Vesa Suominen
Department of Information Studies, University of Oulu, Finland

Linguistic/Semiotic Conditions of Information Retrieval/Documentation in the Light of a Saussurean Conception of Language: "Organising Knowledge" or "Communication Concerning Documents"?

Abstract: Argumentation consists of representation of the basic structuralist concepts of language/semiotic as a two-level form, as a form of expression and here especially form of content, and of application of these concepts to the phenomena of the representation of the contents of documents. On the basis of argumentation the paper questions the notion of "organizing knowledge", is it, or in what sense it is possible to organize knowledge. The paper hinds out some reservations to viewing content representation as organizing knowledge in a strong sense and suggests that instead could be used a notion of (meta)documentation, characterized as communication concerning documents.

1. Introduction

The aspects I introduce in the following should have fundamental conceptual and general-level methodological significance in understanding the phenomena of information retrieval (or documentation, as defined at the end of the paper), and in their research. They also have to do with the self-understanding of the professional practice which should benefit from the results of this research. Further, these aspects are not very much highlighted in the prevailing ways of seeing these phenomena.

2. Language/Semiotic as a Form of Expression and a Form of Content

In Hjelmslevian terms, a semiotic (and also a language) is a form with two levels, namely the levels of content and of expression. A language or semiotic conceived of in this way joins together, and in the same structures two in themselves unstructured, formless or undifferentiated substances ("nebuleuses" as Saussure puts it), the substances of content and of expression. A semiotic as a form with two levels would thus be a two-level system of differences, a differentiation of the undifferentiated on the levels of expression and of content. The conception of a semiotic as a two-level form may be illustrated as in Figure 1.

(undifferentiated, formless) substance of content

| content form |
| expression form |

Figure 1: Language/semiotic as a two-level form

Advances in Knowledge Organization, Vol.6(1998)
Hjelmslev calls "sign" a specific kind of unit in a semiotic. In a sign some particular unit of content and some particular unit of expression are joined together. Words, prepositions, meaning-carrying suffixes, etc. are examples of signs. Hjelmslev (1961, 49-51) stresses that a sign is not a "sign [expression, VS] for something" but a totality of expression and of content. According to Saussure (1915), a language is not "a nomenclature, i.e. a list of terms for corresponding things" (transl. VS, orig. "une nomenclature, c'est à dire une liste des termes correspondant à autant des choses", op. cit., 97) for it is the language that gives identities to the "things" or "chooses" named and expressed. In this sense the specific contents belong to the language (semiotic) as well as the specific expressions.

The 'real thing', i.e., the referent, would be excluded from this conception of semiotic; this does not mean, however, that the 'real thing' is excluded from the reality. Hjelmslev (op. cit., 111) talks about "nominalism" with regard to linguistic theory - without any implications outside this domain.

For my present purposes the core notion of the Hjelmslevian frame is the content form within which particular units of content have their identities (as "cultural units", Eco 1967, 67).

Within the conceptual figure presented, the substances of expression and of content, should be taken as conceptual abstractions. Saussure's notion of substance is, however, substituted in the Hjelmslevian concepts by the notion of "purport" (see below).

3. Dimensions of the Structure

By paradigmatic and syntagmatic Hjelmslev refers to different dimensions of the structuring of expression and of content. Syntagmatic is a dimension where the substance (of content and of expression) is cut into units which occur together, with each other in a syntagm ("chain", Danish "kaedje", in the terminology of Hjelmslev). Hjelmslev calls the syntagmatic relationship a "both-and" relationship. Structuring of substance on the syntagmatic dimension is called partitioning. On the paradigmatic dimension, on the other hand, the substance is cut into a paradigm of alternative units, and thus Hjelmslev characterises the paradigmatic relationship as an "either-or" relationship. The structuring of substance on the paradigmatic dimension is called articulation. (Hjelmslev 1961, def. 26-31, 54, 55, 67 and 68).

