INFORMATIONAL MODELS OF DICTIONARIES I

INFORMATICA 1/91

Keywords: data, dictionary, electronic dictionary, information, informational dictionary, organization, structure, terminology, understanding

Anton P. Železnikar
Volaričeva ul. 8
61111 Ljubljana

A computer-aided dictionary can be comprehended, structured, and organized in three basic ways: as a data-structured information (classical computer program), as an information-oriented structure (advanced electronic dictionary), and as a complex informational entity in the sense of the informational arising capability. All three concepts can substantially surpass the performance of a book form dictionary which always performs as a fixed, unchangeable data entity.

Computer-aided and information-oriented dictionaries can bring into the foreground the most complexly imaginable semantic and pragmatic connectedness of information items. Under these circumstances it is possible to introduce various aids for construction, development, and understanding of information items constituting an advanced and also to the most pretentious user oriented dictionary. Further, the performance of bilingual and multilingual dictionaries can contain the quality of translation of sentences from one language into another on the user request. In monolingual dictionaries, the demands on syntactic, semantic, pragmatic, and stylistic requirements can be covered strictly and efficiently, delivering in any respect improved and corrected texts in written natural languages. This essay stresses the perspectives of an *informational approach* at the design of future electronic dictionaries.

Informacijski modeli slovarjev I. Računalniško podprti slovar je mogoče razumeti, strukturirati in organizirati na tri različne načine: kot podatkovno strukturirano informacijo (klasični računalniški program), kot informacijsko zasnovano strukturo (napreden elektronski slovar) in kot kompleksno informacijsko entiteto v smislu zmogljivosti informacijskega nastajanja. Vsi trije koncepti zmorejo bistveno preseči zmogljivosti knjižnega slovarja, ki predstavlja le fiksirano, nespremenljivo podatkovno entiteto.

Računalniško podprti in informacijsko zasnovani slovarji lahko upoštevajo tudi najbolj kompleksno semantično in pragmatično povezanost informacijskih enot. Pri tem je mogoče vpeljati različne pripomočke za gradnjo, razvoj in razumevanje informacijskih enot, ki sestavljajo napreden in tudi za najzahtevnejšega uporabnika zasnovan slovar. Razen tega pa lahko ima dvojezikoven ali večjezikoven slovar še lastnost prevajanja stavkov iz enega jezika v drugi na zahtevo uporabnika. V enojezikovnih slovarjih je mogoče dosledno in učinkovito izpolniti zahteve po sintaksnih, semantičnih, pragmatičnih in stilističnih značilnostih jezika in tako izboljšati in korigirati (lektorirati) pisana besedila v naravnih jezikih. Ta spis poudarja prednosti *informacijskega pristopa* pri zasnovi prihodnjih elektronskih slovarjev.

1. Introduction

... The semantic system of a language has to do with meanings and thus with the relation between the conventionalized symbols that constitute language and the external reality about which we need to communicate through language. ...

(NCD) 24

What is an informational dictionary and how does it differ from a common, book form dictionary? But even book form dictionaries can differ essentially among each other in their intention, extensiveness, and purpose. For the most pretentious users of dictionaries who search not only for the minimal meaning, translation, or pure concept of a word, but also for several other attributes like word contexts, phrases, idioms, and whole sentences, paragraphs, etc., who investigate informational items in an epistemological and hermeneutic (semiotic and historical) way, a convenient book form dictionary may not be a satisfactory solution at all. Already the semantic and pragmatic conceptual study of words calls for a careful examination of the meanings belonging to various informational items and the studying of origins, going or descending back into the domain of different languages, the modern and the antique ones, in a recurrent meaning-improving manner. To this type of the searching and construction of meaning, the intention of creating a new, innovative, and adequate meaning has to be taken into the conceptualization and design of an informational dictionary. In this respect, the process of using an informational dictionary, extracting the meaning from it, and constructing a resulting meaning of an informational item approaches, in principle, the problem of understanding as information (UI1, UI2).

The tool one would like to have is a particular, let us say informational expert system which could deliver a complex, linguistic (semiotic), inter-linguistic (multi-linguistic), and particular information concerning the possible and variously related meaning of an informational item (word, phrase, sentence). More precisely, this informational tool should perform as a sufficiently sophisticated informational expert dictionary system (IEDS for short) in any respect. In this essay we will develop concepts and symbolic formalism of IEDS's, rooting in the informational algebra (IIA) and suggesting their informational structure and organization in the proposed algebraic way. In the most cases we shall prefer the so-called formal hermeneutic approach of understanding (UI2), which considers the development of the meaning belonging to an informational item (term, headword) through the item's historicity, but giving also the outlook to the so-called direct (non-historical) parallel-cyclic approach of understanding, which can occur at the conferring of a new arising or arisen meaning in an innovative way.

