Abraham Solomonick, Ministry of Education and Culture, Israel

Word - Morpheme balance in dictionary-making

ABSTRACT: Lexicography should be based on the dual dependency of nearly every dictionary entry; word dependence and morpheme dependence. The overt or implied assumption that lexicography deals only with words and their combinations is, therefore, misleading. I will attempt to prove the correctness of this statement for any particular language, since all languages constitute sign-systems, combining the denotational character of the word with its membership in the system, organizing all signs including words, into a comprehensible and coherent whole. One mechanism of such organization, also indispensable for dictionary-making, is morpheme division and subdivision. In practical lexicography every dictionary item is already based on this division.

A head-lexeme (sometimes erroneously called "head-word") has three variations: 1) a word having no run-on continuations in the entry; 2) a pure morpheme; or 3) in most cases – a word/morpheme combination, whereby the word is given an encyclopaedic explanation and the morpheme becomes the foundation for derivatives and their denotational meanings.

1. Word-oriented entries

In terms of English lexicography, a representative English dictionary comprises some entries with head-lexemes based on words or their combinations. Entries like 'and', 'or', 'Alma Mater', 'ad-lib', etc. are self-sufficient either because they have no derivatives in English or because the lexicographer did not find it appropriate to include these derivatives in the entry. The latter decision may be final or held in abeyance; although a particular word may have grammatical extensions, dictionary compiler may consider them so insignificant as to omit them from the dictionary. This is the final variant, admitting no reduction of the lexeme-word into morphemes. On the other hand, morpheme extensions may appear in separete entries of the same dictionary: i.e., 'decor', 'decorate', 'decorous', etc. are presented as different entries, albeit they are clearly derivatives of the same root. Seperate entries are necessary when particular morphemes have subdivisions which are made more comprehensible by using this method. This applies to our example (the head morphemes 'decorate' and 'decorous' have many run-ons attributed to them). Thousands of similar cases abound for different reasons.

The reasons are inconsequential, since even a word serving as a head-lexeme should be considered by a dictionary-maker as a morpheme or morpheme combination, subject to further lexicographic variants either in the same or different dictionary entries.

2. Morpheme-oriented entries

In English dictionaries such entries generally constitute inflection morphemes, since English is primarily an analytical language. Prefixes, suffixes and characteristically English endings can be found in most of to-day's dictionaries as separate entries with proper grammatical explanations.

For languages such as Semitic you will find special entries where head-lexemes comprise root-morphemes. Usually they are verb roots having no denotational meaning, but are the basis for further grammatical derivations, having denotational encyclopaedic explanations. It is important to note that in highly inflectional (mostly in agglutinative) languages it is crucial for the dictionary-maker to select the proper morpheme to serve as the head-lexeme. Thus, in Hebrew lexicography, a long-standing tradition presents verb root-morphemes as the head-lexemes of verbal items. Within these items, the subclass morphemes (like binjanim) take the form of corresponding words together with their meaning. In monolingual dictionaries this explanation is given in Hebrew, whereas translation is the vehicle used in multilingual dictionaries. Nowadays this approach is considered obsolete, because Hebrew learners and even Hebrew speakers find it difficult to extrapolate the root of an unknown verb, encountered in reading or in speech. That is why, nowadays, most bilingual and some monolingual Hebrew dictionaries include binjan-forms as initial head-lexemes. For example, the verb-root   < כ ת ב>   – itself devoid of denotational meaning – heads the entry containing all derivative verbs having to do with "writing":

<div dir="rtl">

< כ ת ב>

</div>

כ ת ב	– ;wrote, was writing (3-rd person singular)
נ כ ת ב	– ;was written
ה ת כ ת ב	– ;exchanged letters
ה כ ת י ב	– ;dictated, obliged
ה כ ת ב	– ;was dictated, was obliged

All these derivatives should be treated as word/morpheme combinations; juxtaposing the properties of words and morphemes. The word properties are numerous: each has denotational meaning/s derived from the extralinguistic world (the main ones are given in the example). Each belongs to specific grammatical categories overtly expressed in dictionaries; each enters different word-combinations, which appear in corresponding entries. On the other hand, it is no accident that the 3-rd person, Singular, Past Tense form is the representative verbal entry – it is a spring-board morpheme for all other verb derivations (tense, number, person et al.). It is not called a "root-morpheme" (since this term was used above in a different context), but rather a "basic-morpheme" for its respective verb paradigm. It is not unusual for a word to correspond in form with its root-morpheme. We shall deal with the problem at length in the next paragraph; for now, this is mentioned in order to emphasize an important point: many dictionary items may be treated equally as words and as root-morphemes. Moreover, all languages contain a continuum of morpheme forms; *the lexicographer's task is to choose the appropriate ones as head-lexemes to represent the entire paradigm.* The above illustration from Hebrew lexico-

graphy shows that, over time, such morpheme forms can be replaced by more appropriate ones.