With Hjelmslev (1961, 106) the general concepts system and process correspond to paradigmatic and syntagmatic. A paradigm is defined as a semiotic system and a syntagm as a semiotic process, i.e. as a system and as a process with the levels of content and of expression. Further, processes are said to manifest systems. Manifestation here means that the process, or syntagm, is made up of units that the system, the paradigm provides for use as a 'store' of alternative units.

For my purposes it is useful to add to the Hjelmslevian concepts one more type of relation which I call syntactic and by which I mean syntagmatic relationships with some specified pattern. I shall call the concepts describing these syntactic relationships categories of syntactic relationships.

4. Arbitrariness of the Content Form, and the Notion of Content Purport

One characteristic element of the structuralist way of thinking is the idea of the arbitrary nature of a sign (or of a semiotic or language). Originally Saussure (1915, 100-102) talks about the arbitrary relationship between expression ("signifiant") and content ("signifié"). This somewhat trivial notion has, however, also been criticised within the tradition of structuralist semiology. Benveniste (1966, 49-55) points out that it is not the relation of
expression and content that is arbitrary. What is arbitrary is the relation of the sign with reality, "that a certain sign, and not some other, be applied to a certain element of reality, and not to some other" (transl. VS, orig. "que tel signe, et non tel autre, soit appliqué à tel élément de réalité, et non à tel autre"). In other words, arbitrariness lies in the fact that a language or a semiotic 'portions' or 'classifies' the reality, things talked about, as well as expression, in some definite and particular way. Weinreich (1966,142) states the arbitrariness of the content form of a language quite clearly:-«The semantic mapping of the universe by a language is, in general, arbitrary, and the semantic map of each language is different from those of all other languages». (Actually, Saussure, 1915, 157, himself also refers to this side of the issue, i.e. that the content-forming function of language helps us to understand the arbitrariness of sign).

In brief, the principle, or moment, of arbitrariness means that a semiotic as a system of differences of contents is not reducible to the reality talked about. A language is to be conceived of as a cultural form on its own, based solely on social agreement and usage.

An important and also interesting Hjelmslevian notion is "purport" (in Danish "mening"). It must be noted that this notion is rather problematic. However, obviously with the notion of purport Hjelmslev intends to refine the Saussurean notion of substance. Hjelmslev's own example is rather lucid regarding the articulation of content (see Figure 2a). The purport of content in the example is the area of content, the particular field of colours, articulated, in the figure, quite concretely, the vertical bar in the figure. The arbitrariness of the content form in a semiotic is illustrated in the different ways in which the bar, i.e. the content purport, is cut into pieces in English and in Welsh. As an example of different partitionings of content Hjelmslev gives syntagms which all express the ignorance of the person using the expressions (and here the speaker's ignorance is the content purport structured, this time partitioned) (see Figure 2b). In this latter example the horizontal dimension would be the purport partitioned.

a) The articulation of purport (Hjelmslev 1961, 53).

```
<table>
<thead>
<tr>
<th>green</th>
<th>gwyrd</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue</td>
<td>glas</td>
</tr>
<tr>
<td>grey</td>
<td></td>
</tr>
<tr>
<td>brown</td>
<td>llwyd</td>
</tr>
</tbody>
</table>
```

b) The partitioning of purport (Hjelmslev 1961, 50; Latin explicans from Mauro's commentaries in Saussure, 1915, 462).

```
jeg ved de ikke (Danish) | EGO SCIO ID NON
I do not know (English) | EGO AG(O) NON SCI(RE)
je ne sais pas (French) | EGO NON SCI(O) PASSUM
en tiedi (Finnish) | EGO-NON-FACIO SCIRE
naluvara (Eskimo) | NON-SCIENS-(SU)M-EGO-ID
```

Figure 2: The articulation and partitioning of purport
In a way the notion of purport can be characterized as an intermediary concept between the notions of substance and form. Substances of expression and of content would be totally undifferentiated. But such undifferentiatedness could be only a theoretical abstraction. So, if substance is a metaphysical, 'absolute' undifferentiated, purport refers to a kind of (functional) medium stage of differentiation of content (and expression). Thus purport could also be characterized as a 'semi-substance' or 'semi-form'. Purport could be illustrated as a 'portion of substance' which in some semiotic is structured in some particular way, in some other in another way - and yet in another perhaps not at all.