In the course of the discourse concerning the problem of understanding at the use of informational dictionaries, several approaches are possible. The most direct approach is the so-called spontaneous-circular understanding as used, for instance, by an inexperienced and linguistically unpretentious user of a dictionary. A professional writer, for instance, will insist to use an extremely hermeneutical method based on elementary spontaneous-circular understanding of an informational item, but also on the mostly considered expertise of understanding. While in the first case, the approach to understanding will be characteristically understanding, using a regular informing of understanding, in the second case, this process will take the form of expertizing, giving attention to the specialized emphases, which cannot only improve the content of meaning but can also enrich or innovate it essentially. In an implicit or explicit form, all these approaches of understanding of informational items of a dictionary will concern the moving from one informational item to another inside the information's semantic and pragmatic nets (maybe mutually connected or not) within an informational dictionary. In some respect, a satisfactory informational dictionary will perform as a »well-connected« (completely connected) net of included informational items. Theoretically, the well-connected net will guarantee the access from an arbitrary informational item of the dictionary to any other concerning item. In general, an arising dictionary will always enlarge and enrich its nets and thus improve its well-connectedness in an informational manner.

The question of building up the process of a dictionary usage goes back to the question, how does a writer or speaker perform in writing or speaking a text. Further, how does a language interpreter or a professional writer search and construct the meaning of words, phrases, and sentences by using single- language, cross-language, and professional (technical) dictionaries, handbooks, lexicons, encyclopedias, scientific, technical, philosophical scriptures, etc. How does someone construct the meaning of an informational item by gathering information from various and semantically different documents, memory, and its own intention? And finally, how does someone invent its own meaning, although considering the usual and existing meaning for an informational item?

The discussion in the previous passage lends our attention to the problem of discourse occurring between a user's metaphysical disposition (as an individual, conscious and unconscious information) and an informational dictionary (as a system information). The moving inside different semantical-pragmatical nets of a dictionary will largely depend on the dialog between the user and the informational expert facilities of the dictionary. With the exception of the problem of understanding and expertness of this bipolar informational game also the problem of discursiveness will come into the focus of our attention.

On the basis of this introductory analysis, the following scheme of our formal informational investigation can be put into the question: What could be a general informational structure of an informational dictionary? Which are the basic questions of the design and usage of a dictionary through the elementary possibilities (approaches) of understanding? Which could be the consequences of the discursive nature of understanding between the user and an informational dictionary (ITD)? How does a dictionary function as a discursive informational expert dictionary system.

What is an informational model? The informational model (IM) of something will be a system of informational formulas describing something as an informational entity (*Železnikar,* IIA). Electronic dictionary, ED for short, will be the term conceptualized by the *Japan Electronic Dictionary Research Institute* (look at OED) and marking a sophisticated project and implementation of dictionaries in natural languages. Thus, an ED is an information-aided or information-oriented structure which essentially surpasses the so-called book form dictionaries and possibly also several performances (functionality) of living human dictionaries. This holds true in particular in cases of inter-lingual dictionaries concerning several natural and ancient languages simultaneously.

In this essay we shall distinguish three main informational models (IM's) of dictionaries, that is, a data-structured dictionary, an information-structured dictionary, and, finally, an informational dictionary.

The second basic aspect of an electronic dictionary we must keep in mind is the question of understanding which concerns an ED, that is, the self-understanding as the quality of the programmed computing system and the so-called other-understanding (understanding of the electronic system users and designers). The system understanding represents the system functionality, its particular performance or intelligent capability in comparison to living and other artificial systems.

2. Data and Information

> ... *No living language stands still, however much we might wish at times that it would.* ...
>
> (NCD) 24

Before we start the discourse on data-structured, information-structured, and informational dictionary, it is recommendable to clear the question concerning the concept of data, i.e. data entity, and that of information, i.e. informational entity, conceptually and in an informationally formal (logically informational) way. So we have to point out clearly the conceptual difference of occurring theoretical distinction between the so-called data and information(al) entities.

What is data as an informationally theoretical entity? How does it inform and how can it be not informed? A written text (fixed on a sheet of paper) appears always as data. It is protected against any change or illegal informational arising. A written text can inform any of its readers, certainly in different manners, depending on the informational circumstances concerning the text and the reader. We agree that a written text cannot be changed in a writing way, thus the copies of the same text perform as equivalent data.

Let data as a particular informational operand be marked by δ. We agree upon the fact that data inform in an open way, that an informational receptor can be informed by data. This openness of data δ possibility is symbolically captured by expression $\delta \models$. On the other hand, we have to consider two further facts concerning the lasting (absolute memorizing, steadiness) of data, roughly as $\delta \models \delta$, and the impossibility of data to be informed (informational unchangeableness), roughly as $\not\models \delta$. The last formula, $\not\models \delta$, is closed in regard to δ because of $\delta \models \delta$, otherwise it is open. Further, formula $\delta \models \delta$ has to be particularized because operator $\models$ is a memorizing operator, that is, a non-arising (fixed) informational operator. Thus, instead of $\delta \models \delta$ we can introduce $\delta \models_m \delta$, where $\models_m$ marks a pure memorizing (data storing) informational operator.

