies the following component: ‘X knows
or believes that Z wants to do Y and that Z does not want to do Y against X’s will’ (in general, this sense of allow occurs in response to a previous request). The negation of the verb does not affect this component: John did not allow his son to go to the party still implies a request – either actual or hypothetical – on the part of John’s son. Consequently, this component constitutes a presupposition in the definition and has to be formulated as follows: ‘knowing or believing that Z wants ...’

It must be stressed that the above three criteria do not fully exhaust all cases of free co-occurrence which must be accounted for in a definition. Further research is needed in this matter.

The following criterion is applicable only in one particular case: that of the definition of a concrete (physical) object or phenomenon. If a semantic component ‘C’ corresponds to an actual characteristic or property of the object, then we will call this characteristic physical, for lack of a better term; let us suppose that this component is not distinctive (see Mel’čuk et al. 1988). Often the question arises of whether or not to include ‘C’ in the definition of the lexeme which designates the object in question. This amounts to deciding whether or not ‘C’ represents encyclopedic information (with regard to the object itself) or linguistic information (with regard to the lexeme designating the object). For instance, should the lexicographic definition of SNOW or SALT mention the white color of these substances? Or would such mention be an intrusion of encyclopedic information into a dictionary of the language? The correct decision is not always easily made, but the lexicographer may have recourse to the following criterion in making this decision.

**Criterion 4.** The definition of L must contain the non-distinctive ‘physical’ component ‘C’ (corresponding to an actual characteristic of the object or phenomenon designated by L), if and only if the language in question possesses expressions which are morphologically related to L and which, in one way or another, use ‘C’; in particular, if the language possesses a lexical unit L, whose signifier has the same stem as L and whose signified includes ‘C’ (or a part of ‘C’).

This criterion may also be called the *criterion of linguistic relevance*. Its importance in justifying non-distinctive semantic components has been noted by Apresjan (1969a: 23) and Wierzbicka (1985: 197ff.).

Some examples follow. The semantic component ‘(partially) hides the sky’ in the definition of CLOUD I is justified by the existence of the lexeme CLOUD II and the phrase CLOUD IN THE SKY, whose definitions contain ‘darkens ... as a cloud I (partially) hides the sky’ (His death was the only cloud in the sky of an otherwise happy year; A cloud darkens the Prime-Ministerial brow). The semantic component ‘violent’ in the definition of STORM 1 is justified by the existence of a derivative lexeme, STORMY 2 (stormy argument), whose meaning includes the idea of violence, as well as by the existence of STORM 2 (storm of anger). Finally, the semantic component ‘white’ in the definition of
SNOW is justified by the existence of SNOWY 'having a beautiful white color', and by the expression white as snow. (In much the same way, the definition of SALT or SUGAR possibly must include an indication of their whiteness, given the expressions salt and pepper [hair] and sugary white; however, the definition of RICE, which refers to an object no less white than salt, sugar, or snow, should not include any mention of whiteness since there are no expressions in English which use this feature: *ricey white, etc.)

Criterion 4 is also used to distinguish lexical connotations from encyclopedic connotations (see Mrček et al. 1984: 35, Condition 5).9

B. Criteria for distinguishing lexical units within a vocabulary

Consider a hypothetical definition which contains two components 'C₁' and 'C₂' (with whatever relation between them). In other words, we have a hypothetical lexical unit L defined as ' ... C₁ ... C₂ ...'. For reasons that I cannot make more precise here, one can ask whether one must keep 'C₁' and 'C₂' in the same definition or divide the definition into two simpler ones, 'A + C₁' and 'A + C₂', and thereby create two different lexical units (instead of a single one). In trying to answer this question, we use the following three criteria.

Criterion 5. A hypothetical lexical unit L ' ... C₁ ... C₂ ...' must not entail ambiguity of the sentences containing it.

This can also be called the ambiguity criterion. Its application can be seen in the following example. The French sentence Jean leur a proposé un voyage 'John proposed a trip to them' can mean two things: either 'Jean leur a communiqué ... qu’il leur rendrait possible un voyage' [= 'offrir'] = 'John informed them that he would make a trip possible' [= 'offer'] or 'Jean leur a communiqué ... qu’il serait souhaitable qu’un voyage ait lieu' [= 'suggérer'] = 'John informed them that it would be desirable for a trip to take place' [= 'suggest']. From this, it follows that the French verb PROPOSER 'propose', the headword for a proposé 'proposed', has at least two different senses: PROPOSER 1 and PROPOSER 2a (see Mrček et al. 1984: 144-45).

