

Lexicography for the Twenty-First Century

Ladislav Zgusta

*University of Illinois
Urbana, Ill.*

For a conference meeting under this title it will not be out of place actually to try to throw a glance towards the future. It undoubtedly would be hubris to attempt vaticinations concerning the whole century to which we are so close; but an estimate of the possibilities and problems that may be expected to be with us in its opening years is perhaps not too audacious or ill-advised an undertaking.

In a general way, one can predict that the importance of lexicography and the volume of lexicographic output will keep growing. Among the reasons for this assumption are the following:

- contacts between speakers of different languages will continue to increase;
- pedagogical needs (in the broadest sense of the word) within communities will continue to increase as well, and
- so will the need of handling constantly greater amounts of information; and, last but not least,
- technical and financial improvements will render it increasingly possible to satiate those needs.

The last clause may need some explanation. The financial situation of any community always has its ups and downs; and we all know that constant complaints are one of the safest means (although not the only one) to effectively get an increased allotment of money from any institution, particularly a public one. These circumstances produce the impression of an intolerable scarcity of funds, particularly in a time of recession, such as at the moment of this writing (winter 1992/3); however, in the long run it is clear that even with the major catastrophes of the two world wars, the extent of cultural activities financed by various governments and private foundations has increased during our century, just as the activities of publishers have; and that there is no reason not to expect this development to continue in future.

One can easily expect that contacts among various social groups such as nations will keep increasing along with the technical facilities which render such contacts constantly easier; that education will continue becoming more extensive and perhaps, let us hope, also more intensive, at least in some areas; and that the handling of information will be one of the most important tasks in running government, business, industry, research, or any other similar activity. Lexicography is, and will increasingly be, involved in all these developments. These favorable expectations are accompanied by propitious developments in the two areas with which lexicography is most intimately connected, namely in linguistics and in computer science.

The development in linguistics which is most favorable and useful to lexicography is that the lexicon has moved from the periphery closer to the focus of linguists' attention. It is true that in some branches of linguistics, such as, for instance, historical linguistics, the interest in the lexicon and its individual items has always been alive, even if only for purposes of etymology and, perhaps, historical semantics. It is also true that European structuralism, such as repre-

4 PAPER OF GUEST LECTURER

sented, for instance, by Lucien Tesnière, developed in the theory of valences one of the most valuable sources of knowledge for lexicography; there probably is no French or German dictionary that would not have, directly or indirectly, gained in its indication of syntactic patterns from the fruits of this research in valences. (See, e.g., Herbst 1984, 1985, 1986.) The most detailed lexicographic analysis of valences is that embodied in the method of Igor Mel'chuk (e.g., Mel'chuk 1981, 1984, 1988).

The Dutch bilingual Dal dictionaries have a syntactic notation of their own, but they belong to this trend of development, which has been very strong in Europe in the last twenty or thirty years (Marello 1989). In America, on the other hand, the development most fruitful for lexicography arrived, after some modest beginnings that started with Chomsky's *Aspects* (1965), only when studies of what is summarily called subcategorization started flourishing. Subcategories, after all, are the first cousins of valences (if once removed and hence younger), because they combine information on the collocational possibilities of a word with its detailed syntactic pattern, and that for every sense of the word in cases of multiple meaning (Grimshaw 1992). The notion, if not the term, is of old standing, and, but for the circumstance that subcategorization, and valences are embedded in different theories, the two notions are not much different. However, considering the manpower of American linguistics, the attention paid to subcategorization cannot but bring many results useful to the lexicographic description. Such a development cannot fail to bring much useful knowledge into the area so successfully opened by Hornby's syntactic patterns (*Advanced learner's* 1948), also as modified and enlarged in scope by Procter (Longman 1978).

The arrival of the learners' dictionary brought about what one could call the user-friendly definition. We all know that as of now, the COBUILD dictionary embodies the greatest effort at definitional user-friendliness. I have discussed elsewhere (Zgusta 1988) the individual features of this user-friendliness. The one feature of the COBUILD style that most obviously leaps to the eye is the indexicality of the text which, among other things, involves the reader as if in direct discourse. When we write in Latin, this style is quite normal; one says, for instance, "si tabulam acutissimis oculis tuis inspicias, doctissime lector, non te fugiet...", that is, "if you look at the map with your very sharp eyes, most learned reader, you will not fail to see...". In English, however, one would say, in a highly impersonal way, something like, "a detailed examination of the map will show...". The "you" of the Latin text, the direct addressing of the reader, suggests something like a direct contact with him. COBUILD does exactly the same thing in the definitions offered (only in English, it is somewhat unusual). For instance: to the entryword **scold** the morphological forms **scolds**, **scolding**, **scolded** and the following definition are indicated: "If you **scold** someone, you speak angrily to them because they have done something wrong."