After this conceptualization of data δ we can adopt the following informational implication:

$$\text{(d1)} \quad \delta \Rightarrow (\delta \models; \delta \models_m \delta; \not\models \delta)$$

In this formula (system of parallel informational formulas), operators $\models$ and $\not\models$ are the most general informational operators in the sense that

$$\text{(d2)} \quad (\delta \models) \Rightarrow (\delta \models_\delta \circ \models);$$
$$(\not\models \delta) \Rightarrow (\models_\delta \circ \models_\delta \delta)$$

Here, the operator composition $\models_\delta \circ \models_\delta$ performs as operator $\models_m$, which is closed (non-informing) to any other informational entity except of the memorizing data δ. Further, a more precise denotation of operator $\models_m$ would be $\models_\delta$ to which the operator composition $\models_\delta \circ \models_\delta$ reduces in a logical way. Finally, the entity δ as data implies

$$\text{(d3)} \quad \delta \Rightarrow (\delta \models; \models_\delta \delta)$$

where $\models$ is a general informational operator, whereas $\models_\delta$ is closed against any other operand except δ and, in this way, represents also a general informational operator of non-informing, that is, operator $\not\models$. Additionally, operator $\models_\delta$ is the operator of preserving or strict memorizing of the data entity δ, that is, particularly structured operator, denoted by $\models_{\delta,m}$.

We see how data δ is an informational exception in regard to regular (arising) informational entities. A question which arises on account of formula (d3) is what could the expression

$$\text{(d4)} \quad \delta \models_\delta; \models_\delta \delta$$

represent. By definition, this data entity cannot inform or cannot be informed by any other informational entity except itself. Such an entity would perform as informational noise or concealment for other informational entities. However, entities (processes) $\delta \models_\delta$ and $\models_\delta \delta$ which constitute (d4) could inform and could be informed, for instance, as the processes $(\delta \models_\delta) \models$; $\models (\delta \models_\delta)$; $(\models_\delta \delta) \models$; $\models (\models_\delta \delta)$; etc. A paragon of such a concept could be, for instance, the so-called Being as information (a universal constant or unity of being which never changes, as taught by the Eleatic and some modern existentialists).

If data is information which as an informational entity during an informational process never changes, information is the entity which ever or always changes. By this qualification, changing data perform as information. But this can never take place in the case of a book as seen from the side of the book itself. A book can neither textually change itself nor can it be textually changed by some other informational means. But this does not hold in the case of an electronic book (dictionary) if the user has the access to the text of the book.

What is information as an informationally theoretical entity and what is a clear distinction in comparison to data? How does information inform and how is it informed? Let information as a particular informational operand be marked by α. We agree upon the fact that information informs in an open way and that by this entity impacted entities can be foreseeable but also unforeseeable. This openness of information α possibility is symbolically captured by $\alpha \models$. On the other hand, we have to consider two further facts concerning the open and spontaneous cycling of information, roughly as $\alpha \models \alpha$, that is the possibility of information to be informed, roughly as $\models \alpha$. The last formula, $\models \alpha$, is unlimitedly open in regard to α as well as in regard to any other informational entity which could impact α. After this conceptualization of information α we can adopt the following informational implication:

(i1) $\alpha \Rightarrow (\alpha \models; \alpha \models \alpha; \models \alpha)$

In this formula (system of parallel informational formulas), operators of the form $\models$ are the most general informational operators in the sense that

(i2) $(\alpha \models) \Rightarrow (\alpha \models_{\alpha} \circ \models)$;
$(\models \alpha) \Rightarrow (\models \circ \models_{\alpha} \alpha)$

As we can see, the concept of information is in some sense controversial to the concept of data. By definition, data never changes or, at least, does not change within a domain of observation. On contrary, information ever changes or, at least, possesses the possibility to change and to be changed by itself and other informational entities. Thus, the so-called Being as a philosophical concept appears as the most universal data, in some respect opposite to the idea of god, which embodies the most pretentiously developing, although self-sufficient informational entity. The advantage of the concept of information lies in the possibility to express ideas being controversial to the concepts themselves.

We can make a direct comparison between the concept of information α and the concept of data δ in the following formal way:

(id1) $\alpha \Rightarrow (\alpha \models; \models \alpha)$; $\delta \Rightarrow (\delta \models)$;
(id2) $\alpha \Rightarrow (\alpha \models \alpha)$; $\delta \Rightarrow (\delta \models_{m} \delta)$

We use operators $\Rightarrow$ and ';' between formulas as informational operators of implication and of 'or/and', respectively. Thus, information α implies $\alpha \models$ or/and $\models \alpha$, i.e. an open informing of α in an outward and inward direction. In this sense, data δ is a sub-concept of information α. However, data δ preserves its form (data information), thus, $\delta \models_{m} \delta$, where $\models_{m}$ is the symbol of an absolute (non-arising) memory operator with the meaning

(d5) $(\delta \models_{m} \delta) \Rightarrow (\delta = \delta)$

The so-called changeable data become information for which formula (d5) holds between two events of their change. On contrary, information can change by itself alone, without any outward influence.