The absence of such ambiguity does not mean that 'C₁' and 'C₂' must be joined in the same definition. This criterion can be used only negatively.

A brief remark is in order here. It is far from easy to distinguish between true ambiguity and simple indeterminacy derived from the vagueness of an utterance. The sentence The prices fell means either that the prices became lower than they were (while still remaining high) or that they became very low; it is vague rather than ambiguous (with fall meaning, in this context, 'become lower or just low'). I do not analyze here this important distinction in detail; let me only note that there is an interesting link between indeterminacy and the presence (in the definition) of the semantic element 'or'.

Criterion 6. If a dictionary entry for a lexical unit L ' ... C₁ ... C₂ ...' contains
two disjoint co-occurrence subsets (whether morphological, syntactic, or lexical) such that one corresponds to component \( C_1 \) and the other to component \( C_2 \), then this is evidence for the division of \( L \) into two lexical units.

This criterion can also be called the criterion of differential co-occurrence. It states that if the actualization of component \( C_1 \) entails a system of co-occurrences which never appear with component \( C_2 \) and vice versa, then the lexicographer must consider dividing the definition \( \ldots C_1 \ldots C_2 \ldots \) into two. For example, it may be the case that \( L \) is not pluralizable or that \( L \) allows a particular complement in the sense of \( C_1 \), but not in the sense of \( C_2 \).

Let us consider, for instance, two senses of the verb ACCEPT, as illustrated by the following sentences: \( \text{John accepted the death of his wife and John accepted their present.} \) Prima facie, ACCEPT in each sentence above can be described by the same general formula: \( \text{\'Facing situation Y \( \{= \text{\textbf{C}_1}\}\}, \text{in particular, when Z offers Y to X \( \{= \text{\textbf{C}_2}\}\), X does what Y demands of him or, at least, does not oppose it.} \) However, there are differences in syntactic co-occurrence for each appearance of \( \text{accepted.} \) In the first sentence, but not in the second, ACCEPT can take a sentential complement instead of the direct object without its original meaning being affected: \( \text{John accepted that his wife had died vs. \*John accepted that they had presented him [with this].} \) In the second sentence, but not in the first, ACCEPT can take a further argument, introduced by \( \text{from: John accepted the present from his friends vs. \*John accepted the death of his wife [\textit{\textbf{from God}?}].} \) According to Criterion 6, we distinguish at least two lexemes in ACCEPT: ACCEPT I for sentences of the first type and ACCEPT II for those of the second type. ACCEPT I is defined by the general formula above, but without \( \text{\textbf{C}_2} \), and ACCEPT II is defined as follows: \( \text{\textbf{Z} offering Y to X, X accepts I Y.} \) Thus, ACCEPT II describes a particular case of the situation described by ACCEPT I.

The criterion of differential co-occurrence has two important properties. First, in the same way as Criterion 5, this criterion may be used only negatively: it can influence the division of a definition, but it does not say anything about the joining of \( \text{\textbf{C}_1} \) and \( \text{\textbf{C}_2} \) in the same definition. Such a situation is common in linguistics. For example, in phonology, while the criterion of minimal pairs prevents the union of two opposed elements, \( X_1 \) and \( X_2 \), in the same X-eme, it does nothing to facilitate the union of \( X_1 \) or \( X_2 \) with other elements in appropriate X-emes.