Naturally, there are words (including verbs) which to bring into direct involvement with the reader may be delicate. In these cases, COBUILD avoids this by returning to the non-indexical, indirect, impersonal type of definition: the division line runs through verbs like **arouse** (**arouses**, **arousing**, **aroused**) which still have the personal type of definition: #4: "If you **are aroused** by something, it makes you sexually excited." This definition still ascribes the action or process to "you", i.e., the innocent reader; but the definitions of real hard-core four-letter verbs and words do not; they are impersonal in the old style.

Another feature of the definitional user-friendliness of COBUILD is a non-pedantic, or even less than precise, dovetailing of some definitions. For instance, **building** is defined as "a structure with a roof and walls"; so far so good, but then **pyramid** is defined as "an ancient stone building...", which brings in the difficulty that a pyramid does not have a roof. Generally

speaking, the editor or definer of COBUILD relies on the user's horse-sense and on his ability to use his own knowledge of the extralinguistic world: the editor supposes that the user will know that while most buildings have roofs, some of them do not, and that the pyramid is one of the latter ones. In more technical wording, he relies on the user's ability to understand the definition of building as the definition of the prototypical building, and on his ability to know that a **pyramid** is not a prototypical building.

If the trend started by the COBUILD dictionary continues, it will have as a consequence a dissociation of lexicography "for human consumption" from lexicography "for computer application". We know that scholars like Amsler (1984 a, b) have used definitions offered in dictionaries for building hierarchical structures of hyperonyms, hyponyms, and (hopefully) cohyponyms. The user-friendly type of definitions offered in COBUILD would exclude any attempt of this kind in future. Another step towards the dissociation of "lexicography for humans" and "lexicography for computers" can be seen in the way COBUILD treats the homonymy of members of different English word classes.

We know that the lexicographer can a) represent the meaning of the entryword as a continuum, that is, he can try to represent the polysemy as a coherent range of meaning; or b) the entry can be compartmentalized, with a strong indication and polarization of the single senses; or c) a middle way between these two extremes can be taken. This is not the place to discuss the relative advantages of these approaches; for our purpose, it is enough to say that COBUILD treats polysemous meaning in as coherent a way as possible: the single senses are polarized, certainly, but they are presented as a coherent whole. Let us take a few examples of some entries (highly reduced and simplified here).

To the entryword **murmur**, the following morphological forms are indicated: **murmurs, murmuring, murmured**. There is the following sequence of definitions: 1 If you **murmur** or **murmur** something, you say it very quietly, so that not many people can hear it. 2 A **murmur** is 2.1 a statement...which can hardly be heard; 2.2 a continuous...sound; 2.3 a quiet complaint; 2.4 an abnormal sound which is made by the sound... 3 If you **murmur**, you complain gently about something. The definitions (and the examples, omitted here) show several things: First, that the form **murmurs** has a double value, standing both for the 3rd pers. sing. pres. of the verb and for the plural of the noun. Second, that the presentation of meaning disregards the morphological differences and tries to present the polysemy in the most coherent way; in this case, first the general meaning, then the more technical one.

This is a short entry with coherent senses. However, the dictionary opts for the same procedure also in cases where no real coherence can be found in the senses themselves. For example: **cow**, forms **cows, cowing, cowed**. Definitions (strongly reduced): 1 A **cow** is... (there follows a list of different animal species for which **cow** is used for the female). 2 If you describe a woman as a **cow**, you mean that she is very unpleasant (an evaluative gloss follows). 3 If someone is **cowed**, they are made afraid... 4 If you say that you could do something **till the cows come home**, you mean that you could keep doing it... Clearly sense #3 has nothing to do with the rest of the entry. (I do not discuss the different etymologies, because COBUILD is a strictly synchronic dictionary.) While I find this confusing, this still is a short entry, easy to survey. However, the entry of **fit** stretches over nearly two columns, having the following arrangement: **fit**, forms **fits, fitting, fitted; fitter, fittest**; the sequence of senses is: #1 - 2 verb, something **fits** in size or shape. #3 If something is a good **fit**... (noun). #4 If you **are fitted** for a particular piece of clothing... #5 - 11 various senses of the verb. #12 Someone or something that is **fit**... (adjective). #13 If you **fit** someone for a particular role... (verb). #14 Someone who **sees fit** or

6 PAPER OF GUEST LECTURER

thinks fit... (phrasal verb). #15 If you say that someone is **fit** to do something... (adjective). #16 Someone who is **fit** is healthy... (adjective). #17 If someone has a **fit**, they suddenly lose consciousness... (noun). #18 - 20 other senses with **fit** 'spasm'. Again, completely disparate senses are presented as a continuum (the suture runs between #16 and 17). As already stated, the etymological consideration (**fit** #1 - 16 and **fit** #17 - 20 are completely different as to their origin, just as **cow** #1, 2, 4 and **cow** #3) is or can be considered irrelevant in the case of a completely synchronic dictionary. The real point is that when one checks users' reaction to this arrangement, as I did with students, they find the arrangement confusing. The excellent and outstanding way in which the grammatical subcategorization of each of these senses is presented seems not to redeem the awkward microstructural arrangement.