3. A General Informational Structure and Organization of a Dictionary

... in English, word order is a dominant factor in determining meaning, while the use of inflectional endings to mark the grammatical function of individual words within a sentence plays a clearly subordinate role ... Other languages show markedly different patterns, such as Latin with its elaborate set of paradigms for nouns, verbs, adjectives, and pronouns and its highly flexible word order. ...

(NCD) 24

In any case, a dictionary is a kind of expert information dedicated to the correct, that is, commonly (socially) understandable speaking, writing, and language forming, creation, and translation. Usually and formally, dictionaries implement implicitly (rarely explicitly) the so called semantic nets, that is to say, networks of connections among various informational items with semantically and pragmatically related meanings.

A semantic net is an informational connectedness of items which originates out of the possibilities of distinct informational items to inform other and to be informed by other informational items. An informational net is determined by an informational system with identified and yet not identified, that is, possible forms of informing. Such as it is, a dictionary implements a complex informational system of informational markers (for instance, words) with their attributes and to them corresponding forms of meaning. But meaning of an item is nothing other than an informational formula with its own meaning which points to the meaning of other items occurring in the formula, in an informationally parallel and cyclic way. Thus, informational markers with their actual and possible meaning create an implicit informational net which arises in the process of searching in a dictionary also through the intention or requirements of the user.

Let us introduce the basic formal terms which constitute the informational structure of a dictionary.

An informational dictionary as a ordered collection of informational items, their attributes, and meanings will be marked by ϑ. A dictionary ϑ informs as a regular information and its informing (understanding) will be marked by $\mathfrak{D}$. In principle,

(g1) $\vartheta, \mathfrak{D} \models \vartheta, \mathfrak{D}$

is a direct, straight-forward (potentially developing) relation which says that a dictionary ϑ informs as such (per se), that is, $\vartheta \models \vartheta$, that it informs a kind of understanding $\mathfrak{D}$ (belonging to the dictionary and/or somebody else), that this process, as a form of understanding, can influence the content of a dictionary in the form $\mathfrak{D} \models \vartheta$ (in case of a dynamic changeable dictionary ϑ), and that any understanding of a dictionary can inform as such (per se) in the form $\mathfrak{D} \models \mathfrak{D}$. The hermeneutical approach to this basic situation would be

(g2) $(\vartheta \models \mathfrak{D}) \models \vartheta$ or also
$(\mathfrak{D} \models \vartheta) \models \mathfrak{D}$

We see how informing $\mathfrak{D}$ assumes the role of understanding of item ϑ and vice versa. In a real case, these basic formulas will appear in a much more complex (reasonably decomposed, developed, and/ or arisen) formal context into which several other components, for instance, various attributes and the meaning of items, will enter.

An informational item (word, clause, phrase, sentence, paragraph, etc.) of a dictionary will be denoted by ω. An informational item (word) informs in a specific way, for instance, within an instantaneous user's metaphysical disposition or as an attributed meaning within the dictionary ϑ, as a form of informing, marked by $\mathfrak{W}$. Formally, we have the following formulas, expressing this part of the definition:

(g3) $\omega, \mathfrak{W} \models \omega, \mathfrak{W}$;
$(\omega \models \mathfrak{W}) \models \omega$; $(\mathfrak{W} \models \omega) \models \mathfrak{W}$;
$\omega \in \vartheta$; $\mathfrak{W} \in \mathfrak{D}$

A possible consequence of formulas (g2) and (g3) is, for instance,

(g4) $(((\omega \models \mathfrak{W}) \models \vartheta) \models \mathfrak{D}) \models \omega$;
$(((\mathfrak{W} \models \omega) \models \mathfrak{D}) \models \vartheta) \models \mathfrak{W}$

In these formulas, informing $\mathfrak{W}$ as a component of the item's understanding already hides the meaning of ω. We see how the complexity of hermeneutic understanding of an item in the dictionary grows with the decomposition (or identification) of the concept.

Entity ω is a marker which marks the corresponding attribute(s) and meaning(s), that is, $\alpha(\omega)$ and $\mu(\omega)$, respectively. In a dictionary ϑ, its item ω always assumes the informational triplet (or in general an n-tuple) $(\omega, \alpha(\omega), \mu(\omega))$. This triplet can be read as an informational correspondence of ω, $\alpha(\omega)$, and $\mu(\omega)$ and expressed formally by

(g5) $(\omega \therefore \alpha(\omega)) \therefore \mu(\omega)$

where $\therefore$ is an informational operator with the meaning 'corresponds'. In general,

(g6) $\omega_1, \dots, \omega_m \in \vartheta$;
$(\omega_i \therefore \alpha(\omega_i)) \therefore \mu(\omega_i)$; $i = 1, \dots, m$

where $\omega_1, \dots, \omega_m$ are elements of dictionary ϑ. The consequence of this concept is

(g7) $(\omega \in \vartheta) \Rightarrow ((\omega \therefore \alpha(\omega)) \therefore \mu(\omega))$;
$((\omega \therefore \alpha(\omega)) \therefore \mu(\omega)) \in \vartheta$

This is certainly only a static definitional expression of the content $((\omega \therefore \alpha(\omega)) \therefore \mu(\omega))$ entering in a dictionary ϑ.