Second, the criterion of differential co-occurrence is not absolute in that the lexicographer can ignore its indications – for reasons that are still not entirely clear. The size and systematicity of the two disjoint subsets of co-occurrences for \( \text{\textbf{C}_1} \) and \( \text{\textbf{C}_2} \) play an especially important role here. If, for the sense of \( \text{\textbf{C}_1} \), there are only one or two special lexical functions that combine with just this sense, while all the other functions combine with both senses \( \text{\textbf{C}_1} \) and \( \text{\textbf{C}_2} \), or with \( \text{\textbf{C}_2} \) only, then the weight of Criterion 6 might not be sufficient to force the division of \( \ldots \text{\textbf{C}_1} \ldots \text{\textbf{C}_2} \ldots \); a restrictive condition on the respective lexical
functions will suffice. The same is true concerning morphological peculiarities. For example, in French, NEIGE I a ‘snow’ has a plural only in such expressions as neiges éternelles, lit. ‘perpetual snows’, les neiges de l’Antarctique ‘the snows of Antarctica’, etc. and in a few similar phrases. If there were no other differences, then this restricted pluralization could be indicated within the single dictionary entry of NEIGE I, rather than postulating, according to the criterion of differential co-occurrence, two different entries (= lexemes), NEIGE I a (without plural) and NEIGE I.b (without singular). 11 Criterion 6 thus does not prescribe a division, but only provides evidence for one.

**Criterion 7.** Criterion 7, the criterion of compatible co-occurrence, must be formulated in two directions: positive and negative.

**Criterion 7a** (positive direction). If a hypothetical unit L’ ... C1 ... C2 ... ‘ permits simultaneous syntactic combination with L1 and L2 (whether both of them be syntactic dependents of L or one the governor of L and the other its dependent), where L1_L2, such that L1 corresponds to ‘C1’ and L2 to ‘C2’, then the definition of L must not be divided into two.

This formulation is a trivial generalization of the so-called Apresjan Criterion (Apresjan 1974: 85), which was postulated to account for a disjunction of components within a definition: if the lexical unit L having a definition of type ‘ ... C1 ... or C2 ... ’ admits the syntactic conjunction of L1 and L2, which are syntactic partners of L such that L1 corresponds to ‘C1’ and L2 to ‘C2’, then the disjunction in the definition of L is justified. For example, in French, BOMBARDER ‘bomb/shell’ means ‘lancer des bombes ou des obus’ = ‘to drop bombs or fire shells’; although bombs are dropped from planes and shells are fired from artillery, BOMBARDER can have a coordinate subject with one noun designating the planes and the other designating the artillery: Depuis deux semaines, les avions et les navires de guerre ennemis bombardaient le port sans répit ‘For two weeks, the enemy airplanes and warships bombed and shelled the port continuously’. (Note that the French sentence is not humorous, something that would obtain if the two component ‘C1’ and ‘C2’ ought not to be within the same definition: compare the subsequent example.)

**Criterion 7b** (negative direction). If a hypothetical unit L’ ... C1 ... C2 ... ‘ does not permit simultaneous syntactic combination with L1 and L2, such that L1 corresponds to ‘C1’ and L2 to ‘C2’, then this is evidence in favor of the division of the definition of L into two.

This is the inverse of the generalized Apresjan Criterion. One cannot join contrasting components under one hypothetical lexical unit without producing a pun. This can be seen in the verbal item PLAY, which can mean both ‘amuse oneself’ and ‘show in public, as theatre’. Combination of the two with simultaneous syntactic manifestation of each leads to a humorous expression: *The child and ‘Hamlet’ were playing this evening, where ‘C1’ is ‘amuse oneself’
('the child') and 'C_2' is 'show in public, as theatre' ('Hamlet'). The lexeme \textsc{play}, therefore, cannot be defined as '...amuse oneself...or show in public, as theatre'.

Criterion 7a (compatible co-occurrence entails the union of 'C_1' and 'C_2' in the same definition) is stronger than Criterion 7b (incompatible co-occurrence is evidence in favor of the division of the definition into two: one with 'C_1' and the other with 'C_2'). Indeed, one cannot describe the compatibility of two co-occurrences (that of 'C_1' and that of 'C_2') under L without declaring L to be a single lexeme including 'C_1' and 'C_2'. But the incompatibility of two co-occurrences may well arise for several reasons other than the simple existence of two different lexemes L_1 'C_1' and L_2 'C_2'. Consequently, the incompatibility of co-occurrences is less conclusive (again 'provides evidence for', not 'prescribes').

\textbf{C. Criteria for ordering lexical units within a vocable}

\textbf{Criterion 8}. If, for two lexical units L_1 and L_2 belonging to the same vocable, the definition of L_2 includes that of L_1 ('L_2' \Rightarrow 'L_1'), then L_1 precedes L_2 within the vocable.