To my mind, it is these grammatical indications of syntactic patterns and the detailed indications of stylistic levels that are eminently good in COBUILD. Whether its definitional and presentational style will find followers outside the area of pedagogical dictionaries for children and young students remains to be seen; I should not be surprised if the old adage 'qui bene distinguit, bene docet' ("He who makes good distinctions teaches well") prevails. If, however, the COBUILD style does find many followers, it will start, as it has already been said, a trend that will dissociate "lexicography for human consumption" from "lexicography for the computer". (Mind that I am not talking here about the computer as the lexicographer's helper in the compilation of dictionaries, but about lexicography as a helper in the handling of information by the computer.)

Another important development that will prove useful to lexicography is prototype theory. The notion of the prototype was introduced into psycholinguistics and cognitive science by Rosch (1977, 1983). For speakers of, at least, the European and many African and Asian languages, the "normal" typical building does indeed have walls and a roof; examples of buildings such as hangars of corrugated iron in the form of a semicircular half of a cylinder are not what occurs to the speaker when a building is mentioned, because they are not prototypical but are somehow on the periphery of the notion. The COBUILD definition of "building" quoted above is based on this prototypical understanding.

It goes without saying that while the term "prototype" and its exact positioning within the framework of a theory are new, the notion itself has not been unknown to lexicographers: any dictionary will give as examples of a bird: a sparrow, a hawk, a colibri, but hardly any dictionary would indicate a turkey; it does not fly, so it is not prototypical. (Naturally, the whole notion is culture-bound: I know from my students from South Africa that for them, the ostrich also is a prototypical bird; a prototype, strongly bound to a culture, is usually called a "stereotype", a terminological usage started by Hilary Putnam.) However this may be, it is of great advantage to lexicography that the study of prototypicality attract attention, because it can be highly useful.

As an example, we can indicate the work done by McKeown (1991) in which she discusses, with good examples, possible improvements of lexicographic definitions. For instance, it would not be good to define the word **meticulous** as "extremely or excessively careful about small details", because the word *careful* prototypically associates with cautiousness about danger; hence, it would be better to define **meticulous** as "extremely or excessively neat and careful about small details", because the word *neat* blocks the association with danger. Another example: it is not so good to define **shrewd** as "having a sharp mind; showing a keen wit; clever; keen; sharp"; a better definition is "able to size up a situation quickly and use it to get what you want", because this formulation explicates the prototypical advantage-oriented component of the meaning of **shrewd**. It is not wrong to define **typical** as "combining or showing the special

characteristics of a group or kind”, but a definition like “describes something that is a good example of a person or thing because it shows what that person or thing is usually like” is said to be better, because it makes the meaning accessible. (User-friendliness again!)

The purpose of this article is not to go into the details of the definitions criticized and proposed, or into the realities of the lexicographic treatment. One can say in general that good dictionaries largely avoid the pitfalls demarcated. True, Webster’s Second has in its definition of *meticulous* the wording “unduly or excessively careful...”, but Gove in Webster’s Third avoided the pitfall by defining “marked by extreme painstaking care...” and Mish in Webster’s Ninth follows suit, defining “marked by extreme or excessive care...” Morris in American Heritage also avoids misunderstanding by defining “1. Extremely careful and precise. 2. excessively concerned with details; overscrupulous.” “Precise” in definition 1 eliminates, it would seem, the possibility of wrong association; the entry is even followed by a paragraph in which synonyms are discriminated and where such a possibility is made even more impossible. Soukhanov in American Heritage Three gives practically the same information. COBUILD has for *meticulous* “A *meticulous* person does things very carefully and with great attention to detail”, where “attention to detail” seems to take care of the danger. No need to go into the details of all the examples; the dictionaries stand the test well. For instance, the user-friendliness implied in McKeown’s suggested definition of *typical* is echoed or rather preceded by the, as usual, user-friendly COBUILD where the motif of a “good example” already occurs: “Something that is typical shows the most usual characteristics of a particular type of person or thing, and is therefore a good example of that type.”

Within prototype theory, nominal phrases are taken as sums or intersections of either the extensional or the intensional meaning of the nouns and attributes combined. For instance: if (in the terms of John Stuart Mill’s logic) the extension of the notion of *British* is the whole class of all the people and things that can be called British, and the extension of the notion of *sportsman* the whole class of people who are sportsmen, then the meaning of *British sportsmen* is the intersection of those two classes: people who are both British and sportsmen. If the intensional meaning of *British* is the bundle of criterial features that determine the Britishness of a person or thing, and the intensional meaning of *sportsman* the corresponding bundle of features that determine whether a person is a sportsman, the intensional meaning of the nominal phrase *British sportsmen* is the sum of the two bundles of criterial features. Within this approach, it is frequently felt that phrases such as *a hand wound* and *a gun wound* cause problems to or are even beyond the scope of prototype theory, because neither their meaning as the sum of their intensionality nor as a sum of their extensionality can capture the difference between a hand wound being a wound in the hand whereas a gun wound is a wound caused by a gun (Hampton 1991). This apprehension is not necessarily correct, at least not in my judgement, if we apply the notion of prototypicality in a flexible way: if we understand the ability of inflicting wounds as part of the prototype of *gun*, then certainly this understanding of prototypicality can play a role in determining the meaning of *gun wound*. That is, associations of a notion, such as in this case its purpose, do belong to the prototype.