The notion of content purport also, in a way, refers to the 'world talked about', to the referent which as such does not belong to the general logic of structuralist thinking. The content purport may, at least in some instances, be an idea, thought or knowledge about the 'thing' or 'world talked about' (i.e., an idea about some particular referent).

5. Documentary Language as a Content Form

The basic claim of my conceptualisation, of the semiology of documentary languages (DLs), is that DLs are semiotics in the Hjelmslevian sense, i.e. forms, or systems of differences of expression and of content, with the arbitrariness of those forms. Especially interesting here is that a DL could structure content in an arbitrary manner, as a cultural form on its own.

The point of my conceptualisation lies in the opposition of two possible ways of thinking about the structure of content in DLs and in content representations. The first - and established way within the research on DLs - is to describe the specific ways in which various given units of content are, or might be, joined, in order to construct complex content representations. Here we could talk about structuring the totality out of the units (or structuring the units into totalities). The second possibility - and the one I am suggesting here - is to describe the specific ways in which content is cut into units in DLs or in content representations. Here, respectively, we could talk about structuring the totality into units (or structuring the units out of the totality).

The words 'syntagmatic' and 'paradigmatic' have belonged to the basic vocabulary of the research on classification and DLs since the 1960's (see e.g. Cros & Gardin & Levy 1964, 40 ff. and Hutchins 1975). Their conceptual content has, however, differed from the content I am going to give them here when applying the Hjelmslevian concepts in a rather straightforward way to conceptualise those phenomena. By paradigmatic relationship I here simply mean a distinction of units of content, an 'either-or' relationship following the characterisation given by Hjelmslev (when speaking about content; naturally there are in DLs paradigmatic relationships, i.e. distinctions of expression, too). In other words, through those paradigmatic relationships, i.e. distinctions of content, certain particular units of content are separated and so given existence and identity as alternatives within a particular DL. By syntagmatic, on the other hand, I mean the relationship between units which appear together, 'both and', units resulting from partitioning of content and forming a syntagm of some kind.

Thus paradigmatic and syntagmatic would be dimensions of structuring (articulation, partitioning) of content, which in itself is regarded as undifferentiated, formless. Here one further type of relationships, the syntactic relationships as a syntagmatic relationship with some specific pattern becomes crucial. In their definiteness syntactic relationships describe some particular and definite patterns of partition of the content, and consequently they indicate dimensions of its articulation.
6. Metasemiotic Systems and Processes of Content Representation

My interpretation of the totality of the relations between content representation and documents as semiotic phenomena is based on the Hjelmslevian notion of metasemioticity. Hjelmslev (1961, 119-120, def. 104) means by metasemiotic a semiotic which "treats of" another semiotic, or a "(scientific) semiotic one or more of whose planes is (are) semiotic(s)".

Reflecting more strictly on how an object semiotic should be conceived of as the content plane of a metasemiotic, it should be said that a metasemiotic is structuring (idea, thought, knowledge about) the object semiotic as a semiotic generally structures the content purports. Consequently, in the metasemiotic relationship there is present a moment of arbitrariness of structuring (the purports of) the content. This means that a metasemiotic does not just duplicate the structures of the object semiotic (or it may do so, but only in a special case), but restructures the object semiotic, including its content form.

Within the structuralist frame, documents, too, and the discourse or "flow" of documents (called in Suominen 1997 "documentary history") may be clearly characterised as semiotic phenomena where some specific expressions are connected with some specific contents. Further, this process may be conceived of as manifesting some semiotic systems where both the specific expressions and the specific contents have their identities. However, the flow of documents, as a semiotic process, is rather complex and heterogeneous, and the systems manifested are various. If documents are literary, obviously they manifest the systems (of expression and of content) of the languages used. Still I claim that there are also some other semiotic systems which are manifested in documents and which are more specifically characteristic of documents. For instance, one could talk about the 'syntax of the content of an article' where the content is partitioned - *grosso modo* - into elements like 'subject', 'point of view', etc. And accordingly, there would be some systems, paradigms of 'subjects' and of 'points of view', manifested in the process - even if they tended to vary and would not be as easily recognised as a system of, say, colours in a natural language can be.