For instance, in the vocable \textsc{speed}, the lexeme \textsc{speed} I.1 precedes the lexemes \textsc{speed} I.3 and II.

\textbf{Criterion 9}. In a vocable having a basic lexical unit L, the lexical units whose meanings are metonymic with regard to L precede the units whose meanings are metaphorical with regard to L.

If 'L_2' \Rightarrow 'L_1' and 'C_3' \Rightarrow 'L_1', such that L_2 is metonymic with regard to L_1, and L_3 is metaphorical with regard to L_1, then L_2 is semantically closer to L_1 than is L_3. This closeness is expressed simply by the order of the entries within the vocable since, for one thing, the sense numbering system in the \textsc{ecd} is not rich enough to accommodate the differences between metonymy and metaphor and, for another, these differences and the relative semantic closeness are not always well-determined nor always sufficiently stable to justify more explicit expression.

Thus, in the vocable \textsc{finger}, the lexeme \textsc{finger} I.a 'terminal part of the hand ...' precedes all other lexemes whose definitions include it; \textsc{finger} II.1 'part of a glove ...' (metonymy) precedes \textsc{finger} II.2 'part of an artifact...' (metaphor). (For other examples, see \textsc{mel'cuk} et al. 1988).

\textbf{Criterion 10}. Within a vocable, the order of lexical units which have metaphorical sense with regard to L must correspond to the order of components which underlay the corresponding metaphor within the definition of L.

For example, in French \textsc{bras} II.2 'arms' as 'parties latérales symétriques d'un artefact ...' = 'lateral symmetric parts of an artifact ...' (\textsc{les bras de l'ancre 'the arms of the anchor'}) precedes \textsc{bras} II.4 'arms/branches' as 'parties allongées d'un cours d'eau [comme si c''était des bras I.1a]' = 'elongated parts of...'
a body of water [as if these were arms I.1a] since in the definition of BRAS I.1a ‘(human) arms’ (= L), the component of localization [‘lateral’] precedes the component of form [‘elongated’].

Criterion 10 has, of course, only psychological or pedagogical value. Its job is, simply, to facilitate the reading of the dictionary, in particular the comparison of vocables within a lexical field.

Criterion 11. The typical lexical units of a lexical field, i.e., those which occur in the majority of its vocables, precede the non-typical lexical units, which appear only in a few of the vocables; this holds only if it does not contradict the preceding criteria.

For example, in French, DOIGT III.2, lit. ‘finger’ = ‘a small amount of liquid’ (un doigt de cognac ‘a drop of cognac’) is placed at the end of its vocable since no other vocable of the lexical field ‘human body parts’ has a similar lexeme: i.e., a lexeme signifying a small amount.

Like Criterion 10, Criterion 11 has psychological value only.

Notes

1 TRANSLATOR’S NOTE: This paper, originally written in French, has been translated into English by William Frawley. The English version, however, is substantially different from the French version; the latter is to appear in Mel’cuk et al. 1988. Since the paper concerns language proper and, in the original, the French language in particular, it necessarily poses certain problems for translation, problems unlike those that appear in, say, the translation of prose or even poetry.

First, as to language proper, because the paper discusses cited linguistic forms and their cited meanings, certain typographical conventions must be established to distinguish use, mention, gloss, etc. The following conventions obtain in this text: lexemes and vocables (see below) are in large capitals; cited linguistic forms are in italics; glosses are in single quotation marks; emphasized expressions and newly introduced terms are in small capitals; quoted material and all other facts of mention, such as ironic expressions, are in single quotation marks.

Second, many of the French examples in the original text, used to illustrate linguistic and lexicographic points, cannot be translated into English felicitously – for obvious reasons. Thus, appropriate English examples have been constructed in many places in the text; at other points, the original French example has been given, with a fairly close translation. In the former case, the task is rather difficult since no ECD of English exists: the present paper, in fact, represents one of the first major attempts to explain the ECD in English with English examples. In the latter case, every attempt has been made to provide comprehensible translation for illustrations that, as it turns out, are better off left in French so that the analysis and explanation are clear.

In short, the reader’s patience is requested for the typography, and his/her indulgence is begged for some of the examples.