Another problem for prototype theory is sometimes seen in a similar difficulty consisting in the difference between nominal phrases such as *criminal act*, which denotes an act that is criminal, and phrases such as *criminal lawyer*, denoting a person who (in the usual understanding) is not (necessarily) a criminal (Shoben 1991). Again, one can observe that language has a mechanism called lexicalization, which stabilizes one meaning of a lexical unit at the expense of several possible ones, but not necessarily to the exclusion of the other possible meanings: a criminal lawyer can specialize in criminal processes, and be a criminal at the same

8 PAPER OF GUEST LECTURER

time. Although a phrase like a *criminal criminal lawyer* is unusual, a slight re-wording shows its possibility: a veritably, truly criminal *criminal lawyer*. Also, a sentence like “He truly is a criminal lawyer” is ambiguous, by any standard, depending on the situational context. Nominal phrases such as *flu virus*, *heat rash*, *picture book*, *lemon peel* are so normal that one does not realize how different the logical relations of their members are: virus that causes flu, rash that is caused by heat, a book that contains pictures, peel from or of a lemon; obviously, language tolerates homonymy not only in lexical meaning but also in the structure of patterns.

Among today’s lexicographers, it is Dirk Geeraerts (1983, 1984, 1985, 1987, 1989) who most strongly recommends prototypicality and related notions as a useful tool for lexicography. He very reasonably discerns several types of lexicography and their relative foci of interest. According to him, encyclopedic and terminological dictionaries are focussed on the extensional meaning; general monolingual dictionaries on the stereotype; and scientifically linguistic, above all historical dictionaries on the prototype. The general correctness of this distribution and of the whole conception leaps to the eye, of course. At least two things, however, should not be forgotten. First, it has already been stated above that the traditional lexicographic treatment of meaning, particularly as presented in modern dictionaries, is not mutually exclusive with the theory of prototypicality: Swanepoel (1991) has shown that quite persuasively in a detailed study in which he meticulously contrasted the advice, or approach, preconized in my ‘Manual’ (Zgusta 1971) with the components of the theory of prototypicality. Second, not only the types of dictionaries but also the individual notions of prototype theory have fuzzy boundaries. (For the critique of the notion itself, see Wierzbicka 1985 p. 340 ff. on “The many senses of the notion ‘prototype’”; see also p. 59 ff., 158 ff. and passim; 1990.) Certainly, the encyclopedia and the terminological dictionary will be interested not so much in the narrow prototype of a notion but rather in its full extension; both of these types of dictionaries must try to delimit the denotative class. However, they do not do it so much by taxative enumeration of the members of the class but rather by the attempt to establish the notion’s intension, by finding the decisive criterial features. Historical dictionaries certainly are mostly interested in the prototypes prevailing in the respective epochs; for instance, the Roman prototype of *virtus* can be seen in the collocation *bellica virtus* “the v. of war”, whereas the Christian prototype of the same word in the collocation *virtutes theologicae* “the theological v.” However, since every word, hence also its prototype, is embedded in culture, every prototype can be considered a stereotype at the same time.

Also, one should consider that one of the main trends of semantic development of standard languages of our day is the increasingly strong terminologization of the lexicon: more and more lexical units are endowed with a terminological, defined meaning. Sometimes, the same expression can have a defined meaning in a technical language, and a general, perhaps prototypical, meaning in general language. For example, the expressions *crime*, *misdemeanor*, and *felony* can be used in a general, approximative, perhaps prototypical way; but they also have an exactly defined meaning in the language of criminology; in the same way, if it is said of someone: “His condition is serious”, the meaning of *serious* will differ depending on whether the speaker is a layman or a physician. And it is the terminological use that encroaches on the general one, not the other way round: for instance, when we have to deal with a possibly underinflated tire, God forbid that we should go by some notions of prototypicality; that may serve as a first warning, but the real decision must be based on the definition of underinflatedness as established in relation to that specific tire.

These can be deemed minor considerations, however, if our main interest is not in the construction or critique of the prototype theory but rather in the assessment of how it can be useful to lexicography. If the latter is the case, the capital thing is that if one considers these

approaches in a broad-minded way, without unnecessary carping at details, there is no doubt that the freshness of this approach and its interest for psychological and cognitive research cannot fail to draw the welcome attention of a new phalanx of researchers to topics that perhaps have not been ignored, but that can be usefully studied from a new perspective. Hence, prototype theory promises to be of great usefulness to lexicography.