The 'semiology of the flow of documents' serves here only as a context for formulating the semiotic problem of DLs and of content representation. Thus I think that the notions presented are sufficient to give an idea of what I mean when speaking about documents as a semiotic process manifesting some semiotic systems. (I assume that more concrete description of the semiotic process of documents could require a richer conceptual apparatus than the basic structuralist notions I am using here).

Now content representation may be defined as a *metasemiotic process manifesting DLs, the object semiotic of which is the flow of documents and the systems manifested there*. The constellation is described in Figure 3.

```
DLs as semiotic systems manifested in
| =============== the semiotic process of content representation ================>
| structuring as content purport

the semiotic systems manifested in
| ====== the semiotic process of documents ======>
```

Figure 3: The metasemiotic relationships of content representation, the semiotic process of...
documents, and the semiotic systems manifested

There is no doubt - it seems to me - that a DL may be conceptualised as a semiotic, as a system of differences of content (and of expression) which as a metasemiotic gives identities to the specific contents of documents in content representation. But we may ask whether it is wise to conceptualise DLs like that, if we also wish to consider their pragmatic functions. In this respect, my principal argument for the semiology of DLs, i.e. that the 'Saussurean conception of language' has something essential to say about DLs, is rather practical, and - as I see it - quite obvious and incontestable. It is the contrast between the syntactic complexity of the content of documents, including documentary history and individual documents, and the relative simplicity of the content form of DLs. In the light of this contrast the content of documents (or knowledge about this content) is in its diversity practically an undifferentiated substance, the 'Saussurean nebuleuse', when faced in the practice of content representation.

The metasemiotic ('meta language') character of DLs has been discussed in the literature on DLs. (see e.g. Gardin 1973 and Hutchins 1975, 33.). Here, however, my point is that documentary history can in itself, not only, for instance, the natural language by which things (often but not always!) are expressed in documents (cf. Hutchins 1975), be conceived of as a semiotic process manifesting some semiotic systems. Thus the object semiotic of a DL would not, strictly speaking, be the natural language or text which documents contain, but rather the flow of documents in itself, especially the ways in which it structures content. Somewhat ironically it may be said that DLs explicate the content forms, the systems of content manifested in documentary history, irrespective of whether such systems did or did not exist. In content representation, and especially when constructing DLs, such systems are nevertheless presumed. In this sense I consider it reasonable to talk about DLs as metasemiotics which have as their object semiotics, i.e. as their content purport (knowledge about) documentary history, where documentary history in itself is conceived of as a semiotic process.

7. Communication Concerning Documents

On the basis of the arguments presented above I can question some prevailing ways of thinking within the field of research. The question is whether content representation may appropriately be seen as "organising knowledge", or if it were more plausible to view it as a particular communicative practice, as communication concerning documents, with the aim of helping people to have access to the documents which might be relevant to their needs and wishes. The answer would have implications with regard to both the self-understanding of the professional practices and the basic methodological positions to be assumed within the research. The ambitious perspective of organising knowledge would be substituted by the more humanities based, more modest, and more practical view of participating in the communication concerning documents. Content representation and especially the DLs manifested there, as viewed in the frame of classical structuralist semiology, especially the function of structuring content, turn out to be a practice and systems giving identities to particular contents. The semiology of DLs implies that the order created within knowledge is not unproblematically reducible to the structures and particular phenomena of knowledge itself. The semiology of the flow of documents further problematises the notion of knowledge in a special way. What documents keep in would not actually be knowledge but content structured in some particular way. The "same" knowledge may in different instances be formed in different ways. Then, knowledge or information as the content of documents is - to be precise - the content purport of documents, something that documents as a semiotic...
process structure through the systems manifested in this process. Thus the basic problem of content presentation would be the problem of assignability of the contents of documents, i.e., how to structure the structures of content manifested in documents. Thus we perhaps should only talk about organising knowledge in a weak sense, in the sense of giving identities for the content elements in documents in terms of another, perhaps quite different system of differences of contents, i.e., in terms of DL conceived of as a system of differences of content, as a culture on its own.