3. Word/morpheme combinations as head-lexemes.

The Hebrew example above is illustrative. In it the form of a word completely coincides with its root (or basic) morpheme. The same phenomenon often appears in English. The morpheme 'find' functions both as a verb and a noun. For all verb derivations we have 'find' as the head-lexeme. What is the nature of this form? Is it an abbreviated infinitive? The Present Simple? In the latter case, which form of the Present Simple does it represent? It is neither one nor the other, but rather the root-morpheme of a great number of words used arbitrarily to represent all these. Incidently, it coincides in form with one or more of these words; and this fact is profitably and efficiently utilized for lexicographic purposes. Immediately following the bold-typed lexeme is an explanation of the various senses in which it is used. It is interesting to note that the explanations are intentionally given in outline form so as to allow them to apply to any grammatical variation of the paradigm. All the usual offshoots of the paradigm are presumed to be known to the user, the exeptions are given in the entry (in brackets or otherwise).

The same treatment is given to nominalisations (nouns, adjectives, etc.) only in this case the derivative suffixes are usually indicated. We refer to both the flections within the same part of speech and those which function to change one part of speech into another. The eagerness to extract all lexicographic advantage from word/morpheme combinations has led to such ingenious designs as run-on entries in different dictionaries. In most, the root-morpheme is not repeated and even the new meaning is omitted. The users are supposed to derive it from the stem + inflection. Moreover, often the root word (here, "root word") does not coincide in form with the new morpheme element within the created combination. In that case, too, the supposition is that the user will be able to perform the necessary subtraction and addition (in fairness, it is worth mentioning that most dictionary-makers include prefaces containing a brief presentation of the rules for proposed transformations). Examples abound and require no further elucidation.

The same state of affairs (even more striking for our analysis), exists in Russian lexicography. All verbs in Russian monolingual dictionaries are given in the infinitive, which is rightly considered to be the best form to represent the entire verb paradigm. The drawback in this approach lies in the fact, that two infinitives exist for each verb: one for perfective action, the second for imperfective action. Both infinitives are given and are cross-referenced; both are supplied with run-on paradigm extensions. These run-ons are different in each case: since the perfective infinitive has no Present Tense, its run-ons are given for Future Tense (the run-ons for the imperfective infinitive are given in Present Tense). To get to real words the user must subtract two or more letters from the initial form, add paradigm endings to the remainder, and sometimes change the core of the root morpheme.

Thus in *sdvigat'* – to move, to replace – the entire prospective paradigm is indicated by *-aju, -aet*, presupposing the shift of the stress in the derivatives and a change of two letters.

In *sdvinut* – a perfective infinitive of the same verb – the inflections belong to the Future Tense conjugation: *-nu, -net.*

The reason why this is done is obvious: to save space. The more voluminous the dictionary, the more entries it comprises and the more necessary is the need for economy of space. Imagine Webster's dictionary with a full inventory of items, including all run-on entries.

4. The Morpheme/Word Continuum

In light of the aforesaid, I suggest the Morpheme/Word Continuum, which stretches from the dictionary word items to the morpheme items via intermediate cases. The word items will be those (some examples have been given earlier) which, having no extensions in the language, are not subject to morpheme division. Examples are 'but', 'and', proper names and some others. One should bear in mind, that each of them can be expanded into a morpheme combination: 'but' 'buts', the adjectival form of any geographic name, etc. Nevertheless, in most cases these are treated as unique words having no derivatives.

The extreme morpheme cases are flection morphemes; they are treated separately in dictionaries. The rest are intermediates occupying a given place on the continuum nearer to one or the other extreme. A lexicographer usually has some options among the word/morpheme combinations selected to represent the whole group of related words. The criteria for choosing this or that alternative are constantly changing; the lexicographer must be aware of options and keep abreast of innovative views regarding their advantages and disadvantages. His choice determines all the run-ons in the dictionary.

Sometimes the criteria for the morpheme/word combinations change drastically influencing whole grammatical categories of the language, as they are presented in the dictionary. I have already cited the shift from verb roots to their binjanim in Hebrew lexicography. This shift involves all Hebrew verbs and completely changes the characteristics of Hebrew dictionaries. An additional example can be drawn from Russian lexicography: of late all the verb items include the first and the third persons singular of the Present or Future Tense conjugation as run-ons to the initial infinitive. Some years ago the run-ons were the first and the second persons singular. The new variant is thought to better represent the whole verbal paradigm. This last example shows, that the morpheme/word continuum influences not only head-lexemes, but also their run-ons.