Australia is developing into a strong center of lexicographic activity, as testified, among other things, by the recent foundation of *Australlex*, the Australian lexicographic society. This is only natural, given the plethora of aboriginal languages in Australia and adjacent areas. Apart from this, important theoretical developments are taking place there as well. For instance, Laughren and Nash (1983) developed a project of a dictionary of Warlpiri, an aboriginal language, in which meaning was described in a quite abstract way, and with indication of in which cases the single actants occur. As examples, one can give the following:

KIPI-RNI (tV): xERG [= an actant in the ergative case] causes yABS [another actant in the absolute case] (small particles, as seeds, stones) to come to be distributed in some desired arrangement by manipulating some entity (zINSTR, e.g., ...piece of bark, basket) in which y is located in such a way that y moves along a circular path beginning and ending in said entity.

PAKA-RNI: xERG causes some adhering substance (typically fat, ochre, charcoal) to come to be thinly distributed over some area of yABS by x coming into contact with said area of surface of y and manipulating said substance in such a way that, at the area of contact, said substance transfers from x to y along a path (zPERL/LOC) coincident with said area on the substance of y.

Wierzbicka (1983 p. 137) objects to this by asking the reader to imagine the reaction of an average dictionary user, such as a school teacher, or a high school student to a definition like this, which undoubtedly one can easily imagine. However, we probably have here another case of different goals pursued by different projects. Laughren and Nash hardly had a high school student in mind when they constructed their model: this is typical M.I.T. stuff, ridiculous if imagined in situational contexts such as those mentioned by Wierzbicka, but moving on a straight trajectory towards a model of definitions that may one day be handled automatically by a computer program.

Wierzbicka (1985) also offers a coherent discussion of her ideas concerning conceptual analysis, in which she proposes a type of definition which is couched in everyday language and based on the understanding of an average speaker of the language. In this way, she clearly steers away from formalized models. However, if the complexity of language is to be captured and described, even the most informal model must be offered highly refined, detailed information. For instance, to get the difference between cups and mugs, it is necessary to compare them as to their purpose, as to the material they are made of, as to their appearance (at top and bottom, on their sides, in their proportions, in respect to their handles), as to their size, as to how they are used, in their relation to saucers, and similarly for other features. Each such analysis would be longer than one printed page, at least, and sometimes much more, and would consist of many descriptive statements. Prototypicality appears to be useful for the analyses, since they are based on the general understanding of the average speaker.

In her later works, Wierzbicka (forthcoming) develops a more formalized model for the description of lexical meaning, consisting in scenarios to be conceived for each meaning and each sense of a polysemous word, in as far as they belong to different concepts; an additional feature of these scenarios is that they are written in a limited language, without using undefined

10 PAPER OF GUEST LECTURER

words (and also imposing some additional limitations). The model will best be explained by two examples. The concept of “punishment” is said to be accurately portrayed by the following scenario:

X punished Y [for Z]

- (a) [Y did Z]
- (b) X thought something like this:
- (c) Y did something bad [Z]
- (d) I want Y to feel something bad because of this
- (e) it will be good if Y feels something bad because of this
- (f) it will be good if I do something to Y because of this
- (g) X did something to Y because of this

For comparison, the concept “revenge” is portrayed by this scenario:

- (a) someone (X) did something bad to someone (Y)
- (b) because of this, Y felt something bad
- (c) after this, Y thought something like this:
- (d) this person (X) did something bad to me
- (e) I want to do the same to this person (X)
- (f) because of this, Y did something bad to X

There is no need to consider every detail of these scenarios (such as, in “revenge” (e), whether the person Y necessarily wishes to do the same thing to X, or only something bad); the important thing is the whole conception of such scenarios. When we consider them from this point of view, we can say that to print them in a dictionary would not be too dissimilar to what Laughren and Nash have to offer. The wording of the scenarios is simpler both in the vocabulary and in the shortness of the less convoluted sentences, but the detailed analysis into the individual elements of meaning and the explicitness of formulation renders the impression of wordiness by the attempt at accuracy. I discussed elsewhere (Zgusta, forthcoming) the consequences of the human mind working in short-cuts, that is, in enthymemes; when talking, we take so much knowledge of the world and of the language for granted that to make all that knowledge with all its implications, presuppositions, etc., explicit causes the resulting description to seem unnatural, because we are not accustomed to such explicitness. However, the scientific requirement of the day is to be as detailed, explicit, and accurate as possible. I do not think that the requirement is produced exclusively by the need to have explicit algorithms for automatic handling of data by an apparatus that knows nothing and can do nothing but handle data by operational rules; it would seem that the logic of the research in itself pushes the researcher in the direction of constantly increasing detailedness, because much is already known and the area and finesse of knowledge keeps increasing. In any case, it is interesting to see that the introduction of everyday language with limited vocabulary into the description of meaning does not necessarily entail its simplification.