While ω belonging to ϑ is a single informational item, the attribute and the meaning of ω, marked by $\alpha(\omega)$ and $\mu(\omega)$, respectively, are, in general, informational sets of items, that is, sets of formulas depending on other informational items belonging to ϑ. In some respect, an informational dictionary appears as an in-itself closed (completed) information.

The attribute of ω is a set of informationally distinct data entities (for instance, pronunciation, grammatical items, the kind and time of ω's origin, graphical explanation, pictorial illustration, etc.). Certainly, the attribute item can have semantic connections to other items in the dictionary.

The meaning of ω is a set of informationally distinct entities expressed in the form of informational formulas of informational operands and operators (as well as parentheses) and ω within ϑ is nothing else than the marker of this set of formulas. How is the meaning of a marker structured and organized into the so-called semantic-pragmatic net within a dictionary?

If we look into a book dictionary, the meaning of an item appears usually as a non-empty collection of phrases, that is composite structures of words. The formally declared meaning of item ω, marked by $\mu(\omega)$, can be informationally divided, in general, as

(g8) $\mu(\omega) = (\pi_1(\omega), \dots, \pi_n(\omega))$

However, formula (g8) is an informationally static expression. In fact, partial meanings of ω, that is, phrases $\pi_1(\omega), \dots, \pi_n(\omega)$, inform the resulting meaning $\mu(\omega)$ of ω and thus instead of formula (g8) the informing of these partial entities to the entire meaning of ω in ϑ is, in general, a more adequate situation:

(g9) $(\pi_1(\omega), \dots, \pi_n(\omega)) \models \mu(\omega)$

This informing simultaneously satisfies the condition that with the exception of the complete inclusion of phrases in the meaning of an item, phrases of the meaning are in no way isolated or mutually and also otherwise independent items, that is,

(g10) $\pi_1(\omega), \dots, \pi_n(\omega) \subset \mu(\omega)$;
(g11) $\pi_1(\omega), \dots, \pi_n(\omega) \models \pi_1(\omega), \dots, \pi_n(\omega)$

In fact, the mutual informing of phrases (g11) produces a resulting phrase of ω's meaning, marked by $\mu_r(\omega)$, also in the user domain, where

(g12)
$(\pi_1(\omega), \dots, \pi_n(\omega) \models \pi_1(\omega), \dots, \pi_n(\omega)) \models \mu_r(\omega)$

or in a more complete form

(g13)
$((\omega \therefore \alpha(\omega)) \therefore$
$(\pi_1(\omega), \dots, \pi_n(\omega) \models \pi_1(\omega), \dots, \pi_n(\omega))) \models \mu_r(\omega)$

It is to stress that operator '$\therefore$' has to be understood as an informationally active operator and not only as a static relation lacking the dynamics of understanding as information. Operator $\therefore$ is a particularized form of the general operator of informing, that is, operator $\models$. Although the proposed schemes (or scenarios) of meaning of an item concern understanding, we have not introduced an explicit indication of understanding yet. This type of formalization we shall develop in some of the following sections expressing the explicit forms of informing belonging to the distinct items of meaning. Thus, we shall proceed to some concrete structures and organizational schemes of an electronic dictionary.

In general, it is possible to introduce several distinct types of understanding concerning a composite informational item ω of a dictionary and particularly the understanding of the meaning, which corresponds to the dictionary item component μ. Various types of understanding can be constructed as parallel-cyclic, hermeneutic, and hermeneutic parallel-cyclic (mixed) modes of understanding.

4. Terminology, Symbols, and Relations Concerning Dictionaries

... The major systems that make up the broad comprehensive system of language itself are four in number: lexicon, grammar, semantics, and phonology. The one that dictionary editors and dictionary users are most directly concerned with is the vocabulary or lexicon, the collection of words and word elements which we put together in various ways to form larger units of discourse: phrases, clauses, sentences, paragraphs, and so forth. ...

(NCD) 23

A dictionary as informational entity will be denoted by ϑ. Particular dictionaries as well as sub-dictionaries of a dictionary will be denoted by subscripted ϑ's. A subscript will be the integer or the letter combination.

Particularly important dictionaries will be, for instance: the word dictionary, marked by ϑ_w; the concept dictionary, marked by ϑ_c; the co-occurrence dictionary, marked by ϑ_{co}; and the bilingual dictionary, marked by ϑ_b. These particular dictionaries appear as information entities within the Japan EDR electronic dictionary (OED).