5. Why is it possible to combine word and morpheme in one unit?

Each word in the language functions as a sign in a linguistic sign-system. It has been well-known since the time of de Saussure, who also indicated the dual character of the linguistic sign. On the one hand, it relates to extralinguistic reality, denoting some of its features or phenomena. Previously we enumerated properties of the linguistic sign which in our word/morpheme dichotomy characterise the word. Prominent among them is its denotational force, its implied meaning of the extra-linguistic.

On the other hand, being included in the language, this sign is provisory to all rules and properties of the same. These rules exist to efficiently process the signs within the

system. One of the main characteristics of linguistic systems is their division of words into morphemes. This is the morphologic property of linguistic signs, allowing us to process initial words into related sign units (inflectional forms, compounds, paradigmatically-connected words). Sometimes it is very easy to differentiate morphemes from words – i.e., flections. Sometimes it is a difficult, almost impossible task; and the most intricate designs (including Bloomfield's free and bound forms, cf. Ullmann 1972, 26-29) fail to function properly. The most difficult part is to differentiate root-morphemes from words, thus providing an opportunity to combine them in the heading and processing of dictionary entries.

6. Why are such obvious things mostly ignored in theory though not in practice?

There are, to my mind, at least two principal reasons. The first and most influential one is that we – professionals and laymen – are still hypnotized by the centuries-old notion that a dictionary is a collection of words or their combinations. This notion is reflected in etymology: we prepare a *Wörterbuch* in Germany, *slovar'* (from 'word') in Russia, מ ל ו ן (again from 'word') in Israel, etc. No wonder that the notion is taken for granted even by theoreticians in our field. Thus, in one of the most influential sources for lexicographers, Hartmann (1983, 3-4) writes: "the dictionary is defined as a 'reference book or list of words (usually in alphabetical order) together with a guide to their meanings, pronunciation, spelling, or equivalents in other languages'. Lexicography (and its practitioner, the lexicographer) is thus concerned with the job of describing all or some of the words of one or more languages in terms of their characteristic features, notably of their meaning." Even though one does encounter the mention of morphemes in dictionary-making, it is considered of casual importance. It is well reflected in the definitions of "dictionary", "lexicography", and "lexicographer" to be found in various influential dictionaries.

The second reason for ignoring morphemes is the fact that the most developed lexicographies are mostly based on analytic and uninflected languages like English and French. Clearly, if there is no pressing need for inflection, there is no urgency to accept initial word/morpheme combination as morpheme construction. I am certain that this problem was discussed (perhaps at length) in many national lexicographies, but the fact remains that at the international level the problem awaits a deep and penetrating formulation worthy of its eminence and scope.

7. Practical and theoretical implications

They are numerous and very important, but I will dwell only on two, which seem to me the most significant: one theoretical and one practical. We ought to give the problem its proper place in theoretical lexicography, adding a new dimension to the whole subject. I wrote only of the main problems; they have numerous implications. For instance, if we choose this or that morpheme-base on which to build the whole paradigm, many questions arise: what criteria should be employed in making a selection from among numer-

ous contenders? (the answer will depend on each particular language as well as on many additional factors); what constitutes run-ons and what form will they take?; what should the user do to achieve the needed result?; how to provide better instruction for use of a dictionary in this respect? Or another problem: which inflected morphemes are to be chosen for separate entries; how should they be marked, etc. ... something that Gabriele Stein did for English dictionaries (1985, 35-44).

To my mind, the most important practical implication for our profession is, that a lexicographer who heads a dictionary-making project should, first and foremost, be linguistically-minded. His main job is not finding the encyclopaedic definitions for se-lected words (this will be done by a specialist in the corresponding field), but rather their proper linguistic representation in the head-lexeme, adjusting the linguistic apparatus following them, choosing the right forms of run-on entries, etc.; I do not claim, that he be excluded from formulating definitions, from determining the list of chosen words, etc.; but the emphasis will be distinctly different.

Bibliography

Hartmann, R.R.K (1983): Lexicography: Principles and Practice. Academic Press, L.-NJ.

Stein, G. (1985): Word Formation in Modern English Dictionaries. In: Dictionaries, Lexicography and Language Learning, ed. by R. Ilson. The British Counsil & Pergamon Press, Oxford.

Ullmann, S. (1972): Semantics. Basil Blackwell, Oxford.