Concerning the developments in computer science that one can reasonably expect in the near future, I do not wish to repeat what I have already said on other occasions (Zgusta 1989, 1991). Nor would it be reasonable to discuss the computer programs either specifically developed for the lexicographer’s use, or those developed for other purposes but useful to the lexicographer; those programs proliferate so quickly that before this will be printed, new programs can be confidently expected. It may suffice to say that, on the one hand, the computer will not, at any foreseeable time, replace human participation in editorial activities such as distributing the senses of polysemous words, formulating definitions, thoughtfully condensing quoted examples, handling metaphorical and other similar expressions, etc. On the other hand,

the computer's present usefulness in many other respects will grow exponentially, and new areas of application of computer programs will be found.

Within the last decade or two, we have observed a huge increase in the number of different types of dictionaries (the emergence of the learner's dictionary being probably the most outstanding case), differentiated by the supposed readership, its level of education, and its purposes in buying the dictionary. This proliferation was largely made possible by the computer, which renders it possible to select subsets of lexical material contained in the database (provided that the single entries in the database are tagged for the respective marker that constitutes the desired subset). Given a database, not necessarily a huge one, but one that corresponds to, or was created in the process of preparing a large dictionary, smaller, specialized dictionaries can be created out of it by the extraction of the respective subsets of data. A further development of this specialization of the dictionaries will require not only a very detailed classification of data in the database being planned and tagged with all imaginable markers, but also a deeper inquiry into the needs and purposes of the various sets of readership. Another field of inquiry, at least as important as the one just mentioned, is the research into the effectiveness of different presentational styles in offering information. It can already be observed that there are users, particularly students, who prefer a more formal presentation of information to a style that relies more on exemplification; e.g., they prefer a rather abstract presentation of syntactic patterns. On the other hand, there are users, also among students, who prefer a more discursive presentation of information. One can imagine that a variety of preference will lead to a modularly constructed on-line or high-capacity personal dictionary, in which the user will be able to select the presentational style. (Let us mention that a huge step in the direction towards such modularity in presentation was taken by the COBUILD dictionary, in spite of its being a printed book, not a computerized dictionary: the user can either gather the grammatical information from the informal, user-friendly wording of the definitions, or he can study the formulaic descriptions of the respective patterns of each sense of a polysemous word printed in the margin of the columns; or he can, in some cases, study the more elaborate grammatical advice offered in separate boxes of information that are inserted in the proper places of the text.) Once the principle of modularity is introduced, there is no reason why it should not pertain to other areas as well: the user should have the possibility to request specifically grammatical, or stylistic, or historical, or encyclopedic, or any other similar information in great detail by selecting the proper module. The solution of all the problems connected with such tasks is largely within the purview of the software and perhaps even hardware specialists. For the lexicographer proper, the main task is to anticipate the many future needs for specialized information and to introduce the necessary classification of the lexical material at the beginning of work on the database.

A particularly difficult task will be the setting up of a database that will be more useful for the preparation of bilingual dictionaries than is usually the case now. In addition to all the classifications mentioned above which are necessary for, for instance, the comparison of the stylistic level of the source language expression and the possible target language equivalent, there is always the primary task of comparing the denotational equivalence or partial equivalence. If the computer is to be useful in that task, it will be necessary to feed into it information about the minutest components of the respective lexical unit's meaning. The value of the completely explicit analyses and descriptions of the lexical meaning, some of which were mentioned above, will then become even more apparent than it is now. Ultimately, one can imagine that lexical units will not be compared between two languages, but that the analysis of a huge number of lexical units in many languages will yield a notional grid which will be used for comparisons of any pair of languages. That this may be quite a distant task if intended to cover the whole lexicon is quite clear; however, if we contemplate this kind of analysis for only

12 PAPER OF GUEST LECTURER

some segments, or subsets, of the total lexicon, namely those belonging to the technical and similar registers, the task is not so overwhelming.

Lexicography can be expected to provide help in all the adjacent fields of human endeavor where information processing plays a crucial role, whether it is information classification and storage or any other task. One of the ultimate tasks is what is traditionally called machine translation. In this area, what has been said above about metaphoricality is equally valid: within the foreseeable future, there is no expectation that there could be a computer program for dealing with any sort of creative language, be it the coinage of (morphologically or semantically) irregular neologisms, metaphoricality, allegory, poetic language, dark hints, complicated conversational implicatures, uncoded slang, puns, etc. On the other hand, there is no reason why one should not expect machine translation, or at least automatically aided translation, of texts couched in a limited set of syntactic patterns and consisting of defined or, at least, well described words, and with equivalence between the languages in question established, to be operative quite soon. A typical case in question is the bureaucratic regulatory language, including normalization of industrial and business procedures; and the most pressing contemporary case is, of course, the European community with its many official, equal-status languages. Indeed, relatively rapid progress in this area can be safely expected; and machine translation of communications in the area of the sciences may follow in the not too distant future. All that has been said above about subcategorization is of immediate usefulness for this task, as a means for disambiguating polysemy, for establishing lexical equivalence, and for recognizing the syntactic pattern in the source language and selecting the proper syntactic pattern in the target language. A more ambitious task is, of course, not to create a program for bilateral machine translation, but to create a system similar to the grid for the analysis and comparison of lexical meaning mentioned above; that is, a system of translation from one natural language into a formalized notional language of a more universal character, and hence into any other language comprised in the system. That may be a more remote goal, but one which will be, as I believe, reached in the not too distant future; I would think in the first quarter of the next century, at least for the limited tasks described above. In any case, in all these activities, lexicography will play a crucial role.