Already particularized dictionaries can be split, for instance, in the following way: the word dictionary ϑ_w into the general vocabulary dictionary ϑ_{gv} and the technical terminology dictionary ϑ_{tt}; the concept dictionary ϑ_c into the concept classifications ϑ_{cc} and the concept descriptions ϑ_{cd}; the co-occurrence dictionary ϑ_{co} into the Japanese co-occurrence dictionary ϑ_{Jco}, the English co-occurrence dictionary ϑ_{Eco}, and the Slovene co-occurrence dictionary ϑ_{Sco}, for instance; the bilingual dictionary ϑ_b into the Japanese-English dictionary ϑ_{JE} and the English-Japanese dictionary ϑ_{EJ} or into the Slovene-English dictionary ϑ_{SE} and the English-Slovene dictionary ϑ_{ES}, etc.

Furthermore, the general vocabulary dictionary ϑ_{gv} can be split into the Japanese general vocabulary dictionary ϑ_{Jgv} and the English general vocabulary dictionary ϑ_{Egv}, or into the Slovene general vocabulary dictionary ϑ_{Sgv} and the English general vocabulary dictionary ϑ_{Egv}, etc. Similarly, the technical terminology dictionary ϑ_{tt} can be split into the Japanese technical terminology dictionary ϑ_{Jtt} and the English technical terminology dictionary ϑ_{Ett}, or into the Slovene technical terminology dictionary ϑ_{Stt} and the English technical terminology dictionary ϑ_{Ett}, etc.

By this surface structure of embedded dictionaries we have obtained a rough structure of an electronic dictionary which can be informationally expressed by the following system of operationally interconnected informational entities:

(D1)
$$\begin{aligned}
&\vartheta = (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b);\\
&\quad \vartheta_w = (\vartheta_{gv}, \vartheta_{tt});\\
&\qquad \vartheta_{gv} = (\vartheta_{Jgv}, \vartheta_{Egv});\\
&\qquad \vartheta_{tt} = (\vartheta_{Jtt}, \vartheta_{Ett})\\
&\qquad \text{or } \vartheta_{gv} = (\vartheta_{Sgv}, \vartheta_{Egv});\\
&\qquad\quad \vartheta_{tt} = (\vartheta_{Stt}, \vartheta_{Ett});\\
&\quad \vartheta_c = (\vartheta_{cc}, \vartheta_{cd});\\
&\quad \vartheta_{co} = (\vartheta_{Jco}, \vartheta_{Eco})\\
&\quad \text{or } \vartheta_{co} = (\vartheta_{Sco}, \vartheta_{Eco});\\
&\quad \vartheta_b = (\vartheta_{JE}, \vartheta_{EJ})\\
&\quad \text{or } \vartheta_b = (\vartheta_{SE}, \vartheta_{ES})
\end{aligned}$$

The next step in this system determination is to define the relations (interconnections, operations, also understanding) among the occurring entities (entries) and the deep structure of any of them, which brings new terms and symbols to the surface.

What are the data entries of all these dictionaries, how are they structured, interconnected, and understood, for instance, in the Japanese EDR electronic dictionaries (OED)?

What is the so-called word entry? The word entry ω is composed of a headword, marked by ω_h, which is the surface representation of a word, its grammatical information ω_g (grammatical features), and the corresponding headconcept ω_{hc}, which is the concept represented by words in the context. Thus,

(h1) $(\omega = (\omega_h, \omega_{hg}, \omega_{hc})) \in \vartheta_w$

is already an inter-dictionary structure. Further, there are relations among headwods ω_h or among headconcepts ω_{hc} which are defined within the word entries ω. Concept entries ω_c define possible relations between headconcepts ω_{hc} with concept relation labels ω_{crl}. Thus,

(h2) $(\omega_c = (\omega_{hc}, \omega_{crl})) \in \vartheta_c$

Co-occurrence entries ω_{co} define all of the possible co-occurrence between headwords ω_h with co-occurrence relation labels ω_{corl}, so,

(h3) $(\omega_{co} = (\omega_h, \omega_{corl})) \in \vartheta_{co}$

Inter-lingual correspondence entries ω_b, in concrete cases, bilingual entries ω_{JE}, ω_{EJ}, ω_{SE}, ω_{ES}, etc., define correspondence between headwords ω_h of different languages, in concrete cases, headwords ω_{hJ}, ω_{hE}, ω_{hS}, etc. For instance,

(h4) $(\omega_{JE} = (\omega_{hJ}, \omega_{hE})) \in \vartheta_{JE}$;
$(\omega_{EJ} = (\omega_{hE}, \omega_{hJ})) \in \vartheta_{EJ}$;
$(\omega_{SE} = (\omega_{hS}, \omega_{hE})) \in \vartheta_{SE}$;
$(\omega_{ES} = (\omega_{hE}, \omega_{hS})) \in \vartheta_{ES}$

etc. Here, ϑ_{JE}, ϑ_{EJ}, ϑ_{SE}, ϑ_{ES}, etc. are bilingual dictionaries.

In formulas (h1), ... , (h4), relations between different entries are implicit. It is possible to make these relations explicit in a general way, with the intention to mark these relations and make them later on completely definite by further particularization of operators and/or composition and decomposition of formulas.