The efforts of communication concerning documents, as opposed to "knowledge organization", I would call documentation (or metadocumentation, for it could be useful to reserve the notion of documentation for a wider use), as opposed to information retrieval. Now a DL as a semiotic may be view as the basis of the communicative competence of documentation, if only communicative competence is conceived of not merely as the mastering of some systems of expressions, but more deeply, as having an appropriate system of (differences of) content in use. The system is appropriate if it is rich enough to provide all the contents and distinctions of content needed. When such a system of contents is created, the "intersubjectivity of possible action-oriented understanding" is "preserved and expanded", as Habermas (1965, 310) describes the interest of sciences which he calls hermeneutical. Basically this communicative competence may be in the heads of persons as well as incorporated in some tools (card catalogues as well as computerised systems).

On remark, however, have to be done yet. Saussure (1915, 155 ff.) characterises a language as a pure system of differences. The units of language (both particular expressions and contents) would be determined in a totally negative fashion, by their 'limiting' neighbours. Thus, the simplest possible system of content would be a distinction of two units of content. They could already limit and thus determine each other. Formally it could be presented as a system with contents A and B. A would be what is not B, and B correspondingly what is not A. However, this kind of determination would require at least some kind of pre-conception of A when determining B, and vice versa. But it would also require something else.

To give a concrete example, A could be for 'ships', and B 'aeroplanes' (as might be the case in the classification of a private collection of a schoolboy interested in sailing and flying). Then, according to the discursively negative determination of the content units, within this ultimately simple semiotic system 'ships' would be everything that is not 'aeroplanes', and vice versa. Within his ultimately simple classification, or semiotic, the schoolboy would probably count as 'ships' things which could in a richer semiotic system be expressed by terms such as 'sea', 'waves', 'icebergs', 'Titanic', etc. And the content of his expression 'aeroplanes' could include things like 'air ports', 'jet engines', 'bombers'.

Some comments are, however, called for. (1) He could have problems with things like 'navigation', 'Battle of Midway', etc., since these would be borderline cases. They could belong to either of the classes. (2) He could not within his semiotic system make a distinction between 'sailing boats' and 'steam ships', or between 'rudder' and 'rig', not to mention 'helm' and 'tiller'. He could, of course, solve both these problems quite easily by refining his semiotic system. However, it should further be noted that (3) even if 'rats' could still belong to the content of 'ships' (rats are on a ship in order to leave it before it sinks), 'ants' would be a real problem, not because it would be a borderline case, but because it is far too remote from the borderline. It is out of the interest of the hypothetical schoolboy and thus not to be classified in terms of his semiotic system. His interest, on the other hand, would come quite close to the notion of content purport. What is articulated, i.e. the content purport, would be the field of his interest. This is basically the point of Giddens (1979, 16), too, when he criticises Saussure's conception of the negative determination of the units of content, referring to Saussure's own 'Geneva-to-Paris train' example. The determination of the units of content
requires something more, a practice, an interest, or perhaps horizon within which it only can be conceived of in terms of distinctions. Within Hjelmslevian terms this can perhaps be grasped by the notion of content purport.

Now, taking into account my notion of (meta)documentation as defined above, the practice, interest or horizon within which the contents become determined would be that of 'communication concerning documents', and consequently, the content purport structured would be knowledge about documents. Thus my view presented here would be heavily document-oriented, as opposed to the prevailing trends within information science, i.e., to creating information systems or "organising knowledge" for the users. To argue for my document-orientation, I here present the following remarks:

(i) The culture of documents (outlined above in a quite sketchy manner in terms of the semiotic systems manifested in the flow of documents) would, after all, be the major problem in helping people to manage with documents.

(ii) Further, the contents of documents, conceived of as manifestations of some culturally existing, and probably historically emerged and developing systems, can hardly be reduced into an "organised knowledge", without ignoring most substantial cultural aspects.

References