REFERENCES

- Advanced Learner's. 1948. *Oxford Advanced Learner's Dictionary*. Ed. A.S. Hornby. Oxford: Oxford University Press. (Second edition 1963; third edition, ed. A.P. Cowie 1974; fourth edition, ed. A.P. Cowie 1989.)
- American Heritage. 1969. *American Heritage Dictionary of the English Language*. Ed. William Morris. Boston: American Heritage Publishing Co. & Houghton Mifflin.
- American Heritage Three. 1992. *The American Heritage Dictionary of the English Language*. Executive editor Anne H. Soukhanov, third edition. Boston, New York, London: Houghton Mifflin.
- Amsler, Robert A. 1984. "Lexical knowledge bases." *Proceeding of the 10th International Conference on Computational Linguistics*, Stanford, California. Morristown, N.J.: COLING.
- . 1984b. "Machine-readable dictionaries." *Annual Review of Information Science and Technology* 19: 161-209.
- Austin, Peter, ed. 1983. "Australian Aboriginal Lexicography." *Australian Linguistics* 15; *Pacific Linguistics Series A* 66. Canberra: The Australian National University.
- Chomsky, Noam. 1965. *Aspects of a Theory of Grammar*. Cambridge: MIT Press.
- COBUILD. 1987. *Collins COBUILD English Language Dictionary*, editor-in-chief John Sinclair, managing editor Patrick Hanks. London and Glasgow: Collins.
- Geeraerts, Dirk. 1982. "Stereotypes en prototypes." *Forum der letteren* 23: 248-258.
- . 1983. "Type en prototype." *Tijdschrift voor tekst- en taalwetenschap* 4: 69-86.
- . 1984. "Dictionary classification and the foundations of lexicography." *I.T.L. Review* 63: 37-63.
- . 1985. "Les donnés stéréotypiques, prototypiques et encyclopédiques dans le dictionnaire." *Cahiers de Lexicologie* 46: 27-43.
- . 1987. "Types of semantic information in dictionaries." Edited by Robert Ilson. *A spectrum of Lexicography. Papers from AILA Brussels 1984*. Amsterdam, Philadelphia: Benjamins: 1-10.
- . 1989. "Principles of monolingual lexicography." Edited by F.J. Hausmann, O. Reichmann, H.E. Wiegand, L. Zgusta. *Wörterbücher - Dictionaries - Dictionnaires. An International Encyclopedia of Lexicography*. Berlin, New York: W. de Gruyter. vol. I: 287-296.
- Grimshaw, Jane. 1992. "Subcategorization and selection." Ed. William Bright. *International Encyclopedia of Linguistics*. Oxford. Oxford Univ. Press. Vol. 4: 90-92 (bibliography).
- Hampton, James. 1991. "The combination of prototype concepts." *Schwanenflugel*. 91-116.
- Herbst, Thomas. 1984. "Bemerkungen zu den Patternsystemen des Advanced Learner's Dictionary und des Dictionary of Contemporary English." Edited by Dierter Goetz and

14 PAPER OF GUEST LECTURER

Thomas Herbst. *Theoretische und praktische Probleme der Lexikographie*. München, Max Hüber. 139-165.

———. 1985. "Das zweisprachige Wörterbuch als Schreibwörterbuch: Informationen zur Syntax in zweisprachigen Wörterbüchern Englisch-Deutsch/Deutsch-Englisch." Edited by Henning Bergenholtz and Joachim Mugdan. *Lexikographie und Grammatik: Akten des Essener Kolloquiums zur Grammatik im Wörterbuch*. (Lexicographica, Series Maior 3.) Tübingen, Max Niemeyer. 308-331.

———. 1986. "A proposal for a valency dictionary of English." Edited by Robert F. Ilson and Michel H.A. Blanc. *A Spectrum of Lexicography: Papers on Lexicography and Lexicology, Papers of the Seventh Congress of Association Internationale de Linguistique Appliquée* (Brussels 1984). Amsterdam: John Benjamins.

Laughren, Mary and David Nash. 1983. "Warlpiri dictionary project: aims, method, organization and problems of definition." *Austin*. 109-133.

Longman. 1978. *Longman Dictionary of Contemporary English*. Editor-in-chief Paul Procter. Harlow: Longman.

———. 1987. *Longman Dictionary of Contemporary English*. Editorial director Della Summers. Second edition. Harlow: Longman.