2 I wish to thank Ju. Apresjan, D. Gaatone, L. Iordanskaja, and N. Pertsov, whose criticisms and judicious remarks have helped me greatly in the writing of this paper.
wish to thank as well W. Frawley, who was instrumental in making it available to the English-speaking public, and F. DiCamilla, J. Lantolf, I. Mackenzie, and D. Verity, who read and commented on the prefinal version of the paper. Finally, I wish to express my gratitude to the University of Delaware, where I was able to spend an unencumbered semester as a Visiting Professor, during which time much of the work was done.

3 Benson et al. (1986: 205–23) offer a survey of ‘twelve basic principles of lexicographic definition’, which, though quite interesting in themselves, are not of a formal nature and therefore cannot be directly compared to what I discuss here (although some commonalities can be seen).

4 Vocabule is a technical neologism; it roughly corresponds to polysemous word.

5 Zholkovsky is the official American spelling of the name of my colleague and friend, now a professor at the University of Southern California; Žolkovskij is the direct transliteration from Cyrillic script and is used in bibliographical references.

6 I am not prepared to give a rigorous definition of the notion ‘non-trivial’ as applied to semantic components; however, I believe that, intuitively, this notion is rather clear. One can add here that, first, in order to be non-trivial, a semantic component must represent a rather important part of the definition (the notion is thus relative to the definition under analysis); second, certain semantic components, very abstract and general, which are specifiable in list form, may always be trivial: ‘and’, ‘to cause that...’; ‘object’, etc.

7 A clarification seems in order here. Would a definition whose definiens contains ‘for boys and girls’ be formally correct? According to feature (2), it would be considered incorrect since the semantic invariant of ‘boys and girls’ has not been rendered explicit: the enumeration ‘boys and girls’ must be replaced by the invariant ‘child’, such that one has ‘for children’.

8 This approach to semantic decomposition thus differs from that of Wierzbicka in the following manner. Wierzbicka begins by looking for semantic primitives, establishing an initial set of them, and justifying them. Only after that does she use them in decompositions. I begin with semantic decomposition of lexemes in terms of other, simpler lexemes, but which are not necessarily primitives; I hope to end up with primitives as the ultimate result of my work, not as the beginning.

9 To distinguish lexical connotations from semantic components (of a definition), the following rule is suggested (Mel’čuk et al. 1984: 36–37). A semantic element ‘S’ is a component in the definition of L if and only if ‘S’ corresponds to an actual constitutive feature of the object in question (i.e., of the object designated by L: for example, the whiteness in SNOW or the hiding of (part of) the sky in CLOUD). Otherwise ‘S’ is a lexical connotation of L. In the latter case, ‘S’ corresponds either to an imaginary trait – ‘slovenly’ for PIG I – or to a real, but non-constitutive, trait – ‘capricious change of direction’ for WIND (notice that the definition of WIND cannot include the component ‘changing direction’ because there can be winds that do not change direction).

10 Note that this is possible but not in the meaning intended.

11 In fact, there are other differences: First of all, we find an important semantic distinction: les neiges tends to denote ‘a space covered with snow and perceived as an element of a landscape in the region X’; hence, les neiges has prepositional government: de + N (les neiges de l’Himalaya), which refers to said region, while la neige has no government. Taking these facts into consideration, we split NEIGE into two lexemes: NEIGE I.a [sg] vs. NEIGE I.b [pl].

12 Green (1969) has proposed the use of conjunction reduction to distinguish the
senses of a given vocable. Conjunction reduction represents a particular (but, perhaps, the most important) case of the phenomenon considered under Criteria 7a and 7b. Green demonstrates for example, that the verb ASK must be divided into a number of distinct lexemes, given, among others, the following incompatibilities:

(i) *I should ask (them) \{ to go and when to go when to go and to go \} ['permission' and 'question'];
(ii) *I should ask them \{ for ten dollars and to dinner to dinner and for ten dollars \} ['request' and 'invitation'].

From (i) and (ii), it follows that we have to distinguish at least four lexemes of ASK:

- ASK [N to V\textsubscript{inf}] 'to ask permission to'
- ASK [N+Wh-word] 'to ask a question'
- ASK [N for N] 'to request'
- ASK [N to N/V\textsubscript{inf}] 'to invite'

References

Mel'čuk, I. A. and A. K. Žolkovskij. 1984. Tolkovo-kombinatornij slovar’ sovremen-