For instance, we can *decompose* the concept dictionary $\vartheta_c = (\vartheta_{cc}, \vartheta_{cd})$ in (D1) which is an informational network structure comprising a set of concept entries $\gamma = (\gamma_n, \gamma_r)$, with nodes γ_n which represent the concepts and arcs γ_r which indicate the relation between one concept and another.

5. Understanding as the Processing Part among the Parallel Dictionaries within an Electronic Dictionary

Understanding in the framework of an ED can be conceived from at least three points of view: as system understanding implemented within the programmed computer system which comprises the ED; as understanding of designers who develop the ED by means of an ED development system and understand also the development system itself; and, at last, as understanding of the ED performance and capability by the ED users who have this system on disposal, however, can still adapt or alter it to some reasonable extent, for instance, by a built-in or outward possibility of learning.

In the process of ED development, the designers and constructors start by a top-down approach of solving the problem of, for instance, the very sophisticated ED not being definitely depicted yet. At the very beginning, they may have in mind the classical data-information system (D1) of the ED, which from this point on can be put into the process of the so-called informational arising, where the occurring entities can be informationally decomposed deeper and deeper to the necessary details. Thus, for instance, the informationally static system (D1) can be put into a formally arising form as

(D2) $\vartheta \models (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b)$;
$\vartheta_w \models (\vartheta_{gv}, \vartheta_{tt})$;
$\vartheta_{gv} \models (\vartheta_{Jgv}, \vartheta_{Egv})$;
$\vartheta_{tt} \models (\vartheta_{Jtt}, \vartheta_{Ett})$

$$\text{or } \vartheta_{gv} \models (\vartheta_{Sgv}, \vartheta_{Egv});$$
$$\vartheta_{tt} \models (\vartheta_{Stt}, \vartheta_{Ett});$$
$$\vartheta_{c} \models (\vartheta_{cc}, \vartheta_{cd});$$
$$\vartheta_{co} \models (\vartheta_{Jco}, \vartheta_{Eco})$$
$$\text{or } \vartheta_{co} \models (\vartheta_{Sco}, \vartheta_{Eco});$$
$$\vartheta_{b} \models (\vartheta_{JE}, \vartheta_{EJ})$$
$$\text{or } \vartheta_{b} \models (\vartheta_{SE}, \vartheta_{ES})$$

System (D2) is a developing formal system which calls for the missing details, that is, for the decomposition possibilities concerning the data structure of the ED on the one side and the system-functional, developmental, and user-friendly understanding on the other side.

From this point on, understanding as a systematic informational approach can be put into the game of system development. As we have seen, hermeneutic and parallel-cyclic approaches of understanding can be applied simultaneously.

What is the built-in quality of understanding in a dictionary? How could it be developed out of basic, static, already listed informational formulas? A dictionary ϑ as a complex informational entity $\vartheta \models (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b)$ possesses a quality (mode) of understanding by which its informational components are informed. Thus, it is possible to put into discourse two basic questions: (1) What is the internal understanding $\mathfrak{D}$ of the dictionary ϑ? (2) How can an electronic dictionary ϑ be used (understood) by a user υ through his/her understanding $\mathfrak{U}$? We see how the problem of understanding can become informationally interweaved from the one and the other side.

At the very beginning of the concept we have merely ϑ as an arising information, as the intention of its development, that is, of its design and implementation. A basic concept ϑ informs and is informed, that is, $\vartheta \Rightarrow (\vartheta \models; \models \vartheta)$. This basic formula can be developed, decomposed, particularized, universalized, etc. in several ways. It can be read in the following way: as soon as the intention of a dictionary ϑ informing is given, it is permitted to transit to the system of formulas $(\vartheta \models; \models \vartheta)$. Thus, the first consequence of this implicative informing is, for instance, $\vartheta \models (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b)$, which from now on becomes the basis, although it possesses a static structure in which understanding of its components is undeveloped, that is, not expressed informationally yet. We conclude that understanding $\mathfrak{D}$ of ϑ is implicit and can be expressed through the development of the concept, by an arising system of formulas $\vartheta \models; \models \vartheta$ where the right $(\vartheta \models)$ and the left informational operands $(\models \vartheta)$ of operator $\models$ are coming into existence, respectively.

In principle, by understanding, informational processes are closed into loops of understanding. The concept of the so-called self-understanding of an informational entity α and understanding of α by the entity β through β's understanding can be the following. Informing of α as its implicit informational process can be marked by $\mathfrak{J}_\alpha$, where $\mathfrak{J}_\alpha$ marks the entire informing of α. Thus, the understanding component of α within its implicit informing $\mathfrak{J}_\alpha$ can be marked by $\mathfrak{A}$ yielding $\mathfrak{A} \subset \mathfrak{J}_\alpha$. Under these circumstances, the understanding part $\mathfrak{A}$ of α's informing $\mathfrak{J}_\alpha$ can be formalized by the system of formulas

$$\text{(D3)} \quad \mathfrak{A} \subset \mathfrak{J}_\alpha; (\alpha \models \mathfrak{A}) \models \alpha$$