Marello, Carla. 1989. *Dizionari bilingui*. Bologna: Zanichelli.

McKcown, Margaret G. 1991. "Learning word meanings from definitions: problems and potential." *Schwanenflugel*. 137-156.

Mel'chuk, Igor A., et al. 1981. "Un nouveau type de dictionnaire: Le Dictionnaire explicatif et combinatoire du français contemporain (Six articles de dictionnaire)." *Cahiers de Lexicologie* 38: 1, 3-34.

———, et al. 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*. Montréal.

Mel'chuk, Igor A. 1988. "Semantic description of lexical units in an explanatory combinatorial dictionary: basic principles and heuristic criteria." *International Journal of Lexicography* 1: 165-188.

Mel'chuk, Igor A. and Ian Mackenzie. 1988. "Crossroads of obstetrics and lexicography: a case study. (The lexicographic definition of the English adjective *pregnant*.)" *International Journal of Lexicography* 1: 71-83.

Mel'chuk, Igor A. and Tilmann Reuther. 1984. "Bemerkungen zur lexikographischen Beschreibung von Phraseologismen und zum Problem unikalener Lexeme (an Beispielen aus dem Deutschen)." *Wiener Linguistische Gazette* 33/34: 19-34.

Mel'chuk, Igor A. and Aleksandr K. Zholkovskij. 1984. *Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka: opyty semantiko-sintaksicheskogo opisanija russkoj leksiki*. Wien: Institut für slawistik der Universität Wien.

Rosch, Eleanor. 1977. "Human categorization." Ed. by Neil Warren. *Studies in Cross-Cultural Psychology*. London: Academic Press. Vol. 1: 1-72.

- . 1983. "Prototype classification and logical classification: the two systems." Ed. Ellin K. Scholnik. *New Trends in Conceptual Representation: Challenges to Piaget's Theory?* Hillsdale, N.J.: Lawrence Erlbaum. 73-86.
- Schwanenflugel, Paula J., ed. 1991. *The Psychology of Word Meanings*. Hillsdale, N.J.: Lawrence Erlbaum.
- Shoben, Edward J. 1991. "Predicating and nonpredicting combinations." *Schwanenflugel*. 117-135.
- Swanepoel, P. H. 1991. "Polisemie in die woordeboek —'n kognitiewe perspektief." Ed. P. Harteveld. *Lexikos* (Afrilex-Reeks 1:1991). Stellenbosch: Buro van die WAT. 221-280.
- Webster's Second. *Webster's New International Dictionary of the English Language*. Second edition, ed. William Allan Neilson. Springfield, Mass.: G. & C. Merriam. Various printings between 1950 and 1960.
- Webster's Third. *Webster's Third New Interantional Dictionary of the English Language*. Ed. Philip Babcock Gove. Springfield, Mass.: Merriam-Webster. Various printings after 1961.
- Webster's Ninth. 1983. *Webster's Ninth New Collegiate Dictionary*. Editor-in-chief Frederick C. Mish. Springfield, Mass.: Merriam-Webster.
- Wierzbicka, Anna. 1983. "Semantics and lexicography: some comments on the Warlpiri dictionary project." *Austin* 135-144.
- . 1985. *Lexicography and Conceptual Analysis*. Ann Arbor, Mich.: Karoma.
- . 1988. "The semantics and lexicography of 'natural kinds'." Edited by Karl Hyldgard-Jensen and Arne Zettersten. *Symposium on Lexicography III*. (Lexicographica, Series Maior 19.) Tübingen: Max Niemeyer. 155-182.
- . 1990. "'Prototypes save': On the uses and abuses of the notion of 'prototype' in linguistics and related fields." Edited by Savas L. Tsohatzides. *Meanings and Prototypes: Studies in Linguistic Cattergorization*. London: Routledge & Kegan Paul. 347-367.
- . forthcoming. "What's the use of 'theoretical lexicography'?" Ed. William Frawley. Special issue of *Dictionaries*.
- Zgusta, Ladislav. 1971. *Manual of Lexicography*. The Hague, Paris: Mouton.
- . 1988. "Pragmatics, Lexicography and Dictionaries of English". *World Englishes* 7:3, 243-253.
- . 1989. "Idle Thoughts of an Idle Fellow; or Vaticinations on the Learners' Dictionary." Ed. Makhan L. Tikhoo. *Learner's Dictionaries: State of the Art*. (Anthology Series 23.) Singapore: SEAMEO Regional Language Centre. 1-9.
- . 1991. "Probable Future Developments in Lexicography (asatyakoçyam)." Edited by F.J. Hausman, O. Reichmann, H.E. Wiegand, L. Zagusta. *Wörterbücher - Dictionaries - Dictionnaires. An International Encyclopedia of Lexicography*. Berlin, New York: W. de Gruyter. Vol. III. 3158-3168.
- . forthcoming. 'Lexicography, its theory, and linguistics.' Ed. William Frawley. Special issue of *Dictionaries*.