Further, if entity β understands α by its understanding $\mathfrak{B}$ as a part of the entire β's informing $\mathfrak{J}_\beta$, then

$$\text{(D4)} \quad \mathfrak{A} \subset \mathfrak{J}_\alpha; (\alpha \models \mathfrak{A}) \models \alpha;$$
$$\mathfrak{B} \subset \mathfrak{J}_\beta; (\beta \models \mathfrak{B}) \models \beta;$$
$$((((\beta; \alpha \models \mathfrak{A}) \models \alpha) \models \mathfrak{B}) \models \beta$$

In this scheme, entity β considers the understanding $\mathfrak{A}$ of α within its own understanding $\mathfrak{B}$. Usually, the understanding of something as an informational process produces a particular information called the meaning μ of something. For instance, in case of self-understanding $\mathfrak{A}$ of α, we have the meaning $\mu_{\mathfrak{A}}(\alpha)$. In case of the so-called other-understanding, the meaning $\mu_{\mathfrak{B}}(\alpha)$ or even $\mu_{\mathfrak{B}}(\mu_{\mathfrak{A}}(\alpha))$, $\mu_{\mathfrak{B}}(\beta, \mu_{\mathfrak{A}}(\alpha))$, etc. can come into existence. In a discourse between entities α and β, system (D4) extends into

(D5) $\mathfrak{A} \subset \mathfrak{I}_\alpha; (\alpha \models \mathfrak{A}) \models \alpha;$
$\mathfrak{B} \subset \mathfrak{I}_\beta; (\beta \models \mathfrak{B}) \models \beta;$
$((((\alpha; \beta \models \mathfrak{B}) \models \beta) \models \mathfrak{A}) \models \alpha;$
$((((\beta; \alpha \models \mathfrak{A}) \models \alpha) \models \mathfrak{B}) \models \beta$

After this discussion, it is possible to answer the question concerning the understanding occurring among parallel dictionaries within an electronic dictionary. The basic formula $\vartheta \models (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b)$ in (D2) and the subsequent formulas of (D2) give a firm orientation in which direction the system design of the system understanding should develop. Here, understanding concerns the necessary processing among dictionaries guarantying the achievement of goals of an ED, that is, to be functional, efficient, user-friendly, and innovative.

The decomposition of $\vartheta \models (\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b)$ can begin by the following steps:

(D6)

(1) $(\vartheta \models \vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b) \Rightarrow$
$(\vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b \models \vartheta_w, \vartheta_c, \vartheta_{co}, \vartheta_b);$

(2) $((\vartheta_w \models \vartheta_c) \Rightarrow$
$(\omega \in \vartheta_w, \gamma \in \vartheta_c)) \Rightarrow$
$(\omega \models_{poi} \gamma; \gamma \models_{del} \mu(\omega));$

(3) $(((\omega \models \mathfrak{W}) \models \gamma) \models \mathfrak{S}) \models \mu(\omega); \ldots$

In formula (2) of (D6), the particularized operators $\models_{poi}$ and $\models_{del}$ are read as »points« (or »marks«) and »delivers«, respectively. Entities $\mathfrak{W}$ and $\mathfrak{S}$ in (3) mark to the word entity ω and to it corresponding concept γ belonging understandings, respectively, informing the meaning $\mu(\omega)$.

From the last system of formulas one can see the course of a dictionary development considering the so-called functions of understanding. In the next sections of the essay we shall point to some further informational details of an electronic dictionary disclosing the structure of some very concrete concepts. In these concepts, problems of monolingual, bilingual and multilingual dictionaries will be treated from the developmental, operational and implementing point of view. Further, a clear conceptual distinction among various concepts of dictionaries has to be developed, for instance, concerning formal models of data-structured (classically programmed), information structured (advanced conceptualized), and the so-called informational dictionaries. In this point, a question of developing tools for these different sorts of dictionaries can arise. Certainly, several questions will remain unanswered, but a new informational way for the development of the most sophisticated and advanced dictionaries will begin to become its specific trace.

References

(OED) *An Overview of the EDR Electronic Dictionary,* Technical Report TR-024, Japan Electronic Dictionary Research Institute, Tokyo (1990).

(NCD) *Webster's Ninth New Collegiate Dictionary,* Merriam-Webster Inc., Publishers, Springfield, Ma (1986)

(ITD) A.P. Železnikar, *An Informational Theory of Discourse I,* Informatica **13** (1989) *4,* 16-37.

(IIA) A.P. Železnikar, *An Introduction to Informational Algebra,* Informatica **14** (1990) *1,* 7-28.

(UI1) A.P. Železnikar, *Understanding as Information (Part One: A General Philosophy and Theory of Understanding as Information),* Informatica **14** (1990) *3,* 8-30.

(UI2) A.P. Železnikar, *Understanding as Information (Part Two: A Formal Theory of Understanding as Information),* Informatica **14** (1990) *4,* 5-30.

A comment. The listed references concern only the first part of the article.